Tony's Research

Exercises Answer Key

Descriptive Statistics

Section 1.0: Introduction and Scales of Measurements
Section 2.1: Central Tendency
Section 2.2: Frequency
Section 2.3: Stem-and-Leaf Display
Section 3.1: Variability
Section 3.2: Boxplot and Outliers
Section 4.1: Standard Score
Section 4.2: Normal Curve
Section 5.1: Pearson Correlation
Section 5.2: Spearman Correlation

Inference Statistics

Section 6.1: Introduction to Hypothesis Testing
Section 6.2: One-Sample Mean for Known Sigma
Section 6.3: One-Sample Mean for Unknown Sigma
Section 7.1: One-Sample Correlation (ρ = 0)
Section 7.1: One-Sample Correlation (ρ = a)
Section 8.1: Two-Sample Means (Independent Samples and Homogeneous Variance)
Section 8.2: Two-Sample Means (Independent Samples and Non-homogeneous Variance)
Section 8.3: Two-Sample Means (Dependent Samples)
Section 9.1: Linear Regression Equations and Plots
Section 9.2: Hypothesis Testing of Linear Regression
Section 10.1: Chi- Square Goodness of Fit Test
Section 10.1: Chi- Square Test for Independence

***Section 1.0*: Basic Math and Scales of Measurements**

Exercises 1.0

1. a. R; b. N; c. R; d. I; e. O; f. I; g. O; h. N
2. a. N; b. I; c. R; d. N; e. I; f. O; g. R
3. a. I_{Low} = 40 - 2.36(v7) = **33.7560** b. I_{High} = 40 + 2.36(v7) = **46.2440**
4. *M* = 50 – 3(0.84) = **47.48**
5. 16 + (15 – 8) * 5 – 2 = 16 + (7 * 5) – 2 = **49**
6. 7^2 + 1*2 = 49 + 2 = **51**
7. 2^3 + 3*2 + 15/3 = 8 + 6 + 5 = **19**
8. Computation table

X	*Y*	*XY*	X^2
5	4	20	25
4	3	12	16
6	2	12	36
3	1	3	9
2	0	0	4
∑X	**∑Y**	**∑XY**	**$\sum X^2$**
20	**10**	**47**	**90**

9. Computation table

X	*X – 12*	*X-11.2*
13	1	1.8
12	0	0.8
8	-4	-3.2
14	2	2.8
9	-3	-2.2
∑X	**∑(X-12)**	**∑(X- 11.2)**
56	**-4**	**0**

10. Given ∑X = 20 and ∑XY = 40
a. $(\sum X)^2 = 20^2$ **= 400**
b. ∑X(XY) = 20 * 40 = **800**
c. *M* = 3 since 2+5+M+7+3 = 20, when M = 3

11. a. 3(0.2347) = **0.7041** ; b. (0.2347)2 = **0.0551**; c. (0.2347)(0.4831) = **0.1134**

12. Given ∑X = 12, ∑Y = 25, ∑XY = 360

$$\frac{\sum XY - \sum X(\sum Y)}{(\sum Y)^2 - \sum Y} = \frac{360 - 12(25)}{25^2 - 25} = \frac{360 - 300}{625 - 25} = \frac{60}{600} = \frac{1}{10} = 0.1$$

Exercises 2.1

1. a. Mean = **1**, median = **1.5**, mode = **8**
   b. Mean = **0.46**, median = **0.395**, mode = **0.34**
   c. Mean = **4.8182**, median = **5**, mode = **1**
2. a. Mean = **2.3832**, median = **2.2361**, mode = **v3**
   b. Mean = **34.375**, median = **33**, mode = **33**
   c. Mean = **0.1**, median = **0**, mode = **1**
3. GPA = 43/16 = **2.69**

Course	Credits	Grade	Cr x Gr
English	3	4	12
Mathematics	4	2	8
Biology	4	3	12
History	2	1	2
Spanish	3	3	9
	∑Cr = 16		∑Gr = 43

4. Becky must get a **94** on the next quiz.

$$\frac{65+80+85+78+78+X}{6} = 80 = \frac{386+X}{6}$$

$80(6) = 386 + X;$ So $X = 480 - 386 = 94$

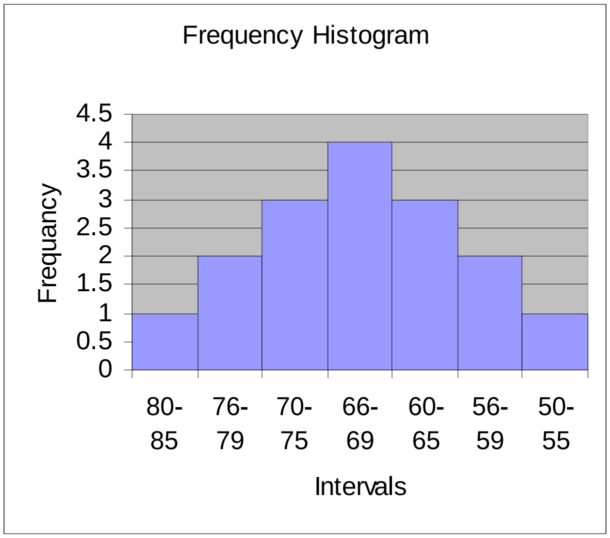
5. The sum of the scores of the 25 students is 1958 (25 x 78) plus 154 from the two other students bring sum to 2104, So new average is 2104/27 = **77.93**

6. Meal 1: Mean = **7**, median = 5, mode = 5
   Meal 2: Mean = 6.88, median = 5, mode = 10
   Meal 3: Mean = 8.57, median = 10, mode = 10
7. Twenty seven student average 80, so their sum is 1600 (27 x 80)
   The sum of 20 students with 90 is 1800; this is more than the sum of all,
   So, the answer is ***no***.
8. The median is the best statistics for measure of CT when the data contains extreme points; since it is the least influenced by outliers or extreme points.
9. a. The median for {8, 10, **11, 12**, 14, 17} is average of 11 and 12, so median is **11.5**
   b. The median for {1, 2, 2, **4**, 4, 5, 6} is the middle score, so median is **4**
10. Histogram: Mean = **13.08**, median = **14**, mode = **14**
11. Data Table: Mean = **11.52**, median = **12**, mode = **12**
12. Data Table: X: Mean = **3**, median = **3**
    Y: Mean = **2.33**, median = **2**
    (X – 3): Mean = **0**, median = **0**

Exercises 2.2

Question 1. For 30 scores, the frequency and cumulative frequency is shown below

Interval	Upper Limit	Freq	Rel Freg	% Freg	Cum Freg
51-60	60	1	0.0333	3.3333	1
61-70	70	5	0.1667	16.6667	6
71-80	80	14	0.4667	46.6667	20
81-90	90	5	0.1667	16.6667	25
91-100	100	5	0.1667	16.6667	30



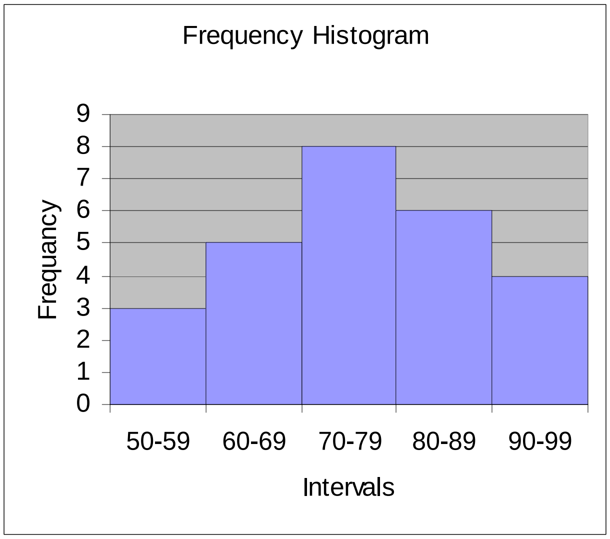
Question 2. From Dataset: Frequency, **mean = 19.05** and **median =19.2**

Interval	Upper Limit	Freq	Rel Freg	% Freg	Cum Freg
15-16	16	2	0.0800	8.0000	2
17-18	18	6	0.2400	24.0000	8
19-20	20	6	0.2400	24.0000	14
21-22	22	10	0.4000	40.0000	24
23-24	24	1	0.0400	4.0000	25

Question 3. From frequency histogram mean = 55.15 = ∑(MP*freq)

Midpoint	Frequency	Rel Freg	Midpt * f
24.5	2	0.0645	1.580645
34.5	4	0.1290	4.451613
44.5	6	0.1935	8.612903
54.5	5	0.1613	8.790323
64.5	8	0.2581	16.64516
74.5	4	0.1290	9.612903
84.5	2	0.0645	5.451613

Question 4. From frequency table: Histogram



Mean = **75.65** = ∑(MP*freq)

Interval	Freq	Cf	Cp	CP %
90-99	4	*26*	*1.0000*	*100.0000*
80-89	6	*22*	*0.8462*	*84.6154*
70-79	**8**	*16*	*0.6154*	*61.5385*
60-69	5	***8***	*0.3077*	*30.7692*
50-59	3	*3*	*0.1154*	*11.5385*

Median = **75.65** and 75 percentile =

$$P_{50\%} = LLi + [(\frac{n_p - C_f}{f_i}) \cdot I] = 69.5 + [(\frac{(13-8}{8}) \cdot 10] = 75.75$$

$$P_{75\%} = LLi + [(\frac{n_p - C_f}{f_i}) \cdot I] = 79.5 + [(\frac{(19.5-16}{6}) \cdot 10] = 85.33$$

Question 5. From frequency histogram: Mean = **66.75** = ∑(MP*freq)

Interval	Freq	Cf	Cp	CP %
80-84	1	*16*	*1.0000*	*100.0000*
75-79	2	*15*	*0.9375*	*93.7500*
70-74	3	*13*	*0.8125*	*81.2500*
65-69	**4**	*10*	*0.6250*	*62.5000*
60-64	3	***6***	*0.3750*	*37.5000*
55-59	*2*	*3*	*0.1875*	*18.7500*
50-54	*1*	*1*	*0.0625*	*6.2500*

Median = 67

$$P_{50\%} = LLi + [(\frac{n_p - C_f}{f_i}) \cdot I] = 64.5 + [(\frac{(8-6}{4}) \cdot 5] = 67$$

Question 6. Mean = **55.64** and Median = **55.75**

Interval	Freq	Cf	Cp	CP %
70-79	4	*26*	*1.0000*	*100.0000*
60-69	6	*22*	*0.8462*	*84.6154*
50-59	8	*16*	*0.6154*	*61.5385*
40-49	5	*8*	*0.3077*	*30.7692*
30-39	3	*3*	*0.1154*	*11.5385*

$$P_{50\%} = LLi + [(\tfrac{n_p - C_f}{f_i}) \cdot I] = 49.5 + [(\tfrac{(13-8}{8}) \cdot 10] = 55.75$$

Question 7. (a) M = **46.4** and (b) M = **16**

$$\text{(a) } M = \frac{\Sigma X}{N} = \frac{232}{5} = 46.4 \text{ (b) } M = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{400}{25} = 16$$

Exercises 2.3

1. Stem and leaf from data

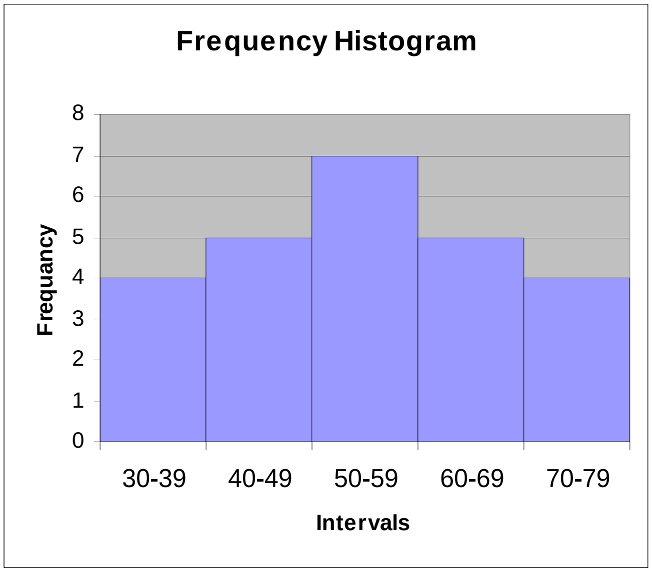
frequency	Stem	Leaf
0	**4**	
1	**5**	**8**
4	**6**	**3579**
12	**7**	**022244666688**
8	**8**	**00025568**
5	**9**	**22358**
0	**10**	

2. Stem and leaf

frequency	Stem	Leaf
2	**15.**	**89**
4	**16.**	**1334**
2	**17.**	**05**
4	**18.**	**1255**
2	**19.**	**24**
5	**20.**	**23399**
3	**21.**	**148**
3	**22.**	**001**

Histogram from stem and leaf

3 Median is 54 (from Midpoint of Stem: 5 | 2 3 3 **4** 5 6 7 (the 50 percentile)



4. Stem and leaf from histogram (midpoint of class intervals)

24. | 5 5
34. | 5 5 5 5
44. | 5 5 5 5 5 5
54. | 5 5 5 5 5
64. | 5 5 5 5 5 5 5 5
74. | 5 5 5 5
84. | 5 5

5. Reconstructed stem and leaf into class interval size of 5

Original	Reconstructed
3 | 0 1 1 1 2 7 8	**3** | 0 1 1 1 2
4 | 2 4 5 6 7 8 8 9	**3** | 7 8
5 | 3 3 3 3 4 5 6 7 8	**4** | 2 4 5
6 | 1 1 2 4 4 5	**4** | 6 7 8 8 9
7 | 1 1 2 2 7 8 9	**5** | 3 3 3 3 4 5
	5 | 6 7 8
	6 | 1 1 2 4 4 5
	6 |
	7 | 1 1 2 2
	7 | 7 8 9

6. Stem and leaf from data:

frequency	Stem	Leaf
0	**3**	
1	**4**	**6**
1	**5**	**7**
1	**6**	**4**
4	**7**	**2457**
8	**8**	**34455668**
5	**9**	**02255**
0	**10**	

7. Stem and leaf from frequency table

Original

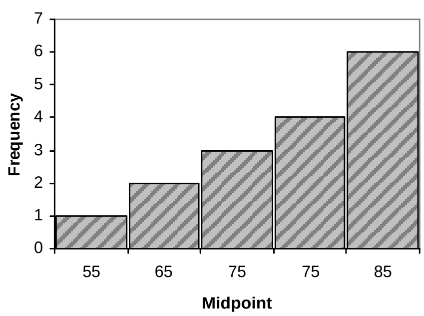
X	Frequency
100	2
80	4
60	7
40	5
20	2

Stem and leaf conversion

2 | 0 0
4 | 0 0 0 0 0
6 | 0 0 0 0 0 0 0
8 | 0 0 0 0
10 | 0 0

8. Stem and leaf and mean, median from histogram

Original Histogram



(a) Stem and leaf display

5 | 5
6 | 5 5
7 | 5 5 5
8 | 5 5 5 5
9 | 5 5 5 5 5 5

MP	Freq	Cf	Cp	CP %
55	*1*	*1*	*0.0625*	*6.2500*
65	*2*	*3*	*0.1875*	*18.7500*
75	*3*	*6*	*0.3750*	*37.5000*
85	*4*	*10*	*0.6250*	*62.5000*
95	*6*	*16*	*1.0000*	*100.0000*

Mean = ∑(MP*Freq) = **82.5**

Median = **85** (by observation, the mean of 8^{th} and 9^{th} ordered score is = (85 + 85)/2 = 85

$P_{50\%} = LLi + [(\frac{n_p - C_f}{f_i}) \cdot I] = 84.5 + [(\frac{(8-6}{4}) \cdot 1] = 85$

Exercises 6.2

1. The standard error is the standard deviation divided by the square root of the sample size, *n*
2. The sampling distribution has mean, μ and standard deviation, $s_M = s/\sqrt{(n)}$
3. As *n* increases, s_M decreases

$$\sigma_M = \frac{4}{\sqrt{64}} = \frac{4}{8} = \frac{1}{2} \quad \sigma_M = \frac{4}{\sqrt{9}} = \frac{4}{3} = 1\frac{1}{3}$$

4. The sampling distribution of means will have a mean equals to μ and standard deviation of s_M.
5. Given $\mu = 40$ and $s = 2$

$$z = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}}$$

n	***M***	***z-score***
25	*35*	-12.5
25	*45*	12.5
16	*35*	-10
16	*45*	10

6. Given $s = 2$ and $\mu = 8$ and sample $M = 6.09$, $n = 11$,
$z =$ **-3.1659** $< z_{CV} = -1.960$. So, **reject H_0: $\mu = 8$**.

7. Given $s = 10$ and $\mu = 69$ and sample $N = 900$:
a. $CI_{95} = 69 \pm 1.960\ (0.3333)$ = **68.35** to **69.65**
b. $CI_{99} = 69 \pm 2.5758\ (0.3333)$ = **68.14** to **69.86**

8. Given $s = 8$ and $\mu = 65$ and sample, $n = 60$, $M = 70$
H_0: $\mu = 65$; H_a: $\mu > 65$, $a = 0.05$
$s_M = 1.0328$, $z_{stat} = 4.8412 > z_{CV} = 2.5758$, **So reject H_0.**

9. Given $s = 15$ and $\mu = 100$ and sample, $n = 49$, $M = 105$
H_0: $\mu = 100$; H_a: $\mu > 100$, $a = 0.05$
$s_M = 2.14$, $z_{stat} = 2.3333$ and $z_{CV} = 1.96$.
Since $z_{stat} = 2.33 > 1.96$. **So, reject H_0.**

10. Given $s = 2.4$ and $\mu = 5.8$ and sample, $n = 36$, $M = 6.4$
H_0: $\mu = 5.8$; H_a: $\mu \neq 5.8$, $a = 0.05$
$s_M = 0.4$, $z_{stat} = 1.50$ and $z_{CV} = \pm 1.96$.
Since $-1.96 <$, $z_{stat} = 1.5 < 1.96$, ***Don't reject H_0.***

11. Given $s = 1.2$ and $\mu = 5.8$ and sample, $n = 36$, $M = 6.0$
H_0: $\mu = 5.8$; H_a: $\mu \neq 5.8$, $a = 0.01$
$s_M = 0.2$, $z_{stat} = 1.0$ and $z_{CV} = \pm 2.5758$.
Since $-2.5758 <$, $z_{stat} = 1.0 < 2.5758$, ***Don't reject H_0.***

12. SAT info: μ = 508 and s = 100
    a. Pr(z < -0.08) = 0.4681 or **46.81%** (**Note. To convert a decimal to %, multiply by 100**)

$$z = \frac{M - \mu}{\sigma} = \frac{500 - 508}{100} = -0.08$$

    b. Pr(z > 1.92) = 1 – Pr(z < 1.92) = 1 – 0.9726 = 0.0274 or **2.74%**

$$z = \frac{M - \mu}{\sigma} = \frac{700 - 508}{100} = 1.92$$

    c. Percent between SAT 400 and 700: Pr(-1.08 < z < 1.92)
    Pr(z < 1.92) = 0.9726 and Pr(z < -1.08) = 0.1401
    So Pr(-1.08 < z < 1.92) = 0.9726 – 0.1401 = 0.8325 = **83.25%**

$$z = \frac{M - \mu}{\sigma} = \frac{400 - 508}{100} = -1.08$$

13. IQ info: μ = 100 and s = 15
    a. Percent with IQ > 140 is Pr(z > 2.67)
    Pr(z > 2.67) = 1 – Pr(z < 2.67) = 1 – 0.9962 = 0.0038 or **0.38%**

$$z = \frac{M - \mu}{\sigma} = \frac{140 - 100}{15} = 2.67$$

    b. Percent with IQ between 60 and 80 is Pr(IQ < 80) – Pr(IQ < 60)
    Pr(z < -1.33) – Pr(z < -2.67) = 0.0918 – 0.0038 = 0.088 or **8.8%**

$$z = \frac{M - \mu}{\sigma} = \frac{60 - 100}{15} = -2.67$$

$$z = \frac{M - \mu}{\sigma} = \frac{80 - 100}{15} = -1.33$$

14. Statewide Statistics: English μ = 75, s = 18 and Math μ = 73, s = 15
    Troy did better, relative to the rest of students taking the exam, on Math since his z-score was higher.

English: $z = \frac{M - \mu}{\sigma} = \frac{82 - 75}{18} = 0.39$

Math: $z = \frac{M - \mu}{\sigma} = \frac{80 - 73}{15} = 0.47$

15. Given μ = 126 and sample s = 15.69, M = 119.5, n = 16
    H_0: μ = 126; H_a: $\mu \neq 126$, a = 0.05 (assumed before hypothesis analysis)
    s_M = 3.82 (used to estimate s_M), z_{stat} = -1.66 and $z_{CV} = \pm 1.96$.
    Since -1.96 <, z_{stat} =-1.66 < 1.96, ***<u>Don't</u> reject H_0.***

16. Find the 60% percentile of population with $\mu = 24$, $s = 2.5$.
    The *z*-score with Pr(z) = 0.60 is ***z* = 0.2533** (Table A1), So ***M* = 24.63**

    $z = 0.2533 = \frac{M - 24}{2.5};\ \text{So } M = 2.5(0.2533) + 24 = 24.63$

In Progress