

Unit 3 Test Review

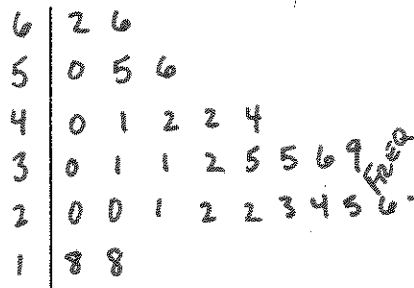
Topics:

- Representing Data
 - Histograms, stem and leaf plots
 - dot plots
- Modeling Data
 - Estimated Line of Best Fit
 - Regression Models
- Correlation
 - Correlation Coefficient
 - Correlation vs. Causation
- Residual Plots
- Measures of Central Tendency
- Analyzing Data Distribution
 - Five Number Summary
 - Box and Whisker Plots
 - Outliers
 - Mean Absolute Deviation
- Frequency Distributions
 - Frequency Marginal Distribution
 - Relative Frequency Distribution
 - Relative Frequency Conditional Distribution

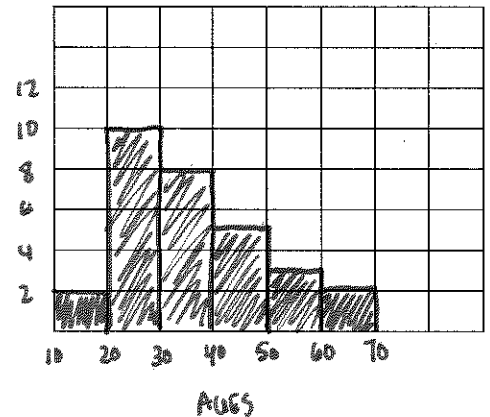
1. For the following set of data, complete the frequency distribution, histogram and stem and leaf plot. The data shows the ages of 30 consumers who bought a certain product advertised on TV:

42 44 62 35 20 39 21 18 24 42
 30 56 20 23 41 40 32 50 31 26
 55 22 31 27 66 18 25 35 36 22

Ages	Frequency
10-19	2
20-29	10
30-39	8
40-49	5
50-59	3
60-69	2



Key: $2|0 = 20$

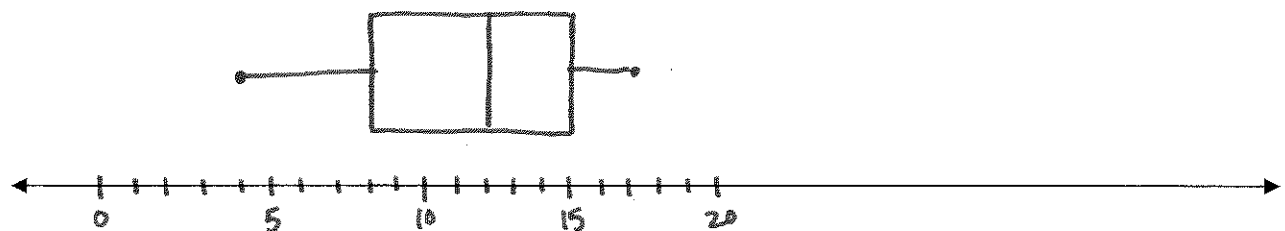


2. Find the five number summary and IQR for the given data set. Create a box and whisker plot to display the data.

4, 5, 8, 10, 12, 12, 13, 15, 16, 17
 5 8 10 12 13 15 17 12 16

Min: 4 Q1: 8 Median: 12 Q3: 15 Max: 17 IQR: 7

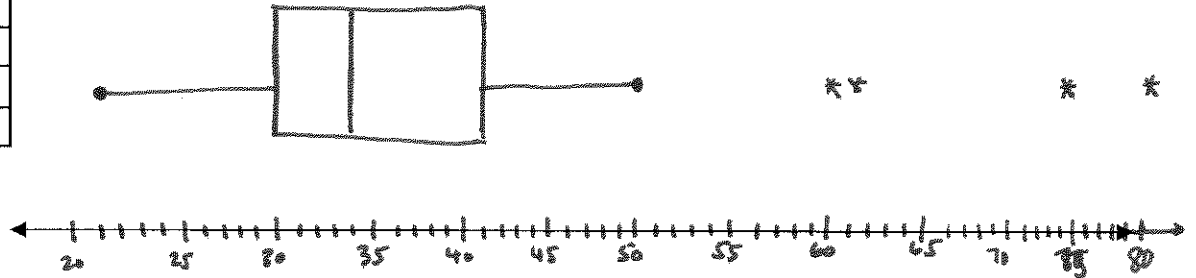
Does the data set contain any outliers? $LF: -2.5$ $UF: 25.5$ NO OUTLIERS



3. The following table lists the ages of actresses when they won their first Oscar. Find the five number summary for the set of data and determine if there are any outliers. Create a box and whisker plot for the data.

21	24	25
26	26	27
30	30	31
33	34	34
34	34	35
35	37	38
39	41	41
44	50	60
61	74	80

MIN: 21 Q_1 : 30 MED: 34 Q_3 : 41 MAX: 80 IQR: 11
 LF: 13.5 UF: 57.5 OUTLIERS: 60, 61, 80
 74



Describe the shape of the distribution: Skewed Right

4. Explain why the mean and MAD are used to describe the center and spread of a symmetrical distribution and the five-number summary is used to describe the center and spread of a skewed distribution. Data that contains skew/outliers pull the mean and MAD away from the center towards the skew/outliers. The median and IQR are so resistant to skew/outliers and will remain in the center of the data set.
5. Given the 18-hole totals for the top golfer's in the men's competition and the women's competition, compare the spread of the data for the two sets using Mean Absolute Deviation.

TOP 6 GOLFER'S SCORES

MEN	WOMEN
67	68
69	70
69	72
71	73
74	74
76	75

Men's
 Mean: 71 MAD: 2.7

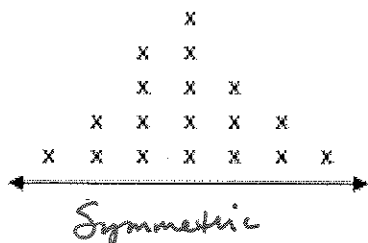
Women's
 Mean: 72 Mad: 2

Which group of golfers was more consistent? Why? Women because the MAD is lower. That means, on average, women's scores are closer to the mean.

6. What does the Mean Absolute Deviation Represent? Explain in your own words. MAD represents the average distance a data value is from the mean. It is a measure of how spread out the data is.

7. For the following dot plots, determine the shape of the distribution. Determine the order in which the mean, median and mode would occur for each distribution.

a.

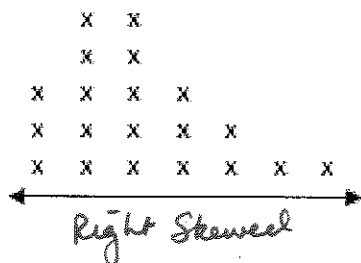


Lowest:

Middle: *MED, MEAN, MODE*

Highest:

b.

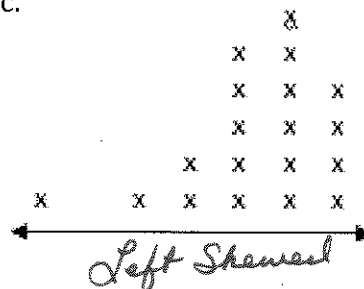


Lowest: *MODE*

Middle: *MEDIAN*

Highest: *MEAN*

c.



Lowest: *MEAN*

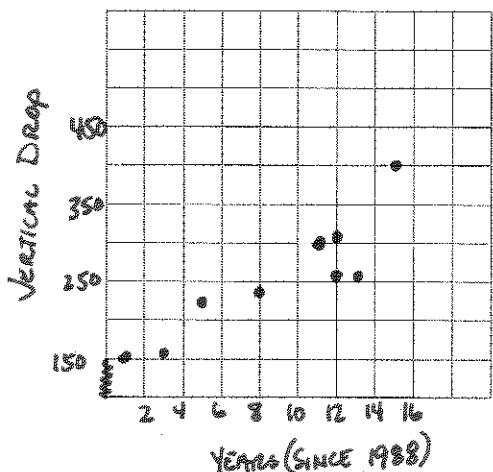
Middle: *MEDIAN*

Highest: *MODE*

8. The table shows the largest vertical drops of nine roller coasters in the United States and the number of year after 1988 that they were opened.

Years Since 1988	1	3	5	8	11	12	12	13	15
Vertical Drop	151	155	225	230	300	306	255	255	400

a. Make a scatter plot for the data.



b. Estimate the equation of a line of best fit.

$$(3, 155) \quad (12, 306)$$

$$y = 16.8x + 104.6$$

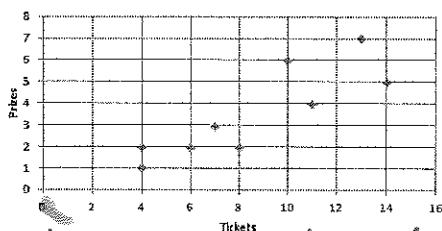
c. Predict the vertical drop of a coaster built in 2018.

$$608.6$$

9. Determine whether each graph shows a positive, negative or no correlation. If there is a positive or negative correlation, describe its meaning in the situation.

a.

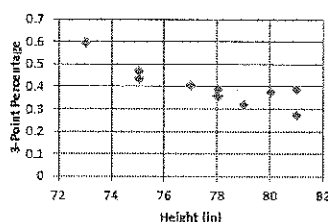
Game Tickets at the Fair



Positive: as the # of tickets increases, the prizes increase

b.

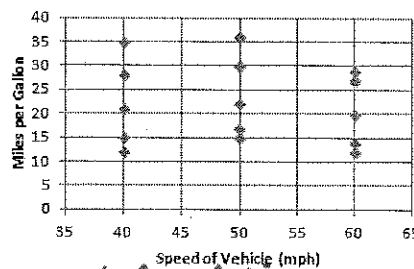
NBA 3-Point Percentage



Negative: as NBA player height increases, 3-pt % decreases.

c.

Gas Mileage



No Correlation.

10. Explain how the correlation coefficient relates to the fit of a model to data. For positive correlation the closer r is to 1, the closer the data is to the model. For negative correlation, the closer r is to -1, the closer the data is to the model.

11. The Body Mass Index (BMI) is a measure of body fat using height and weight. The heights and weights of twelve men with normal BMI are given in the table.

Height(in)	Weight(lb)
62	115
63	124
65	120
37	134
67	140
68	138
68	144
68	152
69	147
72	155
73	168
73	166

a. Use the calculator to find the linear regression equation for the data.

$$y = .856x + 85.906$$

b. Does the correlation coefficient suggest a good linear fit?

$r = .485$ No because r is not close to 1.

c. Predict the normal weight for a man who is 84 inches tall.

~~157.81~~ 157.81 lbs

d. A man's weight is 188 pounds. Use the equation to predict his height.

119.27 in

12. The following data displays the annual salaries, in the thousands, for 10 different people with various levels of education.

Years of Education	5	12	12	14	16	16	18	18	20	24
Annual Salary	23	26	27	48	58	65	95	110	160	250

a. What would be the most appropriate model for the data? Why?

Linear: $y = 12.041x - 100.439$ $r = .867$ Exponential: $y = 6.718(1.157)^x$ $r = .939$ Exponential because r is closer to 1.

b. Use the most appropriate model to predict what someone with 17 years of education would make?

80.151 thousand

c. Use the most appropriate model to predict how many years of schooling it would take to make \$85,000 a year.

Don't do

13. The time to the next eruption of Old Faithful can be predicted by using the duration of the current eruption.

Duration (min)	1.5	2	2.5	3	3.5	4	4.5	5
Interval (min)	48	55	70	72	74	82	93	100

a. Calculate the linear and exponential regression model for the data. (round to one decimal)

$$y = 14.095x + 28.440 \quad y = 38.164(1.218)^x$$

b. Calculate the residuals for the data and create a residual plot.

ON SEPARATE SHEET

c. Based on the residual plot, which regression model is a good fit for the data?

Linear

14. The table below compares the unemployment rate to the graduation rate over the course of several years.

Graduation Rate	73	85	64	81	68	82
Unemployment Rate	6.9	4.1	3.2	5.5	4.3	5.1

a. Graph the residual plot for the data set using a linear and exponential model.

$$\text{LINEAR: } y = .045x + 1.460$$

ON SEPARATE SHEET

$$\text{EXPONENTIAL: } y = 1.929(1.012)^x$$

b. Based on the residual plot, which regression model is a good fit for the data?

Neither

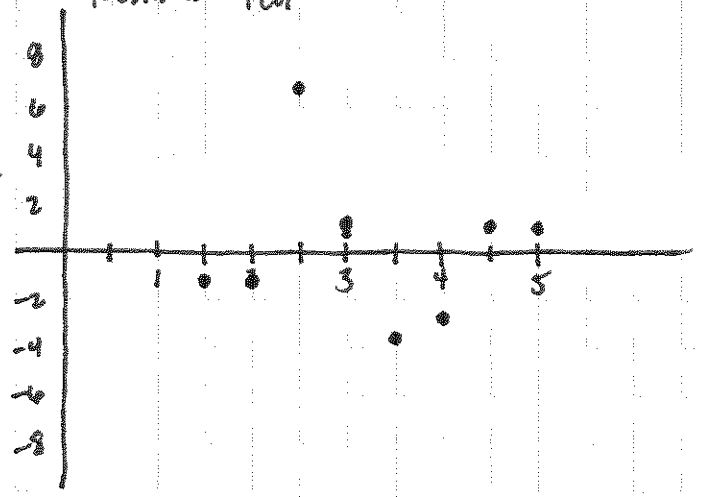
#13

LINEAR	
PREDICTED	RESIDUAL
49.6	-1.6
56.6	-1.6
63.7	6.4
70.7	1.3
77.8	-3.8
84.8	-2.8
91.9	1.2
98.9	1.1

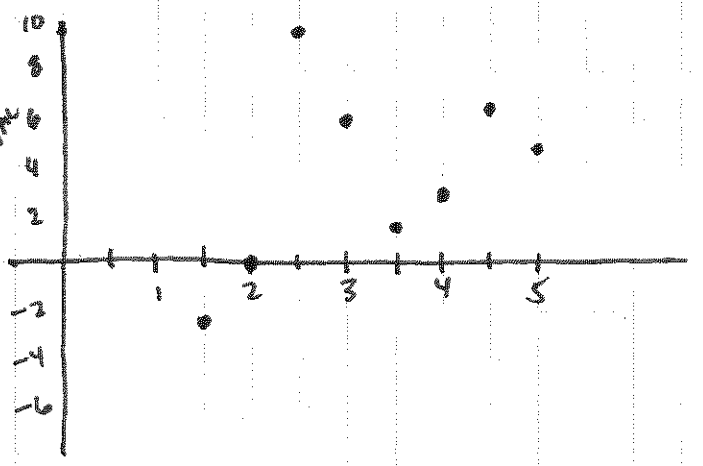
EXPONENTIAL	
PREDICTED	RESIDUAL
50.2	-2.2
55.0	0
60.3	9.7
66.0	6.0
72.3	1.7
79.2	2.8
86.8	6.2
95.1	4.9

LINEAR

RESIDUAL PLOT



EXPONENTIAL



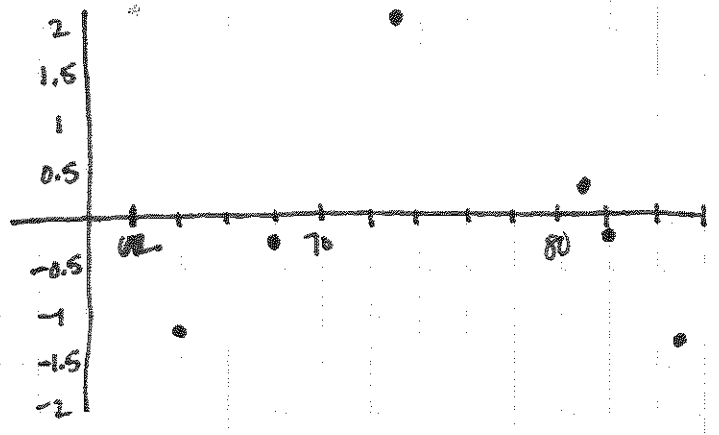
#14

LINEAR	
PREDICTED	RESIDUAL
4.7	2.2
5.3	-1.2
4.3	-1.1
5.1	0.4
4.5	-0.2
4.2	-0.1

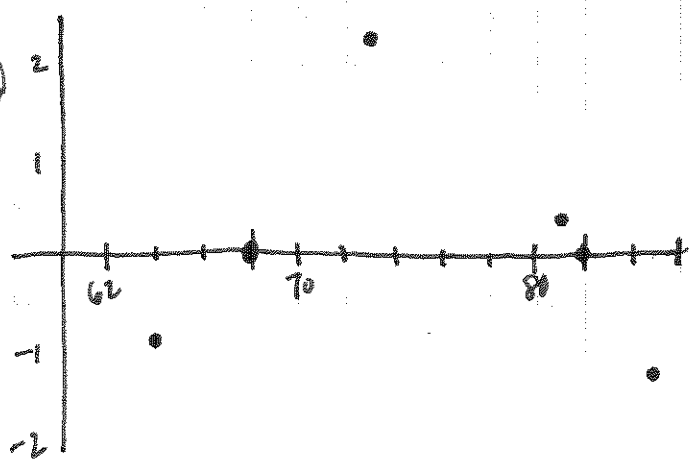
EXPONENTIAL	
PREDICTED	RESIDUAL
4.6	2.3
5.3	-1.2
4.1	-0.9
5.1	0.4
4.3	0
5.1	0

LINEAR

RESIDUAL PLOT



EXPONENTIAL

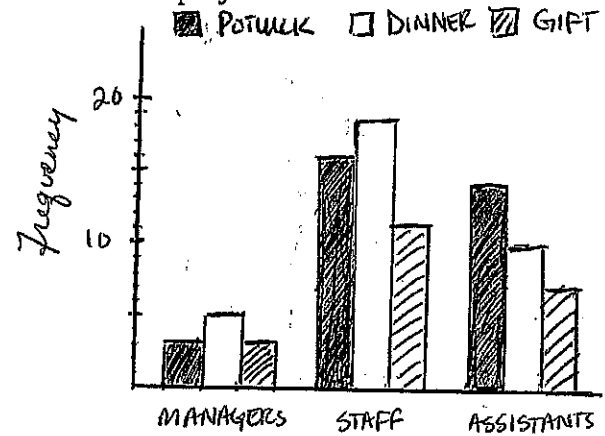


15. The managers, staff, and assistants were given three options for the holiday activity: a potluck, a dinner at a restaurant, and a gift exchange. Five of the 11 managers want a dinner, while 3 want a potluck. Eleven of the 45 staff members want a gift exchange, while 18 want a dinner. Ten of the 32 assistants want a dinner, while 8 of them want a gift exchange.

a. Create a two-way frequency table for the data.

EMPLOYEES	ACTIVITIES			TOTAL
	POTLUCK	DINNER	GIFT	
MANAGERS	3	5	3	11
STAFF	16	18	11	45
ASSISTANTS	14	10	8	32
TOTAL	33	33	22	88

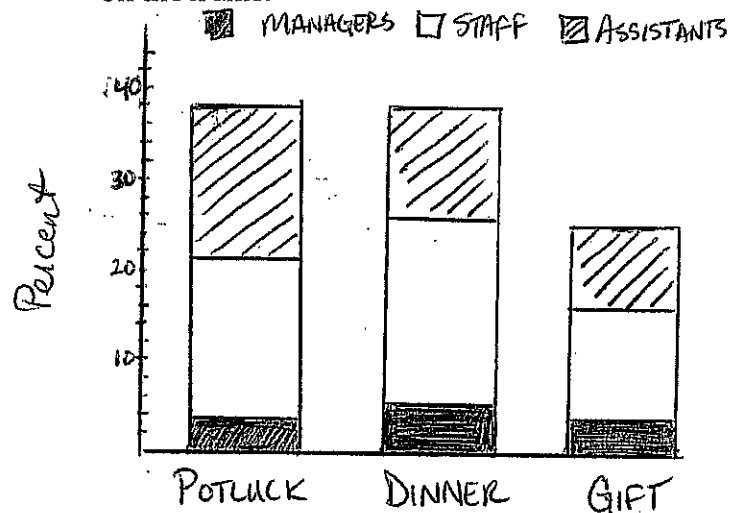
b. Create a multiple bar histogram with the employees on the x axis.



c. Create a relative frequency table for the data.

EMPLOYEES	ACTIVITIES			TOTAL
	POTLUCK	DINNER	GIFT	
MANAGERS	3.4%	5.7%	3.4%	12.5%
STAFF	18.2%	20.5%	12.5%	51.1%
ASSISTANTS	15.9%	11.4%	9.1%	36.4%
TOTAL	37.5%	37.5%	25%	100%

d. Create a stacked bar graph with activity on the x axis.



e. Create two conditional relative frequency distributions: one for the activities and one for the employees.

	POTLUCK	DINNER	GIFT	TOTAL
MANAGER	27.3%	45.5%	27.3%	100%
STAFF	35.6%	40%	24.4%	100%
ASSISTANT	43.8%	31.3%	25%	100%

	POTLUCK	DINNER	GIFT
MANAGER	9.1%	15.2%	13.6%
STAFF	48.5%	54.5%	50%
ASSISTANT	42.4%	30.3%	36.4%
TOTAL	100%	100%	100%

Standards:	S.ID.1	S.ID.2	S.ID.3	S.ID.5	S.ID.6	S.ID.6a	S.ID.6b
	S.ID.6c	S.ID.7	S.ID.8	S.ID.9			