# Assignment 1 - SOLUTIONS

Statistics for Social Science (201-401-DW)
Instructor: Emilie Richer

**Instructions:**

- The assignment is comprised of 10 questions and marked out of a total of **40 marks.**

**[QUESTION 1] (5 marks)**

a. Describe and explain the difference between the mean and the median.

The median is the middle point of a data set. It splits the data in two, with half the data below the median and the other half above.

The mean is computed by summing each data point and dividing by the total number of data points.

b. Make up an example (not from our lectures) in which the median would be the preferred measure of central tendency.

Consider a class with 5 students whose part-time employments generate the following monthly incomes:

$X_1$ = \$500, $x_2$ = \$250, $x_3$ = \$300, $x_4$ = 100, $x_5$ = \$6,000

We find: median = \$300 and mean = \$1430

The median gives a clearer idea of the typical monthly income of the 5 students. One student, an outlier, has a large monthly income which significantly raises the mean and might have us assume that the typical monthly income is higher than is truly the case.

c. Make up an example (not from our lectures) in which the mean would be the preferred measure of central tendency.

Consider the weekly winnings of 5 participants who play a scratch-and-win lottery card once a week:

$X_1$ = \$0, $x_2$ = \$0, $x_3$ = \$0, $x_4$ = 100, $x_5$ = \$100

We find: median = \$0 and mean = \$40

In this case the median does not give very much information about the potential size of lottery winnings, however the mean does.

**[QUESTION 2] (5 marks)**
A researcher is studying the amygdala (a part of the brain involved in emotion). Twenty-five (25) participants in a particular fMRI brain scan study are measured for the increase in activation of their amygdala while they are viewing pictures of violent scenes. The activation increases measured are the following:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.43 | 0.32 | 0.64 | 0.21 | 0.29 | 0.51 | 0.00 | 0.19 | 0.44 | 0.55 | 0.31 |
| 0.27 | 0.50 | 0.40 | 0.47 | 0.30 | 0.19 | 0.75 | 0.31 | 0.10 | 0.15 | 0.27 |
| 0.41 | 0.29 | 0.21 | | | | | | | | |

a. Compute the mean and standard deviation for this sample.

Using a calculator we find sample mean = $\bar{x}$ **= 0.3403** and sample standard deviation = **s = 0.1716**

b. Explain what you have done and what the results mean as if you were explaining to a person who has never had a course in statistics.

The mean is an average which is computed by summing up all the data points and dividing by the total number of data points. In this study we can say that the average increase in activation of the amygdala among the 25 participants is 0.3403.

The standard deviation measures the spread of the data. Roughly speaking, we can say that the "average distance from the mean" in our data set is 0.1716.

**[QUESTION 3] (4 marks)**
Make up two sets of data containing 5 data points each: a. one with the mean greater than the median, b. one with the median and the mean the same.

a. $X_1 = 1$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$

median = 2, mean = 2.2

b. $X_1 = 1$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 3$

median = 2, mean = 2

**[QUESTION 4] (1.5 marks)**
A distribution has a mean of 200 and a standard deviation of 50. A person has a Z-score of 1.26, what is the person's actual score.

$z = (x - \mu)/\sigma$

$1.26 = (x - 200)/50$

$X = (50)(1.26) + 200 = 263$

The person's actual score is **x = 263**

**[QUESTION 5] (6 marks)**
Six months after a divorce, a former wife and husband each take a test that measures divorce adjustment. The higher the score, the better a person has adjusted to divorce. The former wife's score is 63, and the former husband's score is 59. Overall, in the population, the mean score for divorced women on this test is 60 with standard deviation 6, the mean score for divorced men is 55 with standard deviation 4.

Between the former wife and husband who took the test, which of the two has adjusted better to the divorce in relation to other divorced people of the same gender? Explain your answer as if you were explaining to a person who has never had a course in statistics.

We begin by computing the former wife and husband's respective z-scores.

$Z_{husband} = (59-55)/4 = \mathbf{1}$

$Z_{wife} = (63-60)/6 = \mathbf{0.5}$

While the wife's overall adjustment score is higher than the husbands (63 vs. 59), her score is not as high relative to her gender group.

We can see that the wife's score is a half standard deviation from the mean of her gender group and the husband's score is a full standard deviation from the mean of his gender group.

This means that, relative to other men, **the husband has adjusted better** to the divorce than the wife has relative to other women.

**[QUESTION 6] (3.5 marks)**
You are conducting a survey at a College with 800 students, 50 faculty members, and 150 administrators. Each of these 1,000 individuals has a single email address listed in the online campus directory. Suppose you were to select one email address at random. What is the probability it would be the email of:
a. a student;
b. a faculty member;
c. an administrator;
d. a faculty member or administrator;
e. anyone but an administrator?

a. 800/1000 = **0.8**
b. 50/1000 = **0.05**
c. 150/1000 = **0.15**
d. 200/1000 = **0.2**
e. 850/1000 = **0.85**

**[QUESTION 7] (3 marks)**
Consider the two events A = "attending classes regularly" and P = "passing all classes". Would you guess these events to be disjoint? independent?

We would guess that these events are **not disjoint** as they can occur simultaneously.

We would guess that these events are **not independent** because the chance of passing all classes is affected by attendance.

**[QUESTION 8] (6 marks)**
Consider a cegep class with a mean of 71.2, standard deviation of 5.1. The average of sec IV and V compulsory courses for the students in the class is 78 (refer to your notes for the R-score formula).
a. What is the R-score of a student in the class whose grade is 86?
b. Compare and discuss the effects on the R-score of (A) a doubling of the standard deviation and (B) a decreasing by 7 of the students' average of sec IV and V compulsory courses
- for a student with a grade of 86 in the class;
- for a student with a grade of 51 in the class.

**a.**
**ISG** = [HS AVG – 75]/14 = [78-75]/14 = **0.214**
**z** = (x – μ)/σ = (86 – 71.2)/5.1 = **2.9**
**R-score** = (z + ISG + 5) x 5 = (2.9 + 0.214 + 5) x 5 = **40.6**

**b.**

**Student with 86**

R-score = 40.6

Doubling standard deviation

σ = 10.2

new z-score
= (86 - 71.2)/10.2
= 1.45

new R-score
= (1.45 + 0.214 + 5) x 5
= **33.3**

The doubling of standard deviation has the **effect of reducing the r-score.**

Decrease of highschool avg.

new ISG = [71 - 75]/14 = -0.286

new R-score
= (2.9 -0.286 + 5) x 5
= **38.07**

The decrease of highschool avg. deviation has the **effect of reducing the r-score. The effect is less than that of doubling the standard deviation**

**Student with 51**

z = (51 – 71.2)/5.1 = -3.96

R-score
= (-3.96 + 0.214 + 5) x 5
= **6.27**

Doubling standard deviation

σ = 10.2

new z-score
= (51 - 71.2)/10.2
= -1.98

new R-score
= (-1.98 + 0.214 + 5) x 5
= **16.2**

The doubling of standard deviation has the **effect of increasing the r-score.**

Decrease of highschool avg.

new R-score
= (-3.96 -0.286 + 5) x 5
= **3.77**

The decrease of highschool avg. deviation has the **effect of reducing the r-score.**

**[QUESTION 9] (4.5 marks)**
The 268 first year students registered in a Social Science Program at Dawson College fill out a survey regarding their interest in two subjects. The survey and its results are showed below:

| **Subject Interest Survey** |
|---|
| Which of the following subject(s) would you like to study over the course of your College studies (check all boxes that apply): |
| ☐ Biology<br>☐ Mathematics |

| **Results of Subject Interest Survey** |
|---|
| Number of students who checked box: |
| Biology only (67)<br>Mathematics only (81)<br>Biology and Mathematics (34) |

Let M be the event "student checked the mathematics box"
Let B be the event "student checked the biology box"

If one of the 268 Social Science students is picked at random, compute the probability of the following events:

a. $P(M)$
b. $P(B)$
c. $P(M \cap B)$
d. $P(M \cup B)$
e. $P(B|M)$
f. $P(M|B)$

a. $P(M) = (81 + 34)/268 =$ **115/268**
b. $P(B) = (67 + 34)/268 =$ **101/268**
c. $P(M \cap B) =$ **34/268**
d. $P(M \cup B) = P(M) + P(B) - P(M \cap B) = (115 + 101 - 34)/268 =$ **182/268**
e. $P(B|M) = P(M \cap B)/P(M) =$ **34/115**
f. $P(M|B) = P(M \cap B)/P(B) =$ **34/101**

**[QUESTION 10] (1.5 marks)**
Fifty percent of students at a particular college are commuters. Of those, 10% bike to school Find the probability that a student is a commuter and bikes to school.

C = student commutes
B = student bikes to school

$P(C \cap B) = P(B|C) \times P(C) = (0.10)\ (0.5) =$ **0.05**