

Math 247: Final Exam
(125 points)

Name: KEY

Class Time: _____

1. (2 pts) If your score on an exam was the same value as Q3 for the scores for the entire class, which of the following would be true? Circle all correct answers (there may be more than one).

(a) Your score was at the 25th percentile.

(b) Your score was at the 75th percentile.

(c) You scored a 75% on the exam

(d) You scored a 25% on the exam

(e) You did better than 75% of the other people who took the exam.

2. (2 pts) Suppose a baby has a head circumference with a z-score of -2.81. You could tell which of the following from this information (circle the answer).

(a) The baby has an unusually large head

(b) The baby has a head larger than average but not unusually large.

(c) The baby has a head smaller than average but not unusually small.

(d) The baby has an unusually small head.

(e) Can't tell from this information.

3. (4 pts) The histogram displays the frequencies of waiting times, in minutes for 175 patients in a dentist's office.

(a) This distribution is

SKEWED LEFT

SKEWED RIGHT

SYMMETRIC

UNIFORM

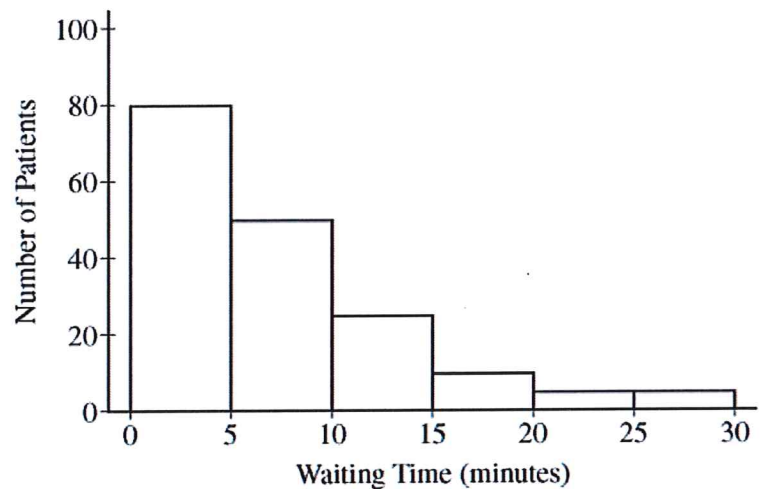
(b) Which would be true for this data set (circle the answer)

(i) The mean is the same as the median

(ii) The mean is greater than the median

(iii) The mean is less than the median

(iv) Can't tell from this information.



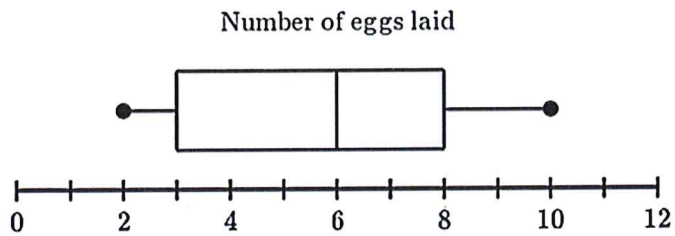
not much understanding here - review this more!

4. (6 pts) Suppose you have a flock of backyard chickens and you gather data for two weeks, seeing how many eggs each chicken laid during that time. The data is summarized in the boxplot below.

2 (a) Find the 5-number summary from the boxplot.

5-number summary:

2, 3, 6, 8, 10



2 (b) Find the IQR for this data set.

$IQR = 8 - 3 = 5$ eggs

2 (c) This data set is which of the following (circle the answer):

strongly skewed left slightly skewed left strongly skewed right slightly skewed right symmetric

-1 for this

5. (4 pts) Use your knowledge of the world to determine whether the following pairs of events are mutually exclusive (ME) or not mutually exclusive

A person lives full time in SLO
A person lives full time in Paso

ME

not ME

can't tell

A student is a business major.
A student is on the basketball team.

ME

not ME

can't tell

6. (6 pts) (a) If you were to flip a coin 3 times, list all the possible outcomes. Use H for heads, and T for tails

TTT HTT HHT HHH
 THT HTH
 TTH THH

(b) Fill in the probability distribution for the number of heads obtained on three flips.

omit

X = number of heads	0	1	2	3
P(X) = probability	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

7. (6 pts) If you have a group of 20 students and 12 of them are women, what are the following probabilities? Express each as a fraction, a decimal, and a percent.

(a) If you choose one student, what is the probability the student would not be a woman?

$P(W^c) = \frac{8}{20} = .4 = 40\%$

(b) If you choose two students **with replacement**, what is the probability both are women?

$P(WW) = \frac{12}{20} \cdot \frac{12}{20} = .36 = 36\%$

add instead of times => -2

(c) If you choose two students **without replacement**, what is the probability both are women?

$P(WW) = \frac{12}{20} \cdot \frac{11}{19} = .347 = 34.7\%$

8. (3 pts) If Minitab or other software gives a P-value as P-value = 0.000, what could you conclude? (Circle all that are correct.)

- (a) The P-value is actually, truly, zero.
- (b) The P-value is not zero but could be something like 0.000002
- (c) The P-value could be reported as $p < .0005$
- (d) The result of the statistical test is significant.

9. (3 pts) Statistical inference includes which of the following:

- (a) Proving something is true with 100% certainty
- (b) Proving something is false with 100% certainty.
- (c) Estimating a population parameter with a level of confidence that is always less than 100%.
- (d) None of the above

10. (3 pts) Which of the following are the same as the "independent variable"?

(Circle all that mean the same)

Response

Factor

Treatment

Outcome

Predictor

11. (8 pts) (a) By hand, using a table, find the standard deviation of the following set of numbers: 1, 5, 6, 6, 12

6

$$\bar{x} = \frac{30}{5} = 6$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
1	-5	25
5	-1	1
6	0	0
6	0	0
12	6	36

$$SSE = 62$$

$$S = \sqrt{\frac{SSE}{n-1}} = \sqrt{\frac{62}{4}} = \sqrt{15.5} = 3.937$$

$$S = 3.937$$

3

3

2 (b) What is the SSE (Sum of the Squared Error) for this data set? (Look at your work from above.)

$$SSE = \underline{62}$$

WEAK!!

12. (10 pts) According to studies done in the 1940s, 29% of people dream in color. Suppose a researcher wanted to check whether this proportion has increased and drew a random sample of 200 people and found 68 reported dreaming in color.

$$\hat{p} = \frac{68}{200} = .34$$

1(a) What is the sample proportion for this problem? Choose the answer with the correct number and the correct notation:

$$\bar{x} = 68$$

$$\hat{p} = 68$$

$$\bar{x} = .34$$

$$\hat{p} = .34$$

$$\hat{p} = .29$$

4(b) By hand, find the 95% confidence interval for the population proportion of people who dream in color.

CI: point estimate \pm margin of error

$$\hat{p} \pm z^* SE_{est}$$

$$.34 \pm 1.96 (.0334)$$

$$.34 \pm .066$$

$$(.34 - .066, .34 + .066)$$

$$(.274, .406) \quad 2$$

$$\hat{p} = .34$$

$$n = 200$$

$$SE_{est} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \sqrt{\frac{(.34)(.66)}{200}}$$

$$= .0334 \quad 2$$

Confidence Level	z^*
80%	1.282
90%	1.645
95%	1.960
99%	2.576

2(c) Interpret the confidence interval in words.

We are 95% confident that the proportion of all people (population of all dreamers) who dream in color is between 27.4% and 40.6%.

3(d) Does the confidence interval support the researcher's idea that the proportion of people dreaming in color has increased? Explain.

No. Since the CI includes 29%, we do not have evidence that the proportion has changed (thus it hasn't increased).

13. (10 pts) A package of Diamond roasted almonds is supposed to contain 6 ounces. If the company wanted to make sure their machines were filling the packages correctly and so wanted to test whether the mean weight was different from 6 ounces, what would the hypotheses be for a 1-Sample t-test? Be sure to use the correct math symbols!

2 (a) Hypotheses (write using symbols and in words)

$H_0: \mu = 6oz$ The average weight of all packages of almonds is 6 ounces (the machine is filling the packages correctly)

$H_a: \mu \neq 6oz$ The average is not 6 ounces (the machine is not filling the packages correctly).

1 (b) This is a (circle the answer) One-Tailed-Test

Two-Tailed Test

Suppose they drew a random sample of 20 packages and found that the sample had a mean of 5.8 ounces with a standard deviation of 0.5 ounces.

4 (c) Are the conditions for a 1-Sample t-Test satisfied? For credit, describe in words how each condition is satisfied and/or assumptions that will have to be made.

1. Samples are random? Yes, stated in problem
 ... and independent? Yes, one package's fill weight does not affect the other

2. Sample size/population? This is a small sample, $n = 20 < 25$
 So we will have to assume the fill weights for all packages are normally distributed.

3. Large population? Pop of all packages $\geq 10(20) = 200$
 It's safe to assume there are at least 200 packages of almonds being filled.

For the compute step, the Minitab results of the hypothesis test are given below.

N	Mean	StDev	SE Mean	T	P
20	5.800	0.500	0.112	-1.79	<u>0.090</u>

2 (d) Interpret the results of the hypothesis test.

$P\text{-value} = .09 > .05 \Rightarrow$ do not reject H_0 .
 There is not evidence to conclude the fill weight is something other than 6 ounces.

1 (e) Based on these results, which would you advise the company? (Circle your answer.)

Don't adjust your machines, they appear to be filling the packages correctly, on average.

Adjust the machines NOW! The machines are clearly not filling the packages correctly!

List = 1 pt

Fulfill = 0 pts

14. (14 pts) A researcher is investigating whether people who consume less sugar have less LDL (bad) cholesterol. She finds the LDL level of a random sample of people who eat a low amount of sugar and the LDL level of another random sample of people who eat a high amount of sugar. The two groups were completely separate with no relationship between them whatsoever.

1 What is the independent variable in this study? Level of sugar consumption

1 What is the dependent variable? LDL levels

2 What type of study was this? Observational Controlled Experiment

1 Choose the best hypothesis test for the study (circle your answer):

One Sample t-Test

Two Sample t-Test

Paired t-Test

One Proportion z-Test

Two Proportions z-Test

2 Write the hypotheses for the test, using math symbols and using words.

$H_0: \mu_{HIGH} = \mu_{LOW}$ The mean LDL level is the same regardless of how much sugar people eat

$H_a: \mu_{HIGH} > \mu_{LOW}$ On average, people with high sugar diet have higher LDL

The 95% confidence interval for the difference between the mean LDL of people with a high-sugar diet and a low-sugar diet is (3.5, 13.4). LDL is measured in mg/DL.

2 Interpret the confidence interval in the context of the problem:

We are 95% confident that the difference in mean levels between all people who eat a high sugar diet and all people with a low sugar diet is between 3.5 and 13.4 mg/DL.

2 Does this interval show that there is a significant difference between LDL levels in people with a low- vs. high-sugar diet? How can you tell?

Yes, there is a significant difference in LDL levels, on average, because the CI does not capture zero, thus the possibility of zero difference is excluded.

2 Can the researcher conclude that the high sugar diet CAUSED the difference in LDL levels? For credit, very briefly explain how you can tell.

No! This is an observational study so we can't conclude cause-and-effect.

1 Describe one possible confounder for this study.

Low sugar \Rightarrow better diet overall (lower fat, salt, etc.)

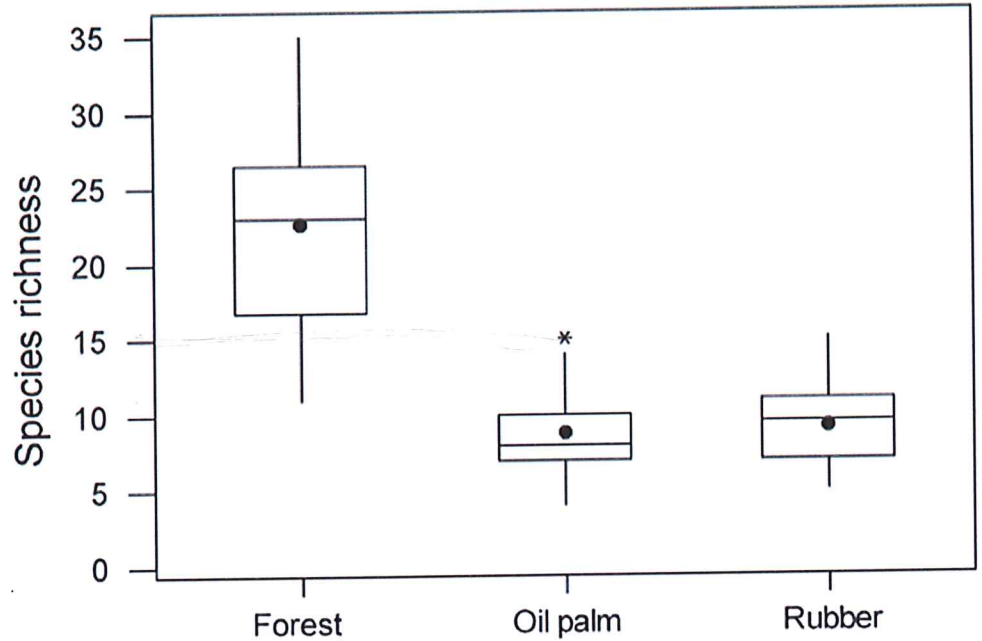
Low sugar \Rightarrow diabetic and on medication

Low sugar \Rightarrow exercise more, less sedentary

High sugar \Rightarrow more stress

15. (10 pts) In studying the environmental impact on bird life in tropical forests where palm oil plantations and rubber tree plantations are located, researchers went to Thailand and recorded bird observations at various sites in old growth forests, oil palm plantations, and rubber tree plantations. The "species richness" was measured by how many different species of birds were observed in a 20-minute period at each site.

The graph below shows the boxplots for each group's data. The dot in each boxplot marks the mean value of the data in each group.



- 3 (a) What hypothesis test could be used to see if there is a significant difference in mean species richness among all three of the groups?

ANOVA (one way)

- 1 (b) Which group showed the most species richness, on average?

Forest group

- 3 (c) Which group's data set had an outlier? Estimate and interpret the value of the outlier.

Oil Palm. 15 birds were seen in one 20 minute session.

- 1 (d) Which group's data set shows the most variability? *Forest group*

- 1 (e) Would it be reasonable to assume the population data for each group is normally distributed? Explain how you can tell from the boxplots.

Yes, each boxplot looks fairly symmetric (Some may say, "no" due to outlier - ok!)

- 2 (f) If you were to perform a Tukey Test comparing the groups, which pair(s) of groups is/are most likely to have a significant difference in species richness, on average (circle all correct answers):

Forest group, Oil Palm group

Forest group, Rubber Tree group

Oil Palm group, Rubber Tree group

16. (13 pts) A study investigated some Pain Management Strategies (PMS) for children. The study used three groups of kids, all recovering from surgery. The kids either 1) listened to an audio book, 2) listened to music, or 3) sat with noise-cancelling ear phones for 30 minutes. The children were then asked to give their pain level on a scale of 1 – 10.

Use the given data to answer the following questions.

Audio Book: 5 6 7 2 6 3 4 8 5 4 Mean = 5.000 StDev = 1.826
 Music: 5 4 4 7 6 4 6 4 3 5 Mean = 4.8000 StDev. = 1.229
 Control: 4 8 7 6 10 6 10 8 5 6 Mean = 7.000 StDev. = 2.000

3 (a) Which variable is the Factor? PMS

Is this variable (circle) QUANTITATIVE or QUALITATIVE?

What are the Factor Levels? Audio Book, Music, Control

2 (b) Which variable is the Response? Pain Level

Is this variable (circle) QUANTITATIVE or QUALITATIVE?

3 (c) What would the hypotheses be if we wanted to conduct an ANOVA test to determine whether there is a significant difference in pain, on average, for the different pain reduction methods?

Write both hypotheses, and write the null hypothesis using both math symbols and words.

$H_0: \mu_{AB} = \mu_M = \mu_C$ On average, all post op kids will have the same amount of pain regardless of strategy - strategies don't help

H_a : At least one mean pain level is different. There is an association between PMS and pain.

2 (d) Would the condition of equal variances be reasonable to assume for this problem? Explain how you can tell.

Yes. The least S.D. is 1.229.
 Two times this is 2.458 which is more than the greatest S.D. of 2.000.
 (No S.D. is more than 2 times any other...)

3 (e) Here is the ANOVA table for the data from Minitab:

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	29.60	14.800	5.02	0.014
Error	27	79.60	2.948		
Total	29	109.20			

What number in the ANOVA table gives the mean variation BETWEEN the groups? 14.8

What number in the ANOVA table gives the mean variation WITHIN the groups? 2.948

Show how the F-value is computed using these numbers:

$$F = \frac{14.8}{2.948} = 5.02$$

17. (12 pts) A sample of 6 households was monitored for one year. The household income (in thousands of dollars) and the amount of power they used (in kilowatts) is given in the table

The Correlation Coefficient and Regression Line Equation are also given.

Income (\$ thousand)	31	40	23	48	195	96
Power (Kilowatts)	5	13	10	15	61	33

Correlation: Income, Power
Correlation Coefficient = 0.991

Regression Line Equation:
Power = - 0.116 + 0.3180 Income

2 (a) Which variable is the predictor: Income
and which is the response: Power

2 (b) What is r? $r = .991$

Judging by the r-value, would the data values be close to the regression line or more widely scattered?

Close Scattered Can't tell

(c) Find the Coefficient of Determination and interpret it in the context of the problem.

2 $r^2 = (.991)^2 = .982$

98.2% of the variation in power use is explained by income level (the regression line).

2 (d) What is the slope of the regression line? .3180

What does the slope mean in terms of income and power? Be specific and use units.

For every \$1 thousand increase in income a household consumes, on average, .3180 KWatts of power.

2 (e) Use the regression equation to predict how much power a household would use annually if their household income was \$100,000.

$\$100,000 \Rightarrow \text{Income} = 100$

$\text{Power} = -.116 + .3180(100)$

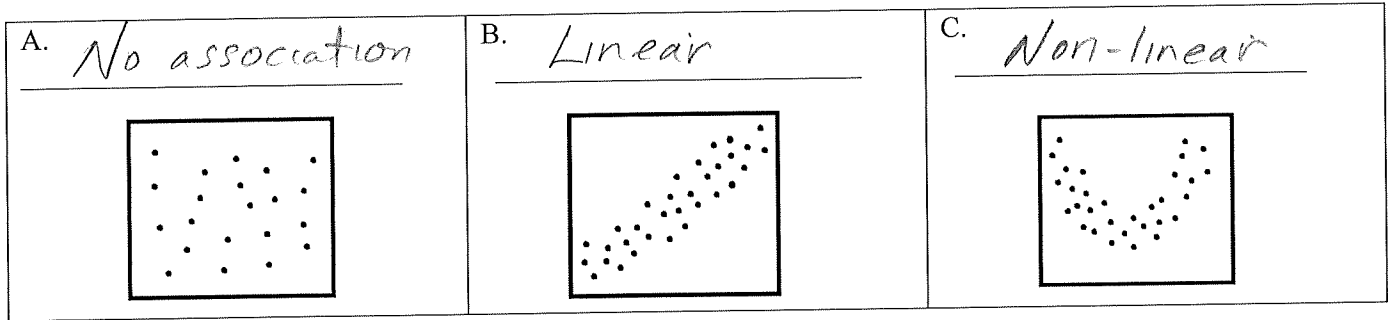
$\underline{\underline{= 31.684 \text{ KWatts}}}$

2 (f) Locate the y-intercept and interpret it. Is it meaningful?

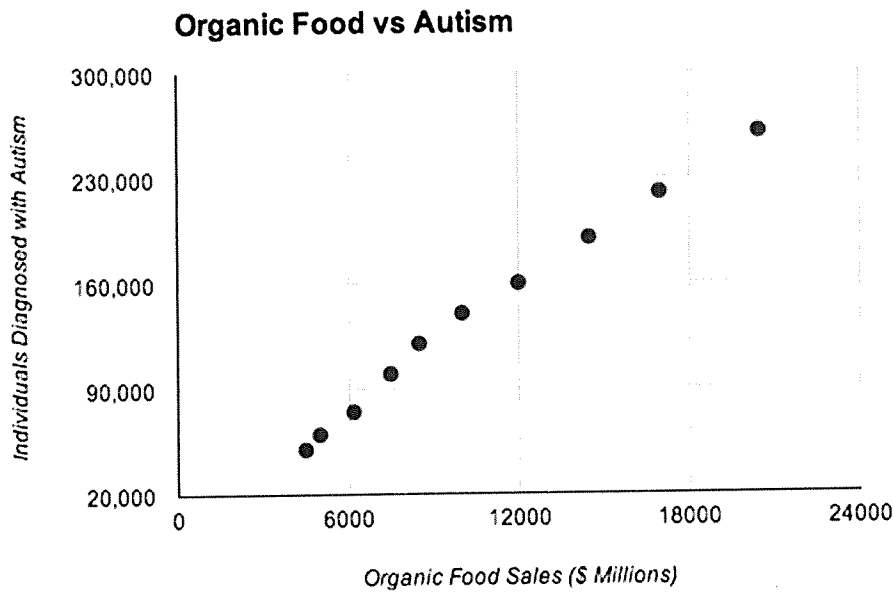
y-intercept = -0.116. This is not meaningful
Zero income isn't meaningful and negative
power use is not meaningful

18. (3 pts) Label each scatterplot with one of the following:

Linear Association, Non-linear Association, No Association



19. (4 pts) The following graph shows the relationship between organic food sales in the U.S. and autism rates.



(a) Circle the correct answer: This data shows that organic food sales and autism rates. . .

- (i) are strongly negatively correlated
- (ii) are weakly negatively correlated
- (iii) have zero correlation
- (iv) are weakly positively correlated.
- (v) are strongly positively correlated

(b) **True** or **false**: Organic food causes autism.

Have a wonderful break and best of luck in your future studies! Please keep me posted on how you're using stats in your continued education!