DAWSON COLLEGE

STATISTICS

CONTENT CHARTS WITH EXERCISES

BY

© GEORGE McARTHUR

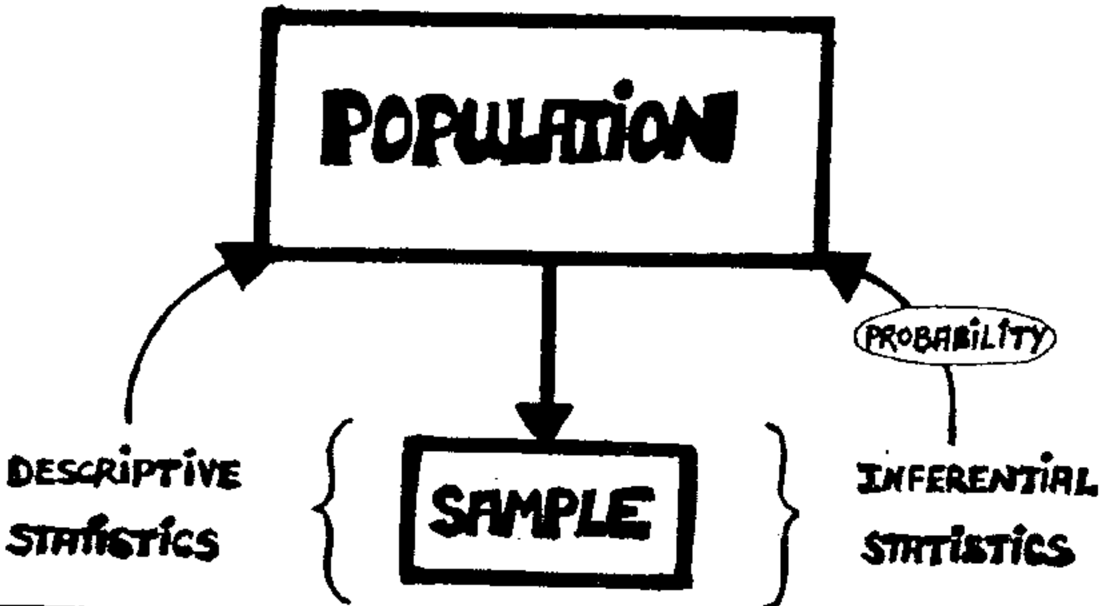
STATISTICS BY MCARTHUR

MCARTHUR
F14(S)
978-1-5501635-7-5
90000
9781550163575

CONTENTS

INTRODUCTORY CONCEPTS

A VIEW OF STATISTICS



A DEFINITION OF STATISTICS

THE COLLECTION AND ANALYSIS OF A <u>SAMPLE</u> IN ORDER TO DESCRIBE, AND MAKE INFERENCES ABOUT, THE <u>POPULATION</u>

INTRODUCTORY CONCEPTS

We consider definitions of the following basic terms:

POPULATION	The whole set of elements under consideration.	
SAMPLE	a subset of a population.	
VARIABLE	The characteristic of interest about each element of a population or sample.	
MEASUREMENT	The name (qualitative) or number (quantitative) assigned to a variable.	
DATA	The measurements from the elements of a sample.	
QUALITATIVE DATA	attribute data Example: Eye-colour	
QUANTITATIVE DATA	DISCRETE	Count data (natural numbers) Example: The number of children in a family
	CONTINUOUS	Calibrated data (real numbers) Example: The height of a person
PARAMETER	a numerical characteristic of a population	
SAMPLE STATISTIC	a (calculated) numerical characteristic of a sample.	

INTRODUCTORY CONCEPTS

PROBABILITY SAMPLES

The most useful samples are PROBABILITY SAMPLES where each element of the population has a non-zero probability (chance) of being selected for the sample. Common PROBABILITY SAMPLES are:

RANDOM SAMPLE	Where each element of the population has an EQUAL probability of being selected for the sample.
SYSTEMATIC SAMPLE	Consists of the SYSTEMATIC selection of every k^{th} one of the listed elements of the population, using a random starting point.
STRATIFIED SAMPLE	Consists of the several random samples chosen from each of several STRATIFICATIONS of the population.
CLUSTER SAMPLE	Consists of the random samples chosen from randomly selected CLUSTERS of the elements of the population

NOTE: When necessary NON-PROBABILITY SAMPLES may be selected based on judgement and/or convenience

LEVELS OF MEASUREMENTS

The 4 LEVELS OF MEASUREMENTS, from the lowest to the highest are:

NOMINAL	Grouping measurements by NAME, but not ordering them Example: eye-colours
ORDINAL	Ordering measurements, but not using equal intervals Example: school grades
INTERVAL	Ordering measurements using equal intervals, but no true zero. Example: temperatures
RATIO	Ordering measurements using equal intervals, and a true zero. Example: heights

DESCRIPTIVE STATISTICS – TOPICS

CLASSIFYING MEASUREMENTS

DESCRIBING MEASUREMENTS BY FREQUENCY DISTRIBUTION TABLES AND GRAPHS

CENTRAL TENDENCY STATISTICS

DESCRIBING THE CENTRAL TENDENCY OF MEASUREMENTS BY CALCULATING

THE MEAN | THE MEDIAN | THE MODE

VARIABILITY STATISTICS

DESCRIBING THE VARIABILITY OF MEASUREMENTS BY CALCULATING

THE VARIANCE | THE STANDARD DEVIATION | THE RANGE

ESTIMATING $\bar{X}$ AND S FROM FREQUENCY TABLES

ESTIMATING THE MEAN, $\bar{X}$, AND THE STANDARD DEVIATION, S, OF A SAMPLE BY USING CLASS MARKS AND FREQUENCIES

THE EMPIRICAL RULE

DESCRIBING THE NORMAL CHARACTERISTICS OF MEASUREMENTS THAT HAVE BELL-SHAPED FREQUENCY DISTRIBUTION GRAPHS

MEASURES OF POSITION (PERCENTILES)

DESCRIBING THE RANKED POSITION OF A MEASUREMENT IN THE ARRAY OF MEASUREMENTS

CLASSIFYING MEASUREMENTS

TO CLASSIFY A SAMPLE OF n MEASUREMENTS (ie OBSERVATIONS), WE CONSIDER:

ARRAY: The ranking (ie ordering) of the observations from the smallest to the largest.

THE NUMBER OF CLASSES	THE WIDTH* OF THE CLASSES
$\approx 1 + 3.3 \log n$	$\approx \dfrac{\text{largest obs.} - \text{smallest obs.}}{\text{number of classes}}$

CLASS DIVISIONS

LIMITS: The smallest (lower limit) and the largest (upper limit) possible observations in a class.

BOUNDARIES: The midpoints between the upper limit of one class and the lower limit of the next class.

CLASS MARKS: The midpoints between the boundaries (or limits) of a class.

CLASS FREQUENCIES

FREQUENCY, f: The number of observations in a class.

RELATIVE FREQUENCY, rf: The frequency of a class, divided by n.

CUMULATIVE FREQUENCY, cf: The number of observations less than the upper boundary of a class.

CUMULATIVE RELATIVE FREQUENCY, crf: The cumulative frequency of a class, divided by n.

FREQUENCY DISTRIBUTION GRAPHS

HISTOGRAMS: bar graphs with horizontal scale of boundaries.

POLYGONS: line graphs with horizontal scale of:
marks for f and rf and upper boundaries for cf and crf

*Actual Class Width = Upper Class Limit (UCL) of a class − UCL of the preceding class

CLASSIFYING MEASUREMENTS – AN ILLUSTRATIVE EXAMPLE

SAMPLE DATA (College Course Grades):

38	32	35	50	60	70	82	93	40	43
45	45	49	97	89	89	70	71	51	65
65	83	85	72	75	68	55	75	59	85
65	78	73	67	62	54	55	60	63	65
87	76	69	62	66	75	58	61	64	57

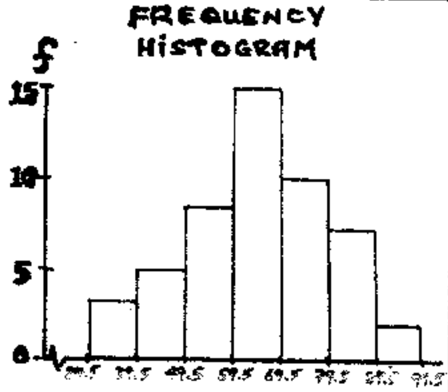
To **CLASSIFY** THE **50** MEASUREMENTS ABOVE WE PROCEED AS FOLLOWS:

① Form the **ARRAY** of the data and determine the **NUMBER** and **WIDTH** of classes.
② Determine the **CLASS LIMITS, BOUNDARIES,** and **MARKS**.
③ Construct the **FREQUENCY DISTRIBUTION TABLE** and draw **GRAPHS**.

ARRAY	32,35,38,40,43,45,45,49,50,51,54,55,55,57,58,59,60,60,61,62,62,63,64,65,65, 65,65,66,67,68,69,70,70,71,72,73,75,75,75,76,78,82,83,85,85,87,89,89,93,97

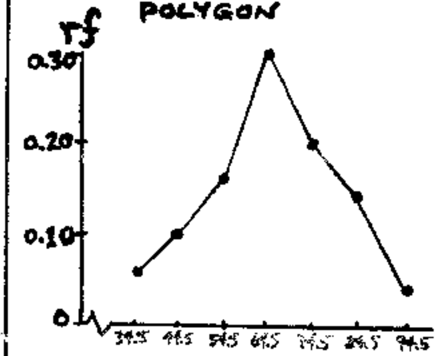
CLASSES	LIMITS	BOUNDARIES	MARKS	f	rf	cf	crf
1	30 – 39	29.5 – 39.5	34.5	3	.06	3	.06
2	40 – 49	39.5 – 49.5	44.5	5	.10	8	.16
3	50 – 59	49.5 – 59.5	54.5	8	.16	16	.32
4	60 – 69	59.5 – 69.5	64.5	15	.30	31	.62
5	70 – 79	69.5 – 79.5	74.5	10	.20	41	.82
6	80 – 89	79.5 – 89.5	84.5	7	.14	48	.96
7	90 – 99	89.5 – 99.5	94.5	2	.04	50	1.00

FREQUENCY HISTOGRAM



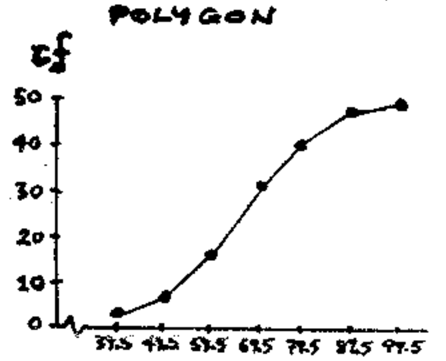
CLASS BOUNDARIES

RELATIVE FREQUENCY POLYGON



CLASS MARKS

CUMULATIVE FREQUENCY POLYGON



UPPER CLASS BOUNDARIES

THE CENTRAL TENDENCY STATISTICS

THE DEFINITION OF THE MEAN, $\overline{X}$

For a sample of **n** observations we calculate the **MEAN, $\overline{X}$,** as:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \quad \text{where} \quad \sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \cdots + X_n$$

$$\text{or} \quad \sum_{i=1}^{n} X_i = \text{"the sum of the } n \text{ observations"}$$

THE DEFINITION OF THE MEDIAN

The **MEDIAN** is the middle observation in the **ARRAY** of an odd number of observations, otherwise it is the mean of the 2 middle observations.

THE DEFINITION OF THE MODE

The **MODE** is the observation that occurs with the greatest frequency.

OUR ILLUSTRATIVE EXAMPLE

Using our sample of 50 grades we note that:

$$\text{MEAN} = \overline{X} = \frac{\sum_{i=1}^{50} X_i}{50} = \frac{38+32+35+\cdots+57}{50} = \frac{3253}{50} = 65.06$$

$$\text{MEDIAN} = \frac{65+65}{2} = \frac{130}{2} = 65$$

$$\text{MODE} = 65$$, with a frequency of 4.

THE VARIABILITY STATISTICS

THE DEFINITION OF THE VARIANCE, S^2

For a sample of n observations we calculate the **VARIANCE, S^2**, as:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}, \text{ where } \overline{X} \text{ is the mean of the observations}$$

THE DEFINITION OF THE STANDARD DEVIATION, S

For a sample of n observations we calculate the **STANDARD DEVIATION, S**, as:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}, \text{ also, } S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}}$$

THE DEFINITION OF THE RANGE

The **DIFFERENCE** between the **LARGEST** and **SMALLEST** observations.

OUR ILLUSTRATIVE EXAMPLE

Using our sample of 50 grades we note that:

$$S = \sqrt{\frac{\sum_{i=1}^{50}(X_i - 65.06)^2}{50-1}} = \sqrt{\frac{\sum_{i=1}^{50} X_i^2 - \frac{\left(\sum_{i=1}^{50} X_i\right)^2}{50}}{49}} = \sqrt{\frac{223179 - \frac{(3253)^2}{50}}{49}}$$

$$\therefore S = \sqrt{235.49} = 15.35 \quad (\text{hence } S^2 = 235.49)$$

and, the **RANGE** $= 97 - 32 = 65$

ESTIMATING $\overline{X}$ AND S FROM FREQUENCY TABLES

THE TWO ESTIMATING CALCULATIONS

Consider n observations classified into k classes, where

f_i is the FREQUENCY of the i^{th} class, for $i = 1, 2, 3, \cdots, k$ and

M_i is the MARK of the i^{th} class, for $i = 1, 2, 3, \cdots, k$.

$$\overline{X} \approx \frac{\sum_{i=1}^{k} f_i(M_i)}{n} \quad \text{and} \quad S \approx \sqrt{\frac{\sum_{i=1}^{k} f_i(M_i)^2 - \frac{\left[\sum_{i=1}^{k} f_i(M_i)\right]^2}{n}}{n-1}}$$

OUR ILLUSTRATIVE EXAMPLE

Using our sample of **50** grades and the **7** classes, we note that:

$$\overline{X} \approx \frac{\sum_{i=1}^{7} f_i(M_i)}{50} = \frac{3(34.5) + 5(44.5) + \cdots + 2(94.5)}{50} = \frac{3255}{50} \approx 65.10$$

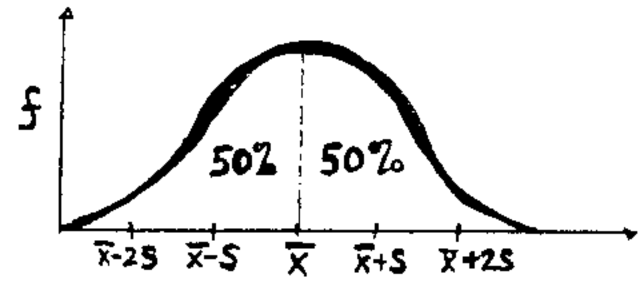
and

$$S \approx \sqrt{\frac{\sum_{i=1}^{7} f_i(M_i)^2 - \frac{\left[\sum_{i=1}^{7} f_i(M_i)\right]^2}{50}}{50-1}}$$

$$\therefore S \approx \sqrt{\frac{3(34.5)^2 + 5(44.5)^2 + \cdots + 2(94.5)^2 - \frac{[3255]^2}{50}}{49}} = \sqrt{\frac{222982.5 - 211900.5}{49}} = 15.04$$

THE EMPIRICAL RULE

For DATA with a "**BELL-SHAPED**" frequency distribution as follows:

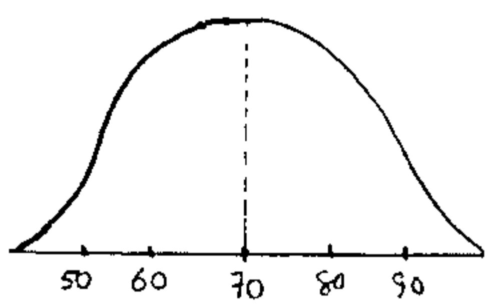


The **EMPIRICAL RULE** applies **NORMALLY** as follows:

$\bar{X} \pm S$ WILL INCLUDE APPROXIMATELY **68%** OF THE DATA
$\bar{X} \pm 2S$ WILL INCLUDE APPROXIMATELY **95%** OF THE DATA

AN ILLUSTRATIVE EXAMPLE

Consider a frequency distribution of test scores that is **BELL-SHAPED**, with $\bar{X} = 70$ and $S = 10$, as follows:



Then, applying the **EMPIRICAL RULE** we note that:

70 ± 10 will include approximately **68%** of the test scores
hence, **68%** of the test scores will be between **60** and **80**.

and

70 ± 20 will include approximately **95%** of the test scores
hence, **95%** of the test scores will be between **50** and **90**.

MEASURES OF POSITION (PERCENTILES)

THE DEFINITION OF PERCENTILES

For a sample of n observations, THE k^{th} PERCENTILE, P_k, is that observation, in the ARRAY of the n observations, such that at most k% of the n observations are less than P_k.

THE DEFINITION OF THE QUARTILES

For a sample of n observations we consider the 3 QUARTILES as:

THE 1^{st} QUARTILE, Q_1, is given by P_{25},
THE 2^{nd} QUARTILE, Q_2, is given by P_{50}, and
THE 3^{rd} QUARTILE, Q_3, is given by P_{75}.

OUR ILLUSTRATIVE EXAMPLE

Using our sample of 50 grades we recall that the ARRAY was:

ARRAY	32, 35, 38, 40, 43, 45, 45, 49, 50, 51, 54, 55, 55, 57, 58, 59, 60, 60, 61, 62, 62, 63, 64, 65, 65, 65, 65, 66, 67, 68, 69, 70, 70, 71, 72, 73, 75, 75, 75, 76, 78, 82, 83, 85, 85, 87, 89, 89, 93, 97

To find P_{15}: consider that $0.15(50) = 7.5$, then P_{15} is the 8^{th} observation in the ARRAY, i.e. 49

Also: $P_{90} = 87$, and

$Q_1 = P_{25} = 55$

$Q_2 = P_{50} = 65$

$Q_3 = P_{75} = 75$

DESCRIPTIVE STATISTICS EXERCISES

(1) Consider the classification of the heights of a sample of 90 Dawson male students shown to the right.

(a) Draw the frequency histogram for the data.

(b) Draw the cumulative frequency polygon also.

HEIGHT (nearest cm.)	NUMBER OF STUDENTS
140 - 149	5
150 - 159	13
160 - 169	49
170 - 179	17
180 - 189	6

(2) Consider the classification of the years of seniority of a sample of 50 CEGEP teachers shown to the right.

(a) Draw the relative frequency polygon for the data.

(b) Draw the cumulative relative frequency polygon.

SENIORITY (YEARS)	NUMBER OF TEACHERS
0 - 2	2
3 - 5	5
6 - 8	11
9 - 11	19
12 - 14	9
15 - 17	4

(3) The following are the miles per gallon obtained with 40 tankfuls of gas:

24.1	25.0	24.8	24.3	24.2	25.3	24.2	23.6	24.5	24.4
24.5	23.2	24.0	23.8	23.8	25.3	24.5	24.6	24.0	25.2
25.2	24.4	24.7	24.1	24.6	24.9	24.1	25.8	24.2	24.2
24.8	24.1	25.6	24.5	25.1	24.6	24.3	25.2	24.7	23.3

(a) Construct the complete Frequency Distribution Table for the data, using 6 classes.

(b) Draw the frequency histogram for the data.

(c) Draw the relative frequency polygon for the data.

(d) Draw the cumulative relative frequency polygon for the data.

(4) From the classification of 20 observations shown here, find:

(a) the number of decimal places the observations have.

(b) the lower and upper boundaries of the 3rd class.

(c) the midpoint of the 4th class.

(d) the frequency of the 2nd class.

(e) the relative frequency of the 4th class.

CLASS	CLASS LIMITS	crf
1	10.0 - 14.9	0.10
2	15.0 - 19.9	0.35
3	20.0 - 24.9	0.75
4	25.0 - 29.9	0.95
5	30.0 - 34.9	1.00

(5) Consider the 9 measurements that are recorded below:

59.6, 53.7, 39.5, 75.5, 26.2, 30.7, 68.8, 51.2, 42.8

Calculate the mean, median, and standard deviation of these 9 measurements.

(6) Calculate the mean, median, mode, and standard deviation for the following sample of class test scores: 41, 18, 25, 30, 49, 35, 11, 50, 36, 46, 34, 36

(7) Consider the following daily profit figures for 20 Montréal newsstands:

81.32	61.47	64.90	70.88	76.02	75.41	64.21
74.92	77.56	58.01	68.05	73.37	76.73	65.43
74.76	76.51	65.10	76.02	75.06	59.41	

Find the mean, median, mode, and standard deviation of the figures.

(8) Consider the numbers of children in 50 completed families given below.

children / family	0	1	2	3	4	5	6	7
# of families	9	7	12	9	5	6	0	2

Regarding the number of children / family, what is the mean, median, and mode

(9) Find the mean, median, mode, and standard deviation for the following classification of the ages of university students:

age	20	21	22	23	24	25
# of students	2	20	35	19	10	6

(10) Consider the following grades from 2 sections of a Statistics course:

SECTION I			
70	80	77	70
63	66	75	68
72	69	65	71
60	67	74	68
72	73	70	70

SECTION II			
90	70	77	74
55	85	82	63
70	61	67	67
70	75	70	79
58	50	72	65

Calculate and COMPARE the mean, median, mode, and standard deviations of the 2 sections.

(11) Given that $\bar{X} = 10$ and $\sum_{i=1}^{25} X_i^2 = 3100$ for a sample of 25 observations, find S.

(12) If the average age of the Dawson basketball team was 18 years, with a standard deviation of 0 years, what could we conclude about the team?

(13) If the scores for a test are changed by:
(a) adding 10 points to each score, and (b) increasing each score by 10%, what effect will these changes have on $\bar{X}$ and S ?

(14) A study of teachers' annual salaries in Québec reported that samples of 100 elementary. 150 high school, and 80 CEGEP teachers averaged \$44000, \$47000, and \$50000 respectively. What is the overall average salary of these teachers?

(15) An elevator is designed for a maximum load of 2000 lbs. Is it overloaded, if at one time it carries 8 women whose mean weight is 123 lbs. and 5 men whose mean weight is 174 lbs. ?

(16) Last term the mean grade for a class of 100 students was 80. If the mean grade of the 60 boys in the class was 70, what was the mean grade of the girls in the class

(17) Consider the classification to the right of the hourly rates (in dollars) charged by 40 Montréal lawyers.
(a) What % of these lawyers charge less than \$50/hr.?
(b) What % of these lawyers charge at least \$60/hr ?
(c) Estimate the mean hourly rate charged by these lawyers.

HOURLY RATES	NUMBER OF LAWYERS
\$40 - \$49	4
50 - 59	7
60 - 69	18
70 - 79	9
80 - 89	2

(18)

CLASS LIMITS	f
3 - 5	2
6 - 8	10
9 - 11	12
12 - 14	9
15 17	7

Consider the table at the left giving the frequency distribution for the number of house sales made last month by 40 Montréal real estate agents. Estimate $\bar{X}$ and S.

(19) Consider the classification to the right of the ages of a sample of 150 NHL players. Estimate $\overline{X}$ and S.

AGES (YEARS)	# OF PLAYERS
17 - 19	5
20 - 22	63
23 - 25	39
26 - 28	24
29 - 31	17
32 - 34	2

(20)

BOOKS LOANED	NUMBER OF DAYS
100 - 119	3
120 - 139	6
140 - 159	13
160 - 179	7
180 - 199	2

Consider the classification to the left of the number of books loaned per day by a college library for a recent month. Estimate $\overline{X}$ and S.

(21) Incomes in a certain city are known to have a bell-shaped distribution. If the mean income is \$18,000 with a standard deviation of \$2,000, then what percentage of the city's wage earners

(a) earn more than \$22,000?
(b) earn less than \$20,000?
(c) earn between \$16,000 and \$22,000?

(22) Recently, 600 students wrote an aptitude test, yielding $\overline{X} = 325$ and $S = 75$. If the distribution of the test scores was bell-shaped, then (approximately) how many of the students

(a) scored less than 250?
(b) scored more than 475?

(23) A tire manufacturer claims that his tires will last an average of 12,400 miles (with $S = 1800$). He guarantees to replace free any tire lasting less than 8800 miles. Assuming a bell-shaped life distribution, if he sells 10,000 of these tires to a large taxi fleet, (approximately) how many will he have to replace free?

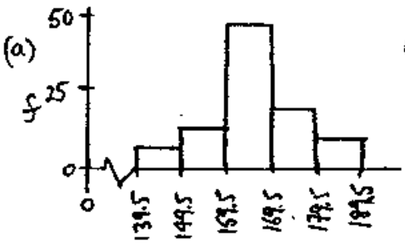
(24) For the data in exercises (3) and (7) above, determine P_{10}, P_{90}, Q_1, Q_2, and Q_3 for each sample.

DESCRIPTIVE STATISTICS EXERCISES — SOLUTIONS

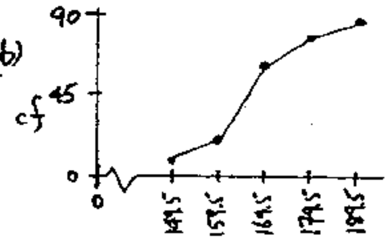
(1) ie

Boundaries	f	cf
139.5-149.5	5	5
149.5-159.5	13	18
159.5-169.5	49	67
169.5-179.5	17	84
179.5-189.5	6	90

(a)



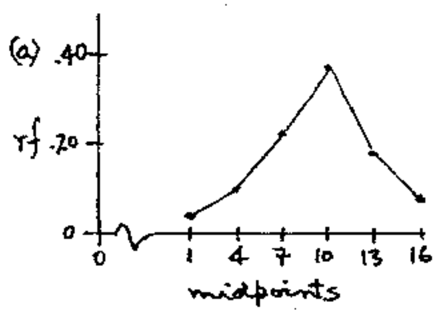
(b)



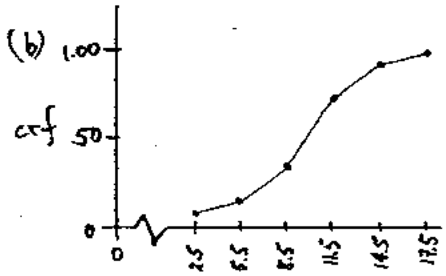
(2) ie

Boundaries	f	rf	cf	crf
-0.5 - 2.5	2	.04	2	.04
2.5 - 5.5	5	.10	7	.14
5.5 - 8.5	11	.22	18	.36
8.5 - 11.5	19	.38	37	.74
11.5 - 14.5	9	.18	46	.92
14.5 - 17.5	4	.08	50	1.00

(a)



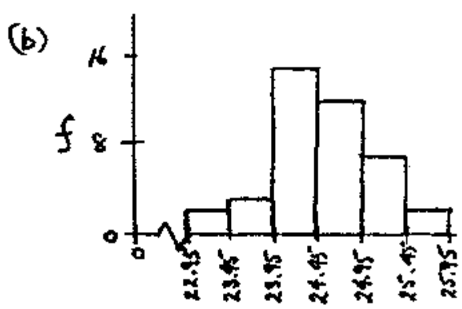
(b)



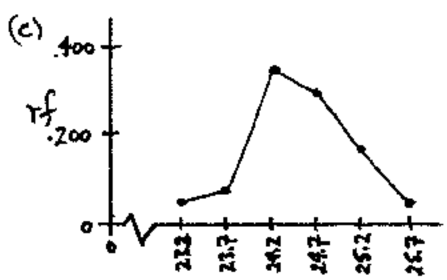
(3) (a)

Limits	Boundaries	Midpoints	f	rf	cf	crf
23.0 - 23.4	22.95 - 23.45	23.2	2	.050	2	.050
23.5 - 23.9	23.45 - 23.95	23.7	3	.075	5	.125
24.0 - 24.4	23.95 - 24.45	24.2	14	.350	19	.475
24.5 - 24.9	24.45 - 24.95	24.7	12	.300	31	.775
25.0 - 25.4	24.95 - 25.45	25.2	7	.175	38	.950
25.5 - 25.9	25.45 - 25.95	25.7	2	.050	40	1.000

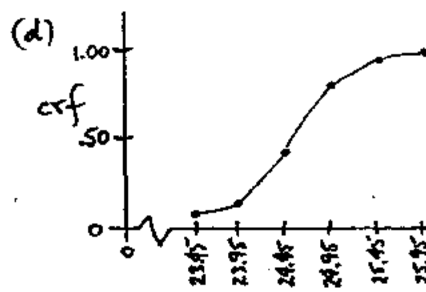
(b)



(c)



(d)



(4)
(a) one
(b) 19.95 - 24.95
(c) 27.45
(d) $(.35-.10)20 = 5$
(e) $.95-.75 = .20$

(5)
$\bar{X} = \frac{448}{9}$
$\therefore \bar{X} = 49.7\dot{7}$
$\tilde{X} = 51.2$
$S = 16.63$
$(\Sigma X^2 = 24512)$

(6)
$\bar{X} = \frac{411}{12} = 34.25$
$\tilde{X} = \frac{35+36}{2} = 35.5$
Mode = 36
$S = 11.92$
$(\Sigma X^2 = 15641)$

(7)
$\bar{X} = \frac{1415.14}{20} = 70.76$
$\tilde{X} = \frac{73.37 + 74.76}{2} = 74.07$
Mode = 76.02
$S = 6.82$
$(\Sigma X^2 = 101{,}015.82)$

(8) $\bar{X} = \dfrac{9(0)+7(1)+12(2)+\cdots+2(7)}{50}$

$\therefore \bar{X} = \dfrac{122}{50} = 2.44$

$\tilde{X} = 2$ and Mode $= 2$

(9) $\bar{X} = \dfrac{2(20)+20(21)+\cdots+6(25)}{92}$ and $S = 1.19$

$\therefore \bar{X} = \dfrac{2057}{92} = 22.36$ $\quad (\Sigma X^2 = 46121)$

$\tilde{X} = 22$ and Mode $= 22$

(10) All 70 except $S_1 = 4.68$ and $S_2 = 9.96$

(11) $\Sigma X_i = 25(10)$

$\therefore \Sigma X_i = 250$

$\therefore S = 5$

(12) All age 18

(13) (a) $\bar{X}$ up 10 pts. S no change (b) both up 10%

EG	1,2,3	11,12,13	1.1,2.2,3.3
$\bar{X}$	2	12	2.2
S	1	1	1.1

(14) $\bar{X} = \dfrac{100(44000)+150(47000)+80(50000)}{330}$

$\therefore \bar{X} = \dfrac{15450000}{330} = \$46818.18$

(15) Σ total $= 8(123)+5(174)$

$= 1854$

$\therefore$ Not overloaded

(16) $\dfrac{\Sigma X_F + 60(70)}{100} = 80$

$\therefore \Sigma X_F = 3800$

$\therefore \bar{X}_F = \dfrac{3800}{40} = 95$

(17) (a) $\dfrac{4}{40} = .10 = 10\%$ (b) $\dfrac{18+9+2}{40} = \dfrac{29}{40} = .725 = 72.5\%$

(c) $\bar{X} \approx \dfrac{4(44.5)+7(54.5)+18(64.5)+9(74.5)+2(84.5)}{40}$

$\therefore \bar{X} \approx \dfrac{2560}{40} = 64$

(18) $\Sigma f_i M_i = 427$ and $\Sigma f_i M_i^2 = 5035$

$\therefore \bar{X} \approx 10.675$

$\therefore S \approx 3.496$

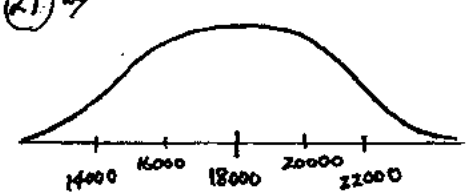
(19) $\Sigma f_i M_i = 3573$ and $\Sigma f_i M_i^2 = 86841$

$\therefore \bar{X} \approx 23.82$ and $S \approx 3.41$

(20) $\Sigma f_i M_i = 4614.5$ and $\Sigma f_i M_i^2 = 700077.75$

$\therefore \bar{X} \approx 148.85$ and $S = 20.97$

(21) ie

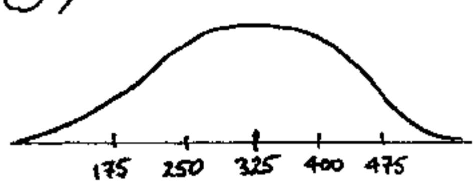


(a) 2.5%

(b) 84%

(c) 34% + 47.5% = 81.5%

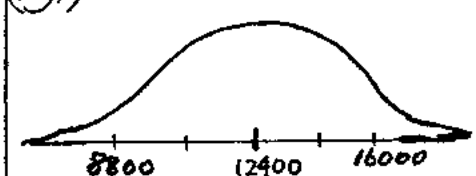
(22) ie



(a) $600 \times (0.16) = 96$

(b) $600 \times (0.025) = 15$

(23) ie



$\therefore$ ANS. $= 10000\left(\dfrac{0.05}{2}\right)$

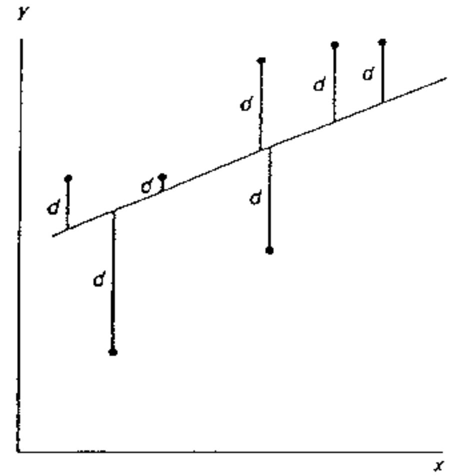
$= 10000(0.025) = 250$

(24) For Exercise (3): $P_{10} = 23.8$, $P_{90} = 25.3$, $Q_1 = 24.1$, $Q_2 = 24.5$, and $Q_3 = 24.9$

For Exercise (4): $P_{10} = 61.47$, $P_{90} = 77.56$, $Q_1 = 65.10$, $Q_2 = 74.76$, and $Q_3 = 76.02$

LINEAR REGRESSION AND CORRELATION

Consider the scatter diagram:



From calculus we minimize $\sum d^2$ to give:

THE LEAST-SQUARES LINE

$$y = a + bx$$

where

$$b = \frac{SS_{xy}}{SS_x}$$ is the slope, and

$$a = \bar{y} - b\bar{x}$$ is the y-intercept

REQUIRED FORMULAS:

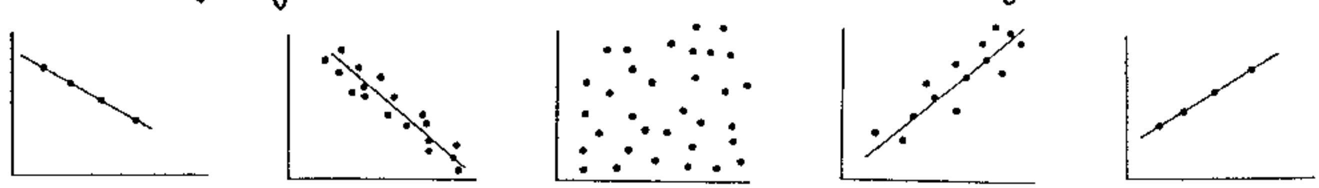
$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}, \quad SS_x = \sum x^2 - \frac{(\sum x)^2}{n}, \quad SS_y = \sum y^2 - \frac{(\sum y)^2}{n}$$

To measure the strength of the LINEAR CORRELATION between x and y we use:

THE LINEAR CORRELATION COEFFICIENT

$$r = \frac{SS_{xy}}{\sqrt{SS_x \, SS_y}}$$

By definition $-1 \le r \le 1$, described as follows:



$r = -1$ | $-1 < r < 0$ | $r = 0$ | $0 < r < 1$ | $r = 1$

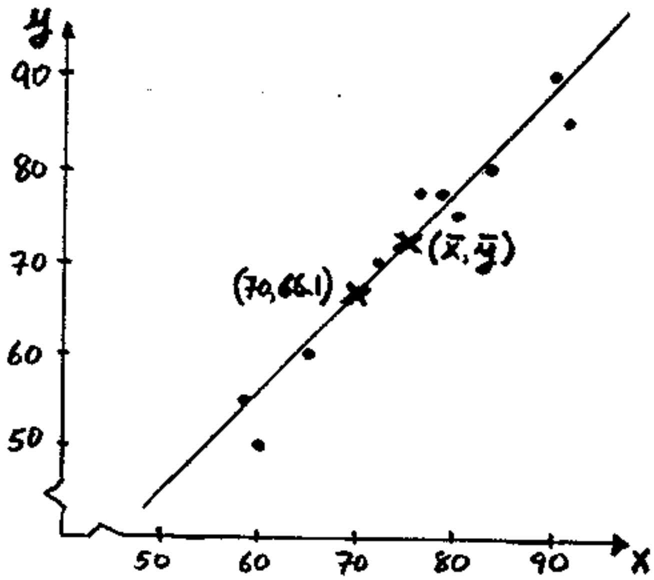
NOTE: We also consider r^2, THE COEFFICIENT OF DETERMINATION, as a measure of the proportion of variation in y that is explained by the regression line (ie. the least-squares line).

LINEAR REGRESSION AND CORRELATION - AN ILLUSTRATIVE EXAMPLE

Consider the bivariate data representing grades for 10 students in a course.

X, MIDTERM GRADE	80	76	65	91	90	60	78	83	58	72
y, FINAL GRADE	75	77	60	85	90	50	77	80	55	70

THE SCATTER DIAGRAM FOR THE DATA



CALCULATIONS

$\Sigma x = 753$, $\Sigma y = 719$, $\Sigma xy = 55463$

$\Sigma x^2 = 57903$, $\Sigma y^2 = 53233$

$\bar{x} = \frac{753}{10} = 75.3$, $\bar{y} = \frac{719}{10} = 71.9$

$SS_x = 57903 - \frac{(753)^2}{10} = 1202.1$

$SS_y = 53233 - \frac{(719)^2}{10} = 1536.9$

$SS_{xy} = 55463 - \frac{(753)(719)}{10} = 1322.3$

then, $b = \frac{SS_{xy}}{SS_x} = \frac{1322.3}{1202.1} = 1.1$ and $a = \bar{y} - b\bar{x} = 71.9 - 1.1(75.3) = -10.9$, hence

THE LEAST-SQUARES LINE
$y = -10.9 + 1.1x$

LINEAR REGRESSION: What final grade is predicted by a midterm grade of 70?

Consider $y = -10.9 + 1.1(70) = 66.1$

THE GRAPH OF $y = a + bx$: Plot the 2 points $(70, 66.1)$ and $(\bar{x}, \bar{y})$ on the scatter diagram above, and draw the line through them.

NOTE: If $x = \bar{x}$, then $y = a + b\bar{x} = (\bar{y} - b\bar{x}) + b\bar{x} = \bar{y}$. Thus $(\bar{x}, \bar{y})$ is always a point on the line. Here $(\bar{x}, \bar{y}) = (75.3, 71.9)$.

THE LINEAR CORRELATION COEFFICIENT:

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{1322.3}{\sqrt{(1202.1)(1536.9)}} = 0.97$$

LINEAR REGRESSION AND CORRELATION EXERCISES

(1) Consider the bivariate data:

x	2	3	3	4	5	5	6
y	7	6	5	4	3	2	1

(a) Draw a scatter diagram for the data.

(b) Find the equation of the least-squares line.

(c) Graph the least-squares line on your scatter diagram.

(d) Find the linear correlation coefficient, r, for the data.

For exercises (2)-(7) below, find <u>the least-squares line</u> and <u>calculate r</u>.

(2)

x	0	1	2	3	4
y	1	3	5	7	9

(3)

x	400	600	500	600	400	500
y	44	47	48	48	43	46

(4)

x	27	22	15	35	30	52	35	55	40	40
y	30	26	25	42	38	40	32	54	50	43

(5)

x	3	5	7	9	15	20	25	27	33	35
y	60	55	57	50	47	44	48	40	35	30

(6)

x	10	15	16	1	4	6	18	12	14	7
y	5	2	1	9	7	8	1	5	3	6

(7)

x	117	92	102	115	87	76	107	108	121	91	113	98
y	3.7	2.6	3.3	2.2	2.4	1.8	2.8	3.2	3.8	3.0	4.0	3.5

(8) Consider the data below giving the weight (in thousands of pounds) x and gasoline mileage (miles per gallon) y for 10 automobiles.

x	2.5	3.0	4.0	3.5	2.7	4.5	3.8	2.9	5.0	2.2
y	40	43	30	35	42	19	32	39	15	44

(a) Find the least-squares line for the data.
(b) Calculate r.
(c) Use the least-squares line to predict the mpg. for a weight of 4.2.

(9) A random sample of 10 male soccer players yielded these heights and weights (in inches and pounds, x and y, respectively):

x	65	67	67	68	69	70	71	72	72	75
y	142	154	143	155	161	158	179	157	181	190

(a) Find the least-squares line for the data.
(b) Calculate r.
(c) What is the predicted weight for a 72-inch tall soccer player?

(10) The midterm grades, x, and the final grades, y, were recorded for 25 students as follows:

x	76	75	66	61	72	56	72	85	63	60	80	25	66	76	89	59
y	81	84	81	62	60	69	82	82	85	53	81	49	84	56	83	70

x	78	68	40	82	77	76	90	76	56
y	65	91	48	86	88	83	94	79	67

(a) Find the least-squares line for the grade data.

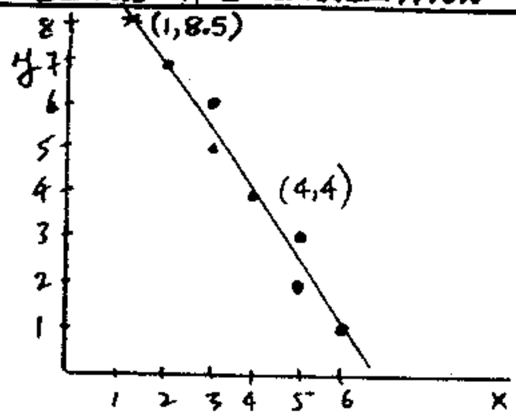
(b) Calculate r.

(c) What final grade is predicted by a midterm grade of 70?

(d) " " " " " " " " " of 90?

LINEAR REGRESSION AND CORRELATION EXERCISES – SOLUTIONS

(1) (a)



(b) $y = 10 - 1.5x$

(c) use $(\bar{x}, \bar{y}) = (4,4)$ and $(1, 8.5)$ say

(d) $r = -0.98$

(2) $y = 1 + 2x$, $r = 1.0$

(3) $y = 36 + 0.02x$, $r = 0.85$

(4) $y = 14.9 + 0.66x$, $r = 0.84$

(5) $y = 60.42 - 0.77x$, $r = -0.95$

(6) $y = 9.75 - 0.49x$, $r = -0.97$

(7) $y = -0.358 + 0.033x$, $r = 0.664$

(8) (a) $y = 70.11 - 10.62x$

(b) $r = -0.95$

(c) 25.51 mpg.

(9) (a) $y = -167.8 + 4.74x$

(b) $r = 0.88$

(c) 173.5 pounds

(10) (a) $y = 31.31 + 0.63x$

(b) $r = 0.68$

(c) ≈ 75.2

(d) ≈ 87.7

PROBABILITY - TOPICS

1 COUNTING TECHNIQUES

THE FUNDAMENTAL PRINCIPLE OF COUNTING

FACTORIAL NOTATION

PERMUTATIONS

COMBINATIONS

2 THE DEFINITION OF PROBABILITY

EVENTS AND SAMPLE SPACES

EQUALLY-LIKELY POSSIBLE OUTCOMES

3 PROBABILITIES OF COMPOUND EVENTS

UNIONS AND INTERSECTIONS

THE ADDITIVE LAW OF PROBABILITY

MUTUALLY EXCLUSIVE EVENTS

CONDITIONAL PROBABILITY

THE MULTIPLICATIVE LAW OF PROBABILITY

INDEPENDENT EVENTS

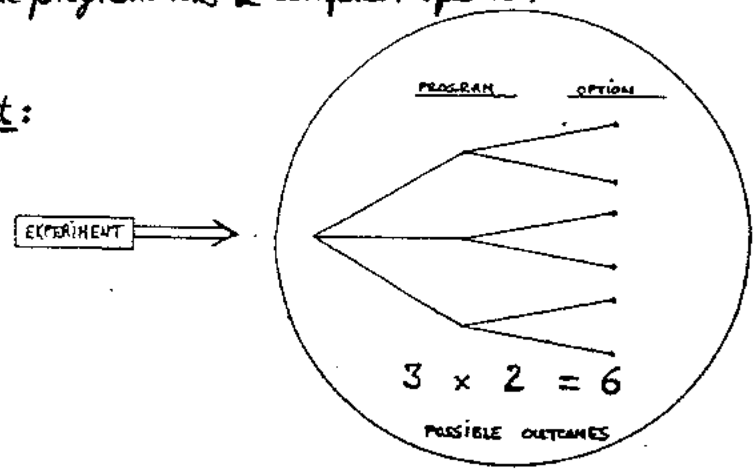
1 Counting Techniques

We first consider the product principle of counting:

The Fundamental Principle of Counting
If an experiment consists of k steps, with $n_1, n_2, n_3, \ldots, n_k$ possibilities, respectively, then there are $n_1 \cdot n_2 \cdot n_3 \cdot \cdots \cdot n_k$ different possible outcomes to the experiment.

<u>Illustration</u>: Consider the experiment of choosing one of 3 university programs, where each program has 2 computer options.

<u>We note that</u>:



The following examples also illustrate the Fundamental Principle of Counting.

<u>EXAMPLE 1.1</u> Consider the experiment of tossing a coin 5 times. How many different possible outcomes are there?

> There are $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$ different possible outcomes.

<u>EXAMPLE 1.2</u> Consider the experiment of rolling 2 dice once. How many different possible outcomes are there?

> There are $6 \cdot 6 = 36$ different possible outcomes.

<u>EXAMPLE 1.3</u> How many different pizzas can be ordered, if one choose from 4 meats, 3 cheeses, and 5 sizes?

> There are $4 \cdot 3 \cdot 5 = 60$ different possible pizzas.

EXAMPLE 1.4 A first prize and a second prize are to be awarded to 2 students in a class of 40 students. How many different ways can this be done?

There are $40 \cdot 39 = 1560$ different ways

EXAMPLE 1.5 How many different orderings of 5 people each are there?

There are $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ different orderings

DEFINITION 1.1 $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$ is called n FACTORIAL, and $0! = 1$

EXAMPLE 1.6 Evaluate (a) $5!$ (b) $10!$

(a) $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ and (b) $10! = 10 \cdot 9 \cdot 8 \cdot \cdots \cdot 3 \cdot 2 \cdot 1 = 3628800$

EXAMPLE 1.7 How many different ways can 7 distinct books be ordered on a shelf, (a) with no restrictions? (b) with the most popular book in the middle?

(a) $7! = 5040$ (b) $6 \cdot 5 \cdot 4 \cdot 1 \cdot 3 \cdot 2 \cdot 1 = 6! = 720$

EXAMPLE 1.8 How many different orderings of 5 people each can be formed from 10 people?

There are $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$ different orderings

DEFINITION 1.2 The number of PERMUTATIONS (or different orderings) of n distinct elements taken r at a time is given by:

$$P(n,r) = \frac{n!}{(n-r)!}$$

EXAMPLE 1.9 Evaluate (a) $P(10,5)$ (b) $P(100,2)$

(a) $P(10,5) = \frac{10!}{(10-5)!} = \frac{10!}{5!} = 30240$ (b) $P(100,2) = \frac{100!}{(100-2)!} = \frac{100!}{98!} = 100 \cdot 99 = 9900$

EXAMPLE 1.10 How many different orderings of 9 batters each can be formed from 15 batters on a baseball team?

There are $P(15,9) = 1816214400$ different 9-man batting orders

EXAMPLE 1.11 How many different orderings of the letters of the word MISSISSIPPI are there?

Let N be the number of different such orderings, then:

$$4!\,4!\,2!\,N = 11!$$

$$N = \frac{11!}{4!\,4!\,2!} = 34650$$

EXAMPLE 1.12 A coin is tossed 5 times. How many different possible outcomes include 3 heads and 2 tails?

Consider the outcome HHHTT, then there are $\frac{5!}{3!\,2!} = 10$ different such outcomes

EXAMPLE 1.13 How many different combinations of 3 letters each can be formed from the letters A, B, C, D, E ?

Consider ABC, ABD, $\cdots$ and let N be the number of different such combinations, then

$$3!\,N = P(5,3)$$

$$N = \frac{P(5,3)}{3!} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!\,2!} = 10$$

DEFINITION 1.3 The number of **COMBINATIONS** of n distinct elements taken r at a time is given by:

$$C(n,r) = \frac{n!}{r!(n-r)!}$$

EXAMPLE 1.14 Evaluate (a) $C(5,3)$ (b) $C(10,5)$

(a) $C(5,3) = \frac{5!}{3!(5-3)!} = \frac{5!}{3!\,2!} = 10$ (b) $C(10,5) = \frac{10!}{5!(10-5)!} = \frac{10!}{5!\,5!} = 252$

EXAMPLE 1.15 How many different 5 card hands can be formed from a deck of 52 distinct playing cards?

There are $C(52,5) = \frac{52!}{5!\,47!} = 2598960$ different 5 card hands

EXAMPLE 1.16 How many different 5 member committees can be formed from a class of 20 students (made up of 12 boys and 8 girls)
(a) with no restrictions? (b) to include 3 boys and 2 girls?

(a) $C(20,5) = \frac{20!}{5!\,15!} = 15504$ (b) $C(12,3)\cdot C(8,2) = \frac{12!}{3!\,9!} \cdot \frac{8!}{2!\,6!}$

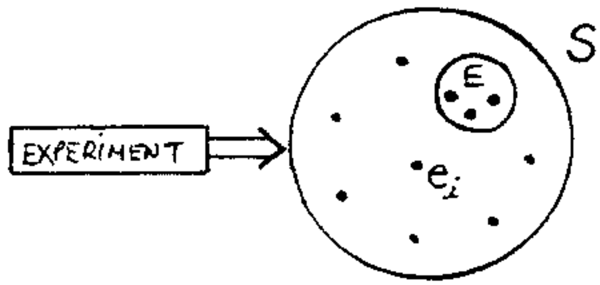
$$= 220 \cdot 28 = 6160$$

2 THE DEFINITION OF PROBABILITY

We now consider the Definition of Probability for the **DISCRETE** case (that is, where the possible outcomes are countable). The **CONTINUOUS** case will be considered later.

DEFINITION 2.1 THE DEFINITION OF PROBABILITY

Consider:



The Event, E, consists of m of the n possible outcomes (e_i's) in the Sample Space, S.

Then, we define Probabilities, $P(e_i)$, for $i = 1, 2, 3, \cdots, n$, such that:

① $P(e_i) \geq 0$ for each i, and

② $\displaystyle\sum_{i=1}^{n} P(e_i) = 1$, whereupon

$$\boxed{P(E) = P(e_1) + P(e_2) + P(e_3) + \cdots + P(e_m)}$$

The following examples illustrate the Definition of Probability.

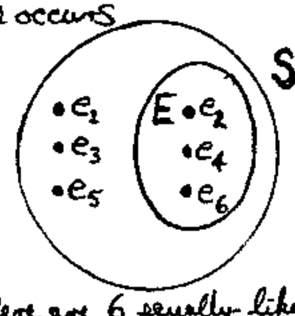
EXAMPLE 2.1 Consider the experiment of rolling a fair die once. What is the probability that an even face occurs?

Let E = the event that an even face occurs

Then, we define $P(e_i) = \frac{1}{6}$ for $i = 1,2,3,4,5,6$ such that:

① $P(e_i) = \frac{1}{6} \geq 0$, for each i , and

② $\displaystyle\sum_{i=1}^{6} P(e_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$



There are 6 equally likely possible outcomes

$$\therefore P(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$$

EXAMPLE 2.2 In a statistics class there are 10 science, 18 social, and 12 technical students. If a student is selected at random from this class, what is the probability that the student will be a science student?

> We note that there are $10+18+12=40$ equally-likely possibilities, hence the required probability is $\frac{10}{40}$.

Such results suggest the following definition:

> DEFINITION 2.2 For an experiment with n equally-likely possible outcomes, the probability of an event, E, consisting of m $(m \le n)$ of these possible outcomes, is given by: $$P(E) = \frac{m}{n}$$

The following examples further illustrate this definition.

EXAMPLE 2.3 A fair coin is tossed 5 times. What is the probability of observing 3 heads and 2 tails?

> There are 32 equally-likely possibilities (see example 1.1) and 10 of them include 3 heads and 2 tails (see example 1.11), hence, by definition 2.2 the required probability is $\frac{10}{32} = 0.3125$

EXAMPLE 2.4 Two fair dice are rolled once. What is the probability of observing a sum of 7?

> There are 36 equally-likely possibilities (see example 1.2) of which 6 give a sum of 7, i.e. (1,6), (6,1), (2,5), (5,2), (3,4), (4,3), hence the required probability is $\frac{6}{36} = 0.1\overline{6}$

EXAMPLE 2.5 A 5-member committee is chosen at random from a class of 20 students. If the class consists of 12 boys and 8 girls, find the probability that the committee will include 3 boys and 2 girls.

> Referring to example 1.15, the required probability is given as:
> $$\frac{\left(\frac{12!}{3!\,9!}\right)\cdot\left(\frac{8!}{2!\,6!}\right)}{\left(\frac{20!}{5!\,15!}\right)} = \frac{6160}{15504} = 0.3973$$

EXAMPLE 2.6 What is the probability that in a group of 5 unrelated people each of them has a different birth month?

Referring to definition 2.2 the required probability is given by:

$$\frac{P(12,5)}{12^5} = \frac{95040}{248832} = 0.382$$

EXAMPLE 2.7 An ordering is selected at random from all of the different possible orderings of the letters of the word MISSISSIPPI. What is the probability that the selected ordering begins and ends with P?

Referring to example 1.10 and the possibility P MISSISSII P, the required probability is given by:

$$\frac{\dfrac{9!}{4!\,4!}}{\dfrac{11!}{4!\,4!\,2!}} = \frac{630}{34650} = 0.018$$

EXAMPLE 2.8 (The 6/49 Lottery problem)

In the 6/49 Lottery, 6 winning numbers are selected at random from the numbers $\{1,2,3,\ldots,49\}$. If you purchase a ticket by marking 6 numbers of your choice from the 49 numbers, then what is the probability that your 6 numbers will include:

(a) all 6 of the winning numbers? (b) only 5 of the winning numbers? (c) only 4 of the winning numbers?

Referring to example 2.5, the required probabilities are:

(a) $$\frac{C(6,6)}{C(49,6)} = \frac{1}{13983816} = 0.000000071$$

(b) $$\frac{C(6,5)\cdot C(43,1)}{C(49,6)} = \frac{258}{13983816} = 0.000018449$$

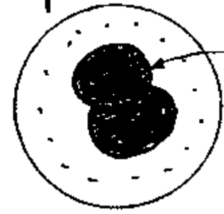
(c) $$\frac{C(6,4)\cdot C(43,2)}{C(49,6)} = \frac{13545}{13983816} = 0.0009686$$

3 PROBABILITIES OF COMPOUND EVENTS

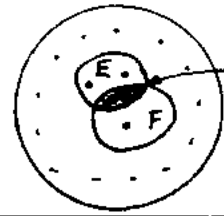
COMPOUND EVENTS (that is, events composed of 2 or more other events) result from the following definition of the two basic event compositions.

DEFINITION 3.1 Consider two events **E** and **F** in a sample space S. Then,

E UNION F, $E \cup F$, is the compound event composed of all the possible outcomes in S that are either in **E OR F**. i.e.



E INTERSECTION F, $E \cap F$, is the compound event composed of all the possible outcomes in S that are both in **E AND F**. i.e.



Definition 3.1 suggests the first of the two laws of Probability.

THE ADDITIVE LAW OF PROBABILITY
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

The following examples illustrate the Additive Law of Probability.

EXAMPLE 3.1 Two fair dice are rolled once. Find $P(E \cup F)$ if:

E = the event that "doubles" occurs, and
F = the event that "a product of 4" occurs.

By example 2.4 we have: $P(E) = \frac{6}{36}$, $P(F) = \frac{3}{36}$, $P(E \cap F) = \frac{1}{36}$, and

By the Additive Law we have: $P(E \cup F) = \frac{6}{36} + \frac{3}{36} - \frac{1}{36} = \frac{8}{36}$

The following definition results in a simplified Additive Law.

DEFINITION 3.2 Two events E and F are called MUTUALLY EXCLUSIVE if $E \cap F$ is empty. In this case, $P(E \cap F) = 0$, and the Additive Law becomes simply:

$$P(E \cup F) = P(E) + P(F)$$

<u>EXAMPLE 3.2</u> A student decides to select a course at random from 9 available courses. If the courses include 4 english, 3 math., and 2 biology courses, then what is the probability that the student will select a math. or biology course?

> Letting E = the event that a math. course is selected, and
> F = the event that a biology course is selected, then
> the required probability is given by $P(E \cup F)$. Since E and F are mutually exclusive, we use the Additive Law as follows:
> $$P(E \cup F) = P(E) + P(F) = \frac{3}{9} + \frac{2}{9} = \frac{5}{9} = 0.56$$

The simplified <u>Additive Law</u> may be extended to 3 or more events as follows:

If the events E_1, E_2, and E_3 are <u>mutually exclusive</u>,
then $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$.

<u>EXAMPLE 3.3</u> A fair coin is tossed 5 times. What is the probability of observing at least 3 heads?

> Letting E_1 = the event of observing 3 heads and 2 tails, and
> E_2 = the event of observing 4 heads and 1 tail, and
> E_3 = the event of observing 5 heads and 0 tails, then
> the required probability is given by $P(E_1 \cup E_2 \cup E_3)$. Since E_1, E_2, and E_3 are <u>mutually exclusive</u>, we use the Additive Law as follows:
> Referring to example 1.12,
> $$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) = \frac{10}{32} + \frac{5}{32} + \frac{1}{32} = \frac{16}{32} = 0.50$$

> <u>DEFINITION 3.3</u> The <u>COMPLEMENT OF EVENT E</u>, $\overline{E}$, consists of all the possible outcomes in S that are not in E.
>
> NOTE: By definition 3.1, $P(E) + P(\overline{E}) = 1$, hence $P(E) = 1 - P(\overline{E})$

<u>EXAMPLE 3.4</u> Two fair dice are rolled once. Find $P(E)$ where:
E = the event that a sum of at least 3 occurs.

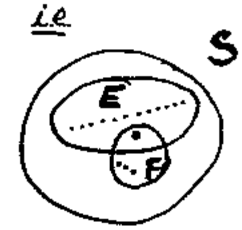
> Consider that $\overline{E}$ consists of the outcome (1,1) only, hence
> $$P(E) = 1 - P(\overline{E}) = 1 - \frac{1}{36} = \frac{35}{36} = 0.97$$

We now consider the following illustration of <u>Conditional Probability</u>.

ILLUSTRATION: A single card is selected at random from a deck of 52 playing cards. <u>GIVEN that</u> the selected card is a diamond, what, then, is the probability that the selected card is an ace?

Intuitively, the answer is $\frac{1}{13}$.

i.e.



Also, if E = the event that the selected card is a diamond, and if F = the event that the selected card is an ace, then we note that the answer equals $\frac{P(E\cap F)}{P(E)} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13}$

This illustration suggests the following definition of <u>Conditional Probability</u>.

DEFINITION 3.4 The **CONDITIONAL PROBABILITY** that an event F will occur, <u>given that</u> an event E has occurred is given by:

$$P(F|E) = \frac{P(E\cap F)}{P(E)}$$

EXAMPLE 3.5 Air Canada finds that 90% of its flights leave on time. However, only 75% of their flights leave AND arrive on time. Hence, what is the probability that an Air Canada flight that leaves on time, will arrive on time?

Letting E = the event that a flight leaves on time, and
F = the event that a flight arrives on time, then
the required probability is given by $P(F|E)$. Using definition 3.3, we have:

$$P(F|E) = \frac{P(E\cap F)}{P(E)} = \frac{0.75}{0.90} = 0.83$$

Definition 3.4 suggests the second of the two Laws of Probability.

THE MULTIPLICATIVE LAW OF PROBABILITY

$$P(E\cap F) = P(E)\cdot P(F|E)$$

The following examples illustrate the Multiplicative Law of Probability.

EXAMPLE 3.6 A box contains 4 good and 3 bad tubes. A tube is selected at random from the box, tested, and set aside. Then another tube is selected and tested. What is the probability that both tubes tested are bad?

> Letting E = the event that the 1st tube selected is bad, and
> F = the event that the 2nd tube selected is bad, then
> the required probability is given by $P(E \cap F)$. By the Multiplicative Law, we have: $P(E \cap F) = P(E) \cdot P(F/E) = \frac{3}{7} \cdot \frac{2}{6} = \frac{2}{14} = 0.14$

EXAMPLE 3.7 In a certain course, 60% of the students do their own assignments. Of these students, 80% will pass the course. However, those students who do not do their own assignments have only a 50% chance of passing the course. What proportion of all the students pass this course?

> Letting E_1 = the event that a student does his own assignment, and
> E_2 = the event that a student does not do his own assignment, and
> F = the event that a student passes this course, then
> the required probability is given by $P(F)$. By the Multiplicative Law, we have:
>
> $$\begin{aligned} P(F) &= P(E_1 \cap F) + P(E_2 \cap F) \\ &= P(E_1) \cdot P(F|E_1) + P(E_2) \cdot P(F|E_2) \\ &= (.60)\,(.80) + (.40)\,(.50) = 0.68 \end{aligned}$$

The following definition results in a simplified Multiplicative Law.

> DEFINITION 3.5 Two events E and F are called **INDEPENDENT** if $P(F|E) = P(F)$.
> In this case the Multiplicative Law becomes simply:
>
> $$P(E \cap F) = P(E) \cdot P(F)$$

EXAMPLE 3.8 A family has 2 children. What is the probability that both are male?

Letting E = the event that the 1st child is a male, and
F = the event that the 2nd child is a male, then
the required probability is given by $P(E \cap F)$. Since $P(F|E) = P(F)$, then E and F are independent, and we use the Multiplicative Law as follows:

$$P(E \cap F) = P(E) \cdot P(F) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

The simplified Multiplicative Law may be extended to 3 or more events as follows:

If the events E_1, E_2, and E_3 are independent,
then $P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2) \cdot P(E_3)$.

EXAMPLE 3.9 A family has 7 children. What is the probability that
(a) 5 are girls and 2 are boys? (b) at least 1 is a girl?
(c) the last child born is the 5th girl?

(a) By independence, the required probability is:

$$\frac{7!}{5!\,2!}\left(\tfrac{1}{2}\right)^5\left(\tfrac{1}{2}\right)^2 = 0.164$$

(b) Referring to definition 3.3 the required probability is:

$$1 - P(0 \text{ girls}) = 1 - \left(\tfrac{1}{2}\right)^7 = 0.992$$

(c) By independence, the required probability is $\left[\frac{6!}{4!\,2!}\left(\tfrac{1}{2}\right)^4\left(\tfrac{1}{2}\right)^2\right]\cdot\left(\tfrac{1}{2}\right) = 0.117$

EXAMPLE 3.10 For a certain operation, a surgeon claims that he is successful 80% of the time. If this is true, what is the probability that, in 10 such operations, he will
(a) be successful exactly 8 times? (b) be successful at least 8 times?

(a) By independence, the required probability is:

$$\frac{10!}{8!\,2!}(.80)^8(.20)^2 = 0.302$$

(b) Referring to example 3.3, the required probability is:

$$\sum_{X=8}^{10} \frac{10!}{X!\,(10-X)!}(.80)^X(.20)^{10-X} = 0.678$$

EXAMPLE 3.11 It is reported that $\frac{2}{3}$rds of all college students will graduate. Hence, consider the probability that more than 100 in a class of 162 of these students will graduate. Give only the expression for this probability.

Referring to example **3.9**, the expression for the required probability is:

$$\sum_{X=101}^{162} \frac{162!}{X!(162-X)!} \cdot \left(\tfrac{2}{3}\right)^X \left(\tfrac{1}{3}\right)^{162-X} \ldots$$

<u>NOTE</u>: Such results are approximated using the <u>Normal approximation to the Binomial</u>, which we consider later.

(In this case the answer is 0.894)

EXAMPLE 3.12 A coach claims that for each game that his team plays, they have a 70% chance of winning, a 20% chance of losing, and a 10% chance of tying. If his claim is correct, what is the probability that the team will complete a 16-game schedule with 11 wins, 3 losses, and 2 ties?

By independence, and example **1.11**, the required probability is:

$$\frac{16!}{11!\;3!\;2!}(.70)^{11}(.20)^3(.10)^2 = 0.069$$

EXAMPLE 3.13 A sample of **500** university students were classified as follows:

	ARTS	SCIENCE	COMMERCE
DEGREE	110	100	90
CERTIFICATE	30	40	30
GRADUATE	10	60	30

(i) Find the probability that a student chosen at random from this sample is a:
(a) Science student (b) Degree and Arts student (c) Graduate or a Science student
(d) Degree student, given he is a commerce student (e) Graduate or Commerce, but not both

(ii) Are the events graduate and commerce independent?

(i) (a) $\frac{100+40+60}{500} = \frac{200}{500}$ (b) $\frac{110}{500}$ (c) $\frac{100}{500} + \frac{200}{500} - \frac{60}{500} = \frac{240}{500}$ (d) $\frac{90/500}{150/500} = \frac{90}{150}$

(e) $\frac{100}{500} + \frac{150}{500} - 2\cdot\frac{30}{500} = \frac{190}{500}$

(ii) check: $P(\text{Grad.} \cap \text{Com.}) \stackrel{?}{=} P(\text{Grad}) \times P(\text{Com.})$ ie $\frac{30}{500} \stackrel{?}{=} \frac{100}{500} \times \frac{150}{500}$ or $0.06 = 0.06$, $\therefore$ YES

PROBABILITY EXERCISES

① A graduating Dawson student has been accepted at 5 universities, each offering him a choice of 3 degree programmes. If, also, he can take each programme with or without a computer science option, then how many different higher education directions can he take?

② Consider a 15 question multiple-choice test. If each question has 4 choices as answer (1 correct, 3 incorrect), then how many different ways can the test be
ⓐ completed? ⓑ completed to include no correct answers?

③ How many lunches consisting of a soup, sandwich, dessert, and a drink are possible if one can select from 4 soups, 3 kinds of sandwiches, 5 desserts, and 4 drin

④ Five people are due to speak at a conference. In how many different orders can they
ⓐ speak? ⓑ speak, if the keynote speaker must speak last.

⑤ Consider the digits $\{1, 2, 3, \ldots, 9\}$. How many different 4-digit numbers can be formed from these digits, if
ⓐ digits may repeat? ⓑ digits may not repeat?
ⓒ digits may not repeat and the numbers must be even?

⑥ Consider a pool of 10 Calculus teachers. How many different ways can teachers be assigned (one to a section) in a Calculus course with
ⓐ 10 sections? ⓑ 5 sections?

⑦ A scrabble player with 7 different letters decides to test all possible 5-letter orderings before playing. If he tests 1 ordering each second, how long will it be before he plays?

(8) How many different ways can 7 men and 2 women sit in a row of 20 seats,
(a) if there are no restrictions?
(b) if the men must sit in the even seats, and the women in the odd seats?

(9) How many different orderings of the letters of the word STATISTICS
(a) are there? (b) begin and end with an S?

(10) How many different orderings of the letters of the word SUCCESSFULLY
(a) are there? (b) are there if the 3 S's must be together in each ordering?
(c) are there, if the word SUCCESS must appear in each ordering?

(11) How many different 7-digit numbers can be formed using the digits of the number 4,221,132?

(12) How many different ways can a 10-question true-false test be
(a) completed? (b) completed to include at least 6 correct answers?

(13) There are 12 points on the circumference of a circle. By joining the points, how many
(a) distinct lines can be formed? (b) distinct triangles can be formed?

(14) In how many ways can 2 different Math. books, 3 different Biology books, and 4 different Psychology books be arranged on a shelf
(a) in any order? (b) if books of the same subject must be together?
(c) if there must be a Math. book at each end?

(15) In how many different teams, consisting of 5 players each, can be chosen from a group of 10 players!

(16) From a group of 5 teachers and 3 students, how many different 4-member committees
(a) can be formed? (b) will include 2 of each?
(c) will include at least 1 of each?

(17) From a deck of 52 playing cards, how many different 5-card hands
(a) are possible? (b) include only cards of the same suit?
(c) include 3 black and 2 red cards? (d) include at least 1 heart?

(18) A student is selected at random from a class containing 7 Arts students, 12 Science students, and 10 Career students. What is the probability that the selected student is
(a) a Career student? (b) a non-Arts student?

(19) An ordering is selected at random from all the possible orderings of the letters PROBABILITY. What is the probability that the selected ordering
(a) is the one above? (b) begins with P and ends with Y?
(c) has the T in the middle?

(20) A fair coin is tossed 7 times. Find the prob. of observing
(a) only heads. (b) 4 heads and 3 tails. (c) more heads than tails.

(21) A balanced die is rolled 5 times. Find the prob. of observing
(a) only faces less than 3. (b) 2 even and 3 odd faces (c) a total sum of 29.

(22) A student guesses at each of the 10 questions on a multiple choice test. If each question has 3 choices (1 correct, 2 wrong), find the prob. that the student
(a) gets all 10 questions wrong. (b) gets at least 1 question correct.
(c) gets 6 correct and 4 wrong. (d) passes the test ($\geq 60\%$).

(23) A 4-member negotiation committee is to be selected at random from a group of 6 union and 6 management people. What is the prob. that the committee will include (a) equal representation from both sides?
(b) more management than union members?

(24) From a shipment of 20 new TV's, 3 are selected at random for testing. The shipment is only accepted if all 3 test out OK. What is the prob. that the shipment will be accepted, if in fact 2 of the TV's in the shipment are defective

(25) A 5-member committee is selected at random from a group of 6 Conservatives, 7 Liberals and 2 NDP'ers. What is the prob that the committee will include
(a) only Liberals? (b) 2 Conservatives and 3 Liberals? (c) both NDP'ers?
(d) at least 1 NDP'er? (e) 2 Conservatives, 2 Liberals and 1 NDP'er?

(26) A pair of fair dice is rolled once.
let E = the event of a sum of 6.
let F = the event of a product of 8.
let G = the event of doubles.

(a) Find $P(E)$, $P(F)$, $P(G)$
(b) Find $P(E \cup F)$, $P(E \cup G)$, $P(F \cup G)$
(c) Find $P(E|F)$, $P(F|E)$, $P(E|G)$, $P(G|E)$, $P(F|G)$, $P(G|F)$.

(27) In a certain course, 80% of the students study. Of these students, 75% will pass the course. However, those students who do not study have only a 50% chance of passing the course. What proportion of all the students pass this course?

(28) The chance of a star hockey player playing in his team's next game is only 30%, due to an injury. If he plays, his team is 90% certain to win, otherwise they are equally-likely to win or not. What is the prob. that the team wins its next game?

(29) A certain problem is assigned to each of 3 students. If the students can independently solve this type of problem with probabilities of $\frac{1}{2}$, $\frac{3}{4}$, and $\frac{2}{5}$, respectively, what is the prob. that:
(a) all 3 students will solve the problem?
(b) the problem will be solved?

(30) A letter is independently chosen at random from each of the words: CHOICE and CHANCE
What is the prob. that the 2 letters chosen are: (a) both C'S? (b) the same?

(31) Consider a family with 6 children. What is the prob. that
(a) at least 1 is a girl? (b) 4 are boys and 2 are girls?
(c) the last child born is the 4th boy?

(32) A trainer claims that his horse has a 70% chance of winning each of the 9 races he is entered in this year. Based on this, what is the prob. that the horse:
(a) will win all 9 races? (b) will lose the first 3 races, and win the last 6 races?
(c) will win 6 races and lose 3 races? (d) will get his 7th win in the last race?

(33) A company reports that there is only a 10% chance that any of its new oil well drillings will actually yield oil. If this is true, what is the prob. that in its next 15 new drillings
(a) none will yield oil? (b) exactly 5 will yield oil? (c) at least 1 will yield oil?

NOTE: Give only the expressions for the probabilities in the following exercises.

(34) If it is true that $\frac{2}{3}$rds of all Montréalers are bilingual, then what is the prob. that a random sample of 50 Montréalers will include: (a) exactly 35 who are bilingual?
(b) more than 40 who are bilingual?

(35) Air Canada claims that 5% of all seats sold are "no-shows". Hence, if they sell 104 seats for a 100 seat flight, what is the prob. that they will have "overbooked"?

(36) A meteorologist claims that for any summer day in Montréal, the chance of sunshine is 60%, of clouds, 30%, and of rain, 10%. Hence, if one spends a 2-week summer vacation in Montréal, what is the prob. that one will experience
(a) 14 sunny days? (b) at least 10 sunny days?
(c) 9 sunny days, 4 cloudy days, and 1 rainy day?

(37) At 10 PM on Saturday nights, 30% of the TV viewers in Canada are watching a CBC station, 25% are watching a CTV station, and 45% are watching a USA station. What is the prob. that a poll of 1000 TV viewers taken at that time will include 300 CBC watchers, 250 CTV watchers, and 450 USA watchers?

PROBABILITY EXERCISES — SOLUTIONS

(1) $5 \times 3 \times 2 = 30$

(2) (a) 4^{15} (b) 3^{15}

(3) $4 \times 3 \times 5 \times 4 = 240$

(4) (a) $5! = 120$ (b) $4! = 24$

(5) (a) $9^4 = 6561$ (b) $9 \cdot 8 \cdot 7 \cdot 6 = 3024$ (c) $8 \cdot 7 \cdot 6 \cdot 4 = 1344$

(6) (a) $10!$ (b) $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30{,}240$

(7) $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2520$ sec. or 42 min.

(8) (a) $20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12$ (b) $(10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4) \times (10 \cdot 9)$

(9) (a) $\frac{10!}{3!3!2!} = 50{,}400$ (b) $\frac{8!}{3!2!} = 3{,}360$

(10) (a) $\frac{12!}{3!(2!)^3} = 9{,}979{,}200$ (b) $\frac{10!}{(2!)^3} = 453{,}600$ (c) $\frac{6!}{2!} = 360$

(11) $\frac{7!}{3!2!} = 420$

(12) (a) $2^{10} = 1024$ (b) $\sum_{r=6}^{10} \frac{10!}{r!(10-r)!} = 386$

(13) (a) $\frac{12!}{2!10!} = 66$ (b) $\frac{12!}{3!9!} = 220$

(14) (a) $9! = 362{,}880$ (b) $3![2! \times 3! \times 4!] = 1728$ (c) $2![7!] = 10{,}080$

(15) $\frac{10!}{5!5!} = 252$

(16) (a) $\frac{8!}{4!4!} = 70$ (b) $\left(\frac{5!}{2!3!}\right) \times \left(\frac{3!}{2!1!}\right) = 30$ (c) $70 - \frac{5!}{4!1!} = 65$

(17) (a) $\frac{52!}{5!47!} = 2{,}598{,}960$ (b) $4 \times \frac{13!}{5!8!} = 5148$ (c) $\left(\frac{26!}{3!23!}\right) \times \left(\frac{26!}{2!24!}\right) = 845{,}000$ (d) $\frac{52!}{5!47!} - \frac{39!}{5!34!} = 2{,}023{,}203$

(18) (a) $\frac{10}{29} = .34$ (b) $\frac{22}{29} = .76$

(19) (a) $\frac{1}{\frac{11!}{2!2!}} = .0000001$ (b) $\frac{\frac{9!}{2!2!}}{\frac{11!}{2!2!}} = .009$ (c) $\frac{\frac{10!}{2!2!}}{\frac{11!}{2!2!}} = .091$

(20) (a) $\frac{1}{2^7} = .0078$ (b) $\frac{\frac{7!}{4!3!}}{2^7} = .2734$ (c) $\frac{\sum_{r=4}^{7} \frac{7!}{r!(7-r)!}}{2^7} = .5000$

(21) (a) $\frac{3^5}{6^5} = .0041$ (b) $\frac{\frac{5!}{2!3!} \times 3^2 \times 3^3}{6^5} = .3125$ (c) $\frac{\frac{5!}{4!1!}}{6^5} = .0006$

(22) (a) $\frac{2^{10}}{3^{10}} = .0173$ (b) $1 - .0173 = .9827$ (c) $\frac{\frac{10!}{6!4!} \times 1^6 \times 2^4}{3^{10}} = .0569$ (d) $\frac{\sum_{r=6}^{10} \frac{10!}{r!(10-r)!} \times 1^r \times 2^{(10-r)}}{3^{10}} = .0750$

(23) (a) $\frac{\left(\frac{6!}{2!4!}\right) \times \left(\frac{6!}{2!4!}\right)}{\frac{12!}{4!8!}} = .456$ (b) $\frac{\frac{6!}{3!3!} \times \frac{6!}{1!5!} + \frac{6!}{4!2!} \times \frac{6!}{0!6!}}{\frac{12!}{4!8!}} = .273$

(24) $= \frac{\frac{18!}{3!15!}}{\frac{20!}{3!17!}} = 0.716$

(25) (a) $\frac{\frac{7!}{2!5!}}{\frac{15!}{5!10!}} = \frac{21}{3003} = .0067$ (b) $\frac{\left(\frac{6!}{2!4!}\right) \times \left(\frac{7!}{3!4!}\right)}{3003} = .1748$ (c) $\frac{\frac{13!}{3!10!}}{3003} = .0956$ (d) $1 - \frac{\frac{12!}{5!8!}}{3003} = .5714$ (e) $\frac{\left(\frac{6!}{2!4!}\right) \times \left(\frac{7!}{2!5!}\right) \times \left(\frac{2!}{1!1!}\right)}{3003} = .2098$

(26) (a) $\frac{5}{36} = .13\dot{8}$, $\frac{2}{36} = .0\dot{5}$, $\frac{6}{36} = .1\dot{6}$ (b) $\frac{5}{36} = .13\dot{8}$, $\frac{10}{36} = .2\dot{7}$, $\frac{8}{36} = .2\dot{2}$ (c) 1, .4, .16, .2, 0, 0

(27) $(.80) \times (.75) + (.20) \times (.50) = .70$

(28) $(.3) \times (.9) + (.7) \times (.5) = .62$

(29) (a) $\left(\frac{1}{2}\right) \times \left(\frac{3}{4}\right) \times \left(\frac{2}{5}\right) = \frac{3}{20}$ (b) $1 - \left[\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{3}{5}\right] = \frac{37}{40}$

(30) (a) $\frac{2}{6} \times \frac{2}{6} = .\dot{1}\dot{1}$ (b) $\frac{2}{6} \cdot \frac{2}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = .1\dot{6}$

(31) (a) $1 - \left(\frac{1}{2}\right)^6 = .984$ (b) $\frac{6!}{4!2!} \times \left(\frac{1}{2}\right)^6 = .234$ (c) $\left[\frac{5!}{3!2!} \times \left(\frac{1}{2}\right)^5\right] \times \left(\frac{1}{2}\right) = .156$

(32) (a) $(.7)^9 = .040$ (b) $(.30)^3 \times (.70)^6 = .003$ (c) $\frac{9!}{6!3!}(.7)^6(.3)^3 = .267$ (d) $\left[\frac{8!}{6!2!}(.7)^6(.3)^2\right] \times (.7) = .208$

(33) (a) $(.90)^{15} = .206$ (b) $\frac{15!}{5!10!}(.10)^5(.90)^{10} = .010$ (c) $1 - .206 = .794$

(34) (a) $\frac{50!}{35!15!} \times \left(\frac{2}{3}\right)^{35} \times \left(\frac{1}{3}\right)^{15}$ (b) $\sum_{r=41}^{50} \frac{50!}{r!(50-r)!} \left(\frac{2}{3}\right)^r \left(\frac{1}{3}\right)^{(50-r)}$

(35) $\sum_{r=0}^{3} \frac{104!}{r!(104-r)!} (.05)^r \cdot (.95)^{(104-r)}$

(36) (a) $(.60)^{14} = .0008$ (b) $\sum_{r=10}^{14} \frac{14!}{r!(14-r)!} \times (.60)^r (.40)^{(14-r)}$ (c) $\frac{14!}{9!4!1!} \times (.60)^9 \times (.30)^4 \times (.10)^1$

(37) $\frac{1000!}{300!\,250!\,450!} \times (.30)^{300} \times (.25)^{250} \times (.45)^{450}$

DISCRETE PROBABILITY DISTRIBUTIONS

The probability function or table that assigns a probability, $P(x)$, to each possible value of a discrete random variable, X, is called

THE DISCRETE PROBABILITY DISTRIBUTION OF X IF:

① $P(x) \geq 0$ for all x, and

② $\sum_{\text{all } x} P(x) = 1$

TWO EXPECTED VALUES

$$\mu = \text{THE MEAN OF } X = E(x) = \sum_{\text{all } x} x \cdot P(x)$$

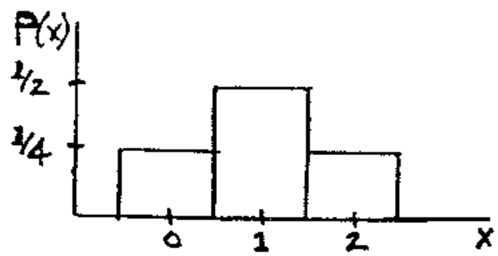
$$\sigma^2 = \text{THE VARIANCE OF } X = E\left[(x-\mu)^2\right] = \sum_{\text{all } x} (x-\mu)^2 \cdot P(x)$$

NOTE: $E\left[(x-\mu)^2\right] = E(x^2) - \mu^2 = \sum x^2 P(x) - \mu^2$

ILLUSTRATIVE EXAMPLE

Consider the random variable, X, described by the number of heads observed when 2 fair coins are tossed:

X	OUTCOMES IN X	P(X)
0	TT	1/4
1	HT and TH	2/4
2	HH	1/4



and $P(x) = \dfrac{C(2,x)}{4}$, for $x = 0, 1, 2$

then $\mu = \sum x \cdot P(x) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1$, and

$$\sigma^2 = \sum (x-\mu)^2 \cdot P(x) = (0-1)^2 \cdot \frac{1}{4} + (1-1)^2 \cdot \frac{2}{4} + (2-1)^2 \cdot \frac{1}{4} = \frac{1}{2}$$

NOTE: also, $\sigma^2 = \sum x^2 P(x) - \mu^2 = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{2}{4} + 2^2 \cdot \frac{1}{4} - 1 = 0 + \frac{1}{2} + 1 - 1 = \frac{1}{2}$

DISCRETE PROBABILITY DISTRIBUTIONS EXERCISES

In exercises ① to ⑤, sketch a histogram of $P(x)$ and find μ and σ^2.

①

X	0	1	2	3	4
$P(x)$	0.1	0.2	0.4	0.2	0.1

②

X	0	1	2	3
$P(x)$	1/8	1/4	1/2	1/8

③

x	10	11	12	13	14
$P(x)$	0.2	0.4	0.2	0.1	0.1

④

X	0	1	2	3	4	5
$P(x)$	0.05	0.3	0.3	0.2	0.1	0.05

⑤

X	0	1	2	3	4
$P(x)$	0.12	0.38	0.30	0.12	0.08

⑥ The number of large ships to arrive at the Port of Montréal on any summer day is a random variable represented by X. Given the probability distribution of X below, find:

ⓐ $P(x \geq 10)$ ⓑ the mean number of ships to arrive.

X	8	9	10	11	12
$P(x)$	0.2	0.3	0.3	0.1	0.1

⑦ The probability distribution for X, the number of house sales per day for a real estate office is:

X	0	1	2	3	4	5	6	7
$P(x)$	0.10	0.20	0.40	0.15	0.10	0.03	0.01	0.01

ⓐ Find $P(x < 3)$ ⓑ μ, and ⓒ σ^2

(8) In the probability distribution below, find $P(0)$, μ, and σ^2.

X	0	1	3	4
$P(X)$		$2/10$	$4/10$	$3/10$

(9) In the probability distribution below, find $P(2)$ and $P(3)$ if $\mu = 1.7$

X	0	1	2	3
$P(X)$	$1/10$	$3/10$		

In exercises (10) to (16) tabulate the probability distribution of X, sketch its histogram, and find μ and σ^2.

(10) $P(X) = \frac{X}{10}$, for $X = 1, 2, 3, 4$

(11) $P(X) = \frac{(5-X)}{10}$, for $X = 1, 2, 3, 4$

(12) $P(X) = \frac{1}{10}$, for $X = 0, 1, 2, ---, 9$

(13) $P(X) = \frac{C(3,X)}{8}$, for $X = 0, 1, 2, 3$

(14) $P(X) = \frac{(X^2-1)}{50}$, for $X = 2, 3, 4, 5$

(15) $P(X) = \frac{3!}{X!(3-X)!}\left(\frac{2}{3}\right)^X\left(\frac{1}{3}\right)^{3-X}$, for $X = 0, 1, 2, 3$

(16) $P(X) = \frac{C(3,X)\,C(2,2-X)}{C(5,2)}$, for $X = 0, 1, 2$

(17) Find $E(X^2)$ for a discrete random variable X if $\mu = 50$ and $\sigma^2 = 100$.

(18) Given the probability distribution $P(x) = \frac{1}{k}$, for $x = 0,1,2,3,4$ find μ and σ^2

(19) Consider the random variable, x, described by the number observed on a single roll of a fair die.
ⓐ Give the probability distribution of X, and ⓑ Find μ and σ^2.

(20) Consider the random variable, x, described by the number of girls in a 3-child family.
ⓐ Give a function for the probability distribution of X.
ⓑ Find μ and σ^2.

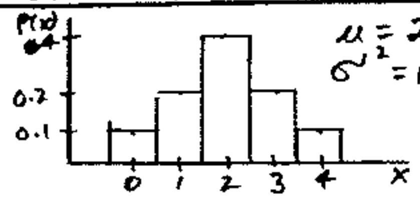
(21) Consider the random variable, x, described by the number of students included when a 3-person committee is chosen at random from a group of 4 students and 5 teachers.
ⓐ Give a function for the probability distribution of X.
ⓑ Find μ and σ^2.

(22) A baseball team is 75% certain to win each game it plays. Consider the random variable, X, described by the number of wins in the team's next 5 games.
ⓐ Give a function for the probability distribution of X.
ⓑ Find μ and σ^2.

(23) Consider the random variable, X, described by the sum of the faces observed when 2 fair dice are rolled once.
ⓐ Tabulate the probability distribution of X.
ⓑ Sketch the histogram for the distribution.
ⓒ Find μ and σ^2.
ⓓ Determine a function for the probability distribution of X.

DISCRETE PROBABILITY DISTRIBUTIONS EXERCISES – SOLUTIONS

(1)



$\mu = 2$

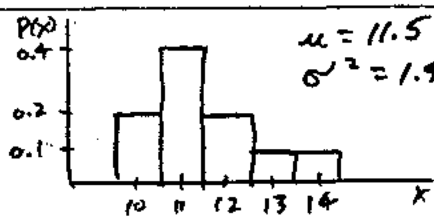
$\sigma^2 = 1.2$

(2)

$\mu = 13/8$

$\sigma^2 = \frac{47}{64}$

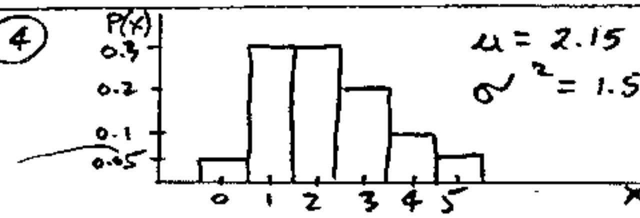
(3)



$\mu = 11.5$

$\sigma^2 = 1.45$

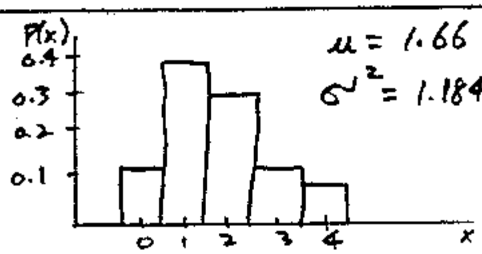
(4)



$\mu = 2.15$

$\sigma^2 = 1.5275$

(5)



$\mu = 1.66$

$\sigma^2 = 1.1844$

(6) (a) $P(X \geq 10) = P(10) + P(11) + P(12)$
$= 0.3 + 0.1 + 0.1$
$= 0.5$

(b) $\mu = 9.6$

(7) (a) $P(X < 3) = P(0) + P(1) + P(2)$
$= 0.10 + 0.20 + 0.40$
$= 0.70$

(b) $\mu = 2.13$ and $\sigma^2 = 1.8131$

(8) $P(0) = 1/10$, $\mu = 2.6$, $\sigma^2 = 1.84$

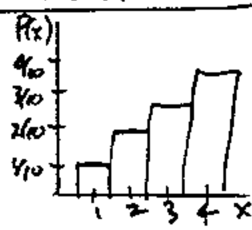
(9) $P(2) = 4/10$ and $P(3) = 2/10$

(10)

X	1	2	3	4
P(X)	1/10	2/10	3/10	4/10

$\mu = 3$

$\sigma^2 = 1$

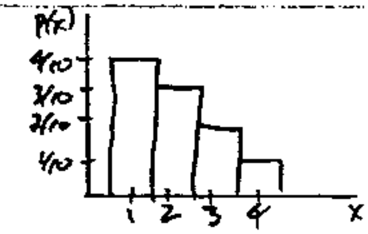


(11)

X	1	2	3	4
P(X)	4/10	3/10	2/10	1/10

$\mu = 2$

$\sigma^2 = 1$

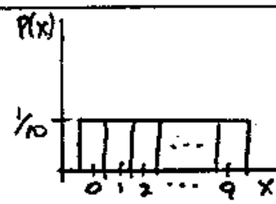


(12)

X	0	1	2	⋯	9
P(X)	1/10	1/10	1/10	⋯	1/10

$\mu = 4.5$

$\sigma^2 = 8.25$

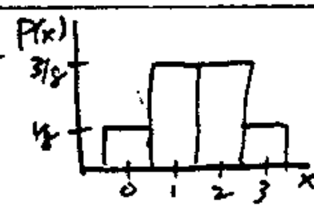


(13)

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

$\mu = 1.5$

$\sigma^2 = 0.75$

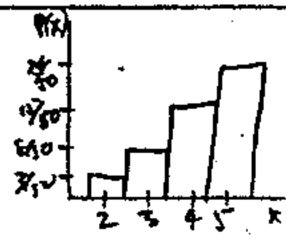


(14)

X	2	3	4	5
P(X)	3/50	8/50	15/50	24/50

$\mu = 4.2$

$\sigma^2 = 0.84$

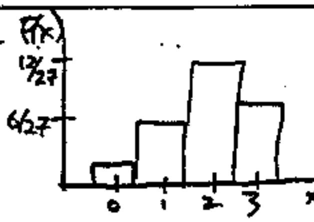


(15)

X	0	1	2	3
P(X)	1/27	6/27	12/27	8/27

$\mu = 2$

$\sigma^2 = 2/3$

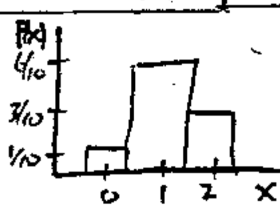


(16)

X	0	1	2
P(X)	1/10	6/10	3/10

$\mu = 6/5 = 1.2$

$\sigma^2 = \frac{9}{25} = 0.36$



(17) $E(X^2) = \sum x^2 P(x) = 2600$

$\because \sigma^2 = 100 = E(X^2) - \mu^2$

$\therefore 100 = E(X^2) - 50^2$

$\therefore 2600 = E(X^2)$

DISCRETE PROB. DISTR. EXERCISES – SOLUTIONS

(18) $\Sigma P(x) = 1$ $\therefore \Sigma \frac{1}{k} = 1$ $\therefore \frac{5}{k} = 1$, $\therefore 5 = k$, hence $\mu = 2$ and $\sigma^2 = 2$

(19) (a) $P(x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$ and (b) $\mu = 3.5$, $\sigma^2 = 2.917$

(20) (a) $P(x) = \dfrac{C(3,x)}{8}$ for $x = 0, 1, 2, 3$

(b) $\mu = 3/2$ and $\sigma^2 = 3/4$

(21) (a) $P(x) = \dfrac{C(4,x)\,C(5,3-x)}{C(9,3)}$ for $x = 0, 1, 2, 3$

(b) $\mu = 12/9$ and $\sigma^2 = 5/9$

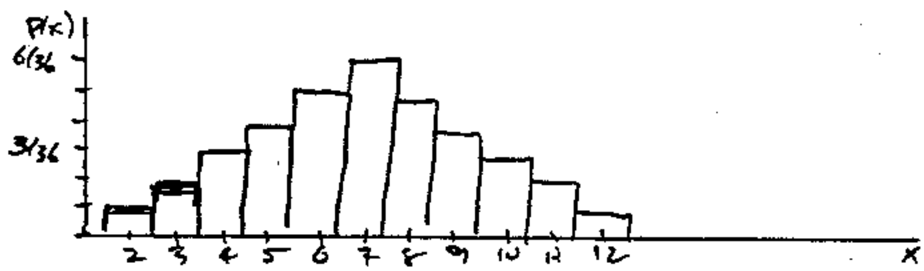
(22) (a) $p(x) = C(5,x)(0.75)^x(0.25)^{5-x}$, for $x = 0, 1, 2, 3, 4, 5$

(b) $\mu = 3.75$ and $\sigma^2 = 0.9375$

(23) (a)

X	2	3	4	5	6	7	8	9	10	11	12
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

(b)



(c) $\mu = 7$ and $\sigma^2 = 5.83$

(d) $P(x) = \dfrac{(6 - |x-7|)}{36}$, for $x = 2, 3, 4, \cdots, 12$

The Binomial Probability Distribution

Consider a **BINOMIAL EXPERIMENT** as follows:

1. It consists of n identical, independent trials.
2. Each trial yields one of 2 possible results, a "success" or a "non-success".
3. The probability of a success, p, and the probability of a non-success, $1-p=q$, are constant from trial to trial and $p+q=1$.

Then the random variable, X, defined by the number of successes in the n trials of the BINOMIAL EXPERIMENT has the:

BINOMIAL PROBABILITY DISTRIBUTION

$$P(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x} \quad \text{for } x = 0,1,2,\cdots,n$$

$$\boxed{B(x;n,p)}$$

AND

$$\textbf{mean} = \mu = E(x) = \sum_{x=0}^{n} x \cdot P(x) = np$$

$$\textbf{variance} = \sigma^2 = E(x-\mu)^2 = \sum_{x=0}^{n} (x-\mu)^2 \cdot P(x) = npq$$

NOTE: $C(n,x) = \frac{n!}{x!\,(n-x)!}$

THE BINOMIAL PROBABILITY DISTRIBUTION - AN ILLUSTRATIVE EXAMPLE

If it is true that $\frac{3}{4}$ ths of all Dawson students go on to university, then consider the random-variable, X, defined by the number of Dawson students who go on to university in a random sample of 10 Dawson students.

THE PROBABILITY DISTRIBUTION OF X: Consider $B(x; 10, \frac{3}{4})$

ie $$P(x) = \frac{10!}{x!(10-x)!}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{10-x} \text{ for } x = 0, 1, 2, \cdots, 10.$$

TO FIND $P(X \le 5)$: Consider that $P(x \le 5) = \sum_{x=0}^{5} \frac{10!}{x!(10-x)!}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{10-x}$

Then from the $B(x; 10, p)$ table in the Appendix with $p = \frac{3}{4} = 0.75$ we have that $P(x \le 5) = 0.07813$

TO FIND $P(X = 5)$: Consider that $P(x=5) = \frac{10!}{5!5!}\left(\frac{3}{4}\right)^{5}\left(\frac{1}{4}\right)^{5}$

and from the table, $$\begin{aligned} P(x=5) &= P(x \le 5) - P(x \le 4) \\ &= 0.07813 - 0.01973 \\ &= 0.05840 \end{aligned}$$

TO FIND $P(x \ge 5)$: Consider that $P(x \ge 5) = \sum_{x=5}^{10} \frac{10!}{x!(10-x)!}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{10-x}$

and from the table, $$\begin{aligned} P(x \ge 5) &= 1 - P(x \le 4) \\ &= 1 - 0.01973 \\ &= 0.98027 \end{aligned}$$

THE MEAN AND VARIANCE OF X:

$$\text{MEAN} = \mu = np = 10\left(\frac{3}{4}\right) = 7.5$$

$$\text{VARIANCE} = \sigma^2 = npq = 10\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = 1.875$$

NOTE: The STANDARD DEVIATION of $X = \sigma = \sqrt{npq} = \sqrt{1.875} = 1.369$

BINOMIAL EXERCISES

① Consider a family of 5 children. Let the random variable, X, be defined as the number of boys in the family. Then: ⓐ State $P(x)$ ⓑ Find $P(x=2)$ ⓒ Verify that $\sum_{all\,x} P(x) = 1$ ⓓ Find μ and σ.

② A resort claims that for each day in July there is a 90% chance of sunshine there. Consider the r.v., X, defined by the number of days of sunshine in a one week vacation at this resort in July. Then find:

ⓐ $P(X=5)$ ⓑ The probability distribution, $P(X)$ ⓒ $E(X)$

③ Consider $B(X; 15, .40)$. Then use the Binomial Tables to find:

ⓐ $P(X \leq 7)$ ⓑ $P(X > 10)$ ⓒ $P(X=6)$ ⓓ $P(4 \leq X \leq 8)$

④ What is the prob. that a family of 10 children will include:
ⓐ less than 4 boys? ⓑ exactly 4 boys? ⓒ at least 6 girls?

⑤ A student guesses at each of the 10 questions on a multiple-choice test. If each question has 5 choices (1 correct, 4 incorrect), then find the prob. that the student:
ⓐ gets none correct. ⓑ gets 6 correct. ⓒ gets less than the expected number correct.

⑥ A coach claims that his team has a 70% chance of winning each of its upcoming games. If the coach is correct, then:

ⓐ What is the prob. that the team will win 5 of their next 10 games?

ⓑ What is the prob. that the team will win more than 10 of their next 15 games?

ⓒ How many games can they expect to win in their next 50 games?

ⓓ How many games must his team play for him to expect 84 wins?

ⓔ Is there a minimum number of games that his team must play to ensure winning at least 1 game?

(7) Out of 2000 families, with 4 children each, how many would you expect to have:
(a) exactly 2 boys? (b) at least 1 boy?

(8) A random sample of 10 batteries is routinely taken from each large shipment of new batteries. The shipment is only accepted if all 10 batteries are good. What is the prob that a shipment will not be accepted if:
(a) 95% of them are good? (b) 80% of them are good?

(9) What is the prob. of observing the 6^{th} even number on the 16^{th} roll of a fair die?

(10) A skier plans to compete in 11 races this winter. If she is 90% certain to win each race, find the prob. that she will:
(a) win all the races. (b) get her 7^{th} win in the last race.

(11) If a World Series baseball team is 75% certain to win each game it plays, what is the prob. that it will win the World Series by being the 1^{st} team to win 4 games in a possible 7 game series?

(12) Two fair dice are tossed repeatedly. Find the prob. that the n^{th} toss will be 1^{st} time that the sum of the dice is a "7 or 11".

(13) (The game of "ODD MAN WINS")
Let n fair coins be tossed ($n \geq 3$). Show that the prob. of observing exactly 1 tail, or exactly 1 head is given by: $\frac{n}{2^{n-1}}$

What is the prob. that the 1^{st} "odd man win" will occur on the k^{th} play, in a game with n players?

(14) A 12-room motel takes 15 reservations for rooms. If experience shows that 30% of all reservations are no-shows, what is the probability that there will not be a room for everyone who shows up (ie that they have "overbooked")?

BINOMIAL EXERCISES – SOLUTIONS

1) (a) $P(x) = \frac{5!}{x!(5-x)!}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x}$ for $x = 0,1,2,3,4,5$ (b) $P(x=2) \Rightarrow \frac{5!}{2!3!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32} = 0.3125$

(c) $\sum_{x=0}^{5} P(x) = \frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32} = \frac{32}{32} = 1$

(d) $\mu = np = 5\left(\frac{1}{2}\right) = 2.5$ and $\sigma = \sqrt{npq} = \sqrt{5\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{1.25} = 1.118$

---

(2) (a) $P(X=5)$ in $B(x; 7, .90)$ (b) $P(x) = \frac{7!}{x!(7-x)!}(.90)^x(.10)^{7-x}$ for $x = 0,1,2,\ldots,7$

$= \frac{7!}{5!2!}(.90)^5(.10)^2$

$= 0.124$

(c) $E(x) = 7 \cdot (.90) = 6.3$

---

(3) $B(x; 15, .40)$

(a) $P(X \le 7) = 0.78690$

(b) $P(x > 10) = P(x \ge 11)$
$= 1 - P(x \le 10)$
$= 1 - 0.99065 = .00935$

(c) $P(x=6) = P(x \le 6) - P(x \le 5)$
$= .60981 - .40322$
$= 0.20659$

(d) $P(4 \le X \le 8) = P(x \le 8) - P(x \le 3) = .90495 - .09050 = 0.81445$

---

(4) $B(x; 10, .50)$

(a) $P(X < 4) = P(x \le 3)$
$= 0.17188$

(b) $P(x=4) = 0.37695$
$- 0.17188$
$= 0.20507$

(c) $P(x \ge 6) = 1 - P(x \le 5)$
$= 1 - 0.62305$
$= 0.37695$

---

(5) $B(x; 10, .20)$

(a) $P(x=0) = 0.10737$

(b) $P(x=6) = .99914$
$- .99363$
$= 0.00551$

(c) $E(x) = 10(.20) = 2$
$\therefore P(x<2) = P(x \le 1) = 0.37581$

---

(6) (a) $B(x; 10, .70)$ $\therefore P(x=5) = P(x \le 5) - P(x \le 4) = .15027 - .04735 = 0.10292$

(b) $B(x; 15, .70)$ $\therefore P(x > 10) = P(x \ge 11) = 1 - P(x \le 10) = 1 - .48451 = 0.51549$

(c) $E(x) = 50 \cdot (.70) = 35$ games (d) i.e. $n \times (.70) = 84$
$\therefore n = 120$ (e) No, no guarantee!!

---

(7) (a) $2000 \times \left[\frac{4!}{2!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2\right]$ (b) $2000 \times \left[1 - \frac{4!}{0!4!}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4\right]$

$= 2000 \times \frac{3}{8}$ $= 2000 \times \frac{15}{16}$

$= 750$ families $= 1875$ families

BINOMIAL EXERCISES - SOLUTIONS

(8) (a) $P(X<10)$ in $B(x;10,.95) = P(X\le 9) = 0.40126$

(b) $P(X<10)$ in $B(x;10,.80) = P(X\le 9) = 0.89263$

(9) Ans. $= [P(X=5)$ in $B(x;15,.50)] \times (.50)$

$= [.15088 - .05923] \times (.50) = [.09165] \times (.50) = 0.045825$

(10) (a) $[P(X=10)$ in $B(x;10,.90)] \times (.90)$

$= [1 - .65132] \times (.90)$

$= [.34868] \times (.90)$

$= 0.313812$

(b) $[P(X=6)$ in $B(x;10,.90)] \times (.90)$

$= [.01280 - .00163] \times (.90)$

$= [.01117] \times (.90)$

$= 0.010053$

(11) Ans. $= (.75)^4 + \left[\frac{4!}{3!1!}(.75)^3(.25)^1\right]\times(.75) + \left[\frac{5!}{3!2!}(.75)^3(.25)^2\right]\times(.75) + \left[\frac{6!}{3!3!}(.75)^3(.25)^3\right]\times(.75)$

$= 0.929$

(12) Ans. $= \left(\frac{28}{36}\right)^{n-1} \times \frac{8}{36} = \frac{8 \times 28^{n-1}}{36^n}$

(13) $P(\text{ODD MAN}) = \frac{n!}{1!(n-1)!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{n-1} + \frac{n!}{1!(n-1)!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{n-1}$

$= n\left(\frac{1}{2}\right)^n + n\left(\frac{1}{2}\right)^n = 2n\left(\frac{1}{2}\right)^n = \frac{n}{2^{n-1}}$

hence Ans. $= \left[1 - \frac{n}{2^{n-1}}\right]^{k-1} \times \left[\frac{n}{2^{n-1}}\right] = \frac{n\,(2^{n-1} - n)^{k-1}}{(2^{n-1})^k}$

(14) Ans. $\Rightarrow P(X\le 2)$ in $B(X;15,0.30) \Rightarrow 0.12683$

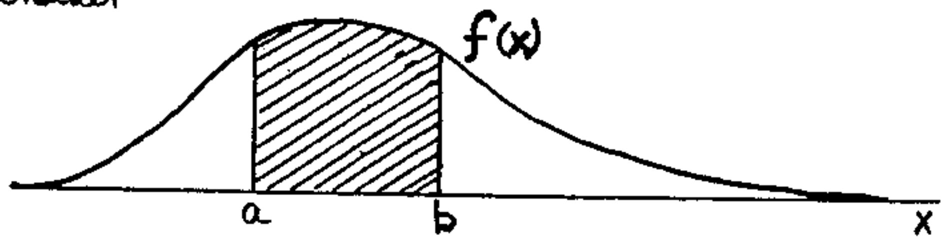
CONTINUOUS PROBABILITY DISTRIBUTIONS

The PROBABILITY (DENSITY) FUNCTION, $f(x)$, giving the probability distribution of a continuous random variable x called THE CONTINUOUS PROBABILITY DISTRIBUTION OF X if:

① $f(x) \geq 0 \quad , \forall x \quad$, and

② $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

Also, consider



then

$$P(a<x<b) = P(a \leq x \leq b) = \int_a^b f(x)\,dx = \text{the shaded area}$$

TWO EXPECTED VALUES

$$\mu = \text{THE MEAN OF } X = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$$

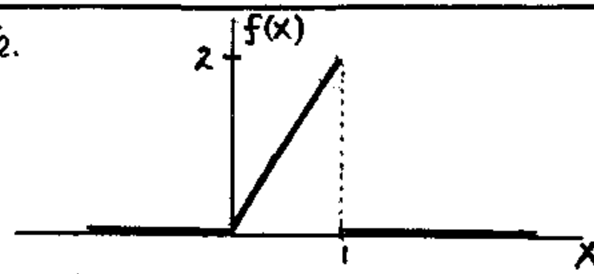
$$\sigma^2 = \text{THE VARIANCE OF } X = E\left[(x-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

NOTE: $E\left[(x-\mu)^2\right] = E(x^2) - \mu^2 = \displaystyle\int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

CONTINUOUS PROBABILITY DISTRIBUTIONS — AN ILLUSTRATIVE EXAMPLE

Consider the function: i.e.

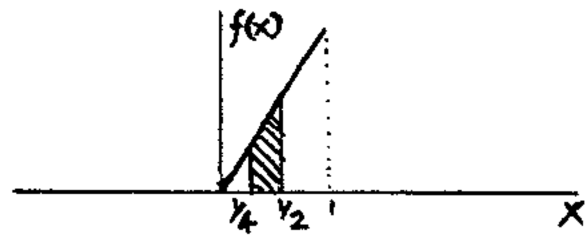
$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$



Then, <u>f(x) is a Continuous Probability Distribution</u>, since:

① $f(x) \geq 0 \quad \forall x$

② $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 2x\,dx = \left[x^2\right]_0^1 = 1^2 - 0^2 = 1$



<u>To find $P(1/4 < x < 1/2)$</u>:

$\Rightarrow P(1/4 \le x \le 1/2)$

$\Rightarrow \displaystyle\int_{1/4}^{1/2} 2x\,dx = \left[x^2\right]_{1/4}^{1/2} = \left(\tfrac{1}{2}\right)^2 - \left(\tfrac{1}{4}\right)^2 = \frac{3}{16} = 0.1875$

<u>The Mean and Variance of X</u>:

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x\,2x\,dx = \int_0^1 2x^2\,dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2(1)^3}{3} - \frac{2(0)^3}{3} = \frac{2}{3} \approx 0.67$$

also,

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2\,2x\,dx = \int_0^1 2x^3\,dx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1^4}{2} - \frac{0^4}{2} = \frac{1}{2}$$

then,

$$\sigma^2 = E(x-\mu)^2 = E(x^2) - \mu^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}, \text{ and } \sigma = \sqrt{\frac{1}{18}} \approx 0.24$$

<u>To find $P(\mu - \sigma < x < \mu + \sigma)$</u>:

$$\Rightarrow \int_{0.67-0.24}^{0.67+0.24} 2x\,dx = \left[x^2\right]_{0.43}^{0.91} = (0.91)^2 - (0.43)^2 \approx 0.64$$

1. Consider the probability density function:

$$f(x) = \begin{cases} cx, & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Find the constant, c.

   (b) Find $P(0.2 < x < 0.5)$.

2. Consider the probability distribution:

$$f(x) = \begin{cases} \frac{x}{2}, & 0 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Verify that $f(x)$ is a probability density function.

   (b) Draw the graph of $f(x)$.

   (c) Find $P(\frac{1}{4} < x < 3/4)$.

   (d) Find μ and σ^2.

   (e) Find $P(\mu - \sigma < x < \mu + \sigma)$.

3. Consider the probability density function:

$$f(x) = \begin{cases} cx^2, & 0 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Find c.

   (b) Draw the graph of $f(x)$

   (c) Find $P(0.5 \le x \le 1)$.

   (d) Find μ and σ^2.

   (e) Find $P(\mu - \sigma < x < \mu - \sigma)$.

   (f) Find $P(\mu - 2\sigma < x < \mu + 2\sigma)$.

(4) Consider the probability distribution:

$$f(x) = \begin{cases} \frac{x^2}{9} & , \quad 0 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify that $f(x)$ is a probability density function.

(b) Draw the graph of $f(x)$.

(c) Find $P(1 < x < 2)$.

(d) Find μ and σ^2.

(e) Find $P(\mu - \sigma < x < \mu + \sigma)$.

(5) Consider the probability density function:

$$f(x) = \begin{cases} k\,x(1-x) & , \quad 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find k.

(b) Find $P(x < 1/4)$.

(c) Find $P(x \geqslant 0.2)$.

(6) Consider the probability density function:

$$f(x) = \begin{cases} \frac{c}{\sqrt{x}} & , \quad 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find c.

(b) Find $P(x \leq 0.25)$.

(c) Find $P(x > \frac{1}{9})$.

(7) Consider the probability distribution:

$$f(x) = \begin{cases} \frac{1}{18}(3+2x) & , 2 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify that $f(x)$ is a probability density function.

(b) Find $P(2 < X < 3)$.

(8) Consider the probability distribution:

$$f(x) = \begin{cases} x & , 0 < x \leq 1 \\ 2-x & , 1 \leq x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify that $f(x)$ is a probability density function.

(b) Find $P(X < 0.5)$.

(c) Find $P(X > 1.3)$.

(d) Find $P(0.2 < X < 1.2)$

(9) Consider the probability distribution:

$$f(x) = \begin{cases} 2e^{-2x} & , x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify that $f(x)$ is a probability density function.

(b) Find $P(2 < X < 4)$.

(10) Consider the probabability density function:

$$f(x) = \begin{cases} kxe^{-x^2} & , x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

then, (a) Find k.

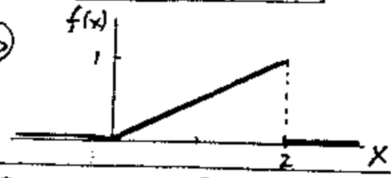
(b) Find $P(1 < X < 2)$.

CONTINUOUS PROBABILITY DISTRIBUTION EXERCISES - SOLUTIONS

(1) (a) $c = 2$ (b) 0.21

(2) (a) $\int_0^2 \frac{x}{2}\,dx \to \left[\frac{x^2}{4}\right]_0^2 \to \frac{2^2}{4} - \frac{0^2}{4} = 1 - 0 = 1$ ✓

(2) (b)

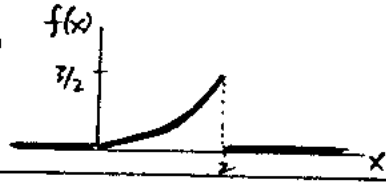


(c) 0.125 (d) $\mu = 4/3$, $\sigma^2 = 2/9$ $\left(\therefore \sigma = \frac{\sqrt{2}}{3}\right)$

(e) $\frac{4\sqrt{2}}{9} = 0.6285$

(3) (a) Consider $c\int_0^2 x^2\,dx = 1$, $\therefore c\left[\frac{x^3}{3}\right]_0^2 = 1$, $\therefore c\left(\frac{2^3}{3} - \frac{0^3}{3}\right) = 1$, $\therefore c \cdot \frac{8}{3} = 1$, $\therefore c = \frac{3}{8}$

(b)

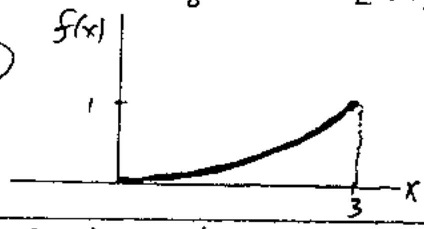


(c) $7/64$ (d) $\mu = 1.5$, $\sigma^2 = 0.15$

(e) 0.67 (f) 0.95

(4) (a) Consider $\int_0^3 x^2/9\,dx \to \left[\frac{x^3}{27}\right]_0^3 \to \frac{3^3}{27} - \frac{0^3}{27} = 1 - 0 = 1$ ✓

(b)



(c) $\frac{7}{27}$ (d) $\mu = 2.25$, $\sigma^2 = 0.3375$

(e) 0.67

(5) (a) Consider $\int_0^1 kx(1-x)\,dx = 1$, $\therefore k\int_0^1 (x - x^2)\,dx = 1$, $\therefore k\left[x^2/2 - x^3/3\right]_0^1 = 1$, $\therefore k\left(\frac{1}{2} - \frac{1}{3}\right) = 1$
$\therefore k\left(\frac{1}{6}\right) = 1$, $\therefore k = 6$

(b) 0.15625 (c) 0.896

(6) (a) Consider $\int_0^4 \frac{c}{\sqrt{x}}\,dx = 1$, $\therefore c\left[2\sqrt{x}\right]_0^4 = 1$, $\therefore c(2\sqrt{4} - 2\sqrt{0}) = 1$, $\therefore c(4) = 1$, $\therefore c = \frac{1}{4}$

(b) $1/4$ (c) $5/6$

(7) (a) Consider $\int_2^4 1/18\,(3 + 2x)\,dx = \left[\frac{1}{18}(3x + x^2)\right]_2^4 = \frac{1}{18}(3\cdot4 + 4^2 - 3\cdot2 - 2^2) = \frac{1}{18}(18) = 1$ ✓

(b) $4/9$

(8) (a) Consider $\int_0^1 x\,dx + \int_1^2 (2-x)\,dx \to \left[x^2/2\right]_0^1 + \left[2x - x^2/2\right]_1^2 = \frac{1}{2} + \frac{1}{2} = 1$ ✓

(b) $1/8$ (c) 0.245 (d) 0.66

(9) (a) Consider $\int_0^\infty 2e^{-2x}\,dx = \left[-e^{-2x}\right]_0^\infty = -e^{-2(\infty)} - (-e^{-2(0)}) = 0 + 1 = 1$ ✓

(b) 0.018

(10) (a) Consider $\int_0^\infty kxe^{-x^2}\,dx = 1$, $\therefore k\left[\frac{-e^{-x^2}}{2}\right]_0^\infty = 1$, $\therefore \frac{k}{2}\left[0 - (-1)\right] = 1$, $\therefore \frac{k}{2}[1] = 1$, $\therefore k = 2$

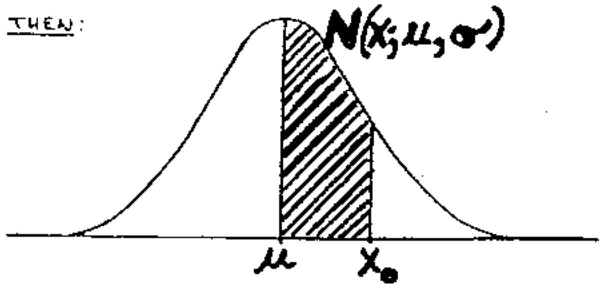
(b) 0.350

THE NORMAL PROBABILITY DISTRIBUTION

THE NORMAL DISTRIBUTION : $N(x;\mu,\sigma)$

Consider a continuous random variable, x, that has a Normal (Probability) Distribution with MEAN, μ, and STANDARD DEVIATION, σ

THEN:



AND:

$P(\mu < x < x_0) =$ shaded area

$P(x = x_0) = 0$

$P(x < \mu) = P(x > \mu) = 0.5000$

$P(\mu - \sigma < x < \mu + \sigma) = 0.6826$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$

Actually, $N(x;\mu,\sigma) = \dfrac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$, $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$, and $-\infty < x < \infty$

NOTE:

① Random variables that are Normally Distributed include: heights, weights, ages, incomes, times, grades and IQ'S.

② Probabilities (i.e. areas) for Normal Distributions are resolved using tabled areas of THE STANDARD NORMAL DISTRIBUTION.

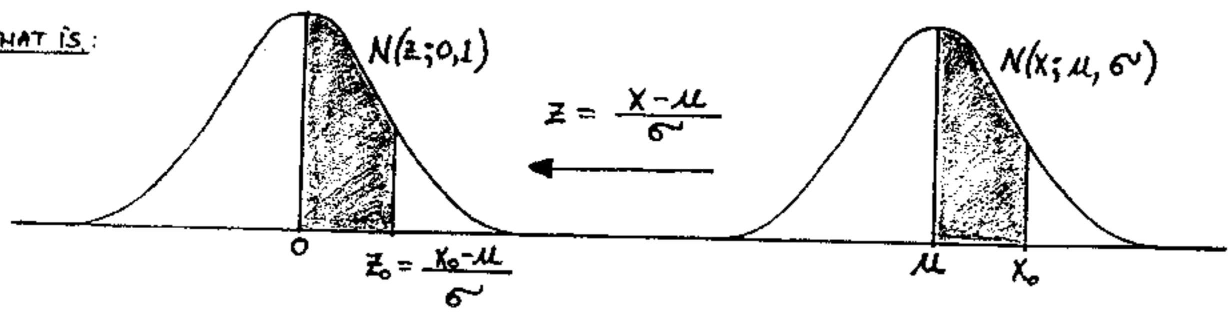
THE NORMAL PROBABILITY DISTRIBUTION – TOPICS
1 THE STANDARD NORMAL DISTRIBUTION : $N(z;0,1)$
2 APPLICATIONS OF THE NORMAL DISTRIBUTION
3 THE NORMAL APPROXIMATION TO THE BINOMIAL

1 THE STANDARD NORMAL DISTRIBUTION

THE STANDARD NORMAL DISTRIBUTION : $N(z;0,1)$

All <u>Normal Distributions</u> can be transformed to <u>The Standard Normal Distribution</u>.

<u>THAT IS</u>:



<u>WHERE</u>:

$$\boxed{P(0<z<z_0) \quad = \quad P(\mu<x<x_0)}$$

<u>NOTE</u>: For different values of z_0, $P(0<z<z_0)$ is available in the $N(z;0,1)$ Table

The following three examples illustrate the uses of the $N(z;0,1)$ Table.

<u>EXAMPLE 1.1</u> Use the $N(z;0,1)$ Table in the Appendix to find:

(a) $P(0<z<1.23)$ (b) $P(-1.00<z<1.00)$

(c) $P(z<1.85)$ (d) $P(z>2.50)$

Using shaded areas to represent the required probabilities, we have:

(a)

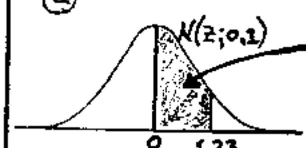


Table $\boxed{.3907}$

$\boxed{\text{See the } N(z;0,1) \text{ Table}}$

(b)

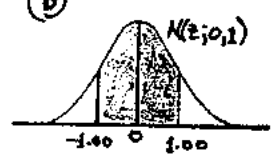


Table $2 \times \boxed{.3413}$ $= 0.6826$

(Recall the Empirical Rule)

(c)

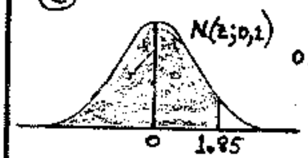


Table $0.5 + \boxed{.4678}$ $= 0.9678$

(d)

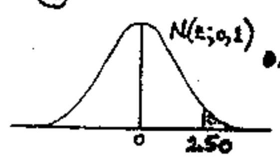


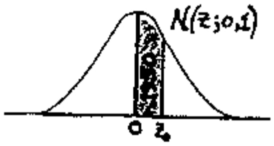
Table $0.5 - \boxed{.4938}$ $= 0.0062$

EXAMPLE 1.2 Use the $N(z;0,1)$ Table to find z_0 if:

ⓐ $P(0 < z < z_0) = 0.1950$ ⓑ $P(z < z_0) = 0.0250$

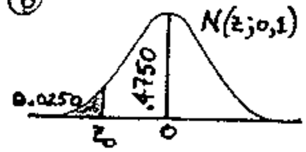
Using shaded areas to represent the given probabilities, we have:

ⓐ



$\therefore z_0 = 0.51$

ⓑ



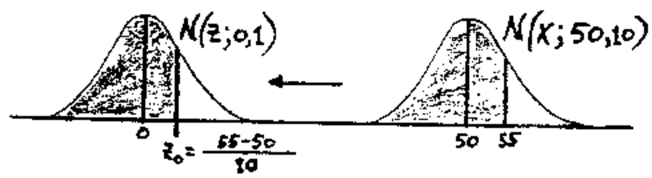
$\therefore z_0 = -1.96$

EXAMPLE 1.3 Consider $N(X;50,10)$. Use the $N(z;0,1)$ Table to find each of the following.

ⓐ $P(X < 55)$ ⓑ $P(X > 75)$

Using shaded areas to represent the required probabilities, we have:

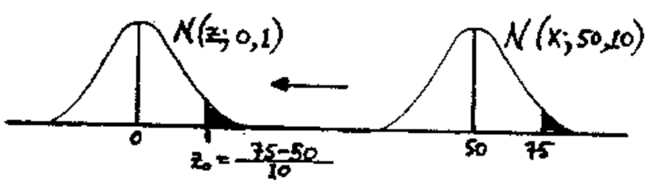
ⓐ



$\therefore z_0 = 0.50$

hence, $P(X < 55) = 0.5 + \boxed{.1915} = 0.6915$

ⓑ



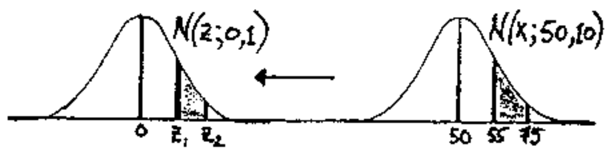
$\therefore z_0 = 2.50$

hence, $P(X > 75) = 0.5 - \boxed{.4938} = 0.0062$

The following examples illustrate some applications of the Normal Distribution.

EXAMPLE **2.1** A manufacturer claims that the life span of its batteries is Normally Distributed with a mean of 50 hours and a standard deviation of 10 hrs. Accordingly, if you buy one of these batteries, what is the probability that it will last between 55 and 75 hours?

Referring to example 1.3, we have:



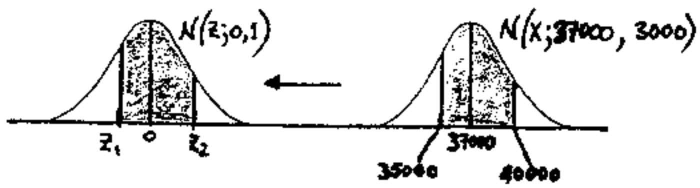
$z_1 = \frac{55-50}{10}$ and $z_2 = \frac{75-50}{10}$

$\therefore z_1 = 0.50$ $\therefore z_2 = 2.50$

hence, $P(55 < x < 75) = \boxed{.4938} - \boxed{.1915} = 0.3023$

EXAMPLE **2.2** It is reported that annual incomes in a certain city are Normally Distributed with a mean of \$37000 and a standard deviation of \$3000. Hence, what proportion of these incomes are between \$35000 and \$40000?

That is:



$z_1 = \frac{35000-37000}{3000}$ and $z_2 = \frac{40000-37000}{3000}$

$\therefore z_1 = -0.67$ $\therefore z_2 = 1.00$

hence, the required proportion is $\boxed{.2486} + \boxed{.3413} = 0.5899$

EXAMPLE 2.3 If the heights of American men are Normally Distributed with a mean of 5.82 feet, and a standard deviation of 0.11 feet, then:

ⓐ Above what height do the tallest 5% of the men fall?

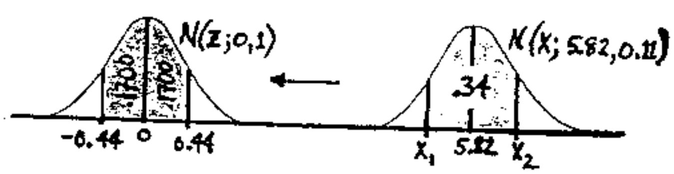
ⓑ Between what two heights do the middle 34% of the men fall?

ⓐ Letting X_0 be the required height, we have:

hence, $1.645 = \dfrac{X_0 - 5.82}{0.11}$

$\therefore X_0 = 6.00$ feet

ⓑ Letting X_1 and X_2 be the required heights, we have:

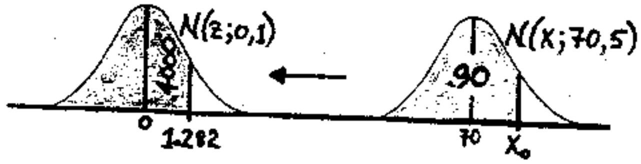


hence, $-0.44 = \dfrac{X_1 - 5.82}{0.11}$ and $0.44 = \dfrac{X_2 - 5.82}{0.11}$

$\therefore X_1 = 5.77$ feet $\quad \therefore X_2 = 5.87$ feet

EXAMPLE 2.4 If students average 70 minutes to complete a test, with a standard deviation of 5 minutes (assuming Normality), when should the test be terminated so that 90% of the students will complete it?

Letting X_0 be the required time, we have:

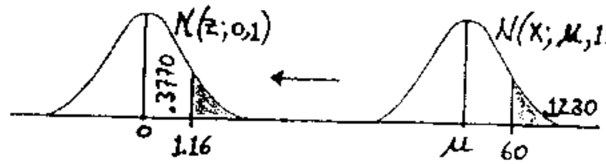


hence, $1.282 = \dfrac{X_0 - 70}{5}$

$\therefore X_0 = 76.41$ minutes

EXAMPLE **2.5** What is the mean age of Montréalers if 12.3 % of them are more than 60 years old? Use a standard deviation of 15 years and assume Normality.

Letting μ be the required mean, we have:

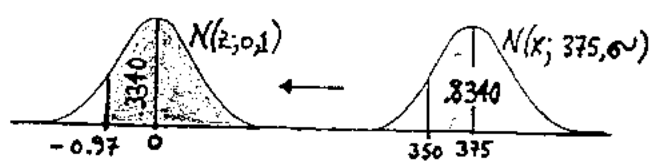


hence, $1.16 = \dfrac{60-\mu}{15}$

$\therefore \mu = 42.6$

EXAMPLE **2.6** Certain aptitude test scores are known to be Normally Distributed, with a mean of 375. If 83.4% of these scores exceed 350, then what is the standard deviation of the scores?

Letting σ be the required standard deviation, we have:

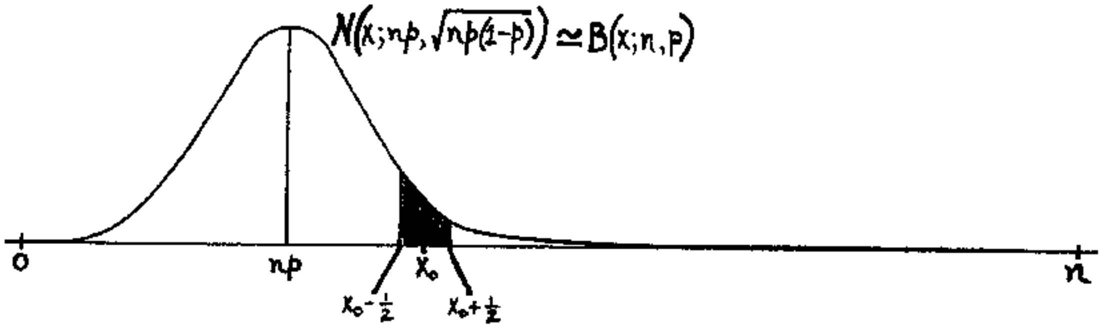


hence, $-0.97 = \dfrac{350-375}{\sigma}$

$\therefore \sigma = 25.8$

We may use $N(x; np, \sqrt{np(1-p)})$ to approximate $B(x; n, p)$ if np and $n(1-p)$ are ≥ 5.

The approximation is best if we make the <u>continuity correction</u> as follows:



THE CONTINUITY CORRECTION
The Normal $P(x_0 - \frac{1}{2} < x < x_0 + \frac{1}{2}) \simeq$ The Binomial $P(x = x_0)$

hence, also

Normal $P(x > k - \frac{1}{2}) \simeq$ Binomial $P(x \geq k)$
Normal $P(x < k + \frac{1}{2}) \simeq$ Binomial $P(x \leq k)$

The following examples illustrate the <u>Normal Approximation to the Binomial</u>.

EXAMPLE 3.1 (a) Use the $B(x; 25, .40)$ Tabled entries to find $P(x \leq 9)$, $P(x = 10)$, and $P(x \geq 11)$.
(b) Use the <u>Normal Approximation to the Binomial</u> to approximate these 3 probabilities.

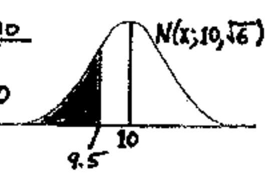
(a) Referring to the $B(x; 25, .40)$ Table in the Appendix, we have:

$P(x \leq 9) = .4246$, $P(x = 10) = .1612$, $P(x \geq 11) = .4142$

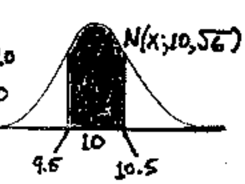
NOTE:	0.4246	+	0.1612	+	0.4142	= 1

(b) Using $N(x; 25(.40), \sqrt{25(.40)(.60)})$ to approximate $B(x; 25, .40)$, we have:

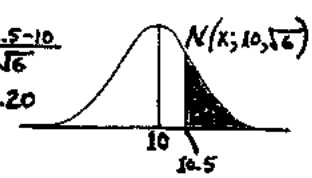
$z = \frac{9.5 - 10}{\sqrt{6}}$
$\therefore z = -0.20$



$\therefore z_1 = -0.20$
$\therefore z_2 = 0.20$



$z = \frac{10.5 - 10}{\sqrt{6}}$
$\therefore z = 0.20$



$\therefore P(x < 9.5) = .5 - \boxed{.0793} = .4207$, $\therefore P(9.5 < x < 10.5) = 2\boxed{.0793} = .1586$, $\therefore P(x > 10.5) = .5 - \boxed{.0793} = .4207$

NOTE:	0.4207	+	0.1586	+	0.4207	= 1

EXAMPLE 3.2 A student guesses at each of 48 questions on a multiple-choice test. If each question has 4 choices (1 correct and 3 incorrect), then:

(a) What is the probability that the student gets exactly 10 questions correct?

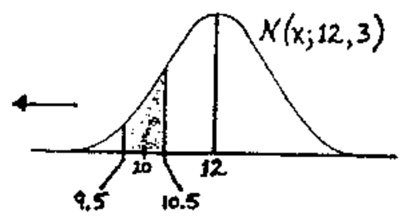
(b) Use the Normal Approximation to the Binomial to approximate this probability.

(a) The required probability is $\frac{48!}{10!\ 38!}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^{38} = 0.1115$

(b) Using $N\left(x;\ 48(\tfrac{1}{4}), \sqrt{48(\tfrac{1}{4})(\tfrac{3}{4})}\right)$ to approximate $B(x;48,\tfrac{1}{4})$, we have:

$z_1 = \frac{9.5-12}{3}$ and $z_2 = \frac{10.5-12}{3}$

$\therefore z_1 = -0.83$ $\quad\therefore z_2 = -0.50$



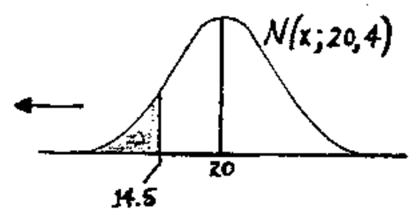
hence, the required probability is approximately $\boxed{.2967} - \boxed{.1915} = 0.1052$

EXAMPLE 3.3 An automaker claims that it has gained a $\frac{1}{5}$th share of the new car market. If this is true, what is the probability that a random sample of 100 new cars sold recently will include less than 15 of their cars?

Using $N\left(x;\ 100(\tfrac{1}{5}), \sqrt{100(\tfrac{1}{5})(\tfrac{4}{5})}\right)$ to approximate $B(x;100,\tfrac{1}{5})$, we have:

$z = \frac{14.5-20}{4}$

$\therefore z = -1.38$



hence, the required probability is approximately $0.5 - \boxed{.4162} = 0.0838$

NOTE: The (Binomial) expression for the actual probability here is:

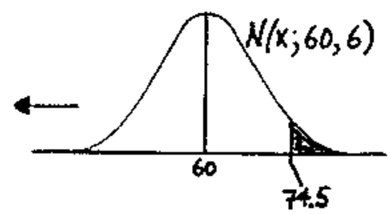
$$\sum_{x=0}^{14} \frac{100!}{x!\,(100-x)!}\left(\frac{1}{5}\right)^{x}\left(\frac{4}{5}\right)^{100-x}$$

EXAMPLE 3.4. Last year a basketball player scored on 40% of his shots. If he continues at that rate this year, then what is the probability that he will score on at least half of his first 150 shots this year?

Using $N\left(x; 150(.40), \sqrt{150(.40)(.60)}\right)$ to approximate $B(x; 150, .40)$, we have:

$$z = \frac{74.5 - 60}{6}$$

$$\therefore z = 2.42$$



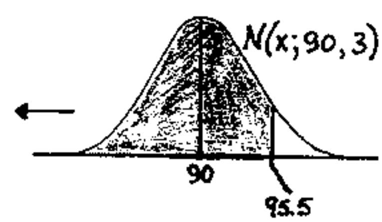
hence, the required probability is approximately $0.5 - \boxed{.4922} = 0.0078$

EXAMPLE 3.5 One hundred reservations are taken for a cruise on a ship that has only 95 cabins. If experience shows that only 90% of the people making such reservations actually show up, then what is the probability that there will be a cabin for each person who shows up?

Using $N\left(x; 100(.90), \sqrt{100(.90)(.10)}\right)$ to approximate $B(x; 100, .90)$, we have:

$$z = \frac{95.5 - 90}{3}$$

$$\therefore z = 1.83$$



hence, the required probability is approximately $0.5 + \boxed{.4664} = 0.9664$

Normal Distribution Exercises

① Use the $N(z;0,1)$ Table to find:

ⓐ $P(0 < z < 0.82)$ ⓑ $P(-1.65 < z < 0)$ ⓒ $P(z < 1.44)$

ⓓ $P(z > -2.01)$ ⓔ $P(z > 0.10)$ ⓕ $P(z < -1.49)$

ⓖ $P(-1.18 < z < 0.55)$ ⓗ $P(1.20 < z < 1.39)$ ⓘ $P(-2.50 < z < -0.89)$

② Use the $N(z;0,1)$ Table to find z_0 if:

ⓐ $P(0 < z < z_0) = 0.4726$ ⓑ $P(z < z_0) = 0.9838$

ⓒ $P(z < z_0) = 0.0668$ ⓓ $P(-z_0 < z < z_0) = 0.8502$

③ Consider $N(x;\mu,10)$. Use the $N(z;0,1)$ Table to find μ if:

ⓐ $P(x < 82.5) = 0.8264$ ⓑ $P(x > 75) = 0.6915$

④ Consider $N(x;50,\sigma^2)$. Use the $N(z;0,1)$ Table to find σ^2 if:

ⓐ $P(x < 57.5) = 0.9332$ ⓑ $P(x < 47.75) = 0.2266$

⑤ Consider $N(x;65,7)$. Use the $N(z;0,1)$ Table to find:

ⓐ $P(x > 75)$ ⓑ $P(60 < x < 75)$ ⓒ $P(55 < x < 60)$

⑥ The time (in minutes) of flights from Montréal to Toronto is Normally distributed with a mean of 55 and a standard deviation of 3. Hence, if you take a flight from Montréal to Toronto, what is the probability that it will take longer than an hour?

⑦ The heights of college males are Normally distributed with a mean of 175 cm and a standard deviation of 4 cm. Hence, what proportion of college males:

ⓐ are between 170 and 180 cm. in height? ⓑ are taller than 185 cm.?

⑧ If it is known that IQ's are Normally distributed with a mean of 100 and a standard deviation of 10, then what percentage of IQ's are:

ⓐ less than 100? ⓑ between 90 and 110?

ⓒ greater than 125? ⓓ between 85 and 95?

(9) The ages of graduating CEGEP students are known to average 19.53 years, with a standard deviation of 0.45 yrs. Assuming that the ages are Normally distributed, then what proportion of graduating CEGEP students:

ⓐ are more than 21 years old? ⓑ are more than 19 years old?

(10) If Statistics grades are Normally distributed with a mean of 70 and a standard deviation of 10, then:

ⓐ what percentage of Statistics students get over 95?

ⓑ what percentage of Statistics students pass (i.e. get at least 60)?

ⓒ what proportion of those who pass, earn a grade over 95?

(11) The time it takes a statistician to walk to work each day is Normally distributed with a mean of 20 minutes, and a standard deviation of 2.5 minutes.

ⓐ If she leaves for work at 7:35 AM, what is the prob. that she will arrive by 8:00 AM?

ⓑ at what time must she leave for work, to give herself a 90% chance of arriving by 8:00 AM?

(12) The daily demand for a particular consumer product is Normally distributed with a mean of 1000 and a standard deviation of 100. If it is decided that the chance of a sellout for any day should not exceed 2.5%, then what should be the minimum inventory each morning?

(13) The distribution of salaries in a certain industry is given as: $N(x; \$25{,}300, \$5000)$.

ⓐ Above what salary do the highest 5% of these salaries lie?

ⓑ Below what salary do the lowest 33% of these salaries lie?

ⓒ Between what salaries do the middle 34% of these salaries lie?

(14) A coffee vending machine fills cups Normally with a standard deviation of 0.3 oz. What should the mean be set at, so that 8 oz. cups will overflow no more than 1% of the time?

(15) The weights of pineapples in a very large shipment are Normally distributed with a mean of 5 lbs. If 2.28% of these pineapples weigh less than 4 lbs., then what percentage of the pineapples weigh between 3.5 and 4 lbs.?

(16) A fair coin is tossed 100 times. What is the probability of observing exactly 57 heads?
 (a) Give the actual Binomial expression representing this probability.
 (b) Use the Normal approximation to the Binomial to estimate this probability.

(17) A student guesses at each of 192 questions on a multiple-choice test. If each question has 4 choices (1 correct, 3 incorrect), then what is the probability that the student will:
 (a) get exactly 48 questions correct? (b) get more than 25 % of the questions correct?

(18) A salesman finds that there is a 40% chance of making a sale to each customer he contacts. Hence, what is the probability of making at least 50 sales if he contacts 150 customers?

(19) A rare blood type occurs in 10% of the population of Canada. What is the prob. that a medical team checking the blood type of 100 Canadians will find more than 15 people with this blood type?

(20) A TV network claims that 20% of all available viewers watch its evening news. If this is true, what is the probability that at most 450 viewers in a poll of 2500 viewers watch its news?

(21) Air Canada finds that only 90% of all persons making reservations actually show up for their flight. Hence, if they take 400 reservations for a flight on a 375 seat plane, what is the prob that there will be more passengers than seats for the flight?

(22) The Plant reports that $2/3^{\text{rds}}$ of all Dawson students plan to go to university. If this is true, what is the probability that less than 100 in sample of 162 Dawson students plan to go to university.

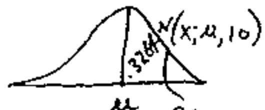
(23) A study estimates that 70% of all Montréalers are bilingual. If this is true, then what is the probability that a random sample of 2100 Montréalers will include between 1428 and 1512 Montréalers, inclusively, who are bilingual?

(24) If the time required for a student to complete an aptitude test is distributed as: $N(X;55,10)$ where X is in minutes, then what is the probability that more than $3/4^{\text{ths}}$ of 100 students who take the test will complete it, if it is terminated after 1 hour?

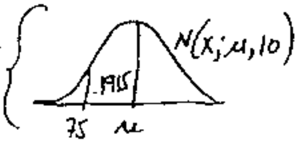
NORMAL DISTRIBUTION EXERCISES – SOLUTIONS

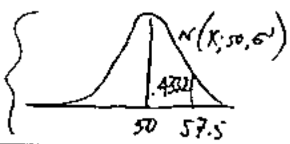
(1) (a) .2939 (b) .4505 (c) .5 + .4251 = .9251 (d) .5 + .4778 = .9778
(e) .5 − .0398 = .4602 (f) .5 − .4319 = .0681 (g) .3810 + .2088 = .5898
(h) .4177 − .3849 = .0328 (i) .4938 − .3133 = .1805

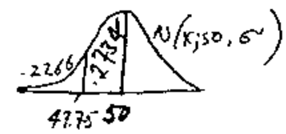
(2) (a) $z_0 = 1.92$ (b) .9838 − .5 = .4838, ∴ $z_0 = 2.14$
(c) .5 − .0668 = .4332, ∴ $z_0 = -1.50$ (d) $\frac{0.8502}{2} = 0.4251$, ∴ $z_0 = 1.44$

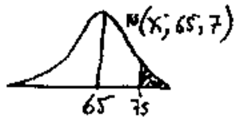
(3) (a)


$\therefore 0.94 = \frac{82.5-\mu}{10}$
∴ μ = 73.1

(b) $\therefore -0.50 = \frac{75-\mu}{10}$
∴ μ = 80


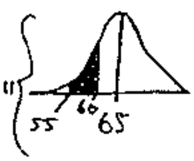
(4) (a) $\therefore +1.50 = \frac{57.5-50}{\sigma}$
∴ 5 = σ


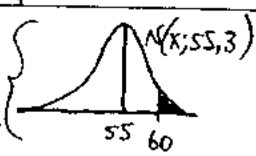
(b) $\therefore -0.75 = \frac{47.75-50}{\sigma}$
∴ σ = 3


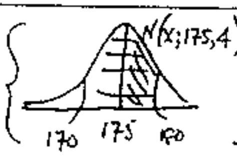
(5) (a)


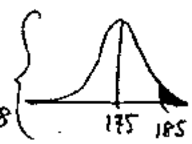
$z = \frac{75-65}{7} = 1.43$
∴ Ans. = .5 − .4236 = .0764

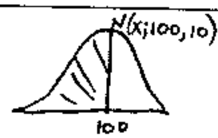
(b) $z_1 = 0.71$
$z_2 = 1.43$
∴ Ans = .2611 + .4236 = .6847

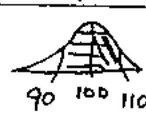

(c) $z_1 = -1.43$
$z_2 = -0.71$
∴ Ans = .4236 − .2611 = .1625


(6) $z = \frac{60-55}{3}$
z = 1.67
∴ Ans = .5 − .4525 = .0475


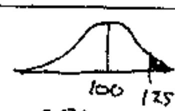
(7) (a) $z = \frac{180-175}{4}$
z = 1.25
∴ Ans = 2(.3944) = .7888


(b) $z = \frac{185-175}{4}$
z = 2.50
∴ Ans. = .5 − .4938 = .0062


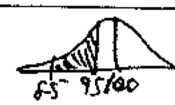
(8) (a) Ans = .50


(b)


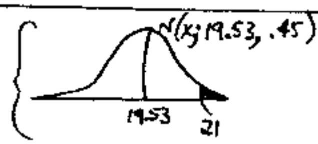
2(.3413) = .6826

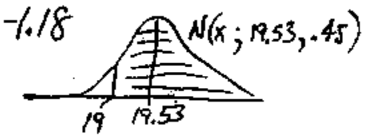
(c)


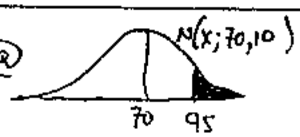
z = 2.50
∴ Ans = .5 − .4938 = .0062

(c)


$z_1 = 1.50$, $z_2 = .50$
∴ Ans = .4332 − .1915 = .2417

(9) (a) $z = \frac{21-19.53}{.45}$
z = 3.27
Ans = .5 − .4995 = .0005


(b) $\therefore z = \frac{19-19.53}{0.45} = -1.18$
Ans = .5 + .3810 = .8810


(10) (a)


$\therefore z = \frac{95-70}{10} = 2.50$
∴ Ans = .5 − .4938 = 0.0062 (0.62%)

(b)


$\therefore z = \frac{60-70}{10} = 1.00$
∴ Ans = .5 + .3413 = 0.8413 (84.13%)

(c) Ans = $\frac{.0062}{.8413}$ = .0074

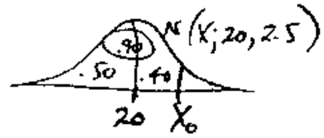
NORMAL DISTRIBUTION EXERCISES – SOLUTIONS

(11) (a)

$\therefore z = \frac{25-20}{2.5} = 2.00$

$\therefore$ Ans $= .5 + .4772 = .9772$

(b)

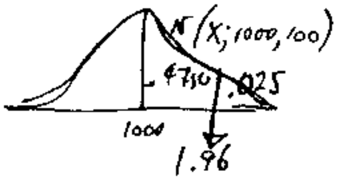


$\therefore 1.282 = \frac{X_0 - 20}{2.5}$

$\therefore 23.205 = X_0$

$\boxed{\text{i.e. } 7{:}36.795 \text{ AM} \simeq 7{:}37 \text{ AM}}$

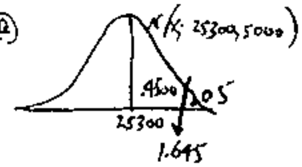
(12)



$\therefore 1.96 = \frac{X_0 - 1000}{100}$

$\therefore 1196 = X_0$

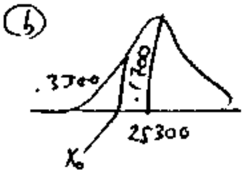
(13) (a)



$\therefore 1.645 = \frac{X_0 - 25300}{5000}$

$\therefore 33525 = X_0$

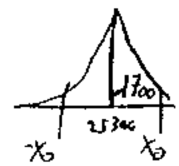
(b)



$\therefore -0.44 = \frac{X_0 - 25300}{5000}$

$\therefore 23100 = X_0$

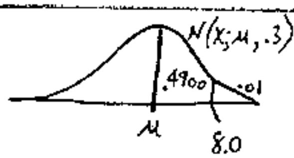
(c)



$\therefore X_0 = 27500$

$(23100 \; X_0 \; 27500)$

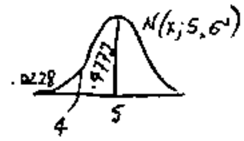
(14)



$\therefore 2.33 = \frac{8 - \mu}{.3}$

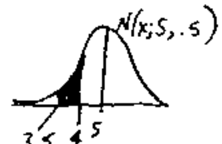
$\therefore \mu = 7.301$

(15)



$\therefore -2.00 = \frac{4-5}{\sigma}$

$\therefore .50 = \sigma$



$\therefore z_1 = \frac{3.5-5}{.5} = -3.00$

$\therefore z_2 = \frac{4-5}{.5} = -2.00$

$\therefore$ Ans $= .4987 - .4772 = .0215$

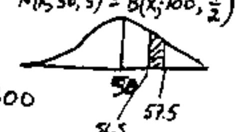
i.e. 2.15%

(16) (a) $\frac{100!}{57!\,43!}\left(\frac{1}{2}\right)^{57}\cdot\left(\frac{1}{2}\right)^{43}$

(b) $z_1 = \frac{56.5-50}{5} = 1.30$

$z_2 = \frac{57.5-50}{5} = 1.50$

$N(X; 50, 5) \simeq B(X; 100, \frac{1}{2})$



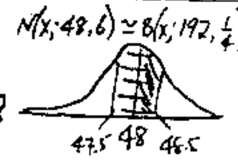
$\therefore$ Ans $= .4332 - .4032 = .0300$

(17) (a) $z = \frac{48.5-48}{6}$

$z = 0.08$

$\therefore$ Ans $= 2(.0319) = .0638$

$N(X; 48, 6) \simeq B(X; 192, \frac{1}{4})$

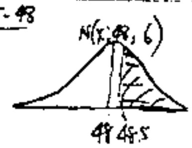


(17) (b) $.25$ of $192 = 48$; $z = \frac{48.5-48}{6}$

$z = 0.08$

$\therefore$ Ans $= .5 - .0319$

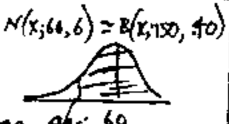
$= 0.4681$



(18) $z = \frac{49.5-60}{6}$

$\therefore z = 1.75$

$N(X; 60, 6) \simeq B(X; 150, .40)$



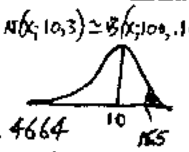
$\therefore$ Ans $= .5 + .4599$

$= .9599$

(19) $z = \frac{15.5-10}{3}$

$\therefore z = 1.83$

$N(X; 10, 3) \simeq B(X; 100, .10)$



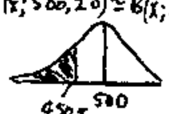
$\therefore$ Ans $= .5 - .4664$

$= .0336$

(20) $z = \frac{450.5-500}{20}$

$z = -2.48$

$N(X; 500, 20) \simeq B(X; 2500, .20)$



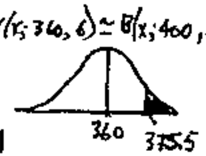
$\therefore$ Ans $= .5 - .4934$

$= .0066$

(21) $z = \frac{375.5-360}{6}$

$\therefore z = 2.58$

$N(X; 360, 6) \simeq B(X; 400, .90)$



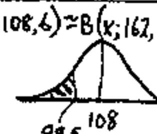
$\therefore$ Ans $= .5 - .4951$

$= .0049$

(22) $N(X; 108, 6) \simeq B(X; 162, \frac{2}{3})$

$z = \frac{99.5-108}{6}$

$z = -1.42$



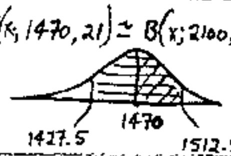
$\therefore$ Ans $= .5 - .4222$

$= .0778$

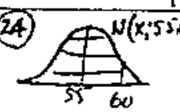
(23) $z = \frac{1512.5-1470}{21} = 2.02$

$\therefore$ Ans $= 2(.4783) = .9566$

$N(X; 1470, 21) \simeq B(X; 2100, .70)$



(24)



$\therefore z = \frac{60-55}{10} = 0.5$

$\therefore 0.5 + .1915$

$\Rightarrow 0.6915$

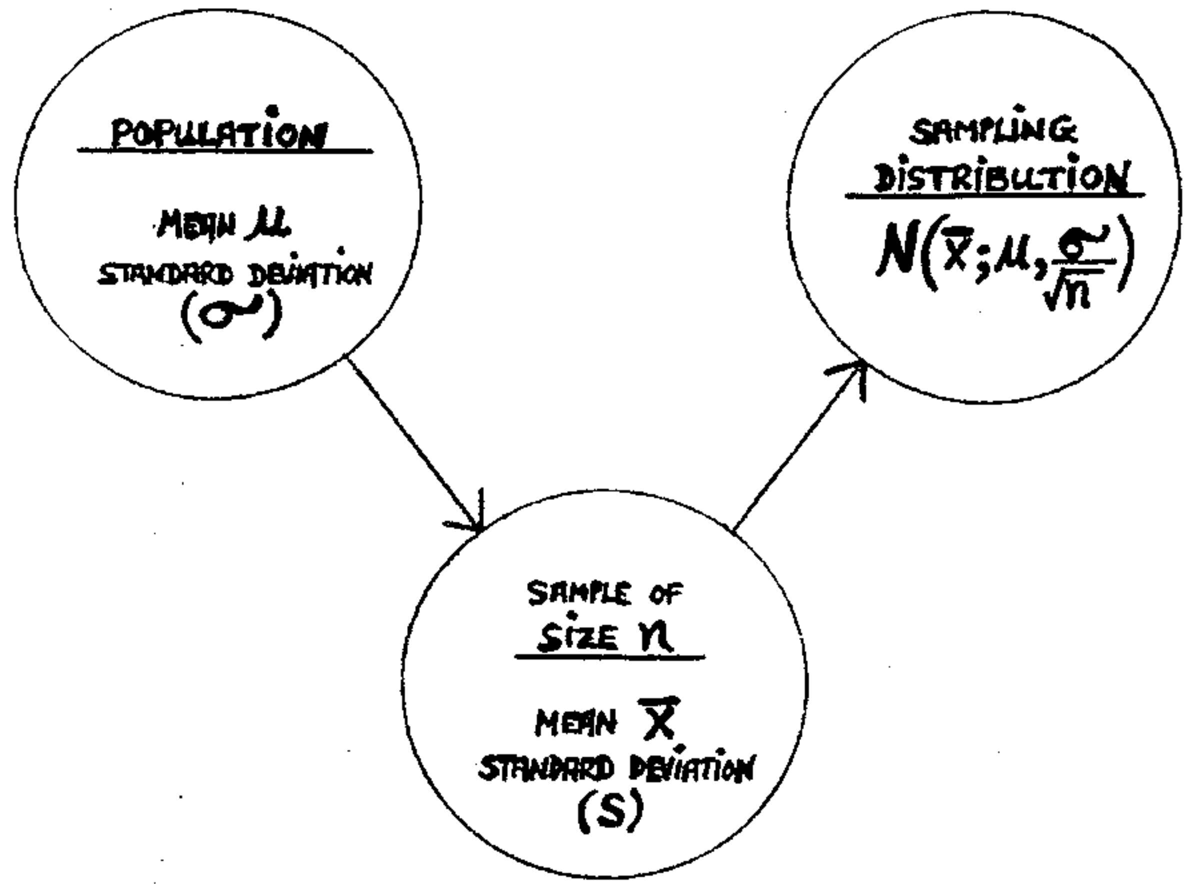
$\therefore P(X \geq 76) \simeq B(X; 100, 0.6915)$

$\simeq N(X; 69.15, 4.62)$

$\therefore z_0 = \frac{75.5 - 69.15}{4.62} = 1.37$

$\therefore$ Ans. $= 0.5 - 0.4147 = 0.0853$

CENTRAL LIMIT THEOREM



	Population σ known		Population σ unknown	
	Population Normal	Population Non-Normal	Population Normal	Population Non-Normal
Sampling Distribution of $\overline{X}$	$N\left(\overline{X};\mu,\frac{\sigma}{\sqrt{n}}\right)$ for all n	$N\left(\overline{X};\mu,\frac{\sigma}{\sqrt{n}}\right)$ for $n > 30$	$N\left(\overline{X};\mu,\frac{S}{\sqrt{n}}\right)$ for $n > 30$ and use the t-distr. for $n \leq 30$	$N\left(\overline{X};\mu,\frac{S}{\sqrt{n}}\right)$ for $n > 30$

THE CENTRAL LIMIT THEOREM — AN ILLUSTRATIVE EXAMPLE

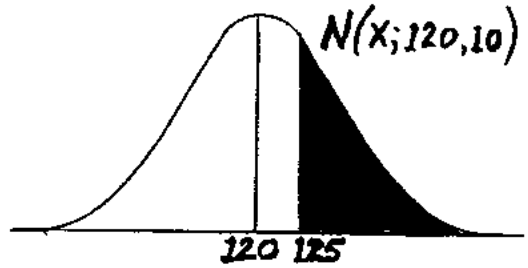
If it is known that the IQ's of Dawson students are Normally Distributed with a mean of 120 and a standard deviation of 10, then:

To find the probability that a randomly selected Dawson student has an IQ greater than 125, consider:

$$\therefore z_0 = \frac{125-120}{10}$$

$$\therefore z_0 = 0.50$$

$$\therefore P(X>125) = 0.5-0.1915 = 0.3085$$



THE (SAMPLING) DISTRIBUTION OF THE MEAN, $\overline{X}$, of a random sample of 36 Dawson students' IQ's is given by THE CENTRAL LIMIT THEOREM as:

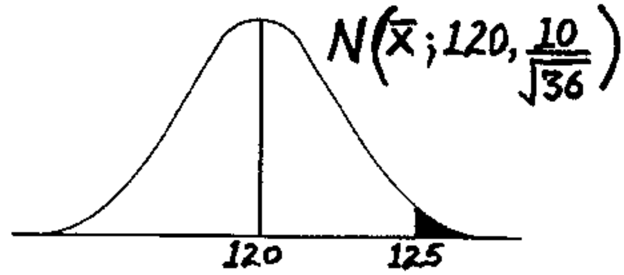
$$N\left(\overline{X};\, 120, \frac{10}{\sqrt{36}}\right)$$

To find the probability that the mean of the sample is greater than 125:

$$\therefore z_0 = \frac{125-120}{\frac{10}{\sqrt{36}}}$$

$$\therefore z_0 = 3.00$$

$$\therefore P(\overline{x}>125) = 0.5-0.4987 = 0.0013$$



THE CENTRAL LIMIT THEOREM EXERCISES

① A random sample of size 49 is selected from a population whose mean is 70, and whose standard deviation is 14. What is the probability distribution of the r.v. $\overline{X}$, the sample mean?

② A random sample of size 36 is selected from a population whose mean is 80 and whose st. dev. is 10. Then find: ⓐ $P(\overline{X} > 80)$, ⓑ $P(\overline{X} > 76)$ and ⓒ $P(77 < \overline{X} < 83)$.

③ Consider the r.v. X that is distributed as: $N(X; 160, 25)$. A random sample of size **16** is selected from this population. Then find:

ⓐ $P(X > 170)$ and $P(\overline{X} > 170)$, ⓑ $P(155 < X < 165)$ and $P(155 < \overline{X} < 165)$.

④ Bell Canada reports that the average duration of long-distance telephone calls in Canada is 12.0 minutes (and the standard deviation is 4.0 minutes. **Hence**,

ⓐ what proportion of the long-distance calls in Canada last longer than 15 minutes? (assume Normality)

ⓑ what is the probability that the average of a random sample of 100 long-distance calls in Canada will exceed 13 minutes?

⑤ If the annual incomes of Montréalers are distributed as $N(X; \$35000, \$5000)$, then find:

ⓐ the proportion of Montréalers whose incomes exceed \$44000.

ⓑ the probability that the mean of a random sample of 64 Montréalers will exceed \$37000 in income.

⑥ A manufacturer claims that it produces vitamin C pills with an average weight of 500 mg and a standard deviation of **20 mg**. If this claim is correct, then find the probability that the mean of a random sample of these pills will weigh more than 496 mg. if the size of the sample is: ⓐ 64 and ⓑ 100.

⑦ The Plant reports that the **mean** age of all CEGEP students is 18.5 years. A random sample of 64 CEGEP students yielded a standard deviation of 1.5 yrs. If the report is correct, then

ⓐ what proportion of all CEGEP students are more than 19 years old (assume normality)?

ⓑ what is the probability that the mean of our random sample will exceed 19 years?

⑧ The scores on an aptitude test are known to be normally distributed with a mean of 475

ⓐ If 33 % of the scores are less than 431, find the standard deviation of the scores.

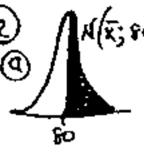
ⓑ Accordingly, what is the probability that the mean score of a random sample of 64 of these scores will exceed 500?

THE CENTRAL LIMIT THEOREM EXERCISES – SOLUTIONS

(1) $N\left(\bar{X}; 70, \frac{14}{\sqrt{49}}\right)$

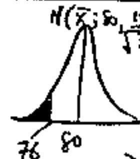
or $N(\bar{X}; 70, 2)$

(2) (a) $N\left(\bar{x}; 80, \frac{10}{\sqrt{36}}\right)$



$\therefore Ans = .50$

(b) $N\left(\bar{x}; 80, \frac{10}{\sqrt{36}}\right)$



$\therefore z_0 = \frac{76-80}{10/6}$

$\therefore z_0 = -2.40$

$\therefore Ans = .5 + .4918 = .9918$

(c) $\therefore z_0 = \frac{83-80}{10/6}$

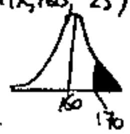
$\therefore z_0 = 1.80$

$N\left(\bar{x}; 80, \frac{10}{\sqrt{36}}\right)$

$\therefore Ans = 2(.4641) = .9282$

(3) (a) $\therefore z_0 = \frac{170-160}{25}$ $N(x; 160, 25)$

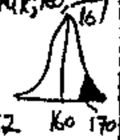
$\therefore z_0 = 0.40$



$\therefore Ans = .5 - .1554 = .3446$

and $\therefore z_0 = \frac{170-160}{25/4}$ $N\left(\bar{x}; 160, \frac{25}{16}\right)$

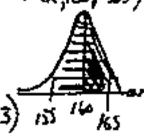
$\therefore z_0 = 1.60$



$\therefore Ans = .5 - .4452 = .0548$

(b) $\therefore z_0 = \frac{165-160}{25}$ $N(x; 160, 25)$

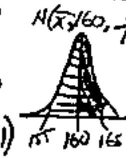
$\therefore z_0 = 0.20$



$\therefore Ans = 2(.0793) = .1586$

and $\therefore z_0 = \frac{165-160}{25/4}$ $N\left(\bar{x}; 160, \frac{25}{16}\right)$

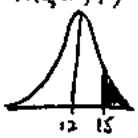
$\therefore z_0 = 0.80$



$\therefore Ans = 2(.2881) = .5762$

(4) (a) $\therefore z_0 = \frac{15-12}{4}$ $N(x; 12, 4)$

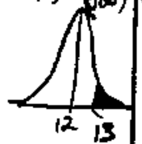
$\therefore z_0 = 0.75$



$\therefore Ans = .5 - .2734 = .2266$

(b) $\therefore z_0 = \frac{13-12}{4/10}$ $N\left(\bar{x}; 12, \frac{4}{\sqrt{100}}\right)$

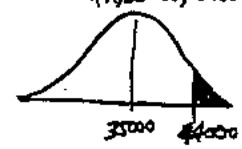
$\therefore z_0 = 2.50$



$\therefore Ans = .5 - .4938 = .0062$

(5) (a) $\therefore z = \frac{44000-35000}{5000}$ $N(X; 35000, 5000)$

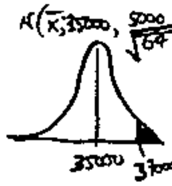
$\therefore z = 1.80$



$\therefore Ans = .5 - .4641 = .0359$

(5) (b) $\therefore z_0 = \frac{37000-35000}{5000/8}$ $N\left(\bar{x}; 35000, \frac{5000}{\sqrt{64}}\right)$

$\therefore z_0 = 3.20$



$\therefore Ans = .5 - .4993 = .0007$

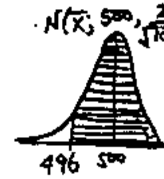
(6) (a) $\therefore z = \frac{496-500}{20/8}$ $N\left(\bar{x}; 500, \frac{20}{\sqrt{64}}\right)$

$\therefore z = -1.60$

$\therefore Ans = .5 + .4452 = .9452$

and (b) $\therefore z_0 = \frac{496-500}{20/10}$ $N\left(\bar{x}; 500, \frac{20}{\sqrt{100}}\right)$

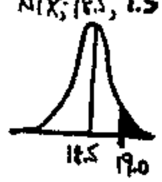
$\therefore z_0 = -2.00$



$\therefore Ans = .5 + .4772 = .9772$

(7) (a) $\therefore z_0 = \frac{19.0-18.5}{1.5}$ $N(x; 18.5, 1.5)$

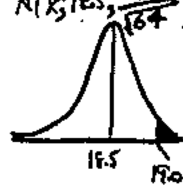
$\therefore z_0 = 0.3\dot{3}$



$\therefore Ans = .5 - .1293 = .3707$

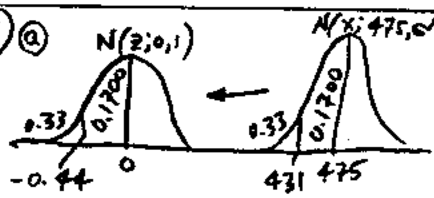
(b) $\therefore z_0 = \frac{19.0-18.5}{1.5/8}$ $N\left(\bar{x}; 18.5, \frac{1.5}{\sqrt{64}}\right)$

$\therefore z_0 = 2.66\dot{6} \approx 2.67$



$\therefore Ans = .5 - .4962 = .0038$

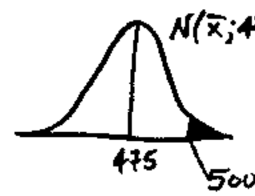
(8) (a) $N(z; 0, 1)$ $N(x; 475, \sigma)$



$\therefore -0.44 = \frac{431-475}{\sigma}$

$\therefore 100 = \sigma$

(b) $\therefore z_0 = \frac{500-475}{100/\sqrt{64}}$ $N\left(\bar{x}; 475, \frac{100}{\sqrt{64}}\right)$



$\therefore z_0 = 2.00$

$\therefore Ans = .5 - 0.4772$

$= 0.0228$

INFERENTIAL STATISTICS – TOPICS

INFERENCES INVOLVING ONE POPULATION

ESTIMATION OF THE MEAN μ $(n > 30)$

HYPOTHESIS TESTING OF THE MEAN μ $(n > 30)$

INFERENCES ABOUT THE BINOMIAL PROPORTION p

INFERENCES ABOUT μ $(n \leq 30)$: THE t – DISTRIBUTION

INFERENCES INVOLVING TWO POPULATIONS

INFERENCES FOR PAIRED DIFFERENCES (DEPENDENT SAMPLES)

INFERENCES FOR THE DIFFERENCE BETWEEN TWO MEANS

- ESTIMATION OF $(\mu_1 - \mu_2)$ FOR LARGE INDEPENDENT SAMPLES
- DIFFERENCE OF MEANS TESTS $(H_0: \mu_1 - \mu_2 = 0)$ AND THE F–TEST

INFERENCES FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

INFERENCES USING THE CHI-SQUARED DISTRIBUTION

MUTINOMIAL EXPERIMENTS

THE CHI-SQUARED GOODNESS OF FIT TEST

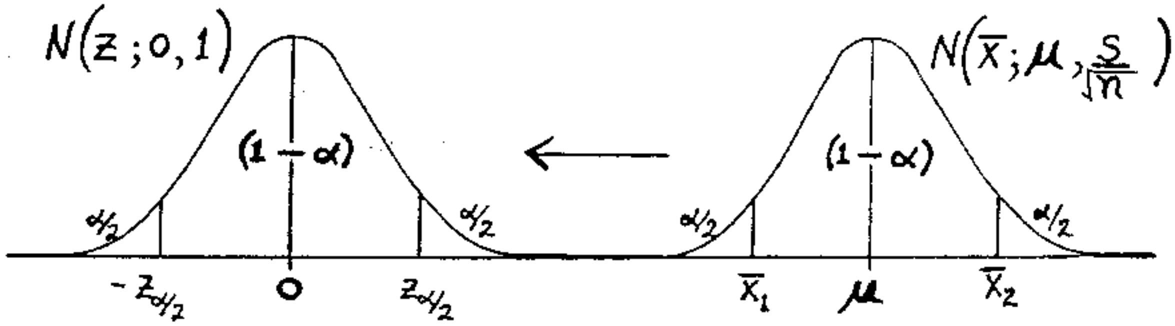
CONTINGENCY TABLES AND THE CHI-SQUARED DISTRIBUTION

THE INDEPENDENT CLASSIFICATIONS TEST

INTRODUCTION TO STATISTICAL INFERENCES: ESTIMATION

ESTIMATION OF μ WHEN $n > 30$

CONSIDER



$\therefore \quad -z_{\alpha/2} = \dfrac{\bar{X}_1 - \mu}{\frac{S}{\sqrt{n}}}$ AND $z_{\alpha/2} = \dfrac{\bar{X}_2 - \mu}{\frac{S}{\sqrt{n}}}$

$\therefore \quad \mu - z_{\alpha/2} \cdot \dfrac{S}{\sqrt{n}} = \bar{X}_1$ AND $\mu + z_{\alpha/2} \cdot \dfrac{S}{\sqrt{n}} = \bar{X}_2$

THEN USING <u>**SAMPLE MEAN**</u> $\bar{X}_0$ AS A <u>**POINT ESTIMATOR**</u> OF μ WE HAVE:

THE $(1-\alpha)$ % CONFIDENCE INTERVAL ESTIMATE OF μ
$$\bar{X}_0 \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

<u>NOTE:</u> $z_{\alpha/2} \cdot \dfrac{S}{\sqrt{n}}$ is called <u>THE BOUND ON THE ERROR OF ESTIMATION</u>, **E**

ESTIMATION OF μ WHEN $n > 30$ - ILLUSTRATIVE EXAMPLE

To estimate the true mean age, μ, of its viewers, a TV show selected a random sample of 36 of its viewers, yielding a sample mean of 28.7 years, with $S = 5.5$ years.

Construct a 95% CONFIDENCE INTERVAL ESTIMATE of μ

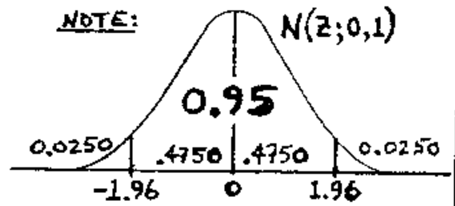
consider $\quad \bar{x}_0 \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$

$\therefore \; 28.7 \pm 1.96 \times \frac{5.5}{\sqrt{36}}$

$\therefore \; 28.7 \pm 1.8$

$\therefore$ 26.9 yrs. to 30.5 yrs.

NOTE:



NOTE: 26.9 is called the LOWER CONFIDENCE LIMIT (LCL), and 30.5 is called the UPPER CONFIDENCE LIMIT (UCL).

also, The Bound on the Error of Estimation, $E = 1.8$ yrs.

To determine the sample size required to estimate μ to within 1.5 years with 95% confidence:

consider $\quad 1.5 = E$

$\therefore \; 1.5 = 1.96 \times \frac{5.5}{\sqrt{n}}$

$\therefore \; \sqrt{n} = \frac{1.96 \times 5.5}{1.5}$

$\therefore \; n = \left[\frac{1.96 \times 5.5}{1.5}\right]^2$

$\therefore \; n \simeq 51.7$, hence use $n = 52$

To determine the sample size required to estimate μ to within $\frac{1}{10}$th of a standard deviation of the ages, with 95% confidence:

let $\quad \frac{1}{10}\cancel{\sigma} = 1.96 \times \frac{\cancel{\sigma}}{\sqrt{n}}$

$\therefore \; n = (1.96 \times 10)^2 \simeq 384.2$, hence use $n = 385$

ESTIMATION OF μ ($n > 30$) EXERCISES

① A consumer agency tested a random sample of 64 new cars, yielding a mean of 33.3 mpg., with a standard deviation of 4.0 mpg.
Construct the 95% confidence interval estimate of the true mean mpg. for these new cars.

② A company, concerned about its postal service, posted 49 letters from its Montréal office, to its Vancouver office. The letters averaged 3.6 days to reach the Vancouver office, with a standard deviation of 1.4 days. Hence, construct a 90% confidence interval estimate of the true average time that it takes the company's letters to go from Montréal to Vancouver.

③ A random sample of 81 Dawson students yielded the following data, in hours of study per week:

ΣX	ΣX^2
1215	21105

Based on this data, construct both a 95% and a 99% confidence interval estimate of the true average number of hours of study per week for Dawson students

④ To estimate the true mean hourly output of a machine, how many hours of output must be sampled if we want our estimate to be accurate to within 2 lbs. with 95% confidence if $S = 12$ lbs.

⑤ A sample of what size is needed to estimate the true population mean to within $\frac{1}{5}$th of a standard deviation with:
ⓐ 90% confidence? ⓑ 99% confidence?

⑥ A random sample of 100 Montreal dentists averaged 35.6 years of age, with a standard deviation of 16 years. Construct a 97% confidence interval estimate of the true mean age of Montreal dentists

(7) A random sample of 49 college textbooks cost, on average, \$55, with a standard deviation of \$8.

(a) Based on this sample, construct a 95% confidence interval estimate of the true mean cost of college textbooks.

(b) Also, what sample size would be required to estimate the true mean cost to within \$1 with 95% confidence?

(8) What sample size is required to estimate the true mean income of a population to within \$100 with 97% confidence, if σ is known to be \$1500?

SOLUTIONS

(1) $33.3 \pm 1.96 \times \frac{4}{\sqrt{64}} \Rightarrow 33.3 \pm 0.98 \Rightarrow 32.32$ mpg. to 34.28 mpg.

(2) $3.6 \pm 1.645 \times \frac{1.4}{\sqrt{49}} \Rightarrow 3.6 \pm 0.329 \Rightarrow 3.271$ days to 3.929 days

(3) $\bar{x} = 15$ & $s = 6$

(i) $15 \pm 1.96 \times 6/\sqrt{81} \Rightarrow 15 \pm 1.31 \Rightarrow 13.69$ hours to 16.31 hours

(ii) $15 \pm 2.575 \times 6/\sqrt{81} \Rightarrow 15 \pm 1.72 \Rightarrow 13.28$ hours to 16.72 hours

(4) $\therefore 2 = 1.96 \times \frac{12}{\sqrt{n}}$

$\therefore n = \left[\frac{1.96 \times 12}{2}\right]^2$

$\therefore n \simeq 138.3$

$\therefore$ use $n = 139$

(5) (a) $\therefore \frac{1}{5}\sigma = 1.645 \times \frac{\sigma}{\sqrt{n}}$

$\therefore n = (1.645 \times 5)^2$

$\therefore n \simeq 67.65$

$\therefore$ use $n = 68$

(b) $\therefore \frac{1}{5}\sigma = 2.575 \times \frac{\sigma}{\sqrt{n}}$

$\therefore n = (2.575 \times 5)^2$

$\therefore n \simeq 165.77$

$\therefore$ use $n = 166$

(6) $35.6 \pm 2.17 \times \frac{16}{\sqrt{100}} \Rightarrow 35.6 \pm 3.47 \Rightarrow 32.13$ years to 39.07 years

(7) (a) $55 \pm 1.96 \times \frac{8}{\sqrt{49}} \Rightarrow 55 \pm 2.24 \Rightarrow$ \$52.76 to \$57.24

(b) $\therefore \$1 = 1.96 \times \frac{8}{\sqrt{n}}$

$\therefore n = \left[\frac{1.96 \times 8}{1}\right]^2$

$\therefore n \simeq 245.86$

$\therefore$ use $n = 246$

(8) $\therefore 100 = 2.17 \times \frac{1500}{\sqrt{n}}$

$\therefore n = \left(\frac{2.17 \times 1500}{100}\right)^2$

$\therefore n \simeq 1059.5$

$\therefore$ use $n = 1060$

HYPOTHESIS TESTING OF μ WHEN $n>30$

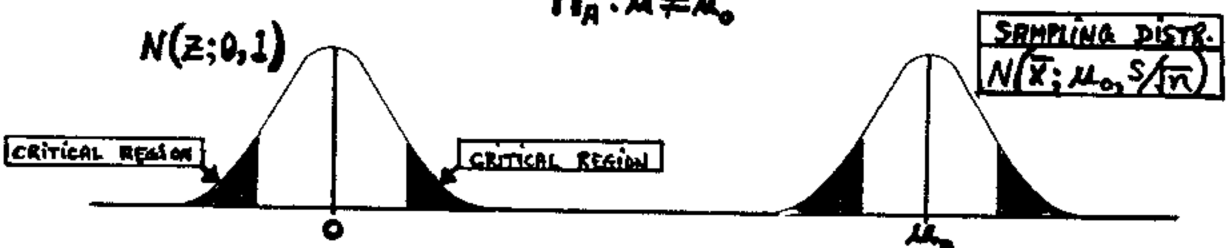
THE NULL HYPOTHESIS, H_0 : Gives the SAMPLING DISTRIBUTION OF $\overline{X}$, from the CLT.

THE ALTERNATE HYPOTHESIS, H_A : Gives the CRITICAL REGION (the rejection region for H_0) required for the Hypothesis Test.

THE TEST STATISTIC : The STANDARD NORMAL VALUE, Z_0, that corresponds to the SAMPLE MEAN, $\overline{X}_0$, and that is used to conduct the Hypothesis Test.

For example, to test THE HYPOTHESIS that μ IS EQUAL TO μ_0, we consider:

$$H_0 : \mu = \mu_0$$
$$H_A : \mu \neq \mu_0$$



THE TEST STATISTIC	←	SAMPLE MEAN
$Z_0 = \dfrac{\overline{X}_0 - \mu_0}{S/\sqrt{n}}$		$\overline{X}_0$

Then : If Z_0 belongs to the CRITICAL REGION, we REJECT H_0, and
If Z_0 does not belong to the CRITICAL REGION, we DO NOT REJECT H_0.

ERRORS IN HYPOTHESIS TESTING

	H_0 TRUE	H_0 NOT TRUE
REJECT H_0	TYPE I ERROR	NO ERROR
DO NOT REJECT H_0	NO ERROR	TYPE II ERROR

ALPHA, α = The probability of a TYPE I ERROR

BETA, β = The probability of a TYPE II ERROR

NOTE: ① α, which gives the SIZE OF THE CRITICAL REGION, is referred to as the LEVEL OF SIGNIFICANCE of the Hypothesis Test.

② The VALUE OF α is always decided upon before the Hypothesis Test is conducted.

HYPOTHESIS TESTING OF μ WHEN $n > 30$ - AN ILLUSTRATION

THE HYPOTHESIS

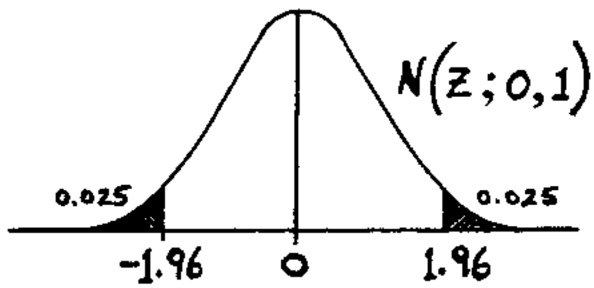
The mean height of Canadian men is 180 cm.

THE HYPOTHESIS TEST

To test the Hypothesis a random sample of **36** Canadian men's heights yielded $\bar{X}_o = 183$ cm and $S = 9$ cm. Based on this, conduct the Hypothesis Test at the $\alpha = 0.05$ level of significance, and make the proper INFERENCE below.

$H_o: \mu = 180$

$H_A: \mu \neq 180$



THE **TEST STATISTIC**

$$Z_o = \frac{\bar{X}_o - \mu_o}{\frac{S}{\sqrt{n}}} = \frac{183-180}{\frac{9}{\sqrt{36}}}$$

$\therefore Z_o = 2.00$

since $2.00 > 1.96$

$\therefore Z_o \in$ CRITICAL REGION

$\therefore$ REJECT H_o, AND

THE INFERENCE

The Hypothesis Test DOES NOT SUPPORT the Hypothesis

That is, the mean height of Canadian men is probably not 180 cm.

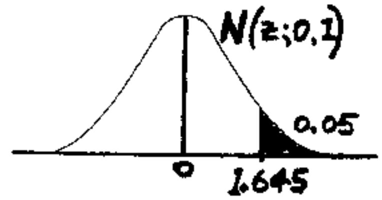
HYPOTHESIS TESTING OF μ WHEN $n > 30$ – ILLUSTRATIVE EXAMPLES

①
The Plant claims that the mean number of hours that Dawson students study per week is <u>more than</u> 15. A random sample of 49 Dawson students yielded a sample mean of 15.9 hours of study per week, with $s = 3.5$ hrs. Conduct a test of hypothesis using $\alpha = 0.05$ to determine whether the sample supports the claim or not.

$H_0: \mu = 15$

$H_A: \mu > 15$

$$z_0 = \frac{15.9 - 15}{3.5/\sqrt{49}}$$



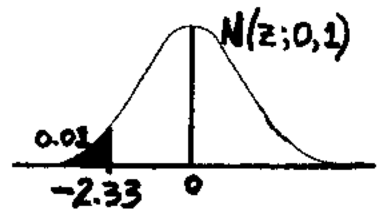
$\therefore z_0 = 1.80 > 1.645$, $\therefore$ <u>YES</u>, it supports the claim.

②
It is reported that the mean weight of NHL players is <u>at least</u> 195 lbs. A random sample of 36 NHL players had a mean weight of 190.2 lbs. with $s = 12.8$ lbs. Conduct a test of hypothesis using $\alpha = 0.01$ to determine whether the sample supports the report or not.

$H_0: \mu = 195$

$H_A: \mu < 195$

$$z_0 = \frac{190.2 - 195}{12.8/\sqrt{36}}$$



$\therefore z_0 = -2.25 \not< -2.33$, $\therefore$ <u>YES</u>, it supports the report.

* HYPOTHESIS TESTING OF μ WHEN $n > 30$ USING A p-VALUE

THE DEFINITION OF A p-VALUE (OR PROBABILITY-VALUE)

The probability that, under H_0, the test statistic is more extreme than its calculated value.

ILLUSTRATIVE EXAMPLES

(1) Using $\alpha = 0.05$, test the hypothesis that the mean grade in statistics is less than 70 if a random sample of 36 grades yielded $\bar{X}_0 = 66$ and $S = 11$.

$H_0: \mu = 70$

$H_a: \mu < 70$

$$z_0 = \frac{66-70}{11/\sqrt{36}}$$

$\therefore z_0 = -2.18$

Hence, the p-value $= P(z < -2.18)$

$= 0.5 - 0.4854$

$= 0.0146$

Then, since $0.0146 < 0.05$ we reject H_0 and support the hypothesis

(2) Using $\alpha = 0.02$, test the hypothesis that the mean age of Montrealers is 40 years if a random sample of 100 Montrealers yielded $\bar{X}_0 = 43.1$ yrs. with $S = 15.5$ yrs.

$H_0: \mu = 40$

$H_a: \mu \neq 40$

$$z_0 = \frac{43.1-40}{15.5/\sqrt{100}}$$

$\therefore z_0 = 2.00$

Hence, the p-value $= 2 \times P(z > 2.00)$

$= 2\,(0.0228)$

$= 0.0456$

Then, since $0.0456 \not< 0.02$ we do not reject H_0 and support the hypothesis.

NOTE: To ensure objectivity, the level of significance, α, should be decided before using a p-value test.

HYPOTHESIS TESTING OF μ WHEN $n > 30$ — EXERCISES

(1) A professor claims that the mean grade in Statistics is 80. A random sample of 36 statistics grades yielded a sample mean of 77.4, with $S = 8.9$. Does this sample support the claim? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(2) Is it true that the average taxi fare in Montreal is less than \$10? Conduct a Test of Hypothesis using $\alpha = 0.05$ if a random sample of 64 fares averaged \$9.55, with $S =$ \$2.05.

(3) A new car tire is advertised as having a mean life of more than 21000 miles. A random sample of 49 of these tires were tested, yielding a sample mean life of $\bar{X}_0 = 21183$ miles, with $S = 800$ miles. Does this result support the advertisement or not? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(4) A computer programmer claims that the average run time of his programs is 5.0 seconds. A random sample of 40 of his programs were run, averaging 5.3 seconds, with $S = 1.1$ seconds. Does this support the claim? Use $\alpha = 0.10$ and conduct a Test of Hypothesis.

(5) McDonald's claim that the average cost of an order in its restaurants is at most \$5.80. A random sample of 64 orders had a sample mean of \$6.30, with a standard deviation of \$2.10. Does this sample support the claim? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(6) A journalist wrote that NBA players averaged at least 80 inches in height. A random sample of 36 player's heights gave $\bar{X} = 78.8$ in. with $S = 3.2$ in. Does this support the journalist or not? Use $\alpha = 0.01$ and conduct a Test of Hypothesis.

(7) The mean batting average for a random sample of 40 major league baseball players was 0.259, with $S = 0.032$. Based on this, can we conclude that the mean batting average for major league baseball players is greater than 0.250? Use $\alpha = 0.025$ and conduct a Test of Hypothesis.

(8) It is believed that Dawson students average less than 40 minutes to travel to school from home. A random sample of 49 Dawson students averaged 37.5 minutes to travel to school from home, with $S = 12$ minutes. Conduct a Test of Hypothesis to see if this sample supports the belief. Use (a) $\alpha = 0.05$ and, (b) $\alpha = 0.10$

(9) A newspaper reports that the average annual income of Québec doctors is at least \$130000. A random sample of 64 Québec doctors had a mean annual income of \$128500, with a standard deviation of \$9000. Conduct a Test of Hypothesis using $\alpha = 0.10$ to determine whether the sample result supports the report or not.

(10) A supermarket advertises that its ground beef contains at most 15% fat, on the average. A government agency finds that the mean fat content for a random sample of 36 of their ground beef packages is 15.9%, with $S = 2.8\%$. Does this result support the advertisement? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(11) A group of 100 fifth graders completed a reading test with a mean score of 103 and a standard deviation of 15. Does this indicate that the the group has superior reading skills, if the national mean score for the test is 100? Conduct a Test of Hypothesis using $\alpha = 0.01$.

(12) To respect anti-pollution guidelines, a refiner controls his refining process by taking a 100 item sample on an hourly basis. Before altering his process, he wants to be 95% certain that the process is different from his target of 200 ppm. of sulphur. If the 1 PM sample gave a mean of 202.1 ppm. of sulphur and a standard deviation of 9.9 ppm., should he alter his process?

(13) A medical journal reports that the mean hemoglobin count for adult males is 160 g/L. Is this report supported, at the $\alpha = 0.10$ level of significance, by a random sample of 100 adult males whose average hemoglobin count was 158.3 g/L., with $S = 9.4$ g/L.? Conduct a Test of Hypothesis.

(14) It is believed that, on average, students take no more than 30 minutes to complete a certain diagnostic test. Test this belief, if a random sample of 50 students who took the test averaged 31.3 minutes to complete it, with $S = 4.6$ minutes. Use $\alpha = 0.015$.

(15) A random sample of 40 patients at a large hospital had an average age of 68.9 years, with $S = 10.2$ yrs. Based on this sample, could the hospital announce that the mean age of its patients is more than 65 years? Conduct a Test of Hypothesis using $\alpha = 0.005$.

(16) Ford advertises that their new cars average more than 35 mpg. A consumer agency tested a random sample of these cars, obtaining $\bar{x} = 35.9$ mpg, with $S = 4.0$ mpg. Conduct a Test of Hypothesis, using $\alpha = 0.05$ to determine whether the data supports the ad, if the sample size was: (a) 64 and (b) 36.

(17) A supplier of bottled water claims that its water source contains no more than 40 mg/L. of sodium, on the average. To test this claim, 50 bottles of this water were selected at random and analyzed for sodium content, giving $\bar{x} = 41.2$ mg/L, with $S = 3.3$ mg/L. Does this refute the supplier? Conduct a Test of Hypothesis using $\alpha = 0.01$.

(18) It is reported that the mean diameter of maple trees in a mature sugarbush is at least 30 cm. A random sample of 81 of the trees yielded $\bar{x} = 28.6$ cm, with $S = 5.2$ cm. Does this contradict the report? Conduct a Test of Hypothesis with $\alpha = 0.01$.

(19) A soft drink manufacturer produces cans of soda with a mean of 25 mg. of caffeine per can and standard deviation of 3 mg./can. The production process is adjusted when the mean caffeine per can becomes significantly more than 25 mg. Quality control consists of regular random samples of 36 cans each, that are analyzed for caffeine content, followed by a Test of Hypothesis using $\alpha = 0.01$. Above what sample mean is the production process adjusted?

(20) To test the hypothesis that μ is more than 75, a random sample of size 49 is selected, yielding $S = 8$. What is the value of α if it is decided to reject H_0 when the sample mean exceeds 77?

*(21) Use $\alpha = 0.05$ and the p-value approach to test the hypothesis that the mean grade in statistics is more than 75 if a random sample of 49 grades yielded $\bar{x}_0 = 78$ and $S = 13$.

*(22) Use $\alpha = 0.05$ and the p-value approach to test the hypothesis that the mean age of professors is 55 years if a random sample of 100 professors yielded $\bar{x}_0 = 57.2$ years with $S = 10.5$ years.

HYPOTHESIS TESTING OF μ WHEN $n > 30$ – EXERCISE SOLUTIONS

(1) $H_0: \mu = 80$
$H_a: \mu \neq 80$
$z_0 = \frac{77.4 - 80}{8.9/\sqrt{36}} = -1.75 \not< -1.96$
∴ YES, supports claim

(2) $H_0: \mu = 10$
$H_a: \mu < 10$
$z_0 = \frac{9.55 - 10}{2.05/\sqrt{64}} = -1.76 < -1.645$
∴ YES, it's true

(3) $H_0: \mu = 21000$
$H_a: \mu > 21000$
$z_0 = \frac{21183 - 21000}{800/\sqrt{49}} = 1.60 \not> 1.645$
∴ No, doesn't support

(4) $H_0: \mu = 5.0$
$H_a: \mu \neq 5.0$
$z_0 = \frac{5.3 - 5.0}{1.1/\sqrt{40}} = 1.72 > 1.645$
∴ NO, does not support

(5) $H_0: \mu = 5.80$
$H_a: \mu > 5.80$
$z_0 = \frac{6.30 - 5.80}{2.10/\sqrt{64}} = 1.90 > 1.645$
∴ NO, does not support

(6) $H_0: \mu = 80$
$H_a: \mu < 80$
$z_0 = \frac{78.8 - 80}{3.2/\sqrt{36}} = -2.25 \not< -2.33$
∴ YES, does support

(7) $H_0: \mu = 0.250$
$H_a: \mu > 0.250$
$z_0 = \frac{0.259 - 0.250}{0.032/\sqrt{40}} = 1.78 \not> 1.96$
∴ No, can not conclude

(8) $H_0: \mu = 40$
$H_a: \mu < 40$
$z_0 = \frac{37.5 - 40}{12/\sqrt{49}} = -1.46$
(a) $-1.46 \not< -1.645$, No, does not
(b) $-1.46 < -1.282$, YES, does support

(9) $H_0: \mu = 130000$
$H_a: \mu < 130000$
$z_0 = \frac{128500 - 130000}{9000/\sqrt{64}} = -1.33 < -1.282$
∴ No, does not support

(10) $H_0: \mu = 15$
$H_a: \mu > 15$
$z_0 = \frac{15.9 - 15}{2.8/\sqrt{36}} = 1.93 > 1.645$
∴ No, does not support

(11) $H_0: \mu = 100$
$H_a: \mu > 100$
$z_0 = \frac{103 - 100}{15/\sqrt{100}} = 2.00 \not> 2.33$
∴ No, not superior

(12) $H_0: \mu = 200$
$H_a: \mu \neq 200$
$z_0 = \frac{202.1 - 200}{9.9/\sqrt{100}} = 2.12 > 1.96$
∴ YES, should alter process

(13) $H_0: \mu = 160$
$H_a: \mu \neq 160$
$z_0 = \frac{158.3 - 160}{9.4/\sqrt{100}} = -1.81 < -1.645$
∴ NO, does not support

(14) $H_0: \mu = 30$
$H_a: \mu > 30$
$z_0 = \frac{31.3 - 30}{4.6/\sqrt{50}} = 2.00 \not> 2.17$
∴ YES, supports belief

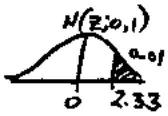
(15) $H_0: \mu = 65$
$H_a: \mu > 65$
$z_0 = \frac{68.9 - 66}{10.2/\sqrt{40}} = 2.42 \not> 2.575$
∴ NO, could not announce

(16) $H_0: \mu = 35$, $H_a: \mu > 35$
(a) $z_0 = \frac{35.9 - 35}{4/\sqrt{64}} = 1.80 > 1.645$, ∴ YES
(b) $z_0 = \frac{35.9 - 35}{4/\sqrt{36}} = 1.35 \not> 1.645$, ∴ NO

(17) $H_0: \mu = 40$
$H_a: \mu > 40$
$z_0 = \frac{41.2 - 40}{3.3/\sqrt{50}} = 2.57 > 2.33$
∴ YES, this refutes the supplier

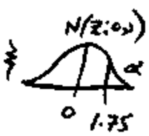
(18) $H_0: \mu = 30$
$H_a: \mu < 30$
$z_0 = \frac{28.6 - 30}{5.2/\sqrt{81}} = -2.42 < -2.33$
∴ YES, this contradicts report

(19) Consider
$H_0: \mu = 25$
$H_a: \mu > 25$ and



and solve for $\bar{x}$:
$$\frac{\bar{x} - 25}{3/\sqrt{36}} = 2.33$$
∴ $\bar{x} = 26.165$ mg/can

(20) Consider
$z_0 = \frac{77 - 75}{8/\sqrt{49}} = 1.75$;



∴ $\alpha = 0.5 - 0.4599$
∴ $\alpha = 0.0401$

*(21) $H_0: \mu = 75$
$H_a: \mu > 75$
$z_0 = \frac{78 - 75}{13/\sqrt{49}} = 1.62$
∴ p-value $= 0.0526 \not< 0.05$
∴ No, does not support

*(22) $H_0: \mu = 55$
$H_a: \mu \neq 55$
$z_0 = \frac{57.2 - 55}{10.5/\sqrt{100}} = 2.10$
∴ p-value $= 2(0.0179)$
$= 0.0358 < 0.05$
∴ No, does not support

INFERENCES ABOUT THE BINOMIAL PROPORTION p

IF $B(x;n,p)$ CAN BE APPROXIMATED* BY $N(x;np,\sqrt{npq})$

THEN $N(x;np,\sqrt{npq}) \longrightarrow N\left(\frac{x}{n};p,\sqrt{\frac{pq}{n}}\right)$ AND

WITH THE NUMBER OF SUCCESSES, X_0, IN A SAMPLE OF n BINOMIAL TRIALS, WE HAVE

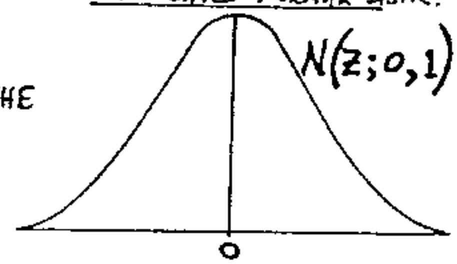
THE TEST OF HYPOTHESIS FOR $H_0: p = p_0$ $(\therefore q_0 = 1 - p_0)$

THE TEST STATISTIC

$$z_0 = \frac{\frac{X_0}{n} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

HAS THE

STANDARD NORMAL DISTR.



AND

THE $(1-\alpha)\%$ CONFIDENCE INTERVAL ESTIMATE OF p

USING $\hat{p} = \frac{X_0}{n}$ AS A POINT ESTIMATOR OF p $(\therefore \hat{q} = 1 - \hat{p})$:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

NOTE: $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$, *the* ERROR OF ESTIMATION, E, *is also called* THE MARGIN OF ERROR *or, simply,* THE SAMPLING ERROR.

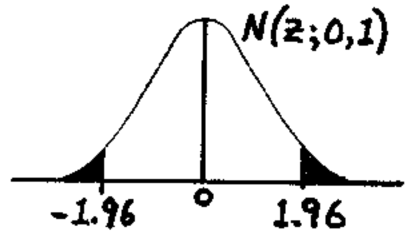
* WHERE np AND nq ARE ≥ 5

INFERENCES ABOUT THE BINOMIAL PROPORTION p – ILLUSTRATIVE EXAMPLES

① Using $\alpha = 0.05$, test The Hypothesis that 60 % of CEGEP students are female, if a random sample of 250 CEGEP students included 136 females.

$$H_0:\ p = 0.60$$
$$H_A:\ p \neq 0.60$$

$$Z_o = \frac{\frac{x_o}{n} - p_o}{\sqrt{\frac{p_o q_o}{n}}} = \frac{\frac{136}{250} - 0.60}{\sqrt{\frac{0.60(1-0.60)}{250}}}$$



$\therefore Z_o = -1.81 \not< -1.96$, $\therefore$ YES, The Hypothesis is supported.

② Construct a 95 % CONFIDENCE INTERVAL ESTIMATE OF p if 100 trials of a binomial experiment yielded 12 successes.

consider $\frac{12}{100} \pm 1.96 \sqrt{\frac{\frac{12}{100}\left(1-\frac{12}{100}\right)}{100}}$

$\therefore\ 0.12 \pm 0.064$

$\therefore\ 0.056$ to 0.184

NOTE: Here the sampling error is ± 0.064 or $\pm 6.4\%$

③ If a preliminary (small) binomial experiment found that $p \approx 0.15$, then how large a sample is required to estimate p to within 2 % with 95 % confidence?

consider $0.02 = 1.96 \sqrt{\frac{0.15(1-0.15)}{n}}$

$\therefore\ n = \left(\frac{1.96}{0.02}\right)^2 [0.15(1-0.15)]$

$\therefore\ n = 1224.5$, hence $n = 1225$ is required.

④ How large a sample is required to estimate p to within 2 % with 95 % confidence, if no preliminary value of p is available?

consider $0.02 = 1.96 \sqrt{\frac{0.50(1-0.50)}{n}}$

$\therefore\ n = \left(\frac{1.96}{0.02}\right)^2 [0.50(1-0.50)]$

$\therefore\ n = 2401$

INFERENCES ABOUT THE BINOMIAL PROPORTION p — EXERCISES

1. The Gazette reports that 70% of Montrealers are bilingual. It is known that a random sample of 600 Montrealers included 396 who were bilingual. Does this sample support the report? Conduct a Test of Hypothesis using $\alpha = 0.05$.

2. It is believed that each human birth is equally-likely to be male or female. If a random sample of 100 recent births included only 41 females, would this support the belief or not? Conduct a Test of Hypothesis with $\alpha = 0.05$.

3. If 150 digits selected at random from a random digit table included 71 digits less than 4, can we conclude that 40% of the digits in the table are less than 4? Use $\alpha = 0.05$ to conduct a Test of Hypothesis.

4. A coin is tossed 400 times. Conduct a Test of Hypothesis using $\alpha = 0.05$ to determine whether the coin is fair or not if:

   (a) 219 heads are observed, and (b) 220 heads are observed.

5. Are voters equally divided over an issue if a poll of 1000 of them included 528 for and 472 against? Conduct a Test of Hypothesis with $\alpha = 0.10$.

6. Can a salesman claim that he is successful in more than 1/4 of his sales contacts, if he was successful in 122 of a sample of 432 contacts? Conduct a Test of Hypothesis using $\alpha = 0.10$.

7. A political party claims that more than 2/3 of all citizens support their platform. A survey of 4050 citizens included 2754 who supported their platform. Does this survey agree with the claim? Conduct a Test of Hypothesis with $\alpha = 0.025$.

8. A baseball player's lifetime batting average is 0.300. If in his next 60 at bats he has only a 0.217 batting average, would this indicate a significant decline in his batting average? Conduct a Test of Hypothesis with $\alpha = 0.05$.

9. A new drug is found to be effective 371 times in 400 trial applications. Based on this, can a researcher state, with 95% confidence, that this new drug is effective in more than 90% of its applications?

(10) Last year, a TV show had a 30% share of the viewers in its time slot. This year, a random survey of 2100 viewers in the show's time slot included 681 who had watched the show. Is this a significant increase in the share of viewers? Conduct a Test of Hypothesis using $\alpha = 0.01$.

(11) The Apple Growers Association reports that last year in Québec, 15% of the apple trees were damaged by frost. This year, only 13% of the trees in a random sample of 2000 trees were damaged by frost. Is this significantly less than last year? Conduct a Test of Hypothesis using $\alpha = 0.01$.

(12) A candidate for office believes that he will receive at least 40% of the votes. An advance poll of 2400 voters indicated that 918 will vote for the candidate. Does this poll support the belief? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(13) A new medical treatment is claimed to work in more than 70% of its trials.

(a) If the treatment worked 1500 times in 2100 trials, would the claim be supported at the $\alpha = 0.05$ level of significance?

(b) What is the minimum number of times that the treatment would have to work in 2100 trials in order to support the claim using $\alpha = 0.05$?

(14) A random sample of 192 emergency cases at a Montreal hospital included 56 serious ones requiring admittance.

(a) Conduct a Test of Hypothesis using $\alpha = 0.05$ of the report that at most 25% of this hospital's emergency cases are serious ones.

(b) Construct a 95% confidence interval estimate of the true percentage of this hospital's emergency cases that are serious ones.

(15) Construct a 95% confidence interval estimate for the true proportion of seeds that germinate, if in a random sample of 400 seeds that were planted, 340 of them germinated.

(16) A random sample of 100 college students included 17 who wore eyeglasses. Based on this, construct a 90% confidence interval estimate for the true proportion of college students who wear eyeglasses.

(17) In a random sample of 80 students who entered CEGEP 2 years ago, 33 graduated on time. Construct the 99% confidence interval estimate for the true proportion of CEGEP students who graduate on time.

(18) A random sample of 500 teenagers included 120 who smoked cigarettes. Based on this, construct a 97% confidence interval estimate for the true proportion of teenagers who smoke cigarettes.

(19) Construct a 98% confidence interval estimate for the proportion of fire alarms in Montreal that prove to be false alarms, if a recent random sample of 60 Montreal fire alarms included 21 false alarms.

(20) A preliminary poll indicated that 65% of ~~the~~ voters preferred a certain candidate. Based on this, how large a sample would be required to estimate the true percentage of voters who prefer the candidate to within 1% with 99% confidence?

(21) A preliminary study indicated that 12% of men were colour-blind. Accordingly, how large a sample of men would be required to estimate the true percentage of men who are colour-blind to within 2% with 95% confidence?

(22) It is thought that about 70% of Montreal commuters use the Metro. Based on this estimate, how large a sample of Montreal commuters would be required to estimate the true percentage who use the Metro to within 3% with 97% confidence?

(23) Historically, about 80% of English Montrealers have read The Gazette daily. Hence, how large a sample is required to estimate the current percentage of English Montrealers who read The Gazette daily to within 4% with 90% confidence.

(24) How large a sample is required to estimate a binomial proportion p to within 5% with 90% confidence, if no preliminary value of p is available?

(25) How large a sample must be polled to estimate the percentage of citizens who support a new government policy to within 2% with 98% confidence?

INFERENCES ABOUT THE BINOMIAL PROPORTION p − EXECISE SOLUTIONS

(1) $H_0: p = 0.70$, $H_a: p \neq 0.70$

$$z_0 = \frac{\frac{396}{600} - 0.70}{\sqrt{\frac{0.70(1-0.70)}{600}}} = -2.14 < -1.96$$

∴ NO, does not support

(2) $H_0: p = 0.50$, $H_A: p \neq 0.50$

$$z_0 = \frac{\frac{41}{100} - 0.50}{\sqrt{\frac{0.50(1-0.50)}{100}}} = -1.80 \nless -1.96$$

∴ YES, supports belief

(3) $H_0: p = 0.40$, $H_A: p \neq 0.40$

$$z_0 = \frac{\frac{31}{150} - 0.40}{\sqrt{\frac{0.40(1-0.40)}{150}}} = 1.83 \ngtr 1.96$$

∴ YES, we can conclude

(4) $H_0: p = 0.50$, $H_A: p \neq 0.50$

(a) $z_0 = \frac{\frac{219}{400} - 0.50}{\sqrt{\frac{0.5(1-0.5)}{400}}} = 1.90 \ngtr 1.96$, ∴ YES fair

(b) $z_0 = \frac{\frac{220}{400} - 0.50}{\sqrt{\frac{0.5(1-0.5)}{400}}} = 2.00 > 1.96$, ∴ No. not fair

(5) $H_0: p = 0.50$, $H_A: p \neq 0.50$

$$z_0 = \frac{\frac{528}{1000} - 0.50}{\sqrt{\frac{0.50(1-0.50)}{1000}}} = 1.77 > 1.645$$

∴ No, not equally divided

(6) $H_0: p = 0.25$, $H_A: p > 0.25$

$$z_0 = \frac{\frac{122}{432} - 0.25}{\sqrt{\frac{0.25(1-0.25)}{432}}} = 1.56 > 1.282$$

∴ YES, he can claim

(7) $H_0: p = 2/3$, $H_A: p > 2/3$

$$z_0 = \frac{\frac{2754}{4050} - 2/3}{\sqrt{\frac{2/3(1-2/3)}{4050}}} = 1.80 \ngtr 1.96$$

∴ NO, can not claim

(8) $H_0: p = 0.300$, $H_A: p < 0.300$

$$z_0 = \frac{0.217 - 0.300}{\sqrt{\frac{0.30(1-0.30)}{60}}} = -1.40 \nless -1.645$$

∴ NO, not a significant decline

(9) $H_0: p = 0.90$, $H_A: p > 0.90$

$$z_0 = \frac{\frac{371}{400} - 0.90}{\sqrt{\frac{0.90(1-0.90)}{400}}} = 1.83 > 1.645$$

∴ YES, he can state

(10) $H_0: p = 0.30$, $H_A: p > 0.30$

$$z_0 = \frac{\frac{681}{2100} - 0.30}{\sqrt{\frac{0.30(1-0.30)}{2100}}} = 2.43 > 2.33$$

∴ YES, is a significant increase

(11) $H_0: p = 0.15$, $H_A: p < 0.15$

$$z_0 = \frac{0.13 - 0.15}{\sqrt{\frac{0.15(1-0.15)}{2000}}} = -2.50 < -2.33$$

∴ YES, significantly less

(12) $H_0: p = 0.40$, $H_A: p < 0.40$

$$z_0 = \frac{\frac{918}{2400} - 0.40}{\sqrt{\frac{0.40(1-0.40)}{2400}}} = -1.75 < -1.645$$

∴ NO, does not support

(13) (a) $H_0: p = 0.70$, $H_A: p > 0.70$

$z_0 = \frac{\frac{1500}{2100} - 0.70}{\sqrt{\frac{0.70(1-0.70)}{2100}}} = 1.43 \ngtr 1.645$, ∴ NO

(b) $\frac{\frac{x_0}{2100} - 0.70}{\sqrt{\frac{0.70(1-0.70)}{2100}}} = 1.645$, ∴ $x_0 \approx 1504.5$ & $x_0 = 1505$

(14) (a) $H_0: p = 0.25$, $H_A: p > 0.25$

$$z_0 = \frac{\frac{56}{192} - 0.25}{\sqrt{\frac{0.25(1-0.25)}{192}}} = 1.33 \ngtr 1.645$$

∴ YES, supports the report

(14) (b) $\frac{56}{192} \pm 1.96\sqrt{\frac{\frac{56}{192}(1-\frac{56}{192})}{192}}$

∴ 0.292 ± 0.064

ie 0.228 to 0.356

(15) $\frac{340}{400} \pm 1.96\sqrt{\frac{\frac{340}{400}(1-\frac{340}{400})}{400}}$ ∴ 0.85 ± 0.035, 0.815 to 0.885

(16) $\frac{17}{100} \pm 1.645\sqrt{\frac{\frac{17}{100}(1-\frac{17}{100})}{100}}$ ∴ 0.17 ± 0.06, 0.11 to 0.23

(17) $\frac{33}{80} \pm 2.575\sqrt{\frac{\frac{33}{80}(1-\frac{33}{80})}{80}}$ ∴ 0.4125 ± 0.1417, 0.2708 to 0.5542

(18) $\frac{120}{500} \pm 2.17\sqrt{\frac{\frac{120}{500}(1-\frac{120}{500})}{500}}$ ∴ 0.24 ± 0.041, 0.199 to 0.281

(19) $\frac{21}{60} \pm 2.33\sqrt{\frac{\frac{21}{60}(1-\frac{21}{60})}{60}}$ 0.35 ± 0.143, 0.207 to 0.493

(20) $0.01 = 2.575\sqrt{\frac{0.65(1-0.65)}{n}}$

∴ $n = \left(\frac{2.575}{0.01}\right)^2[0.65(1-0.65)]$

∴ $n \approx 15084.6$, hence 15085

(21) $0.02 = 1.96\sqrt{\frac{0.12(1-0.12)}{n}}$

∴ $n = \left(\frac{1.96}{0.02}\right)^2[0.12(1-0.12)]$

∴ $n \approx 1014.2$, hence 1015

(22) $0.03 = 2.17\sqrt{\frac{0.70(1-0.70)}{n}}$

∴ $n = \left(\frac{2.17}{0.03}\right)^2[0.70(1-0.70)]$

∴ $n \approx 1098.7$, hence 1099

(23) $0.04 = 1.645\sqrt{\frac{0.80(1-0.80)}{n}}$

∴ $n = \left(\frac{1.645}{0.04}\right)^2[0.80(1-0.80)]$

∴ $n \approx 270.6$, hence 271

(24) $0.05 = 1.645\sqrt{\frac{0.5(1-0.5)}{n}}$

∴ $n = \left(\frac{1.645}{0.05}\right)^2[0.5(1-0.5)]$

∴ $n \approx 270.6$, hence 271

(25) $0.02 = 2.33\sqrt{\frac{0.5(1-0.5)}{n}}$

∴ $n = \left(\frac{2.33}{0.02}\right)^2[0.5(1-0.5)]$

∴ $n \approx 3393.1$, hence 3394

INFERENCES ABOUT μ $(n \leq 30)$: THE t DISTRIBUTION

IF THE POPULATION IS NORMAL

THEN

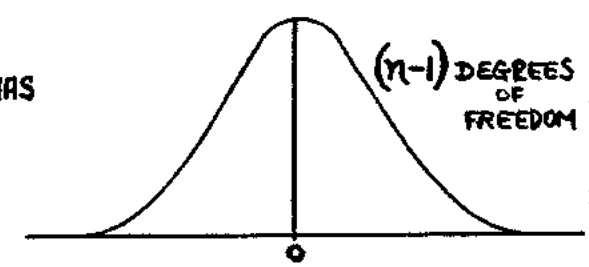
HYPOTHESIS TESTING OF μ WHEN $n \leq 30$

THE TEST STATISTIC

$$t_0 = \frac{\bar{X}_0 - \mu_0}{\frac{S}{\sqrt{n}}}$$

HAS

THE t DISTRIBUTION



AND

ESTIMATION OF μ WHEN $n \leq 30$

THE $(1-\alpha)\%$ CONFIDENCE INTERVAL ESTIMATE OF μ

IS GIVEN BY

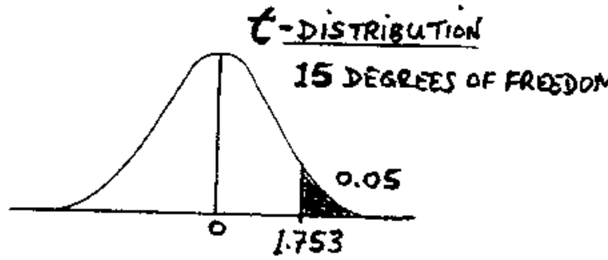
$$\bar{X}_0 \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

INFERENCES ABOUT μ ($n \leq 30$): THE t-DISTRIBUTION – ILLUSTRATIVE EXAMPLES

①

Air Canada reports that the mean number of passengers on its Montreal to Toronto flights is more than 120. A random sample of 16 recent such flights yielded a sample mean $\bar{X}_o = 123.3$ passengers, with $S = 7.7$ passengers. Does this sample support the report or not? Conduct a Test of Hypothesis using $\alpha = 0.05$.

$H_o: \mu = 120$
$H_A: \mu > 120$



$$t_o = \frac{\bar{X}_o - \mu_o}{S/\sqrt{n}} = \frac{123.3 - 120}{7.7/\sqrt{16}}$$

$\therefore t_o = 1.71 \not> 1.753$, $\therefore$ No, it does not support the report.

②

To estimate the protein content (in grams) of a popular nutrition bar, 9 bars were selected at random, and their protein contents were:

8.1, 8.5, 7.7, 7.9, 8.3, 8.0, 8.6, 8.9, 8.7

Construct a 90% CONFIDENCE INTERVAL ESTIMATE for the true mean protein content of the bars.

First, we calculate $\bar{X}_o = 8.30$ grams and $S = 0.40$ grams, then consider

$$\bar{X}_o \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

, where $t_{\alpha/2} = t_{0.05} = 1.860$ with $(9-1) = 8$ DEGREES OF FREEDOM

$\therefore 8.30 \pm 1.860 \frac{0.40}{\sqrt{9}}$

$\therefore 8.30 \pm 0.25$

$\therefore$ 8.05 grams to 8.55 grams

INFERENCES ABOUT μ ($n \leq 30$): THE t-DISTRIBUTION – EXERCISE SOLUTIONS

(1) $H_0: \mu = 25$, $H_A: \mu \neq 25$

$t_0 = \dfrac{25.8 - 25}{1.2/\sqrt{9}} = 2.00 \not> 2.306$

$\therefore$ YES, supports ad

(2) $H_0: \mu = 3$, $H_A: \mu \neq 3$

$t_0 = \dfrac{3.18 - 3}{0.21/\sqrt{10}} = 2.71 > 2.262$

$\therefore$ No, not accurate

(3) $H_0: \mu = 100$, $H_A: \mu > 100$

$t_0 = \dfrac{104 - 100}{10/\sqrt{25}} = 2.00 \not> 2.492$

$\therefore$ No, not higher

(4) $H_0: \mu = 500$, $H_A: \mu < 500$

$t_0 = \dfrac{488 - 500}{32/\sqrt{16}} = -1.50 \not< -1.341$

$\therefore$ No, does not support

(5) $50 \pm 2.131 \times \dfrac{10}{\sqrt{16}}$

$\therefore 50 \pm 5.3275$

$\therefore$ 44.6725 min. to 55.3275 min

(6) $0.53 \pm 2.571 \cdot \dfrac{0.0559}{\sqrt{6}}$

$\therefore 0.53 \pm 0.059$

$\therefore$ 0.471 to 0.589 karats

(7) $\bar{x}_0 = \dfrac{521.25}{25} = \$20.85$

$s = \sqrt{\dfrac{10989.56 - \frac{(521.25)^2}{25}}{24}} = 2.25$

(a) $20.85 \pm 2.064 \times \dfrac{2.25}{\sqrt{25}}$

$\therefore 20.85 \pm 0.929$

$\therefore$ \$19.92 to \$21.78

(b) $20.85 \pm 1.711 \times \dfrac{2.25}{\sqrt{25}}$

$\therefore 20.85 \pm 0.77$

$\therefore$ \$20.08 to \$21.62

(8) $\bar{x}_0 = 9.75$ & $s = 4.83$

$9.75 \pm 3.499 \times \dfrac{4.83}{\sqrt{8}}$

$\therefore 9.75 \pm 5.98$

$\therefore$ 3.77 hrs. to 15.73 hrs.

(9) $H_0: \mu = 50$, $H_A: \mu > 50$

$t_0 = \dfrac{52.3 - 50}{4/\sqrt{9}} = 1.73 \not> 1.860$

$\therefore$ No, does not support

(10) $H_0: \mu = 8$, $H_A: \mu \neq 8$

$t_0 = \dfrac{7.5 - 8}{0.8/\sqrt{16}} = -2.50 < -2.131$

$\therefore$ No, not properly

(11) $\bar{x}_0 = 55$ & $s = 4.42$

$55 \pm 1.860 \times \dfrac{4.42}{\sqrt{9}}$

$\therefore 55 \pm 2.74$

$\therefore$ 52.26 yrs. to 57.74 yrs.

(12) $\bar{x}_0 = 43.75$, $s = 2.5$; $H_0: \mu = 45$, $H_A: \mu < 45$

$t_0 = \dfrac{43.75 - 45}{2.5/\sqrt{16}} = -2.00 < -1.753$

$\therefore$ YES, it is true

(13) (a) $H_0: \mu = 35$, $H_A: \mu < 35$

$t_0 = \dfrac{33.86 - 35}{3/\sqrt{9}} = -1.14 \not< -1.860$

$\therefore$ YES, supports the ad

(13) (b) $33.86 \pm 2.306 \times \dfrac{3}{\sqrt{9}}$

$\therefore 33.86 \pm 2.306$

$\therefore$ 31.55 mpg to 36.17 mpg.

(14) $H_0: \mu = 18$, $H_A: \mu > 18$

$t_0 = \dfrac{18.4 - 18}{1.5/\sqrt{25}} = 1.33 > 1.318$

$\therefore$ YES, supports hypothesis

(15) $\bar{x}_0 = 4.34$ & $s = 0.88$

$4.34 \pm 3.707 \times \dfrac{(0.88)}{\sqrt{7}}$

$\therefore 4.34 \pm 1.23$

$\therefore$ 3.11 to 5.57 ppm of CO

(16) $\bar{x}_0 = 498$ & $s = 2.828$

$\rightarrow 498 \pm 2.131 \times \dfrac{2.828}{\sqrt{16}}$

$\therefore 498 \pm 1.51$

$\therefore$ 496.49 mm. to 499.51 mm.

(17) $\bar{x}_0 = 12.00$ & $s = 4.45$

$\Rightarrow 12 \pm 2.821 \times \dfrac{4.45}{\sqrt{10}}$

$\therefore 12 \pm 3.97$

$\therefore$ 8.03 hrs. to 15.97 hrs.

INFERENCES FOR PAIRED DIFFERENCES (DEPENDENT SAMPLES)

Consider a random sample of n data pairs.
If the **PAIRED DIFFERENCES**, d, between the pairs are normally distributed with MEAN μ_d, then:

$$t = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}}$$ has the **t-DISTRIBUTION** with $(n-1)$ DEGREES OF FREEDOM

where $\bar{d} = \dfrac{\sum d}{n}$ and $S_d = \sqrt{\dfrac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$

hence

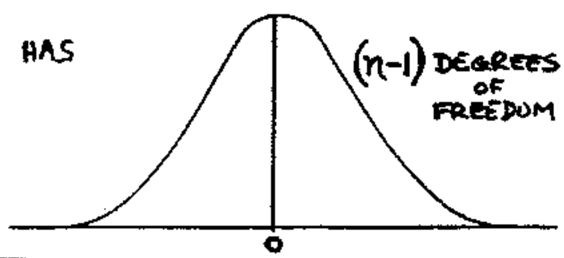
THE TEST OF HYPOTHSIS FOR $H_0: \mu_d = \mu_0$

THE TEST STATISTIC

$$t_0 = \frac{\bar{d} - \mu_0}{S_d/\sqrt{n}}$$

THE t DISTRIBUTION

HAS



and

THE $(1-\alpha)$% CONFIDENCE INTERVAL ESTIMATE OF μ_d

$$\bar{d} \pm t_{\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$$

INFERENCES FOR PAIRED DIFFERENCES (DEPENDENT SAMPLES) - AN ILLUSTRATIVE EXAMPLE

Consider the data pairs for 10 students' exam results given below:

STUDENT	MID-TERM EXAM M	FINAL EXAM F	$d = M - F$
1	65	64	1
2	58	56	2
3	71	69	2
4	85	82	3
5	90	83	7
6	75	76	-1
7	88	88	0
8	66	60	6
9	50	45	5
10	95	91	4

$\therefore \Sigma d = 29, \ \Sigma d^2 = 145$

$\therefore \bar{d} = \frac{\Sigma d}{n} = \frac{29}{10} = 2.9$

and

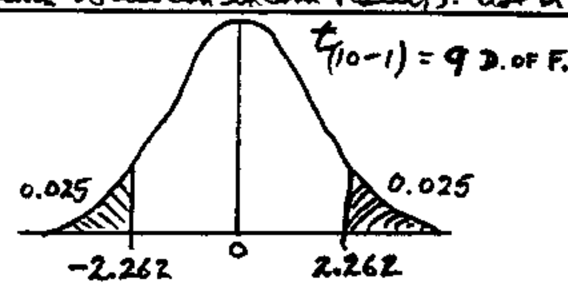
$$S_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}}$$

$$\therefore S_d = \sqrt{\frac{145 - \frac{(29)^2}{10}}{10-1}} = 2.6$$

Test THE HYPOTHESIS that their is NO DIFFERENCE between exam results. Use $\alpha = 0.05$

$H_0: \mu_d = 0$

$H_A: \mu_d \neq 0$



THE TEST STATISTIC

$$t_0 = \frac{\bar{d} - 0}{S_d/\sqrt{n}} = \frac{2.9}{2.6/\sqrt{10}}$$

$\therefore t_0 = 3.53 > 2.262$, $\therefore$ YES, there is a significant difference.

Construct the 95% CONFIDENCE INTERVAL ESTIMATE OF μ_d.

Consider $\quad \bar{d} \pm t_{\alpha/2} \times \frac{S_d}{\sqrt{n}}$

$\therefore \quad 2.9 \pm 2.262 \times \frac{2.6}{\sqrt{10}}$

$\therefore \quad 2.9 \pm 1.86$

ie $\quad 1.04$ to 4.76

INFERENCES FOR PAIRED DIFFERENCES (DEPENDENT SAMPLES) — **EXERCISES**

(1) Consider the amount of wear recorded below for 2 types of tires. (randomly assigned to the rear wheels of each car)

CAR	TIRE A	TIRE B
1	10.6	10.2
2	9.8	9.4
3	12.3	11.8
4	9.1	9.3
5	8.8	8.3

(a) Is there a significant difference in wear? Conduct a T. of H. using $\alpha = 0.05$.

(b) Construct a 95% Confidence Interval Estimate of the true mean difference in wear.

(2) Consider the number of sit-ups that 10 people could do, before and after a fitness course.

BEFORE	29	22	25	29	26	24	31	46	34	28
AFTER	30	26	25	35	33	36	32	54	50	43

(a) Is there a significant increase in the number of situps? Use $\alpha = 0.01$

(b) Construct a 99% confidence interval estimate of the mean difference, μ_d.

(3) Consider the weights in pounds of 8 women before and after a diet.

WOMAN	BEFORE	AFTER
A	127	122
B	130	120
C	114	116
D	139	132
E	150	144
F	147	138
G	167	155
H	153	152

(a) Can we conclude that the diet worked? Conduct a T. of H. using $\alpha = 0.01$.

(b) Construct a 95% confidence interval estimate of the mean difference in weights.

(4) Consider the test scores for 10 students before and after a tutoring session.

BEFORE	75	62	67	70	55	59	60	64	72	59
AFTER	77	65	68	72	62	61	60	67	75	68

(a) Is there a significant improvement? Conduct a T. of H. using $\alpha = 0.05$.

(b) Construct a 90% confidence interval estimate for the mean improvement in scores.

(5) The number of points scored by half for a random sample of 10 games by a college basketball team are recorded below. Can we conclude that there is a significant difference in scoring between halves? Use $\alpha = 0.05$

GAME	1	2	3	4	5	6	7	8	9	10
1ST HALF	29	26	25	25	25	24	26	26	30	31
2nd HALF	25	26	25	25	24	23	27	25	29	30

(6) The monthly sales for a real estate office were recorded below for the last 2 years. Construct a 95% C.I.E. for the mean difference in monthly sales for the office between years.

FIRST YEAR	10	13	18	12	9	8	14	12	17	20	7	11
SECOND YEAR	5	9	13	17	4	5	11	14	13	18	7	12

(7) The blood sugar readings for 12 people with diabetes were recorded before and after a medication regime. Can we conclude that the medication lowers blood sugar by more than 10 points? Conduct a T. of H. using $\alpha = 0.05$

BEFORE	128	145	115	156	129	147	130	127	120	145	124	135
AFTER	117	133	108	133	111	129	122	124	103	129	116	117

(8) Consider the cholesterol values below (in mmol/l) of 7 adult males both before and after a new lipid drug treatment regime.

BEFORE	AFTER
6.2	6.0
4.5	3.5
7.3	5.0
3.8	4.7
6.8	5.3
7.0	6.0
6.6	5.4

(a) Did the drug lower cholesterol values significantly? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(b) Construct a 98% confidence interval estimate for the mean difference between before and after cholesterol values.

INFERENCES FOR PAIRED DIFFERENCES (DEPENDENT SAMPLES) – EXERCISE SOLUTIONS

(1) let $d = A - B$ $\therefore \bar{d} = 0.32$ & $S_d = 0.295$

(a) $H_0: \mu_d = 0$, $H_A: \mu_d \neq 0$

$t_0 = \frac{0.32}{0.295/\sqrt{5}} = 2.43 \not> 2.776$

$\therefore$ No, not a signif. diff.

(b) $0.32 \pm 2.776 \times \frac{(0.295)}{\sqrt{5}}$

$\therefore 0.32 \pm 0.37$

ie -0.05 to 0.69

---

(2) let $d = A - B$ $\therefore \bar{d} = 7$ $\therefore S_d = 5.79$

(a) $H_0: \mu_d = 0$, $H_A: \mu_d > 0$

$t_0 = \frac{7}{5.79/\sqrt{10}} = 3.82 > 2.821$

$\therefore$ YES, significant increase

(b) $7 \pm 3.250 \times \frac{5.79}{\sqrt{10}}$

$\therefore 7 \pm 5.95$

$\therefore 1.05$ to 12.95

---

(3) let $d = B - A$ $\therefore \bar{d} = 6$ $\therefore S_d = 4.66$

(a) $H_0: \mu_d = 0$, $H_A: \mu_d > 0$

$t_0 = \frac{6}{4.66/\sqrt{8}} = 3.64 > 2.998$

$\therefore$ YES, diet worked

(b) $6 \pm 2.365 \times \frac{4.66}{\sqrt{8}}$

$\therefore 6 \pm 3.9$

$\therefore 2.1$ to 9.9 lbs.

---

(4) let $d = A - B$ $\therefore \bar{d} = 3.2$ $\therefore S_d = 2.74$

(a) $H_0: \mu_d = 0$, $H_A: \mu_d > 0$

$t_0 = \frac{3.2}{2.74/\sqrt{10}} = 3.69 > 1.833$

$\therefore$ YES, signif improvement

(b) $3.2 \pm 1.833 \times \frac{2.74}{\sqrt{10}}$

$\therefore 3.2 \pm 1.59$

$\therefore 1.61$ to 4.79

---

(5) let $d = 1^{st} - 2^{nd}$ $\therefore \bar{d} = 0.8$ $\therefore S_d = 1.32$

$H_0: \mu_d = 0$, $H_A: \mu_d \neq 0$ $\therefore t_0 = \frac{0.8}{1.32/\sqrt{10}} = 1.92 \not> 2.262$, $\therefore$ No, not signif. diff.

---

(6) let $d = 1^{st} - 2^{nd}$ $\therefore \bar{d} = 1.92$ $\therefore S_d = 3.23$

hence, $1.92 \pm 2.201 \times \frac{3.23}{\sqrt{12}}$

$\therefore 1.92 \pm 2.05$ $\therefore -0.13$ to 3.97

---

(7) let $d = B - A$ $\therefore \bar{d} = 13.25$ $\therefore S_d = 5.966$

$H_0: \mu_d = 10$, $H_A: \mu_d > 10$ $\therefore t_0 = \frac{13.25 - 10}{5.56/\sqrt{12}} = 1.89 > 1.796$, $\therefore$ YES, lowers by more than 10

---

(8) let $d = B - A$ $\therefore \bar{d} = 0.90$ $\therefore S_d = 1.01$

(a) $H_0: \mu_d = 0$, $H_A: \mu_d > 0$

$t_0 = \frac{0.90 - 0}{1.01/\sqrt{7}} = 2.36 > 1.943$

$\therefore$ YES, lowers significantly

(b) $0.90 \pm 3.143 \times \frac{1.01}{\sqrt{7}}$

$\therefore 0.90 \pm 1.20$

$\therefore -0.30$ to 2.10

ESTIMATION OF $(\mu_1-\mu_2)$ FOR LARGE INDEPENDENT SAMPLES*

CONSIDER

POPULATION 1 WITH MEAN μ_1	POPULATION 2 WITH MEAN μ_2
RANDOM SAMPLES OF SIZE n_1 WITH MEANS $\overline{X}_1$ (VARIANCE S_1^2)	RANDOM SAMPLES OF SIZE n_2 WITH MEANS $\overline{X}_2$ (VARIANCE S_2^2)

THEN, THE EXTENDED CENTRAL LIMIT THEOREM GIVES THE <u>SAMPLING DISTRIBUTION</u> OF $(\overline{X}_1-\overline{X}_2)$ AS:

$$N\left[(\overline{X}_1-\overline{X}_2);\ (\mu_1-\mu_2),\ \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

HENCE, WE HAVE:

THE $(1-\alpha)\%$ CONFIDENCE INTERVAL ESTIMATE OF $(\mu_1-\mu_2)$

$$(\overline{X}_1-\overline{X}_2) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

* THE SAMPLES ARE INDEPENDENTLY DRAWN AND BOTH $n_1>30$ AND $n_2>30$.

ESTIMATION OF $(\mu_1-\mu_2)$ FOR LARGE INDEPENDENT SAMPLES – ILLUSTRATIVE EXAMPLE

To estimate the difference between the mean mathematic's grades of female and male CEGEP students, 2 random samples of grades were selected, yielding the following data:

	n	$\bar{x}$	s
FEMALES, ♀	50	80.1	7.9
MALES, ♂	40	77.7	8.2

Construct a 97% CONFIDENCE INTERVAL ESTIMATE of $(\mu_{♀}-\mu_{♂})$.

consider $$(\bar{X}_1-\bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$\therefore\ (80.1-77.7) \pm 2.17\sqrt{\frac{7.9^2}{50}+\frac{8.2^2}{40}}$$

$$\therefore\ 2.4 \pm 3.71$$

$$\therefore\ -1.31 \text{ to } 6.11$$

ESTIMATION OF $(\mu_1-\mu_2)$ FOR LARGE INDEPENDENT SAMPLES — EXERCISES

(1) Construct a 90% Confidence Interval Estimate (C.I.E.) of $\mu_1-\mu_2$ given the following sample data:

	n	$\bar{x}$	s
POPULATION 1 :	125	15.5	2.7
POPULATION 2 :	115	14.3	3.0

(2) Construct a 95% C.I.E. for $\mu_1-\mu_2$ given the following data:

	n	$\bar{x}$	s
POPULATION 1 :	50	5.2	1.1
POPULATION 2 :	50	7.6	1.3

(3) Construct a 99% C.I.E. for $\mu_A-\mu_B$ given the following data:

	n	$\bar{x}$	s
POPULATION A :	50	57.5	6.2
POPULATION B :	60	54.4	10.6

(4) Construct a 95% C.I.E. for $\mu_{II}-\mu_{I}$ given the following data:

	n	Σx	$\Sigma(x-\bar{x})^2$
POPULATION I	32	2253	32462
POPULATION II	35	5157	58600

(5) Construct a 99% C.I.E. for $\mu_1-\mu_2$ given the following data:

	n	Σx	Σx^2
POPULATION 1	36	278.4	2163.76
POPULATION 2	42	310.8	2332.26

(6) Construct a 90% C.I.E for the true difference between the mean hours of study per week for science and arts students based on:

	n	$\bar{x}$	s^2
SCIENCE	36	15.5 hrs.	66
ARTS	36	12.1 hrs.	78

(7) To estimate the difference in average life-spans of women and men in Canada, $\mu_W - \mu_M$, a researcher sampled 50 women and 75 men. The women averaged 81.1 years of life, with $s = 5.2$ yrs., while the men averaged 76.5 yrs, with $s = 6.3$ yrs. Construct a 95% C.I.E for $\mu_W - \mu_M$.

(8) Two types of auto tires (100 of each) were road tested until they wore-out, yielding: $\bar{x}_1 = 26400$ miles, $s_1^2 = 1440000$, $\bar{x}_2 = 25100$ mi., and $s_2^2 = 1960000$. Construct a 99% C.I.E. for $\mu_1 - \mu_2$.

(9) A random sample of 50 Toronto salaries last year yielded a mean of \$56500, with $s =$ \$6200, while a r.s. of 50 Montreal salaries last year yielded a mean of \$49900, with $s =$ \$5800. Construct a 95% C.I.E. for the difference of mean salaries, $\mu_T - \mu_M$.

SOLUTIONS

(1) $(15.5-14.3) \pm 1.645\sqrt{\frac{2.7^2}{125} + \frac{(3.0)^2}{115}}$

$\therefore 1.2 \pm 0.61$

$\therefore 0.59$ to 1.81

(2) $(5.2-7.6) \pm 1.96\sqrt{\frac{1.1^2}{50} + \frac{1.3^2}{50}}$

$\therefore -2.4 \pm 0.47$

$\therefore -2.87$ to -1.93

(3) $(57.5-54.4) \pm 2.575\sqrt{\frac{6.2^2}{50} + \frac{10.6^2}{60}}$

$\therefore 3.1 \pm 4.18$

$\therefore -1.08$ to 7.28

(4) $\bar{x}_I = 70.4$, $s_I^2 = 1047.2$

$\bar{x}_{II} = 147.3$, $s_{II}^2 = 1723.5$

$(147.3 - 70.4) \pm 1.96\sqrt{\frac{1047.2}{32} + \frac{1723.5}{35}}$

$\therefore 76.9 \pm 17.75$

$\therefore 59.15$ to 94.65

(5) $\bar{x}_1 = 7.7$, $s_1^2 = 0.31$

$\bar{x}_2 = 7.4$, $s_2^2 = 0.79$

$(7.7 - 7.4) \pm 2.575\sqrt{\frac{0.31}{36} + \frac{0.79}{42}}$

$\therefore 0.3 \pm 0.43$

$\therefore -0.13$ to 0.73

(6) $(15.5-12.1) \pm 1.645\sqrt{\frac{66}{36} + \frac{78}{36}}$

$\therefore 3.4 \pm 3.29$

$\therefore 0.11$ to 6.69 hours

(7) $(81.1-76.5) \pm 1.96\sqrt{\frac{5.2^2}{50} + \frac{6.3^2}{75}}$

$\therefore 4.6 \pm 2.03$

$\therefore 2.57$ to 6.63 years

(8) $(26400-25100) \pm 2.575\sqrt{\frac{1440000}{100} + \frac{1960000}{100}}$

$\therefore 1300 \pm 473.8$

$\therefore 826.2$ to 1773.8 miles

(9) $(56500-49900) \pm 1.96\sqrt{\frac{6200^2}{50} + \frac{5800^2}{50}}$

$\therefore 6600 \pm 2353.31$

$\therefore$ \$4246.69 to \$8953.31

Difference of Means Tests $(H_0 : \mu_1 - \mu_2 = 0)$

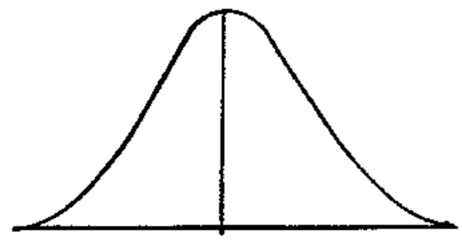
Large Independent Samples ①

THE TEST STATISTIC

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

HAS THE

$N(z;0,1)$ DISTRIBUTION



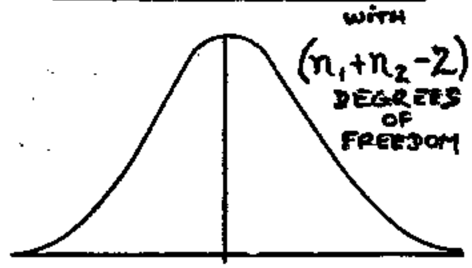
Small Independent Samples
(WITH POPULATIONS NORMAL AND $\sigma_1^2 = \sigma_2^2$) ②

THE TEST STATISTIC

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

HAS THE

t- DISTRIBUTION WITH (n_1+n_2-2) DEGREES OF FREEDOM



WHERE $$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$$

NOTE: ① BOTH $n_1 > 30$ AND $n_2 > 30$

② TO TEST $H_0: \sigma_1^2 = \sigma_2^2$ WE MAY USE THE F-DISTRIBUTION (SEE NEXT SECTION)

DIFFERENCE OF MEANS TESTS ($H_0: \mu_1-\mu_2=0$) – ILLUSTRATIVE EXAMPLES

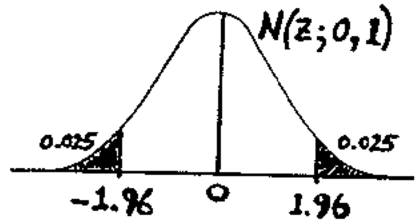
①

Using $\alpha = 0.05$ and the data below test the Hypothesis that there is no significant difference of mean ages between the citizens of Montreal and the citizens of Toronto.

	n	$\bar{x}$	s
MONTREAL	500	44.4 yrs.	15.2 yrs.
TORONTO	600	42.5 yrs.	18.7 yrs.

$H_0: \mu_1-\mu_2=0$

$H_A: \mu_1-\mu_2 \neq 0$



$$z_0 = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{44.4-42.5}{\sqrt{\frac{15.2^2}{500}+\frac{18.7^2}{600}}}$$

$\therefore z_0 = 1.86 \not> 1.96$, NO, not a significant difference.

②

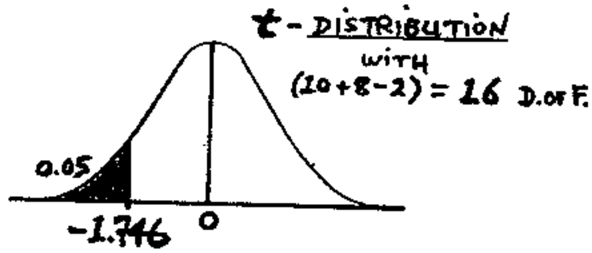
Consider the data below concerning the costs of college textbooks.

	n	$\bar{x}$	s
MATH. BOOKS	10	\$73.50	\$2.90
CHEMISTRY BOOKS	8	\$76.30	\$3.50

assuming that both populations are NORMAL and $\sigma_1^2 = \sigma_2^2$, use $\alpha = 0.05$ to test the Hypothesis that Math. books cost less, on average, than Chemistry books

$H_0: \mu_1-\mu_2=0$

$H_A: \mu_1-\mu_2<0$



$$t_0 = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{73.50-76.30}{\sqrt{8.94\left(\frac{1}{10}+\frac{1}{8}\right)}}$$

$\therefore t_0 = -1.97 < -1.746$, $\therefore$ YES, $\mu_M < \mu_C$.

WHERE $S_p^2 = \dfrac{(10-1)\,2.90^2 + (8-1)\,3.50^2}{(10+8-2)} = 8.94$

DIFFERENCE OF MEANS TESTS ($H_0: \mu_1 - \mu_2 = 0$) – EXERCISES

(1) Is the mean amount of vitamin C in pill A equal to the mean amount of vitamin C in pill B? Conduct a difference of means test using $\alpha = 0.05$ and the data below.

	SAMPLE SIZE	SAMPLE MEAN	SAMPLE VARIANCE
PILL A	100	498 mg	84
PILL B	100	495 mg	60

(2) It is believed that the domestic flights of 2 Canadian airlines average the same distance. Use $\alpha = 0.01$ to test this belief with the data below.

	n	$\bar{x}$	s
AIRLINE I	100	1925 miles	40 miles
AIRLINE II	100	1913 miles	30 miles

(3) A nutritionist wishes to compare the effectiveness of 2 different reducing diets. The following data was obtained from 2 independent random samples of dieters.

	DIET 1	DIET 2
SAMPLE SIZE	60	40
MEAN WEIGHT LOSS	10.7 lbs.	9.0 lbs.
SAMPLE VARIANCE	30	20.

Are the diets significantly different? Test using $\alpha = 0.05$.

(4) Statistics Canada reports that the average annual income of Torontonians is higher than the average annual income of Montrealers. Conduct a Test of Hypothesis to determine whether the data below supports the report or not. Use $\alpha = 0.05$.

	n	$\bar{x}$	s^2
TORONTO	36	\$45300	688000
MONTREAL	36	45000	312000

(5) Two types of batteries are advertised as having the same mean life. Test the ad using $\alpha = 0.05$ and the results below.

	n	$\bar{x}$	s^2
BATTERY TYPE A	100	30 days	36
BATTERY TYPE B	100	32 days	28

DIFFERENCE OF MEANS TESTS ($H_0: \mu_1 - \mu_2 = 0$) — EXERCISES

(6) It is claimed that the mean weight of Grade A eggs is greater than that of Grade B eggs. Test the claim using $\alpha = 0.05$ and the following data.

	n	$\bar{X}$	S^2
GRADE A	36	3.10 oz.	0.43
GRADE B	36	2.80 oz.	0.57

(7) Is it true that the mean annual income of Ontario MD's is higher than the mean annual income of Québec MD'S? Conduct a Test of Hypothesis using $\alpha = 0.01$ and the results below.

	n	$\bar{X}$	S
ONTARIO MD'S	49	\$151625	\$6000
QUEBEC MD'S	49	\$149050	\$8000

(8) Ford claims that their new cars have better brakes than Chevrolet's new cars. Test the claim using $\alpha = 0.05$ if random samples of 64 of each type of car were tested for braking distance required to stop at 40 mph., yielding these results.

	FORD	CHEVROLET
MEAN	115 feet	118 feet
VARIANCE	77	92

(9) It is believed that there is a significant difference in the study habits of science and arts students. Do the following results (in hours of study per week) support the belief or not? Conduct a Test of Hypothesis using $\alpha = 0.10$.

	SCIENCE	ARTS
n	36	36
$\bar{X}$	15.5 hours	12.1 hours
S^2	66	78

(10) According to the data below, can we conclude that the average adult American is taller than the average adult Canadian? Test with $\alpha = 0.05$.

	AMERICAN	CANADIAN
SAMPLE SIZE	36	36
SAMPLE MEAN HEIGHT	178.1 cm	169.7 cm
SAMPLE VARIANCE	414	370

DIFFERENCE OF MEANS TESTS ($H_0: \mu_1 - \mu_2 = 0$) — EXERCISES

(11) Consider that 250 random water temperature recordings from each ocean yielded these results:

	$\overline{X}$	S
ATLANTIC OCEAN	12.2 °C.	5 °C.
PACIFIC OCEAN	13.3 °C.	6 °C.

Can we conclude that the Pacific Ocean is warmer, on average, than the Atlantic Ocean? Conduct a difference of means test using $\alpha = 0.01$.

(12) Based on the data below, is it true that the hemoglobin count of adult females is significantly less, on the average, than that of adult males? Conduct a Test of Hypothesis using $\alpha = 0.01$.

	n	$\overline{X}$	S
♀	33	13.8	2.3
♂	35	15.2	2.5

(13) Is there a significant difference in the total number of runs scored per game this season between the National and American baseball leagues? Use $\alpha = 0.03$ and the sample results below to conduct a Test of Hypothesis.

	n	$\overline{X}$	S
NATIONAL LEAGUE	40	7.2	3.1
AMERICAN LEAGUE	40	8.6	2.8

Note: In the following exercises, assume that both populations are NORMAL and that $\sigma_1^2 = \sigma_2^2$.

(14) Based on the results below, is there a significant difference between the mean ages of Québec and Ontario residents? Test using $\alpha = 0.05$.

	n	$\overline{X}$	S^2
QUEBEC	9	45.2 yrs.	20
ONTARIO	16	42.3 yrs.	15

(15) Does car A have higher average mileage per gallon (mpg.) than car B? Conduct a difference of means test with the results below and use $\alpha = 0.05$.

	CAR A	CAR B
n	9	9
$\overline{X}$	35.7 mpg	33.9 mpg
S^2	4.1	4.9

(16) To determine whether caffeine affects test scores, consider the following:

(I) Caffeine was given to an experimental group of 10 students before a test, and their scores gave: $\bar{x} = 82$ and $s^2 = 12$.

(II) A placebo was given to a control group of 11 students before the test, and their scores gave: $\bar{x} = 78$ and $s^2 = 10$.

Is there a significant difference in the average scores? Test with $\alpha = 0.01$.

(17) A journal reports that vegetarians live longer, on the average, than non-vegetarians. Is this report supported by the data below? Use $\alpha = 0.05$.

	n	$\bar{x}$	s^2
VEGETARIANS	9	80 yrs.	34
NON-VEGETARIANS	9	76 yrs.	30

(18) Do the results below indicate that there is a significant difference in the mean summer temperatures at noon between Toronto and Montreal? Conduct a difference of means test using $\alpha = 0.01$.

	n	$\bar{x}$	s
TORONTO	16	26.9°C.	1.8°C.
MONTREAL	10	25.4°C.	2.1°C.

(19) Use $\alpha = 0.05$ and the data below to test whether there is a significant difference in the average grades between 2 sections of a college course.

	SECTION 1	SECTION 2
n	7	13
Σx	514	1062
Σx^2	37867	86900

(20) Statistics grades are known to have a standard deviation of 10. A random sample of 16 Career Statistics grades yielded $\bar{x} = 70$. What minimum sample mean of a random sample of 9 Science Statistics grades would be required to infer, at the $\alpha = 0.01$ level, that Science Statistics grades are higher, on the average, than Career Statistics grades?

(21) (optional) Construct the 95% confidence interval estimate of $(\mu_1 - \mu_2)$ for the data to the right.

	n	$\bar{x}$	s
POP. 1	12	110.2	16.2
POP. 2	10	100.9	17.0

DIFFERENCE OF MEANS TESTS ($H_0: \mu_1 - \mu_2 = 0$) – EXERCISE SOLUTIONS

(1) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{498 - 495}{\sqrt{\frac{84}{100} + \frac{60}{100}}} = 2.50 > 1.96$$

∴ NO, not equal

(2) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{1925 - 1913}{\sqrt{\frac{40^2}{100} + \frac{30^2}{100}}} = 2.40 \not> 2.575$$

∴ YES, average the same

(3) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{10.7 - 9.0}{\sqrt{\frac{30}{60} + \frac{20}{40}}} = 1.70 \not> 1.96$$

∴ NO, not significantly diff.

(4) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$z_0 = \frac{45800 - 45000}{\sqrt{\frac{688000}{36} + \frac{312000}{36}}} = 1.80 > 1.645$$

∴ YES, supports report

(5) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{30 - 32}{\sqrt{\frac{36}{100} + \frac{28}{100}}} = -2.50 < -1.96$$

∴ NO, does not support the ad.

(6) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$z_0 = \frac{3.1 - 2.8}{\sqrt{\frac{0.43}{36} + \frac{0.57}{36}}} = 1.80 > 1.645$$

∴ YES, supports the claim

(7) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$z_0 = \frac{151675 - 149050}{\sqrt{\frac{6000^2}{49} + \frac{8000^2}{49}}} = 1.80 \not> 2.33$$

∴ NO, not true

(8) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 < 0$

$$z_0 = \frac{115 - 118}{\sqrt{\frac{72}{64} + \frac{92}{94}}} = -1.85 < -1.645$$

∴ YES, better brakes on Fords

(9) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{15.5 - 12.1}{\sqrt{\frac{66}{36} + \frac{78}{36}}} = 1.70 > 1.645$$

∴ YES, supports belief

(10) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$z_0 = \frac{178.1 - 169.7}{\sqrt{\frac{814}{36} + \frac{330}{36}}} = 1.80 > 1.645$$

∴ YES, American's taller ave.

(11) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 < 0$

$$z_0 = \frac{12.2 - 13.3}{\sqrt{\frac{5^2}{250} + \frac{6^2}{250}}} = -2.23 \not< -2.33$$

∴ NO, not warmer on ave.

(12) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 < 0$

$$z_0 = \frac{13.8 - 15.2}{\sqrt{\frac{2.3^2}{33} + \frac{2.5^2}{33}}} = -2.40 < -2.33$$

∴ YES, it is true

(13) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$z_0 = \frac{7.2 - 8.6}{\sqrt{\frac{3.1^2}{40} + \frac{2.8^2}{40}}} = -2.12 \not< -2.17$$

∴ NO, not a signif. diff.

(14) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$t_0 = \frac{45.2 - 42.3}{\sqrt{S_p^2\left(\frac{1}{9} + \frac{1}{16}\right)}} = 1.70 \not> 2.069$$

$\left(S_p^2 = \frac{8(20) + 15(15)}{23} = 16.74\right)$ ∴ NO, no signif. diff.

(15) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$t_0 = \frac{35.7 - 33.9}{\sqrt{S_p^2\left(\frac{1}{9} + \frac{1}{9}\right)}} = 1.80 > 1.746$$

$\left(S_p^2 = \frac{8(4.1) + 8(4.9)}{16} = 4.5\right)$ ∴ YES, A higher ave

(16) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$t_0 = \frac{82 - 78}{\sqrt{S_p^2\left(\frac{1}{10} + \frac{1}{11}\right)}} = 2.77 \not> 2.861$$

$\left(S_p^2 = \frac{9(12) + 10(10)}{19} = 10.95\right)$ ∴ NO, not signif. diff

(17) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$

$$t_0 = \frac{80 - 76}{\sqrt{S_p^2\left(\frac{1}{9} + \frac{1}{9}\right)}} = 1.50 \not> 1.746$$

$\left(S_p^2 = \frac{8(34) + 8(30)}{16} = 32\right)$ ∴ NO, does not support

(18) $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$

$$t_0 = \frac{26.9 - 25.4}{S_p^2\left(\frac{1}{16} + \frac{1}{10}\right)} = 2.51 \not> 2.797$$

$\left(S_p^2 = \frac{15(1.1)^2 + 9(2.1)^2}{24} = 3.18\right)$ ∴ NO, not a signif. diff.

(19) $\bar{X}_1 = 73.43$ $\bar{X}_2 = 81.69$

$S_1^2 = 20.786$ $S_2^2 = 11.897$

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$$t_0 = \frac{73.43 - 81.69}{\sqrt{S_p^2\left(\frac{1}{7} + \frac{1}{13}\right)}} = -4.57 < -2.101$$

∴ YES, signif. diff.

$\left(S_p^2 = \frac{6(20.786) + 12(11.897)}{18} = 14.86\right)$

(20) $S_p^2 = \frac{8(10)^2 + 15(10)^2}{23} = 100$

∴ consider

$$\frac{\bar{X} - 70}{\sqrt{100\left(\frac{1}{9} + \frac{1}{16}\right)}} = 2.500$$

∴ $\bar{X} = 80.416$

(21) (optional) We consider

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

for $n_1 + n_2 - 2$ D. of F.

∴ $(110.2 - 100.9) \pm 2.086\sqrt{274.37\left(\frac{1}{12} + \frac{1}{10}\right)}$, with $12 + 10 - 2 = 20$ D. of F.

9.3 ± 14.8

-5.5 to 24.1

STATISTICAL INFERENCES: THE F DISTRIBUTION

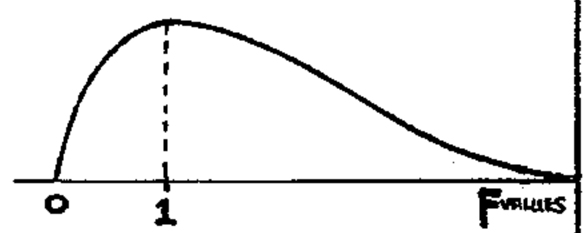
IF 2 NORMAL POPULATIONS HAVE EQUAL VARIANCES, THEN:

THE TEST STATISTIC

$$F_0 = \frac{S_1^2}{S_2^2}$$

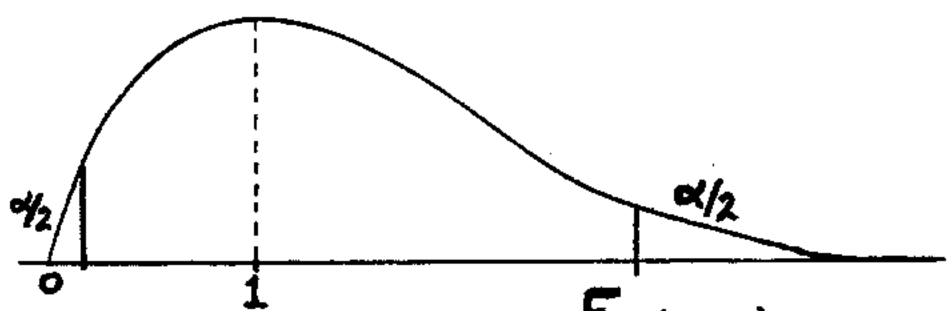
HAS

THE F DISTRIBUTION



AND

WE CONSIDER (USING $S_1^2 > S_2^2$)



WHERE $v_1 = n_1 - 1$,
$v_2 = n_2 - 1$

EXAMPLE: FROM THE F-TABLE, $F_{.025(7,5)} = 6.8531$

STATISTICAL INFERENCES : THE F DISTRIBUTION – INTRODUCTORY EXAMPLES

① Consider :

	Normal Pop I	Normal Pop. 2
Sample Size	$n_1 = 10$	$n_2 = 8$
Sample Variance	$S_1^2 = 7.14$	$S_2^2 = 3.21$

Can we conclude that the 2 Population's have equal variances? Conduct a Test of Hypothesis using $\alpha = .05$.

② Based on the data below, is there a significant difference in the true variance in weights (g) of pennies and dimes? Conduct an "F-test" using $\alpha = .05$.

	Pennies	Dimes
n	9	16
S^2	.0020	.0006

③ The closing prices of 2 common stocks were recorded for 10 working days, yielding the following data:

$\Sigma X_1 = 604.95$ and $\Sigma X_2 = 638.10$

$\Sigma X_1^2 = 36{,}597.62$ $\Sigma X_2^2 = 40{,}717.52$

Is there a significant difference in the variances of the 2 closing prices? Conduct a Test of Hypothesis using $\alpha = .05$

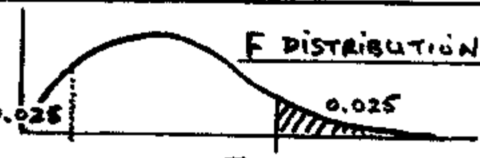
④ Based on the data below, is there a significant difference in the variances of heights (cm) between males and females? Conduct an F-test with $\alpha = .05$

	male	female
n	20	8
S^2	13.92	29.65

STATISTICAL INFERENCES: THE F DISTRIBUTION — INTRODUCTORY EXAMPLE SOLUTIONS

(1) $H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$



$F_{0.025}(9,7) = 4.8232$

$$F_0 = \frac{7.14}{3.21} = 2.22 \ngtr 4.8232$$

∴ YES, they have equal variances

NOTE: All F-tests are inherently 2-tailed, but by placing the largest S^2 on the top of the F_0 ratio, we need only consider right-tail critical values for $\alpha/2$.

(2) $H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$

$F_{0.025}(8,15) = 3.1987$

$$F_0 = \frac{0.0020}{0.0006} = 3.33 > 3.1987$$

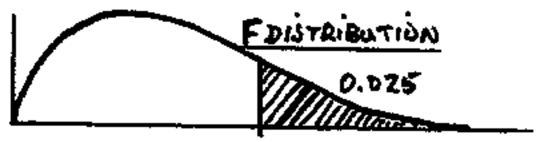
∴ YES, there is a significant difference

(3)
$$S_1^2 = \frac{36597.62 - \frac{(604.95)^2}{10}}{9} \quad \text{and} \quad S_2^2 = \frac{40717.52 - \frac{(638.10)^2}{10}}{9}$$

$\therefore S_1^2 = 0.13$ $\qquad$ $\therefore S_2^2 = 0.04$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$



$F_{0.025}(9,9) = 4.0260$

$$F_0 = \frac{0.13}{0.04} = 3.25 \ngtr 4.0260$$

∴ No, there is not a significant difference

(4) $H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$

$F_{0.025}(7,19) = 3.0509$

$$F_0 = \frac{29.65}{13.92} = 2.13 \ngtr 3.0509$$

∴ No, there is not a significant difference

INFERENCES FOR THE DIFFERENCE OF TWO PROPORTIONS

Consider 2 independent **BINOMIAL** experiments* and their **NORMAL** approximations:

$$B(x_1; n_1, p_1) \sim N\left(x_1; n_1p_1, \sqrt{n_1p_1q_1}\right) \rightarrow N\left(\frac{x_1}{n_1}; p_1, \sqrt{\frac{p_1q_1}{n_1}}\right)$$

and

$$B(x_2; n_2, p_2) \sim N\left(x_2; n_2p_2, \sqrt{n_2p_2q_2}\right) \rightarrow N\left(\frac{x_2}{n_2}; p_2, \sqrt{\frac{p_2q_2}{n_2}}\right)$$

Then, the sampling distribution of $\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)$ is given as:

$$N\left[\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right); (p_1 - p_2), \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}\right]$$

hence

THE $(1-\alpha)\%$ CONFIDENCE INTERVAL ESTIMATE OF $(p_1 - p_2)$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

where $\hat{p}_1 = \frac{x_1}{n_1}$ with $\hat{q}_1 = 1 - \hat{p}_1$ and $\hat{p}_2 = \frac{x_2}{n_2}$ with $\hat{q}_2 = 1 - \hat{p}_2$

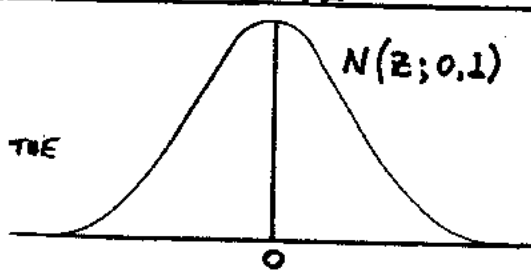
and

THE TEST OF HYPOTHESIS FOR $H_0: p_1 - p_2 = 0$

THE TEST STATISTIC

$$z_0 = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

HAS THE



where $p_1 = p_2 = \hat{p}$ and $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ with $\hat{q} = 1 - \hat{p}$

* Such that n_1p_1, n_2p_2, n_1q_1, and n_2q_2 are each ≥ 5.

INFERENCES FOR THE DIFFERENCE OF TWO PROPORTIONS – AN ILLUSTRATIVE EXAMPLE

Consider the results below of 2 polls concerning the question of support for a certain political party.

	MEN	WOMEN
NUMBER OF SUPPORTERS	120	70
NUMBER POLLED	150	100

then $\hat{p}_1 = \frac{120}{150} = 0.8$, $\hat{q}_1 = 1 - 0.8 = 0.2$ and $\hat{p}_2 = \frac{70}{100} = 0.7$, $\hat{q}_2 = 1 - 0.7 = 0.3$

and $\hat{p} = \frac{120+70}{150+100} = \frac{190}{250} = 0.76$ and $\hat{q} = 1 - \hat{p} = 1 - 0.76 = 0.24$

Construct the 90% CONFIDENCE INTERVAL ESTIMATE OF $(p_1 - p_2)$:

consider $$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$\therefore \quad (0.8 - 0.7) \pm 1.645\sqrt{\frac{(0.8)(0.2)}{150} + \frac{(0.7)(0.3)}{100}}$$

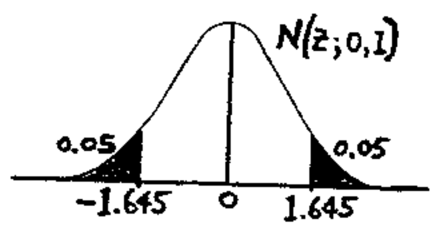
$$\therefore \quad 0.1 \pm 0.093$$

ie $$0.007 \text{ to } 0.193$$

Test the Hypothesis that there is NO DIFFERENCE between p_1 and p_2 (use $\alpha = 0.10$)

$H_0: p_1 - p_2 = 0$

$H_A: p_1 - p_2 \neq 0$



$$z_0 = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} = \frac{(0.8 - 0.7)}{\sqrt{\frac{(0.76)(0.24)}{150} + \frac{(0.76)(0.24)}{100}}}$$

$\therefore z_0 = 1.81 > 1.645$, $\therefore$ YES, there is a significant difference.

INFERENCES FOR THE DIFFERENCE OF TWO PROPORTIONS – EXERCISES

(1) Consider the data below for 2 independently drawn samples of students:

	SCIENCE STUDENTS	ARTS STUDENTS
NUMBER WHO PASSED	75	85
NUMBER IN SAMPLE	100	100

(a) Construct a 90% Confidence Interval Estimate (C.I.E.) for the difference of proportions between science and arts students who passed.

(b) Conduct a Test of Hypothesis to determine if there is a significant difference between the proportions who passed. Use $\alpha = 0.10$.

(2) A candidate for political office polls 1000 men and 1000 women voters. If 459 of the men and 388 of the women said that they would vote for the candidate, then:

(a) Construct a 99% C.I.E. for the difference of proportions between men and women who would vote for the candidate.

(b) Is the proportion of men who would vote for the candidate greater than the proportion of women? Conduct a Test of Hypothesis (T. of H.) using $\alpha = 0.05$.

(3) Conduct a T. of H. with $\alpha = 0.10$ to determine if the proportions of car owners for Dawson and Vanier students are different.

	NUMBER SAMPLED	NUMBER OF CAR OWNERS
DAWSON	500	30
VANIER	600	28

(4) Consider the data below for 2 samples of high school seniors.

	SAMPLE SIZE	SMOKERS
FEMALE	500	215
MALE	500	170

Is the proportion of female smokers greater than the proportion of male smokers? Conduct a T. of H. using $\alpha = 0.01$.

(5) A TV ratings survey of 200 men and 200 women found that 29 % of the men and 24 % of the women viewed a certain TV show. Based on this survey, was there a significant difference in the percentages of men and women who viewed the show? Conduct a T. of H. using $\alpha = 0.05$

(6) Of 250 Democrats polled, 120 supported a new U.S. policy, while 105 of 200 Republicans supported it. Is there a significant difference in the proportions who support the policy? Use $\alpha = 0.05$ and conduct a T. of H.

(7) An annual survey of 17000 seniors showed that 19.7% of them visited a hospital in 1995 and 21.3% in 1996. Construct a 95% C.I.E. for the difference in percentage of visits between years.

(8) A survey of 1000 Montrealers and 1000 Torontonians revealed that 34 % of the Montrealers were hockey fans, while only 29 % of the Torontonians were hockey fans. Is this a significant difference? Conduct a T. of H. using $\alpha = 0.05$.

(9) Consider the data below concerning computers from 2 manufacturers.

MANUFACTURER	SAMPLE SIZE	PROPORTION REQUIRING SERVICE
A	75	0.15
B	75	0.09

Construct a 95% Confidence Interval Estimate for $(P_A - P_B)$.

(10) Is there a difference in the proportions of left-handers by gender? Use the data below and conduct a T. of H. with $\alpha = 0.05$

	NUMBER SAMPLED	NUMBER OF LEFT-HANDERS
♂	90	9
♀	50	9

(11) Consider the sample data below concerning the bilingual status of Dawson students by gender.

	FEMALE, ♀	MALE, ♂
NUMBER SAMPLED	180	175
NUMBER WHO ARE BILINGUAL	63	49

(a) Construct a 97% Confidence Interval Estimate for $(P_♀ - P_♂)$.

(b) Conduct a T. of H. to determine if there is a significant difference between $P_♀$ and $P_♂$, using $\alpha = 0.10$.

(12) Is there a significant difference in the proportions of CEGEP and university students who play sports? Use the following data to conduct a T. of H. using $\alpha = 0.05$.

	CEGEP	UNIVERSITY
X	45	45
n	200	250

INFERENCES FOR THE DIFFERENCE OF TWO PROPORTIONS – EXERCISE SOLUTIONS

1 (a) $\hat{p}_1 = \frac{75}{100} = 0.75, \therefore \hat{q}_1 = 0.25$

$\hat{p}_2 = \frac{85}{100} = 0.85, \therefore \hat{q}_2 = 0.15$

$\therefore (0.75 - 0.85) \pm 1.645\sqrt{\frac{(0.75)(0.25)}{100} + \frac{(0.85)(0.15)}{100}}$

$\therefore -0.10 \pm 0.092$

$\therefore -0.192$ to 0.008

(b) $\hat{p} = \frac{75+85}{200} = \frac{160}{200} = 0.80, \therefore \hat{q} = 0.20$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{(0.75 - 0.85)}{\sqrt{\frac{(0.80)(0.20)}{100} + \frac{(0.80)(0.20)}{100}}} = 1.77 > 1.645$$

$\therefore$ YES, there is a significant difference

2 (a) $\hat{p}_1 = \frac{459}{1000} = 0.459, \therefore \hat{q}_1 = 0.541$

$\hat{p}_2 = \frac{388}{1000} = 0.388, \therefore \hat{q}_2 = 0.612$

$\therefore (0.459 - 0.388) \pm 2.575\sqrt{\frac{(0.459)(0.541)}{1000} + \frac{(0.388)(0.612)}{1000}}$

$\therefore 0.071 \pm 0.057$

$\therefore 0.014$ to 0.128

(b) $\hat{p} = \frac{459+388}{2000} = \frac{847}{2000} = 0.4235, \therefore \hat{q} = 0.5765$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 > 0$

$$z_0 = \frac{(0.459 - 0.388)}{\sqrt{\frac{2(0.4235)(0.5765)}{1000}}} = 3.21 > 1.645$$

$\therefore$ YES, $p_{♂} > p_{♀}$

3 $\hat{p} = \frac{30+28}{1100} = 0.053, \therefore \hat{q} = 0.947$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{\left(\frac{30}{500} - \frac{28}{600}\right)}{\sqrt{\frac{(0.053)(0.947)}{500} + \frac{(0.053)(0.947)}{600}}} = 0.96 \ngtr 1.645$$

$\therefore$ NO, not different

4 $\hat{p} = \frac{215+170}{1000} = 0.385, \therefore \hat{q} = 0.615$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 > 0$

$$z_0 = \frac{\left(\frac{215}{500} - \frac{170}{500}\right)}{\sqrt{\frac{2(0.385)(0.615)}{500}}} = 2.92 > 2.33$$

$\therefore$ YES, $p_{♀} > p_{♂}$

5 $\hat{p} = \frac{58+48}{400} = 0.265, \therefore \hat{q} = 0.735$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{(0.29 - 0.24)}{\sqrt{\frac{2(0.265)(0.735)}{200}}} = 1.13 \ngtr 1.96$$

$\therefore$ NO, no significant diff.

6 $\hat{p} = \frac{120+105}{450} = 0.5, \therefore \hat{q} = 0.5$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{\left(\frac{120}{250} - \frac{105}{200}\right)}{\sqrt{\frac{(0.5)(0.5)}{250} + \frac{(0.5)(0.5)}{200}}} = -0.95 \nless -1.96$$

$\therefore$ NO, not a significant difference

7 $\hat{p}_1 = 0.197, \therefore \hat{q}_1 = 0.803$; $\hat{p}_2 = 0.213, \therefore \hat{q}_2 = .787$

$\therefore (0.197 - 0.213) \pm 1.96\sqrt{\frac{(.197)(.803)}{17000} + \frac{(.213)(.787)}{17000}}$

$\therefore -0.016 \pm 0.0086$

$\therefore -0.0246$ to -0.0074

8 $\hat{p} = \frac{340+290}{2000} = 0.315, \therefore \hat{q} = 0.685$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{(0.34 - 0.29)}{\sqrt{\frac{2(0.315)(0.685)}{1000}}} = 2.41 > 1.96$$

$\therefore$ YES, significant diff.

9 $\therefore (0.15 - 0.09) \pm 1.96\sqrt{\frac{(.15)(.85)}{75} + \frac{(.09)(.91)}{75}}$

$\therefore 0.06 \pm 0.104$

$\therefore -0.044$ to 0.164

10 $\hat{p} = \frac{9+9}{140} = 0.129, \therefore \hat{q} = 0.871$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{\left(\frac{9}{90} - \frac{9}{50}\right)}{\sqrt{\frac{(.129)(.871)}{90} + \frac{(.129)(.871)}{50}}} = 1.36 \ngtr 1.96$$

$\therefore$ NO, no difference in $p_{♂}$ and $p_{♀}$

11 (a) $\hat{p}_1 = \frac{63}{180} = 0.35, \therefore \hat{q}_1 = 0.65$

$\hat{p}_2 = \frac{49}{175} = 0.28, \therefore \hat{q}_2 = 0.72$

$\therefore (0.35 - 0.28) \pm 2.17\sqrt{\frac{(0.35)(0.65)}{180} + \frac{(0.28)(0.72)}{175}}$

$\therefore 0.07 \pm 0.107$

$\therefore -0.037$ to 0.177

(b) $\hat{p} = \frac{63+49}{355} = \frac{112}{355} = 0.32, \therefore \hat{q} = 0.68$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{(0.35 - 0.28)}{\sqrt{\frac{(0.32)(0.68)}{180} + \frac{(0.32)(0.68)}{175}}} = 1.41 \ngtr 1.645$$

$\therefore$ NO, no significant difference in $p_{♀}$ and $p_{♂}$

12 $\hat{p} = \frac{45+45}{200+250} = \frac{90}{450} = 0.20$

$\therefore \hat{q} = 1 - 0.20 = 0.80$

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z_0 = \frac{\left(\frac{45}{200} - \frac{45}{250}\right)}{\sqrt{(0.20)(0.80)\left(\frac{1}{200} + \frac{1}{250}\right)}} = 1.19 \ngtr 1.96$$, $\therefore$ NO, no significant difference

THE MULTINOMIAL PROBABILITY DISTRIBUTION

CONSIDER THE MULTINOMIAL EXPERIMENT:

1. It consists of n identical, independent trials.
2. Each trial yields one of k $(k \geq 3)$ possible results.
3. The probability p_i $(i = 1,2,3,\cdots,k)$ of each possible result is constant from trial to trial and, $\sum_{i=1}^{k} p_i = 1$

Then the probability that we will observe O_i results of each type in the n trials $(\therefore \sum_{i=1}^{k} O_i = n)$ is given by the:

MULTINOMIAL PROBABILITY DISTRIBUTION

$$P(O_1, O_2, O_3, \cdots, O_k) = \frac{n!}{O_1!\,O_2!\,O_3!\cdots O_k!}\, p_1^{O_1} p_2^{O_2} p_3^{O_3} \cdots p_k^{O_k}$$

AND

We expect $E_i = np_i$ results of each type in the n trials

THE MULTINOMIAL PROBABILITY DISTRIBUTION — ILLUSTRATIVE EXAMPLES

① A random sample of 150 digits is selected from a table of (random) digits, yielding:

	DIGITS									
	0	1	2	3	4	5	6	7	8	9
O_i	9	18	12	17	20	21	16	10	13	14

If the table is truly random, then the probability of observing the O_i results of each type above is given by:

$$P(9,18,12,\cdots,14) = \frac{150!}{9!\,18!\,12!\cdots 14!}\times\left(\frac{1}{10}\right)^{9}\times\left(\frac{1}{10}\right)^{18}\times\left(\frac{1}{10}\right)^{12}\times\cdots\times\left(\frac{1}{10}\right)^{14}$$

②

A coach claims that the probabilities that his team will win, lose, or draw any game are .70, .20, and .10 respectively. If this is true, then the probability that his team will complete a 32-game schedule with 20 wins, 7 losses, and 5 draws is given by:

$$P(20,7,5) = \frac{32!}{20!\,7!\,5!}\times(.70)^{20}\times(.20)^{7}\times(.10)^{5}$$

THE CHI-SQUARED GOODNESS OF FIT TEST

Consider the **MULTINOMIAL** Experiment with n trials, and k possible results per trial, each with the probability

$p_i \; , \; i = 1,2,3,\cdots,k$. Then

We may observe O_i results of each type, while

We expect $E_i = n p_i$ results of each type, and

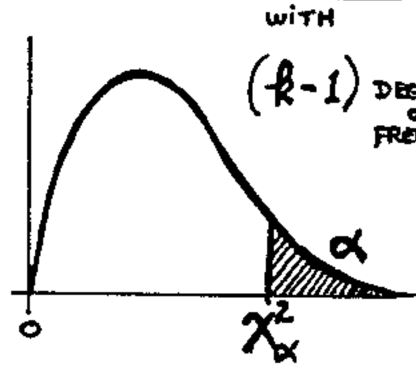
conduct a

GOODNESS OF FIT TEST

THE TEST STATISTIC

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

HAS THE **CHI-SQUARED DISTR.** WITH $(k-1)$ DEGREES OF FREEDOM



WITH $E_i \geq 5 , \; \forall i$

NOTE: The Goodness of Fit Test is inherently 2-tailed, but we include all of α in a single critical region to the right, since χ_0^2 is always positively large when critical.

THE CHI-SQUARED GOODNESS OF FIT TEST — AN ILLUSTRATIVE EXAMPLE

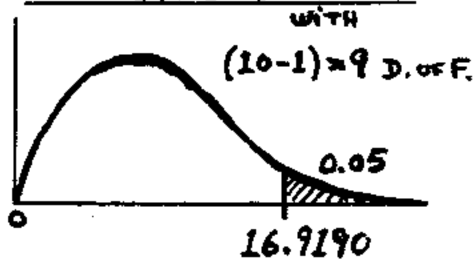
If 150 digits selected at random from a table of (random) digits yielded the observed results below, could we conclude that the table is truly random?

DIGITS	0	1	2	3	4	5	6	7	8	9
OBSERVED	9	18	12	17	20	21	16	10	13	14

Conduct <u>The Goodness of Fit Test</u> using $\alpha = 0.05$.

$$\left.\begin{array}{l} H_0: p_i = \frac{1}{10}, \text{ for } i = 1, 2, 3, \cdots, 10 \\ H_A: \text{NOT AS IN } H_0 \end{array}\right\} \therefore E_i = 150 \cdot \frac{1}{10} = 15, \forall i$$

CHI-SQUARED DISTR. WITH



$$\chi_0^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

$$\therefore \chi_0^2 = \frac{(9-15)^2}{15} + \frac{(18-15)^2}{15} + \cdots + \frac{(14-15)^2}{15}$$

$$\therefore \chi_0^2 = 10.0 \not> 16.9190, \therefore$$ <u>YES</u>, table is truly random.

THE CHI-SQUARED GOODNESS OF FIT TEST — EXERCISES

(1) A single die is rolled 60 times, yielding the following:

POSSIBLE RESULTS	1	2	3	4	5	6
OBSERVED RESULTS	15	10	5	10	5	15

(a) If the die is fair, give the expression for the probability of the observed results.

(b) Conduct a Chi-Squared Goodness of Fit Test using the observed results above to determine whether the die is fair or not. Use $\alpha = 0.05$.

(2) A random sample of 100 digits selected from a table of (random) digits yielded:

	DIGITS									
	0	1	2	3	4	5	6	7	8	9
O_i	6	12	10	8	14	10	13	10	7	10

(a) If the table is truly random, give the expression for the actual probability of the results above.

(b) Conduct a χ^2 Goodness of Fit Test using the results above, and $\alpha = 0.05$, to determine whether the table is truly random or not.

(3) Are wins equally distributed over post positions in horse races? Conduct a Goodness of Fit Test using $\alpha = 0.05$ and the data below.

	POST POSITIONS							
	1	2	3	4	5	6	7	8
WINS	29	19	18	25	17	10	15	11

(4) Do the results below indicate that students had a preference at registration for certain of the 7 sections of a Math. course that were offered? Conduct a Goodness of Fit Test using $\alpha = 0.05$.

	SECTIONS						
	1	2	3	4	5	6	7
STUDENTS REGISTERED	18	12	25	23	8	19	14

(5) Are CEGEP students equally-likely to have each of the 12 possible birth-months? Conduct a Goodness of Fit Test using $\alpha = 0.05$ and the birth data below from a random sample of 400 CEGEP students.

	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
BIRTHS	38	31	42	46	28	31	24	29	33	36	27	35

(6) Historically a real estate agent has made 40% of his sales in the first quarter of the year, and 20% in each of the other 3 quarters. Last year he made 100 sales, distributed by quarter as follows:

	1ST QUARTER	2nd QUARTER	3rd QUARTER	4th QUARTER
NUMBER OF SALES	50	10	30	10

Conduct a Goodness of Fit Test to determine whether there was a significant change in his quarterly distribution of sales last year. Use $\alpha = 0.01$.

(7) A random sample of 80 families, each with 4 children, yielded the frequency distribution:

NUMBER OF BOYS	FREQUENCY
0	7
1	17
2	25
3	25
4	6

Is the model distribution $B(x; 4, \frac{1}{2})$ a good fit of this data?

Conduct a Goodness of Fit Test using $\alpha = 0.05$.

(8) Is the model distribution $N(x; 50, 10)$ a good fit of the following frequency distribution?

Class boundaries	f
27.95 – 34.95	30
34.95 – 41.95	60
41.95 – 48.95	120
48.95 – 55.95	150
55.95 – 62.95	90
62.95 – 69.95	30
69.95 – 76.95	20

Conduct a Goodness of Fit Test using $\alpha = 0.05$.

(9) A pair of dice are rolled 360 times. Do the following results support the claim that the dice are fair (ie give the proper distribution of sums)? Test with $\alpha = 0.05$.

SUM	2	3	4	5	6	7	8	9	10	11	12
O_i	8	19	28	43	50	64	46	37	32	21	12

(10) According to Mendelian inheritance theory, offspring of a certain seed crossing should be coloured red, black, and white in the ratio 9:3:4. If an experiment yielded 70, 36, and 38 offspring of these colours respectively, would this support the theory? Test using $\alpha = 0.05$.

(11) In the past an NHL player scored 30% of his goals in Jan., 20% in each of Dec. and March, and 10% in each of Oct., Nov. and Feb. Last year he scored 50 goals as follows:

MONTH	OCT.	NOV.	DEC.	JAN.	FEB.	MAR.
O_i	6	6	6	20	7	5

Was there a significant change in scoring distribution last year? Use $\alpha = 0.10$.

(12) Historically a salesman has made 40% of his sales in the spring, 40% in the winter, and 10% in each of summer and fall. Last year the seasonal breakdown of 200 of his sales was:

SPRING	SUMMER	FALL	WINTER
75	15	25	85

Does this indicate a significant change in his seasonal sales distribution? Test with $\alpha = 0.05$

(13) Do the census report percentages concerning the languages of Montrealers and the sample results to the right agree? Test with $\alpha = 0.01$.

LANGUAGE	CENSUS %	SAMPLE RESULTS
FRENCH	41%	502
ENGLISH	38%	480
ITALIAN	10%	127
SPANISH	6%	56
GREEK	3%	40
OTHERS	2%	10

(14) It is believed that the heights of Canadian men are normally distributed with a mean of 5.85 feet and a standard deviation of 0.20 feet. Does the following frequency distribution for a random sample of Canadian mens' heights support the belief?

HEIGHT IN FEET	FREQUENCY
5.25 – 5.50	40
5.50 – 5.75	280
5.75 – 6.00	430
6.00 – 6.25	220
6.25 – 6.50	30

Conduct a Goodness of Fit Test using $\alpha = 0.10$.

THE INDEPENDENT CLASSIFICATIONS TEST – AN ILLUSTRATIVE EXAMPLE

Consider the 2 × 3 **CONTINGENCY TABLE** consisting of the **OBSERVED RESULTS** from an opinion poll of 200 people concerning a new law.

GENDER	OPINIONS: FOR	AGAINST	UNDECIDED	
♂	50	40	10	100
♀	60	30	10	100
	110	70	20	200

Can we make the inference that **OPINIONS ARE INDEPENDENT OF GENDER**? Conduct **THE INDEPENDENT CLASSIFICATIONS TEST** using $\alpha = 0.05$.

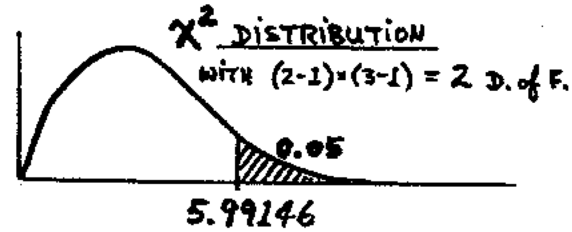
H_0 : opinions are independent of gender.

H_A : opinions are not independent of gender.

Consider the **ROW** and **COLUMN TOTALS** entered above, then compute:

$E_{11} = \frac{100 \times 110}{200} = 55$	$E_{12} = \frac{100 \times 70}{200} = 35$	$E_{13} = \frac{100 \times 20}{200} = 10$
$E_{21} = \frac{100 \times 110}{200} = 55$	$E_{22} = \frac{100 \times 70}{200} = 35$	$E_{23} = \frac{100 \times 20}{200} = 10$

$$\chi_0^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$



$$\therefore \chi_0^2 = \frac{(50-55)^2}{55} + \frac{(40-35)^2}{35} + \dots + \frac{(10-10)^2}{10}$$

$$\therefore \chi_0^2 = 2.34 \not> 5.99146,$$ **YES**, opinions are independent of gender.

THE INDEPENDENT CLASSIFICATIONS TEST – EXERCISES

① Is a team's distribution of wins, losses, and ties, independent of the season? Conduct The Independent Classification Test with $\alpha = 0.05$ and the data below.

	WINS	LOSSES	TIES
LAST SEASON	56	11	13
THIS SEASON	60	5	15

② Determine whether the proportions of students passed and failed are independent of the teacher. Use $\alpha = 0.05$ and the following results to conduct an Independent Classifications Test.

	TEACHER A	TEACHER B	TEACHER C
PASSED	50	47	56
FAILED	5	14	8

③ Determine whether the proportion of right-handers is independent of sex. Use $\alpha = 0.05$ and the sample data below to conduct a test of hypothesis.

	RIGHT-HANDERS	LEFT-HANDERS
♂	2780	311
♀	3281	300

④ Is the proportion of Dawson students who go on to university independent of their sector? Conduct an Independent Classifications Test with $\alpha = 0.05$ and the tabled results below.

	SCIENCE SECTOR	ARTS SECTOR
GO TO UNIVERSITY	75	65
DO NOT GO TO UNIVERSITY	25	35

⑤ Use the results below to test whether the educational level completed is independent of the area of birth, at the $\alpha = 0.01$ level.

		AREA OF BIRTH	
EDUCATION	RURAL	TOWN	CITY
HIGH SCHOOL	75	54	12
COLLEGE	64	106	28
UNIVERSITY	28	82	51

(6) Are the opinions about a recent federal government initiative the same in Montreal and Toronto? Test using $\alpha = 0.05$ and the poll results below.

OPINIONS

	FOR	AGAINST	UNDECIDED
MONTREAL	28	15	7
TORONTO	20	21	9

(7) Two political polls, of 100 voters each, revealed the following preferences.

PREFERENCES

	PARTY A	PARTY B	PARTY C	PARTY D
1ST POLL	31	21	9	39
2ND POLL	41	15	15	29

Has there been a significant change in preferences since the 1ST poll? Conduct an Independent Classifications Test using $\alpha = 0.05$.

(8) A certain TV show is believed to be equally popular with viewers of either sex. Is this belief supported by the results below from a survey of 400 viewers? Test with $\alpha = 0.01$.

	WATCH THE SHOW	DON'T WATCH THE SHOW
♀	55	95
♂	65	185

(9) Can we conclude that the distribution of grades is the same at Dawson and Vanier based on the following data? Test using $\alpha = 0.05$.

GRADES

	A	B	C	D	E
DAWSON	15	25	50	5	5
VANIER	10	15	60	10	5

(10) Are there equal proportions of females and males at each of the education levels according to the table below? Test using $\alpha = 0.10$.

	ELEMENTARY	HIGH SCHOOL	CEGEP	UNIVERSITY
FEMALE	58	64	68	75
MALE	42	36	32	25

(11) Do apartment dwellers and home owners have the same opinion of a new city tax proposal, based on the data below? Test using $\alpha = 0.05$.

	IN FAVOUR	OPPOSED
APARTMENT DWELLERS	237	63
HOME OWNERS	141	59

(12) Are the feelings of citizens about their city the same for the 3 cities below? Conduct an Independent Classifications Test using $\alpha = 0.05$.

	FEELINGS OF CITIZENS		
	POSITIVE	NEGATIVE	NEUTRAL
MONTREAL	153	61	86
TORONTO	86	56	58
VANCOUVER	67	15	18

(13) Do the 4 machines (see below) perform the same? Test using $\alpha = 0.05$.

	NON-DEFECTIVE	ACCEPTABLE	DEFECTIVE
MACHINE A	35	48	17
MACHINE B	30	54	16
MACHINE C	33	46	21
MACHINE D	28	56	16

(14) Using the results of a political poll shown below, test whether age and party affiliation are independent. Use $\alpha = 0.05$.

		PARTY AFFILIATION		
		LIBERAL	CONSERVATIVE	NDP
AGE	< 25	18	36	46
	25–40	22	48	50
	> 40	44	92	44

(15) Consider the results below from a consumer study of the performance of 4 competing brands of toothpaste

	BRAND A	BRAND B	BRAND C	BRAND D
NO CAVITIES	9	13	17	11
1-5 CAVITIES	63	70	85	82
>5 CAVITIES	28	37	48	37

Test the hypothesis that the incidence of cavities is independent of the brand of toothpaste. Use $\alpha = 0.05$.

(16) Is the distribution of annual income independent of the city? Test with $\alpha = 0.05$ and the sample data tabled below.

	MONTREAL	TORONTO	VANCOUVER
< \$30000	57	27	47
\$30000 – \$50000	45	25	31
> \$50000	23	16	12

(17) Use the table below to test whether the distribution of scoring by quarter has been the same for the 3 decades. Use $\alpha = 0.05$.

		SCORING		
DECADE	1ST QUARTER	2ND QUARTER	3RD QUARTER	4TH QUARTER
70's	54	51	46	32
80's	150	144	90	112
90's	96	105	64	56

(18) Do the 5 regions of Canada have the same population proportions of urban, suburban and rural people? Test with $\alpha = 0.01$ and the results below.

	URBAN	SUBURBAN	RURAL
BRITISH COLUMBIA	45	45	10
WESTERN CANADA	70	50	30
ONTARIO	100	75	25
QUEBEC	65	55	30
MARITIMES	45	40	15

(19) Consider the following distribution of 100 Science and 100 Career grades. Can we conclude that the distribution of grades is the same for each sector? Conduct a test of hypothesis with $\alpha = 0.10$.

	90-100	80-89	70-79	60-69	0-59
SCIENCE	15%	25%	30%	15%	15%
CAREERS	10%	15%	40%	25%	10%

(20) To guage public opinion with regard to a proposed recreational facility, 2 polls were conducted, one by a polling agency, and the other by a local newspaper. Do the results to the right show that opinions are the same for each poll? Test with $\alpha = 0.05$.

	POLLING AGENCY	LOCAL NEWSPAPER
FAVOUR	52%	45%
AGAINST	36%	40%
UNDECIDED	12%	15%
SAMPLE SIZE	250	200

THE INDEPENDENT CLASSIFICATIONS TEST - EXERCISE SOLUTIONS

(1) H_0: Distr. indep. of season
H_A: Not independent

	W	L	T	
LAST	56 (58)	11 (18)	13 (14)	80
THIS	60 (58)	5 (8)	15 (14)	80
	116	16	28	160

(NOTE: E_{ij}'s are in brackets)

$\therefore \chi_o^2 = 2.53 \not> 5.99146$

$\therefore$ YES, independent

(2) H_0: Proportion of pass/fail indep. of teacher
H_A: Not independent

	A	B	C	
PASS	50 (46.75)	47 (51.85)	56 (54.40)	153
FAIL	5 (8.25)	14 (9.15)	8 (9.60)	27
	55	61	64	180

$\therefore \chi_o^2 = 4.84 \not> 5.99146$

$\therefore$ YES, independent

(3) H_0: Proportion of rt-handers indep. of sex
H_A: Not independent

	R	L	
♂	2780 (2807.93)	311 (283.06)	3091
♀	3281 (3253.06)	300 (327.93)	3581
	6061	611	6672

$\therefore \chi_o^2 = 5.65 > 3.84146$

$\therefore$ NO, not independent

(4) H_0: University prop. indep. of sector
H_A: Not independent

	SCIENCE	ARTS	
UNIVERSITY	75 (70)	65 (70)	140
NOT UNIVERSITY	25 (30)	35 (30)	60
	100	100	200

$\therefore \chi_o^2 = 2.38 \not> 3.84146$

$\therefore$ YES, independent

(5) H_0: Educ. level indep. of birth area
H_A: Not independent

	RURAL	TOWN	CITY	
H.S.	75 (47.1)	54 (68.2)	12 (25.2)	141
COLL.	64 (66.1)	106 (95.8)	28 (36.0)	198
UNIV.	28 (53.8)	82 (77.9)	51 (29.3)	161
	167	242	91	500

$\therefore \chi_o^2 = 58.63 > 13.2767$

$\therefore$ NO, not independent

(6) H_0: Opinions the same in each city
H_A: Not the same

	FOR	AGAINST	UNDECIDED	
MONTREAL	28 (24)	15 (18)	7 (8)	50
TORONTO	20 (24)	21 (18)	9 (8)	50
	48	36	16	100

$\therefore \chi_o^2 = 2.58 \not> 5.99146$

$\therefore$ YES, the same

(7) H_0: No change in preferences
H_A: Change in preferences

	A	B	C	D	
1st POLL	31 (36)	21 (18)	9 (12)	39 (34)	100
2nd POLL	41 (36)	15 (18)	15 (18)	29 (34)	100
	72	36	24	68	200

$\therefore \chi_o^2 = 5.36 \not> 7.81473$

$\therefore$ No change

(8) H_0: Equally popular by sex
H_A: Not equally popular by sex

	WATCH	DON'T WATCH	
♀	55 (45)	95 (105)	150
♂	65 (75)	185 (175)	250
	120	280	400

$\therefore \chi_o^2 = 5.08 \not> 6.6349$

$\therefore$ YES, equally popular by sex

(9) H_0: Distribution of grades is the same
H_A: Not the same

	A	B	C	D	E	
DAWSON	15 (12.5)	25 (20)	50 (55)	5 (7.5)	5 (5)	100
VANIER	10 (12.5)	15 (20)	60 (55)	10 (7.5)	5 (5)	100
	25	40	110	15	10	200

$\therefore \chi_o^2 = 6.08 \not> 9.48773$

$\therefore$ YES, the same

(10) H_0: Proportions the same at each level
H_A: Proportions not the same

	ELEM.	H.S.	CEGEP	UNIV.	
♀	58 (66.25)	64 (66.25)	68 (66.25)	75 (66.25)	265
♂	42 (33.75)	36 (33.75)	32 (33.75)	25 (33.75)	135
	100	100	100	100	400

$\therefore \chi_o^2 = 6.83 > 6.25139$

$\therefore$ NO, not the same

THE INDEPENDENT CLASSIFICATION TEST - EXERCISE SOLUTIONS

(11) H_0: Both have the same opinion
H_A: They don't have the same opinion

	(+)	(−)	
APT.	237 (226.8)	63 (73.2)	300
HOME	141 (151.2)	59 (48.8)	200
	378	122	500

$\therefore \chi_0^2 = 4.7 > 3.84146$

∴ No, not the same

(12) H_0: Feelings the same in each city
H_A: Feelings not the same

	(+)	(−)	ϕ	
MTL.	153 (153)	61 (66)	86 (81)	300
TOR.	86 (102)	56 (44)	58 (22)	200
VAN.	67 (51)	15 (22)	18 (27)	100
	306	132	162	600

$\therefore \chi_0^2 = 17.0 > 9.48773$

∴ No, feelings not the same

(13) H_0: The machines perform the same
H_A: The machines don't perform the same

	NON-DEF.	ACC.	DEF.	
A	35 (31.5)	48 (51.0)	17 (17.5)	100
B	30 (31.5)	54 (51.0)	16 (17.5)	100
C	33 (31.5)	46 (51.0)	21 (17.5)	100
D	28 (31.5)	56 (51.0)	16 (17.5)	100
	126	204	70	400

$\therefore \chi_0^2 = 3.23 \not> 12.5916$

∴ YES, perform the same

(14) H_0: Age and party affiliation are indep.
H_A: Not independent

	LIB.	CONS.	NDP	
< 25	18 (21.0)	36 (44.0)	46 (35.0)	100
25-40	22 (25.2)	48 (52.8)	50 (42.0)	120
> 40	44 (37.8)	92 (79.2)	44 (63.0)	180
	84	176	140	400

$\therefore \chi_0^2 = 16.53 > 9.48773$

∴ No, not independent

(15) H_0: Incidence of cavities indep. of brand
H_A: Not independent

	A	B	C	D	
0	9 (10)	13 (12)	17 (15)	11 (13)	50
1-5	63 (60)	70 (72)	85 (90)	82 (78)	300
> 5	28 (30)	37 (36)	48 (45)	37 (39)	150
	100	120	150	130	500

$\therefore \chi_0^2 = 1.91 \not> 12.5916$

∴ YES, independent

(16) H_0: Income independent of city
H_A: Not independent

	MTL.	TOR.	VAN.	
< \$30000	57 (57.86)	27 (31.48)	47 (41.66)	131
\$30000-50000	45 (44.61)	25 (24.27)	31 (32.12)	101
> \$50000	23 (22.53)	16 (12.25)	12 (16.22)	51
	125	68	90	283

$\therefore \chi_0^2 = 3.65 \not> 9.48773$

∴ YES, independent

(17) H_0: Distribution of scoring by quarter the same
H_A: Not the same

	1st	2nd	3rd	4th	
70's	54 (54.9)	51 (54.9)	46 (36.6)	32 (36.6)	183
80's	150 (148.8)	144 (148.8)	90 (99.2)	112 (99.2)	496
90's	96 (96.3)	105 (96.3)	64 (64.2)	56 (64.2)	321
	300	300	200	200	1000

$\therefore \chi_0^2 = 7.79 \not> 12.5916$

∴ YES, the same

(18) H_0: The regions have the same proportions
H_A: Not the same

	URBAN	SUB.	RURAL	
B.C.	45 (46.4)	45 (37.9)	10 (15.7)	100
W.C.	70 (69.6)	50 (56.8)	30 (23.6)	150
ONT.	100 (92.9)	75 (75.7)	25 (31.4)	200
QUE.	65 (69.6)	55 (56.8)	30 (23.6)	150
MAR.	45 (46.4)	40 (37.9)	15 (15.7)	100
	325	265	110	700

$\therefore \chi_0^2 = 10.23 \not> 20.0902$

∴ YES, the same

(19) H_0: Distribution of grades same for each sector
H_A: Distribution not the same by sector

	90-100	80-89	70-79	60-69	0-59	
SCIENCE	15 (12.5)	25 (20)	30 (35)	15 (20)	15 (12.5)	100
CAREERS	10 (12.5)	15 (20)	40 (35)	25 (20)	10 (12.5)	100
	25	40	70	40	25	200

$\therefore \chi_0^2 = 8.42 > 7.77944$

∴ NO, not the same

(20) H_0: Opinions the same for each poll
H_A: Opinions not the same for each poll

	POLL	NEWSPAPER	
FAVOUR	130 (122.22)	90 (97.78)	220
AGAINST	90 (94.44)	80 (75.56)	170
UNDECIDED	30 (33.33)	30 (26.67)	60
	250	200	450

$\therefore \chi_0^2 = 2.33 \not> 5.99146$

∴ YES, the same

APPENDIX

STATISTICAL TABLES

CUMULATIVE BINOMIAL PROBABILITY TABLES
– FOR $n = 10, 15, 25$ AND SELECTED p'S

THE STANDARD NORMAL PROBABILITY TABLE

CRITICAL VALUES OF t- DISTRIBUTION TABLE

CRITICAL VALUES OF F- DISTRIBUTION TABLE

CRITICAL VALUES OF χ^2- DISTRIBUTION TABLE

STATISTICAL FORMULAS

FORMULAE 1

FORMULAE 2

Cumulative Binomial Probability Table

$B(x;10,p)$

p	.05	.10	.20	.25	.30	.40
P(X = 0)	0.59874	0.34868	0.10737	0.05631	0.02825	0.00605
P(X ≤ 1)	0.91386	0.73610	0.37581	0.24403	0.14931	0.04636
P(X ≤ 2)	0.98850	0.92981	0.67780	0.52559	0.38278	0.16729
P(X ≤ 3)	0.99897	0.98720	0.87913	0.77588	0.64961	0.38228
P(X ≤ 4)	0.99994	0.99837	0.96721	0.92187	0.84973	0.63310
P(X ≤ 5)	1.00000	0.99985	0.99363	0.98027	0.95265	0.83376
P(X ≤ 6)	1.00000	0.99999	0.99914	0.99649	0.98941	0.94524
P(X ≤ 7)	1.00000	1.00000	0.99992	0.99958	0.99841	0.98771
P(X ≤ 8)	1.00000	1.00000	1.00000	0.99997	0.99986	0.99832
P(X ≤ 9)	1.00000	1.00000	1.00000	1.00000	0.99999	0.99990
P(X ≤ 10)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

AND

p	.50	.60	.70	.75	.80	.90	.95
P(X = 0)	0.00098	0.00010	0.00001	0.00000	0.00000	0.00000	0.00000
P(X ≤ 1)	0.01074	0.00168	0.00014	0.00003	0.00000	0.00000	0.00000
P(X ≤ 2)	0.05469	0.01229	0.00159	0.00042	0.00008	0.00000	0.00000
P(X ≤ 3)	0.17188	0.05476	0.01059	0.00351	0.00086	0.00001	0.00000
P(X ≤ 4)	0.37695	0.16624	0.04735	0.01973	0.00637	0.00015	0.00000
P(X ≤ 5)	0.62305	0.36690	0.15027	0.07813	0.03279	0.00163	0.00006
P(X ≤ 6)	0.82812	0.61772	0.35039	0.22412	0.12087	0.01280	0.00103
P(X ≤ 7)	0.94531	0.83271	0.61722	0.47441	0.32220	0.07019	0.01150
P(X ≤ 8)	0.98926	0.95364	0.85069	0.75597	0.62419	0.26390	0.08614
P(X ≤ 9)	0.99902	0.99395	0.97175	0.94369	0.89263	0.65132	0.40126
P(X ≤ 10)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Cumulative Binomial Probability Table

$B(x; 15, p)$

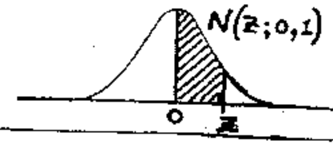
p	.05	.10	.20	.25	.30	.40
$P(X=0)$	0.46329	0.20589	0.03518	0.01336	0.00475	0.00047
$P(X \le 1)$	0.82905	0.54904	0.16713	0.08018	0.03527	0.00517
$P(X \le 2)$	0.96380	0.81594	0.39802	0.23609	0.12683	0.02711
$P(X \le 3)$	0.99453	0.94444	0.64816	0.46129	0.29687	0.09050
$P(X \le 4)$	0.99939	0.98728	0.83577	0.68649	0.51549	0.21728
$P(X \le 5)$	0.99995	0.99775	0.93895	0.85163	0.72162	0.40322
$P(X \le 6)$	1.00000	0.99969	0.98194	0.94338	0.86886	0.60981
$P(X \le 7)$	1.00000	0.99997	0.99576	0.98270	0.94999	0.78690
$P(X \le 8)$	1.00000	1.00000	0.99922	0.99581	0.98476	0.90495
$P(X \le 9)$	1.00000	1.00000	0.99989	0.99921	0.99635	0.96617
$P(X \le 10)$	1.00000	1.00000	0.99999	0.99988	0.99933	0.99065
$P(X \le 11)$	1.00000	1.00000	1.00000	0.99999	0.99991	0.99807
$P(X \le 12)$	1.00000	1.00000	1.00000	1.00000	0.99999	0.99972
$P(X \le 13)$	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997
$P(X \le 14)$	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
$P(X \le 15)$	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

p	.50	.60	.70	.75	.80	.90	.95
$P(X=0)$	0.00003	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$P(X \le 1)$	0.00049	0.00003	0.00000	0.00000	0.00000	0.00000	0.00000
$P(X \le 2)$	0.00369	0.00028	0.00001	0.00000	0.00000	0.00000	0.00000
$P(X \le 3)$	0.01758	0.00193	0.00009	0.00001	0.00000	0.00000	0.00000
$P(X \le 4)$	0.05923	0.00935	0.00067	0.00012	0.00001	0.00000	0.00000
$P(X \le 5)$	0.15088	0.03383	0.00365	0.00079	0.00011	0.00000	0.00000
$P(X \le 6)$	0.30362	0.09505	0.01524	0.00419	0.00078	0.00000	0.00000
$P(X \le 7)$	0.50000	0.21310	0.05001	0.01730	0.00424	0.00003	0.00000
$P(X \le 8)$	0.69638	0.39019	0.13114	0.05662	0.01806	0.00031	0.00000
$P(X \le 9)$	0.84912	0.59678	0.27838	0.14837	0.06105	0.00225	0.00005
$P(X \le 10)$	0.94077	0.78272	0.48451	0.31351	0.16423	0.01272	0.00061
$P(X \le 11)$	0.98242	0.90950	0.70313	0.53871	0.35184	0.05556	0.00547
$P(X \le 12)$	0.99631	0.97289	0.87317	0.76391	0.60198	0.18406	0.03620
$P(X \le 13)$	0.99951	0.99483	0.96473	0.91982	0.83287	0.45096	0.17095
$P(X \le 14)$	0.99997	0.99953	0.99525	0.98664	0.96482	0.79411	0.53671
$P(X \le 15)$	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Cumulative Binomial Probability Table

B(x;25,p)

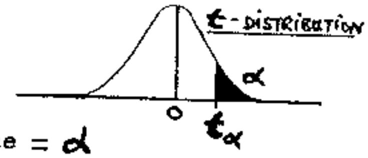
p	.05	.10	.20	.25	.30	.40
P(X = 0)	0.27739	0.07179	0.00378	0.00075	0.00013	0.00000
P(X ≤ 1)	0.64238	0.27121	0.02739	0.00702	0.00157	0.00005
P(X ≤ 2)	0.87289	0.53709	0.09823	0.03211	0.00896	0.00043
P(X ≤ 3)	0.96591	0.76359	0.23399	0.09621	0.03324	0.00237
P(X ≤ 4)	0.99284	0.90201	0.42067	0.21374	0.09047	0.00947
P(X ≤ 5)	0.99879	0.96660	0.61669	0.37828	0.19349	0.02936
P(X ≤ 6)	0.99983	0.99052	0.78004	0.56110	0.34065	0.07357
P(X ≤ 7)	0.99998	0.99774	0.89088	0.72651	0.51185	0.15355
P(X ≤ 8)	1.00000	0.99954	0.95323	0.85056	0.67693	0.27353
P(X ≤ 9)	1.00000	0.99992	0.98267	0.92867	0.81056	0.42462
P(X ≤ 10)	1.00000	0.99999	0.99445	0.97033	0.90220	0.58577
P(X ≤ 11)	1.00000	1.00000	0.99846	0.98927	0.95575	0.73228
P(X ≤ 12)	1.00000	1.00000	0.99963	0.99663	0.98253	0.84623
P(X ≤ 13)	1.00000	1.00000	0.99992	0.99908	0.99401	0.92220
P(X ≤ 14)	1.00000	1.00000	0.99999	0.99979	0.99822	0.96561
P(X ≤ 15)	1.00000	1.00000	1.00000	0.99996	0.99955	0.98683
P(X ≤ 16)	1.00000	1.00000	1.00000	0.99999	0.99990	0.99567
P(X ≤ 17)	1.00000	1.00000	1.00000	1.00000	0.99998	0.99879
P(X ≤ 18)	1.00000	1.00000	1.00000	1.00000	1.00000	0.99972
P(X ≤ 19)	1.00000	1.00000	1.00000	1.00000	1.00000	0.99995
P(X ≤ 20)	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999
P(X ≤ 21)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
P(X ≤ 22)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
P(X ≤ 23)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
P(X ≤ 24)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
P(X ≤ 25)	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000



THE STANDARD NORMAL PROBABILITY TABLE

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

(For $z = 4.0$ and $z = 5.0$, the areas are .49997 and .4999997)



Critical Values of Student's t-Distribution:

degrees of freedom	(One Tail) Level of Significance = α .10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

(9) A meterologist claims that the true mean annual rainfall in Montreal is more than 50 inches. A random sample of 9 recent years averaged 52.3 inches of rain, with $S = 4.0$ inches. Does this support the claim? Conduct a Test of Hypothesis with $\alpha = 0.05$.

(10) A vending machine is designed to discharge exactly 8 ounces of coffee into a cup. To test whether the machine is operating properly, 16 cups were selected at random, yielding $\bar{x} = 7.5$ g. and $S = 0.8$ g. Is the machine operating properly? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(11) Construct a 90% confidence interval estimate for the mean age of judges, if a random sample of 9 judges yielded (age in years): 47, 50, 54, 55, 55, 56, 57, 60, 61.

(12) Is it true that Dawson students take less than 45 minutes to travel to school? Conduct a Test of Hypothesis using $\alpha = 0.05$ and the data to the right for the travel times (minutes) of a sample of 16 students.

Σx	Σx^2
700	30718.75

(13) GM advertises that their new cars average at least 35 mpg. A random sample of 9 such cars averaged 33.86 mpg., with $S = 3.0$ mpg.
(a) Test the ad using $\alpha = 0.05$ (b) Construct a 95% confidence interval estimate of μ.

(14) A random sample of 25 CEGEP students had an average age of 18.4 years, with $S = 1.5$ yrs. Based on this, test the hypothesis that the true mean age of CEGEP students is more than 18 years. Use $\alpha = 0.10$.

(15) Seven measurements of the air quality in an office yielded these results in ppm. of CO: 3.7, 5.9, 4.5, 4.9, 3.2, 4.2, 4.0
Construct a 99% confidence interval estimate of the true mean level of CO (in ppm) in the office.

(16) Construct a 95% confidence interval estimate for the true mean diameter of steel pipes made by a certain manufacturer, 16 of whose pipes were measured, yielding (in mm.): $\Sigma x = 7968$, $\Sigma x^2 = 3968184$.

(17) A random sample of 10 Dawson students revealed the following numbers of homework hours per week: 12, 11, 8, 5, 20, 15, 17, 9, 10, 13.
Construct a 98% confidence interval estimate for the true mean number of homework hours per week for Dawson students.

(1) A chocolate bar is advertised as having an average weight of 25.0 grams. To test this ad, 9 bars were randomly selected and weighed, yielding $\bar{X}_0 = 25.8$ gms and $S = 1.2$ gms. Based on these results, conduct a Test of Hypothesis using $\alpha = 0.05$.

(2) Ten trials with a 3-minute egg-timer yielded $\bar{X}_0 = 3.18$ minutes and $S = 0.21$ min. Is the timer accurate? Conduct a Test of Hypothesis using $\alpha = 0.05$.

(3) A random sample of 25 Geminis yielded an average IQ of 104, with $S = 10$. Do Geminis have a higher IQ than the average person (whose IQ is 100)? Conduct a Test of Hypothesis using $\alpha = 0.01$.

(4) A paint manufacturer claims that, on the average, a gallon of its paint will cover at least 500 square feet. Using $\alpha = 0.10$ test this claim if a random sample of 16 gallons of this paint covered an average of 488 sq. ft. with $S = 32$ sq. ft.

(5) At Dawson's last registration, 16 students were timed. They averaged 50 minutes, with a standard deviation of 10 min. to complete their registration. Construct a 95% confidence interval estimate for the true mean time that it took students to complete their last registration at Dawson.

(6) Construct a 95% confidence interval estimate for the true mean weight of diamonds cut by a new machine if a random sample of 6 diamonds cut by the machine averaged 0.53 karats with $S = 0.0559$ karats.

(7) A random sample of 25 new textbooks from the bookstore yielded:

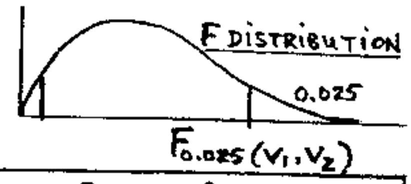
$$\Sigma X = \$521.25 \quad \text{and} \quad \Sigma X^2 = \$^2 10989.56$$

Based on these results, construct a 95% and a 90% confidence interval estimate of the true average cost of new textbooks.

(8) Eight randomly selected Dawson students were each asked to list the number of hours that they watched TV per week. The results were:

12, 7, 16, 10, 11, 5, 2, 15

Construct the 99% confidence interval estimate for the mean number of hours of TV watched per week by Dawson students.

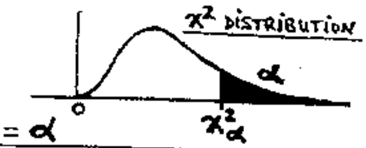


Critical Values of Snedecor's F Distribution:

v_2 \ v_1	1	2	3	4	5	6	7	8	9
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473
4	12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047
5	10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6810
6	8.8131	7.2598	6.5988	6.2272	5.9876	5.8197	5.6955	5.5996	5.5234
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8994	4.8232
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4332	4.3572
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1971	4.1020	4.0260
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8800
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9024	2.8077	2.7313
24	5.7167	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309
28	5.6096	4.2205	3.6264	3.2863	3.0625	2.9027	2.7820	2.6872	2.6106
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519
60	5.2857	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344
120	5.1524	3.8046	3.2270	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217
∞	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136

V_1 = df of Numerator

V_2 = df of Denominator



Critical Values of the Chi-Square Distribution:

degrees of freedom	Level of Significance = α				
	.10	.05	.01	.005	.001
1	2.70554	3.84146	6.63490	7.87944	10.828
2	4.60517	5.99146	9.21034	10.5966	13.816
3	6.25139	7.81473	11.3449	12.8382	16.266
4	7.77944	9.48773	13.2767	14.8603	18.467
5	9.23636	11.0705	15.0863	16.7496	20.515
6	10.6446	12.5916	16.8119	18.5476	22.458
7	12.0170	14.0671	18.4753	20.2777	24.322
8	13.3616	15.5073	20.0902	21.9550	26.125
9	14.6837	16.9190	21.6660	23.5894	27.877
10	15.9872	18.3070	23.2093	25.1882	29.588
11	17.2750	19.6751	24.7250	26.7568	31.264
12	18.5493	21.0261	26.2170	28.2995	32.909
13	19.8119	22.3620	27.6882	29.8195	34.528
14	21.0641	23.6848	29.1412	31.3194	36.123
15	22.3071	24.9958	30.5779	32.8013	37.697
16	23.5418	26.2962	31.9999	34.2672	39.252
17	24.7690	27.5871	33.4087	35.7185	40.790
18	25.9894	28.8693	34.8053	37.1565	42.312
19	27.2036	30.1435	36.1909	38.5823	43.820
20	28.4120	31.4104	37.5662	39.9968	45.315
21	29.6151	32.6706	38.9322	41.4011	46.797
22	30.8133	33.9244	40.2894	42.7957	48.268
23	32.0069	35.1725	41.6384	44.1813	49.728
24	33.1962	36.4150	42.9798	45.5585	51.179
25	34.3816	37.6525	44.3141	46.9279	52.618
26	35.5632	38.8851	45.6417	48.2899	54.052
27	36.7412	40.1133	46.9629	49.6449	55.476
28	37.9159	41.3371	48.2782	50.9934	56.892
29	39.0875	42.5570	49.5879	52.3356	58.301
30	40.2560	43.7730	50.8922	53.6720	59.703
40	51.8051	55.7585	63.6907	66.7660	73.402
50	63.1671	67.5048	76.1539	79.4900	86.661
60	74.3970	79.0819	88.3794	91.9517	99.607
70	85.5270	90.5312	100.425	104.215	112.317

FORMULAE 1

DESCRIPTIVE STATISTICS

$$\bar{x} = \frac{\sum x}{n} \qquad \bar{x} \approx \frac{\sum x \cdot f}{n}$$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \qquad s \approx \sqrt{\frac{\sum x^2 f - \frac{(\sum x \cdot f)^2}{n}}{n-1}}$$

LINEAR REGRESSION AND CORRELATION

$$SS_x = \sum x^2 - \frac{(\sum x)^2}{n}, \quad SS_y = \sum y^2 - \frac{(\sum y)^2}{n}, \quad SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$y = a + bx, \quad b = \frac{SS_{xy}}{SS_x}, \quad a = \bar{y} - b\bar{x}, \text{ and } r = \frac{SS_{xy}}{\sqrt{SS_x \, SS_y}}$$

PROBABILITY

$$P(n,r) = \frac{n!}{(n-r)!} \qquad C(n,r) = \frac{n!}{r!\,(n-r)!}$$

$P(E \cup F) = P(E) + P(F) - P(E \cap F)$ $\qquad$ $P(E \cap F) = P(E) \cdot P(F/E)$

$P(E \cup F) = P(E) + P(F)$ if E and F are mutually exclusive $\qquad$ $P(E \cap F) = P(E) \cdot P(F)$ if E and F are independent

PROBABILITY DISTRIBUTIONS

DISCRETE: $\mu = E(x) = \sum x\,P(x)$ $\qquad$ CONTINUOUS: $\mu = E(x) = \int_{-\infty}^{\infty} x\,f(x)\,dx$

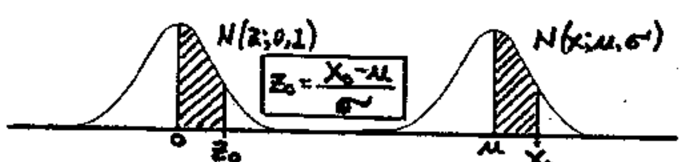
$$\sigma^2 = E[(x-\mu)^2] = E(x^2) - \mu^2$$

THE BINOMIAL PROBABILITY DISTRIBUTION: $B(x;n,p)$

$$P(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x} \text{ for } x = 0, 1, 2, \cdots, n \text{ and } q = 1-p$$

where $\mu = np$ and $\sigma = \sqrt{npq}$

THE NORMAL PROBABILITY DISTRIBUTION: $N(x;\mu,\sigma)$



$$\boxed{P(0 < z < z_0) = P(0 < x < x_0)}$$

$B(x;n,p) \approx N(x;\,np,\,\sqrt{npq})$

$$z_0 = \frac{(x_0 \pm \tfrac{1}{2}) - np}{\sqrt{npq}}$$

$N(\bar{x};\,\mu,\,s/\sqrt{n})$

$$z_0 = \frac{\bar{x}_0 - \mu}{s/\sqrt{n}}$$

FORMULAE 2

INFERENCES INVOLVING ONE POPULATION

$$\bar{x}_o \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{and} \quad z_o = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$$

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}, \quad \hat{p} = \frac{x_o}{n}, \hat{q} = 1 - \hat{p} \quad \text{and} \quad z_o = \frac{\frac{x_o}{n} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

$$\bar{x}_o \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{and} \quad t_o = \frac{\bar{x}_o - \mu_o}{s/\sqrt{n}}, \ (n-1) \text{ D. of F.}$$

INFERENCES INVOLVING TWO POPULATIONS

$$\bar{d} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{and} \quad t_o = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}, \ (n-1) \text{ D. of F.}$$

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and} \quad z_o = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t_o = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ where } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}, \ (n_1+n_2-2) \text{ D. of F.}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \quad \text{and} \quad z_o = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

where $\hat{p}_1 = \frac{x_1}{n_1}$, and $\hat{p}_2 = \frac{x_2}{n_2}$
$\hat{q}_1 = 1 - \hat{p}_1$, and $\hat{q}_2 = 1 - \hat{p}_2$

where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$
$\hat{q} = 1 - \hat{p}$

INFERENCES USING THE CHI-SQUARED DISTRIBUTION

GOODNESS OF FIT: $\chi_o^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$, with $E_i = np_i$, and $(k-1)$ D. of F.

INDEPENDENT CLASSIFICATIONS: $\chi_o^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, with $E_{ij} = \frac{r_i \times c_j}{n}$, and $(r-1)\times(c-1)$ D. of F.

All Coursepacks are Final Sale

STATISTICS BY MCARTHUR

MCARTHUR

978-1-5501635-7-5

90000

9781550163575

For more information concerning bookstore return policy, hours of operation, or to leave a comment:

http://www.dawsoncollege.qc.ca/our-services/bookstore