2.12 $\bar{x} = \sum x/n = 1429/14 = \underline{\$102.07}$

2.13
a. $\bar{x} = \sum x/n = 367/10 = \underline{36.7}$
b. ranked data: 10 12 20 20 30 35 37 50 53 100
$d(\tilde{x}) = (n+1)/2 = (10+1)/2 = 5.5\text{th}$; $\tilde{x} = \underline{32.5}$
c. the mean is pulled towards the higher values due to the 100
d. $\bar{x} = \sum x/n = 267/9 = \underline{29.7}$; $d(\tilde{x}) = (n+1)/2 = (9+1)/2 = 5\text{th}$; $\tilde{x} = \underline{30}$
e. mean

2.14 mode = $\underline{2}$

2.15
a. midrange = (L+H)/2 = (138+349)/2 = 487/2 = $\underline{243.5}$
b. 138 and 349 are data, 100 and 350 are contest limits, not data.

2.16
a. $\bar{x} = \sum x/n = (2+4+7+8+9)/5 = 30/5 = \underline{6.0}$
b. $d(\tilde{x}) = (n+1)/2 = (5+1)/2 = 3\text{rd}$; $\tilde{x} = \underline{7}$
c. no mode, no value repeats
d. midrange = (H+L)/2 = (9+2)/2 = 11/2 = $\underline{5.5}$

2.17
a. The data value x = 45 is 12 units above the mean; therefore the mean must be 33.
b. The data value x = 84 is 20 units below the mean; therefore the mean must be 104.

2.18
a. "Only zero and positive values can occur." Or "The smallest value would be "zero" or "All non-zero values are positive"
b. There would be "no variation." That is "all values would be the same."
c. The existence of any variation in the data, that is, at least one value is different from the others.

2.19
a. range = H - L = 9 - 2 = $\underline{7}$
b. 1st: find mean, $\bar{x} = \sum x/n = 30/5 = 6$

	x	$x - \bar{x}$	$(x - \bar{x})^2$
	2	-4	16
	4	-2	4
	7	1	1
	8	2	4
	9	3	9
$\sum$	30	0	34

$s^2 = \sum(x-\bar{x})^2/(n-1)$

$= 34/4 = \underline{8.5}$

c. $s = \sqrt{s^2} = \sqrt{8.5} = 2.915 = \underline{2.9}$

2.20 a. 1st: find mean, $\overline{x} = \sum x/n = 104/15 = 6.9$

x	$x - \overline{x}$	$(x - \overline{x})^2$
4	-2.9	8.41
5	-1.9	3.61
5	-1.9	3.61
6	-0.9	0.81
6	-0.9	0.81
6	-0.9	0.81
7	0.1	0.01
7	0.1	0.01
7	0.1	0.01
7	0.1	0.01
8	1.1	1.21
8	1.1	1.21
8	1.1	1.21
9	2.1	4.41
11	4.1	16.81
$\sum$ 104	+0.5*	42.95

$s^2 = \sum(x-\overline{x})^2/(n-1)$

$= 42.95/14$

$= 3.0679$

$= \underline{3.1}$

*The 0.5 is due to the round-off error introduced by using $\overline{x} = 6.9$ instead of 6.933333.

b.

x	x^2
4	16
5	25
5	25
6	36
6	36
6	36
7	49
7	49
7	49
7	49
8	64
8	64
8	64
9	81
11	121
$\sum$ 104	764

$SS(x) = \sum x^2 - ((\sum x)^2/n)$

$= 764 - ((104)^2/15)$

$= 764 - 721.0667 = 42.93333$

$s^2 = SS(x)/(n-1)$

$= 42.9333/14 = 3.0667 = \underline{3.1}$

c. $s = \sqrt{s^2} = \sqrt{3.0667} = 1.751 = \underline{1.8}$

2.21 Set 1: $\bar{x}$ = 250/5 = 50

x	$x - \bar{x}$	$(x - \bar{x})^2$
45	-5	25
80	30	900
50	0	0
45	-5	25
30	-20	400
250	0	1350

Set 2: $\bar{x}$ = 250/5 = 50

x	$x - \bar{x}$	$(x - \bar{x})^2$
30	-20	400
80	30	900
35	-15	225
30	-20	400
75	25	625
250	0	2550

	$\sum x$	$\sum(x - \bar{x})$	$\sum(x - \bar{x})^2$	Range
Set 1:	250	0	1350	50
Set 2:	250	0	2550	50

The two sets have the same mean and same range; Set 2 is more dispersed.

2.22 The statement is incorrect. The standard deviation can never be negative. There has to be an error in the calculations or a typographical error in the statement.

2.23 a. 91 is in the 44th position from the Low value of 39
91 is in the 7th position from the High value of 98

b. nk/100 = (50)(20)/100 = 10.0; therefore $d(P_{20})$ = 10.5th from L
P_{20} = (64+64)/2 = <u>64</u>

nk/100 = (50)(35)/100 = 17.5; therefore $d(P_{35})$ = 18th from L
P_{35} = <u>70</u>

c. nk/100 = (50)(20)/100 = 10.0; therefore $d(P_{80})$ = 10.5th from H
P_{80} = (88+89)/2 = <u>88.5</u>

nk/100 = (50)(5)/100 = 2.5; therefore $d(P_{95})$ = 3rd from H
P_{95} = <u>95</u>

2.24 ranked data:

1.4 2.3 2.4 2.6 2.6 2.7 2.7 2.8 2.8 2.9 2.9 2.9 3.0 3.1 3.1 3.2 3.3 3.4 3.5
3.5 3.6 3.7 3.7 3.9 3.9 4.0 4.0 4.0 4.1 4.1 4.2 4.2 4.2 4.4 4.4 4.5 4.6 4.6
4.6 4.7 4.8 4.8 4.8 4.9 5.2 5.2 5.5 5.6 5.7 6.5 7.0 13.3

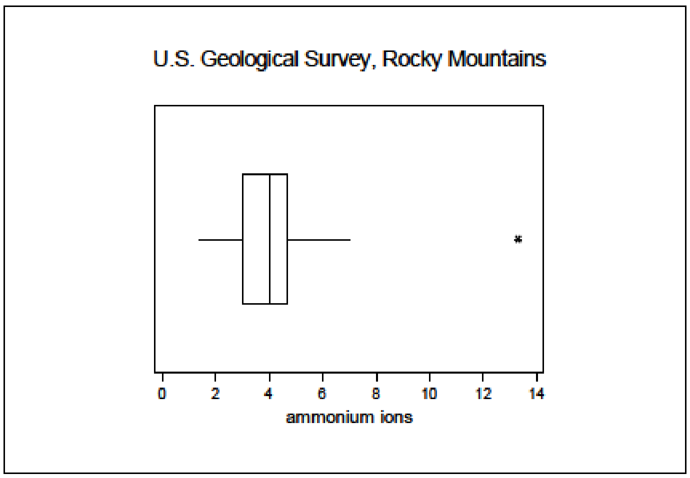
a. nk/100 = (52)(25)/100 = 13; therefore $d(Q_1) = 13.5^{th}$
   Q1 = (3.0+3.1)/2 = 3.05
b. d(median) = (52+1)/2 = 26.5^{th}, Q_2 = median = (4.0+4.0)/2 = 4.0

c. Q_3 = (4.6+4.7)/2 = 4.65

d. midquartile = (3.05+4.65)/2 = 3.85

e. nk/100 = (52)(30)/100 = 15.6; therefore $d(P_{30}) = 16^{th}$
   P_{30} = 3.2
f. 5-number summary: 1.4, 3.05, 4.0, 4.65, 13.3

g.



2.25 z = (x - mean)/st.dev.

a. for x = 54, z = (54 - 74.2)/11.5 = <u>-1.76</u>
b. for x = 68, z = (68 - 74.2)/11.5 = <u>-0.54</u>
c. for x = 79, z = (79 - 74.2)/11.5 = <u>0.42</u>
d. for x = 93, z = (93 - 74.2)/11.5 = <u>1.63</u>

2.26 If z = (x - mean)/st.dev; then x = (z)(st.dev) + mean

a. for z = 0.0, x = (0.0)(20.0) + 120 = <u>120</u>
b. for z = 1.2, x = (1.2)(20.0) + 120 = <u>144.0</u>
c. for z = -1.4, x = (-1.4)(20.0) + 120 = <u>92.0</u>
d. for z = 2.05, x = (2.05)(20.0) + 120 = <u>161.0</u>

2.27		English	Math
	a.	(30-20.6)/6.0 = 1.57	(30-21.0)/5.1 = 1.76
	b.	(23-20.6)/6.0 = 0.40	(23-21.0)/5.1 = 0.39
	c.	(12-20.6)/6.0 = -1.43	(12-21.0)/510 = -1.76

d. The relative size of the difference from the mean and the standard deviation caused the reversal of position.

e. Eng. z = 0.90, Math z = 0.98, Read. z = 0.75, Science z = 1.06, Composite z = 1.00. Therefore Science, it has the highest positive z-score.

2.28 a. ≈ 68%

b. ≈ 95%

c. ≈ 99.7%

2.29 The interval 22,500 to 37,500 represents the mean plus or minus three standard deviations.

a. If the distribution is normal, then approximately 99.7% of the distribution is contained within the interval.

b. If nothing is known about the shape of the distribution, then we can be sure that at least 89% of the distribution is contained within the interval.

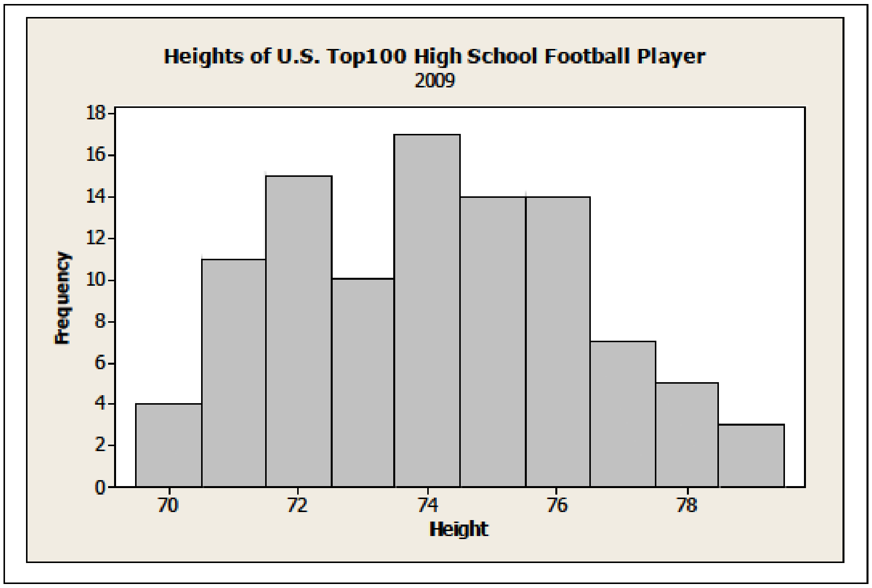
2.30 a. 50%

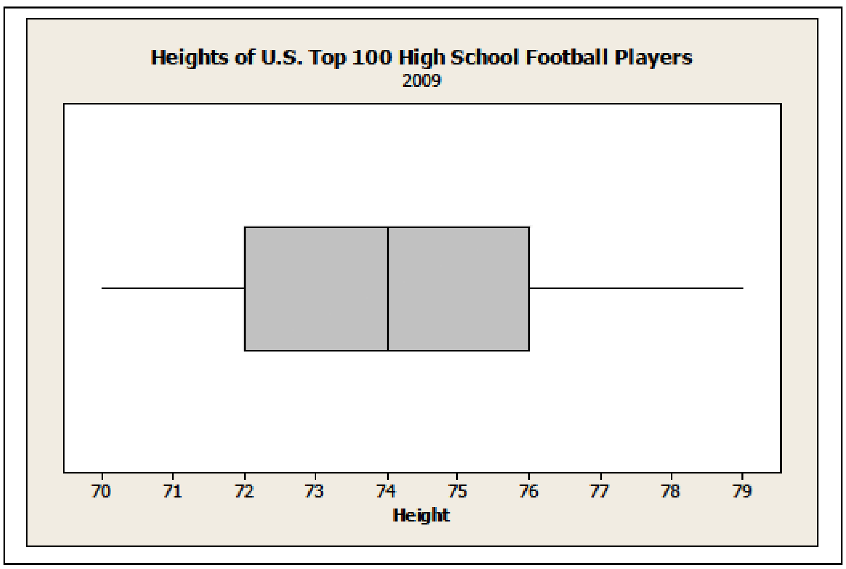
b. 0.50 – 0.34 = 0.16 = 16%

c. 0.50 + 0.34 = 0.84 = 84%

d. 0.34 + 0.475 = 0.815 = 81.5%

2.31 a.





b. $\bar{x} = 74.09$, $s = 2.292$

c. Ranked data:

70	70	70	70	71	71	71	71	71	71	71	71	71	71	71	72	72	72	72	72	72
72	72	72	72	72	72	72	72	72	73	73	73	73	73	73	73	73	73	73	74	74
74	74	74	74	74	74	74	74	74	74	74	74	74	74	74	75	75	75	75	75	75
75	75	75	75	75	75	75	75	76	76	76	76	76	76	76	76	76	76	76	76	76
76	77	77	77	77	77	77	77	78	78	78	78	78	79	79	79					

d. $\overline{x} \pm 1s = 74.09 \pm (2.292)$ or 71.798 to 76.382
70% of the data (70/100) is between 71.798 and 76.382.

$\overline{x} \pm 2s = 74.09 \pm 2(2.292) = 74.09 \pm 4.584$ or 69.506 to 78.674
97% of the data (97/100) is between 69.506 and 78.674.

$\overline{x} \pm 3s = 74.09 \pm 3(2.292) = 74.09 \pm 6.876$ or 67.214 to 80.966
100% of the data (100/100) is between 67.214 and 80.966.

e. The empirical rule says approximately 68%, 95%, and 99.7% of the data are within one, two, and three standard deviations, respectively; the 70%, 97% and 100% do somewhat agree with the rule; based solely on this information, the distribution can be considered 'approximately' normal.

f. Chebyshev's theorem says at least 75%, and 89%, of the data are within two, and three standard deviations, respectively; 97%, and 100% both satisfy the theorem.

g. The graphs indicate an approximately normal distribution. The histogram is mounded in the center and the boxplot's middle 50% is just about centered between the two whiskers. Both show a slight skewness to the right.

h. The points are following the straight line except at the two extremes. The p-value is less than 0.005 indicating that the data is not normally distributed. This result is contrary to the results found in part (e), thereby encouraging various forms of testing for normality.