STAT 145 Final Exam Review
Part II Practice

All interpretations MUST be in the context of the problem.
When statistical technology can be used, it should be used.

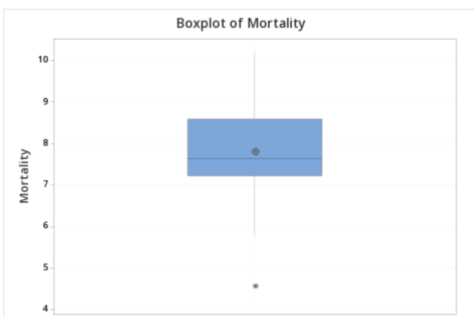
SOLUTIONS

The following data can be found in the Week 12 STAT 145 Data File under sheet: **Infant Mortality**

1. Infant mortality rate is concerned with infant deaths during the first year of life. Generally, the infant mortality rate provides the number of deaths per 1000 live births. A study listed the 24 developed nations by their infant mortality rate. The table below provides the results:

A. Construct a boxplot showing any outliers for the infant mortality rates.

INFANT MORTALITY
Boxplot of Mortality



B. In the boxplot of this data, any mortality rates below _____ OR above _____ would be considered outliers. Fill in the blank; HOW YOUR WORK as you calculate the outlier fences.

INFANT MORTALITY
Descriptive Statistics: Mortality

Statistics

Variable	N	N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	IQR
Mortality	24	0	7.823	1.355	4.590	7.235	7.640	8.593	10.190	1.358

Q1 – 1.5(IQR)
7.235 – 1.5(1.358) = 5.198 is the lower fence

Q3 + 1.5(IQR)
8.593 + 1.5(1.358) = 10.63 is the upper fence

Infant mortality rates lower than 5.198 and higher than 10.63 are considered outliers.

The following data can be found in the Week 12 STAT 145 Data File under sheet: Computer Sales

There are two sheets---one is unstacked data and one is stacked computer sales data.

2. The number of computers <u>Carl</u> sold at his computer store job each month last year:

    51, 17, 25, 39, 7, 49, 62, 41, 20, 6, 43, 13

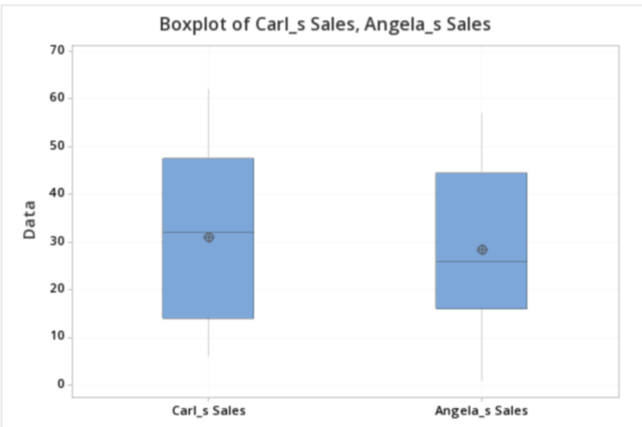
    The number of computers <u>Angela</u> sold at her computer store job each month last year:

    34, 47, 1, 15, 57, 24, 20, 11, 19, 50, 28, 37

    A. Build comparative boxplots. Find the mean value for each person.

COMPUTER SALES

Boxplot of Carl_s Sales, Angela_s Sales



Mean Carl: 31.08

Mean Angela: 28.58

COMPUTER SALES

Descriptive Statistics: Carl_s Sales, Angela_s Sales

Statistics

Variable	N	N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	IQR
Carl_s Sales	12	0	31.08	18.76	6.00	14.00	32.00	47.50	62.00	33.50
Angela_s Sales	12	0	28.58	16.91	1.00	16.00	26.00	44.50	57.00	28.50

B. Compare the box plots (compare shape, center and spread)

Shape: Carl’s computer sales data is somewhat symmetric (mean is close to median) OR one might call it slightly skewed left]

whereas Angela’s distribution is slightly skewed right (mean > median).

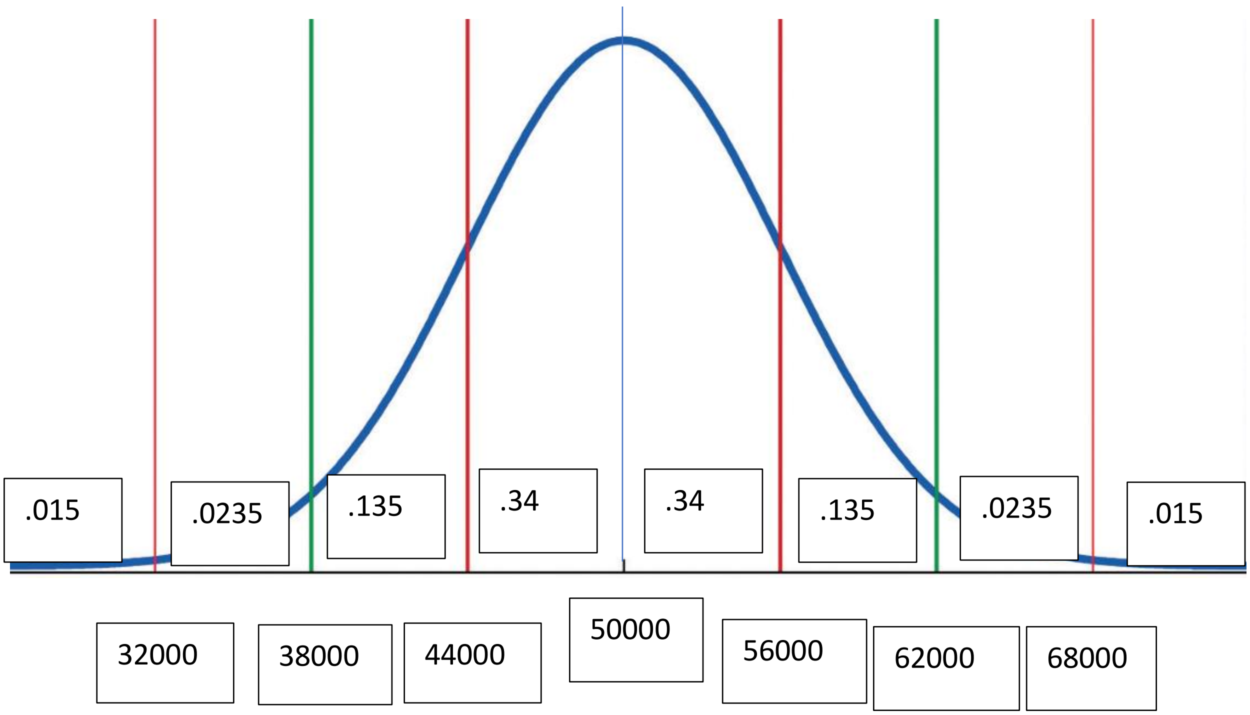
Center: Comparing median values, Carl appears to have the higher computer sales, on average.

Spread: Angela’s data is less spread out (smaller IQR).

3. The manufacturer of *Drive On Us* tires states that, "The lifetimes of our tires follow a bell-shaped distribution with mean 50,000 miles and standard deviation 6,000 miles".

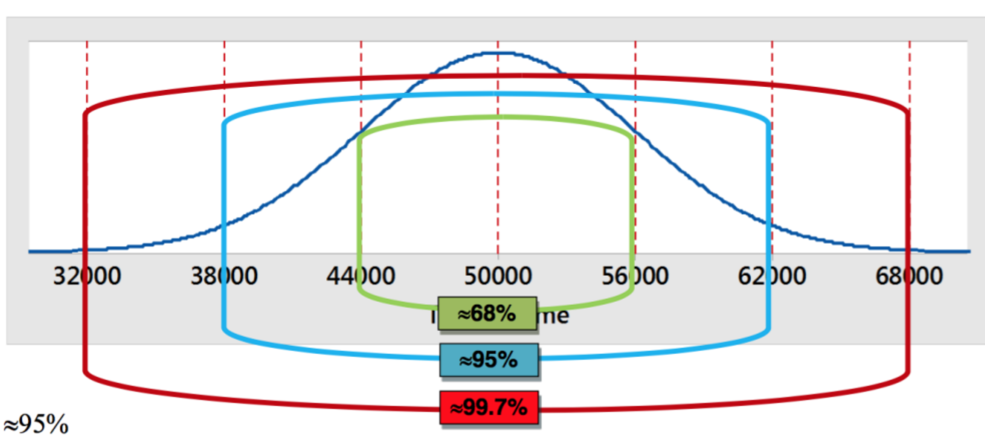
A. Draw a picture of the empirical rule intervals for the lifetimes. Include a number line with values, and label the percentages.

Draw a picture of the empirical rule for this data – **label the miles and percentages.**



OR

A



B. ≈95%

C. ≈16%

D. 56000 miles

E. Yes, a lifetime of 36780 miles would be considered unusually low because its z-score is –2.203 and any z-score beyond +2 or –2 is considered unusual.

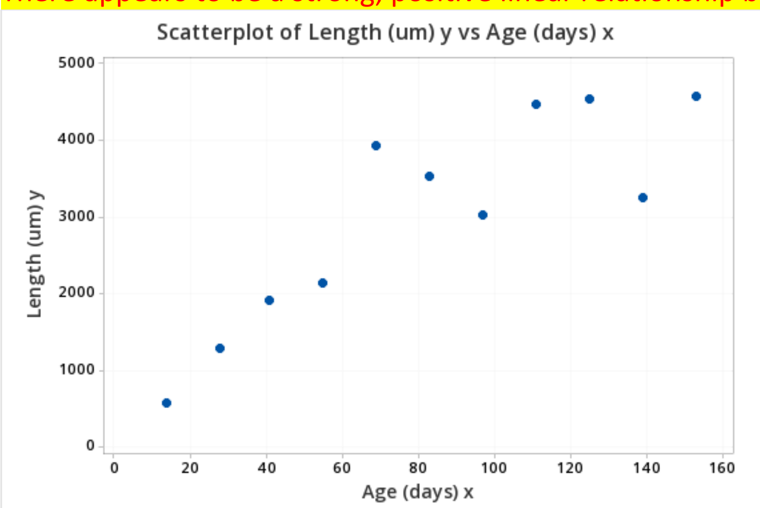
The following data can be found in the Week 12 STAT 145 Data File under sheet: Fish Length

4. Researchers are studying the growth of a species of fish. The fish are placed in a tank immediately after birth. At regular intervals, a fish is chosen at random and its length measured.

X = Age (days)	Y = Length (um)
14	590
28	1305
41	1915
55	2140
69	3920
83	3535
97	3030
111	4465
125	4530
139	3257
153	4566

A. Construct a scatterplot. Compute and interpret the correlation value.

There appears to be a strong, positive linear relationship between the age of the fish and the length.



Correlations

	Age (days) x
Length (um) y	0.864

B. Provide the equation of the regression line for this data.

Regression Equation

Length (um) y = 892 + 25.62 Age (days) x

C. Write a sentence to interpret the slope of your regression line.

For each one unit increase in age, the length of the fish increases by 25.62 um, on average.

D. Compute and interpret R^2.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
726.022	74.67%	71.86%	61.97%

74.67% of the variation in fish length is being explained by the linear model.

E. What length would you predict for a fish who is 50 days old? SHOW YOUR WORK and clearly state your answer.

Length = 25.62(50) + 892 = 2173
The equation predicts 2173 um for a 50-day old fish.

The following data can be found in the Week 12 STAT 145 Data File under sheet: Calories

5. People like to eat chocolate chip cookies, and love to believe that there aren't that many calories in each serving. A researcher at Consumer Reports would like to convince consumers that there are, in fact, more than 150 calories in a serving of chocolate chip cookies. She randomly selected both fresh-baked (from a boxed cookie mix such as Duncan Hines and Pillsbury) and packaged (such as Pepperidge Farm and Nabisco) cookies to obtain the following calorie information.

158	157	151	143	135	151	154	143
173	152	145	161	160	168	158	

A. Does the researcher's sample provide sufficient evidence to support the idea that the mean calories in a serving of chocolate chip cookies is greater than 150? Show the complete testing process.

Population

The variable is calories in a serving of chocolate chip cookies and it is numerical.

Goal: test to see if the true mean calories is more than 150 calories.

Method

Ho: $\mu = 150$

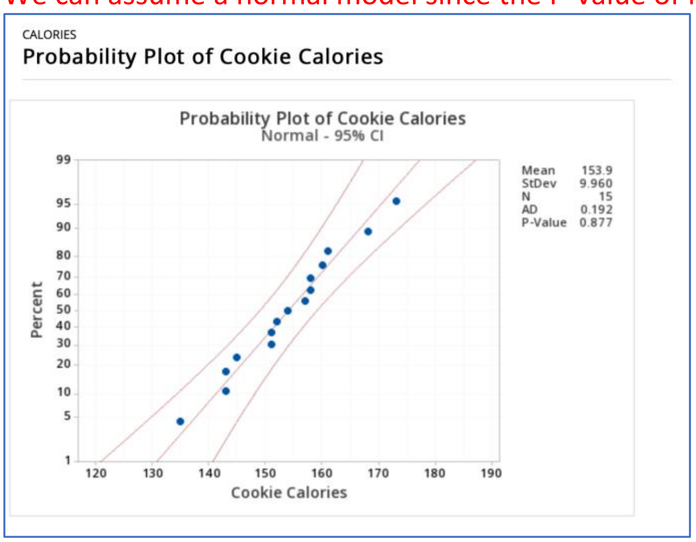
Ha: $\mu > 150$

Alpha = .05

T-curve

Sample

We can assume a normal model since the P-value of NPP is greater than 0.05. (The NPP must be shown).



CALORIES

One-Sample T: Cookie Calories

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
15	153.93	9.96	2.57	149.40

μ: population mean of Cookie Calories

Test

Null hypothesis H_0: μ = 150
Alternative hypothesis H_1: μ > 150

T-Value	P-Value
1.53	0.074

Results
T = 1.53 with df = 14
The sample mean is 1.53 standard errors above the hypothesized mean of 150 calories.
P-value = .074
Assuming the true mean calories is 150, there is a .074 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion

The P-value is not less than alpha = 0.05. We cannot reject the null.
At the 5% level of significance, the sample data DOES NOT provide sufficient evidence to say that the true mean calories in a serving of chocolate chip cookies is more than 150.
The researcher was wrong. There is no evidence that there are more calories.

B. There is a small chance that your decision was incorrect. If so, what type of error would this be? EXPLAIN.

Type II error could have occurred since we did not reject the null.

6. Red/green color blindness causes problems in distinguishing reds and greens. Many studies have been conducted with Caucasian males, among whom 8% have red-green color blindness. However, there is limited information on this phenomenon among other ethnicities. A researcher in China believes the percentage of red-green color for Asian males is different than the 8% for Caucasians. She conducts a study on 200 randomly selected Asian men, and finds that 9 have red-green color blindness.

   A. Does the Chinese researcher's sample provide sufficient evidence to support the idea that the proportion of red-green color blindness among Asians is different from 8%? Show the complete testing process.

<u>Population</u>
The variable is whether or not a person is red-green colorblind and it is categorical.
p = the true proportion of red-green colorblind among Asian males
Goal: test to see if the true proportion of red-green colorblind among Asian males is different than .08.

<u>Method</u>
Ho: p = .08
Ha: p ≠ .08
Alpha = .05
Z-curve

<u>Sample</u>
We can assume a normal model since (n)(p)(1 – p) = (200)(.08)(1 - .08) = 14.72 (greater than 10).

FISH LENGTH
Test and CI for One Proportion

Method

p: event proportion
Normal approximation method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% CI for p
200	9	0.045000	(0.016270, 0.073730)

Test

Null hypothesis H_0: p = 0.08
Alternative hypothesis H_1: p ≠ 0.08

Z-Value	P-Value
-1.82	0.068

Results
Z = -1.82
The sample proportion is 1.82 standard errors BELOW the hypothesized proportion of .08.
P-value = .068
Assuming the true proportion is 0.08, there is a .068 probability of getting a sample proportion at least as extreme as the one we got from sampling.

Conclusion

The P-value is not less than alpha = 0.05. We cannot reject the null.
At the 5% level of significance, the sample data DOES NOT provide sufficient evidence to say that the true proportion of red-green colorblind among Asian males is different than .08.
We do not have evidence that Asian males have a different proportion of color-blindness compared to Caucasians.

B. There is a small chance that your decision was incorrect. If so, what type of error would this be? EXPLAIN.

Type II error since we cannot reject the null.

7. Researchers surveyed a random sample of Americans asking whether or not they had student loan debt in excess of $25,000. The 95% CI for the proportion of all Americans with student load debt in excess of $25,000 is (.21, .39).

A. Write a concluding sentence for the confidence interval result.

We are 95% confident that the true proportion of all Americans with student loan debt in excess of $25,000 is between .21 and .39.

B. The U.S. Department of Education recently stated that "Thirty percent of Americans have student loan debt that is greater than $25K". Based on the given confidence interval, is this statement reasonable? EXPLAIN.
Yes, since .30 is in the confidence interval.

C. Based on the given confidence interval, can you be 95% confident that the percentage of Americans with student loan debt in excess of $25,000 is less than 25%? EXPLAIN.

No, since the entire CI is not less than .25.

D. Determine the values for the point estimate and the margin of error.

The point estimate is (.21 + .39) /2 = .30
The margin of error is .30 - .21 = .09

The following data can be found in the Week 12 STAT 145 Data File under sheet: Shelf Life

8. The shelf life of packaged food depends on many factors. Dry cereal is considered to be a moisture sensitive product (no one likes soggy cereal!), with the shelf life primarily determined by moisture content. A study of one particular brand of cereal examined Time on Shelf (days) and Moisture Content (%). The data can be found in the Week 12 Data File under: Shelf Life.

Time on Shelf (X)	0	3	6	8	10	13	16	20	24	27	30	34	37	41
Moisture Content (Y)	2.8	3.0	3.1	3.2	3.4	3.4	3.5	3.1	3.8	4.0	4.1	4.3	4.4	4.9

A. Which of the following is the definition of correlation? Choose one. (ANSWER: 3)

1. Measure of the cause-and-effect relationship between X and Y
2. Strength of the relationship between any two variables
3. Strength and direction of the linear relationship between two numerical variables
4. Measure of the indirect relationship between two categorical variables

B. What is the value of the correlation for this data? __.953__

Correlations

	Time on Shelf (x)
Moisture Content (y)	0.953

C. Compute the least squares regression line for Moisture Content and Time on Shelf.

Regression Equation

Moisture Content (y) = 2.7855 + 0.04462 Time on Shelf (x)

E. Write a sentence to interpret the slope of the line from part C (use variable names and units)

For each day increase in the time on shelf, the moisture content increases by .04462 %, on average.

E. Write a sentence to interpret the coefficient of determination (R^2)

90.75% of the variation in moisture content is being explained by this regression line.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.196246	90.75%	89.98%	87.87%

F. A line with a "good" fit has a value of R^2 close to ______. EXPLAIN.

100%. The closer to 100% the better since we want the percentage of variation in the y-variable that is being explained to be very high.

9. A medical researcher is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication. A random sample of 41 patients results in mean pulse rate of 84.2, and a standard deviation s = 1.3.

    1. Build and interpret a 95% confidence interval for the mean pulse rate. Show the statistical output and interpret in the context of the problem.

SHELF LIFE

One-Sample T

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
41	84.200	1.300	0.203	(83.790, 84.610)

μ: population mean of Sample

We are 95% confident that the true mean pulse rate of all patients who take the new medication is between 83.79 and 84.61 beats per minute.

    2. Once completed, determine if it is reasonable to say that the mean pulse rate is 82 beats per minute. Explain your answer.

No, it is not reasonable to claim that it is 82 beats per minute since 82 is not in the CI.