

Name Ken

**MATH 119 EXAM ONE**  
**Fall 2016**  
**Form A**

**Show all work on this exam form. Free response questions REQUIRE that you show supporting work to get full credit.**

Please round your answers to four digits after the decimal when possible. Make sure to BOX your final answers.

**All questions are worth 4 points unless noted otherwise.**

A

**Use the following to answer problems 1-4**

The following is a list of points scored by the players on my fantasy football team (including those on the bench, but excluding one Monday night player) in Week 3 of this season.

*Note: I clearly did not win.*

0	3	4.8	5.2	6	6.2	6.5
6.9	7.5	8.1	8.6	9.8	11.1	12.9

1. Calculate and label the 5-number summary for the data.

min	Q1	med	Q3	max
0	5.2	6.7	8.6	12.9

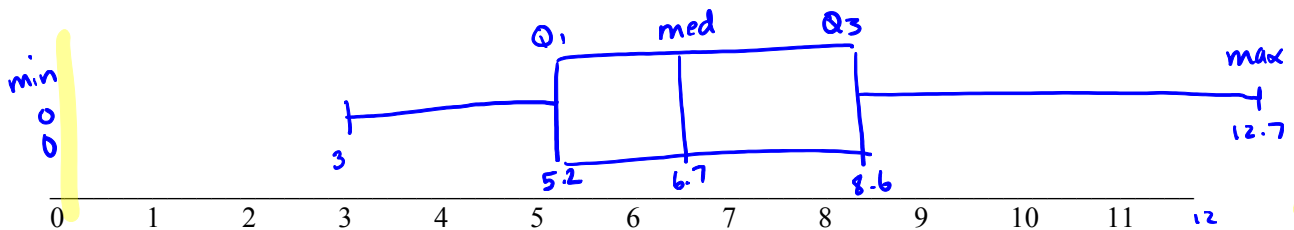
2. Are there any outliers in the data? Show all work and answer the question.

$$IQR = Q3 - Q1 = 8.6 - 5.2 = 3.4$$

$$\text{lower fence} = Q1 - 1.5(IQR) = 5.2 - 1.5(3.4) = 0.1$$

$$\text{upper fence} = Q3 + 1.5(IQR) = 8.6 + 1.5(3.4) = 13.7$$

3. (5 points) Construct and label a modified boxplot for the data.



4. Identify the best measure of center and spread for the fantasy football points above, based on the distribution. Do NOT calculate. Then justify your choice of measures in a sentence.

Center: median

Spread: IQR

Justification:

the outlier at 0 may influence the mean and standard deviation, so we should use the resistant measures of median and IQR.

A

5. To answer the question of whether cardiovascular fitness (as measured by time to exhaustion running on a treadmill (min)) is related to an athlete's performance in a 20-km ski race (min), regression analysis was performed, resulting in the following output:

Least squares regression line:  $\hat{y} = 88.8 - 2.33x$

$R^2 = 61.4\%$

What is the correlation coefficient for ski time and time to exhaustion?

$$r = -\sqrt{.614} = -0.7836$$

6. A researcher wishes to study the relationship between the level of background noise and mental concentration. The treatment (noise level) will have three levels: no noise, low-intensity noise, and high-intensity noise. The subjects are divided into three groups, and each group is to receive one of the treatments. He has available to him, a set of 60 female volunteers and a set of 90 male volunteers. What experimental design strategy would help him eliminate the introduction of gender as a confounding variable?

A. Stratified sampling  
B. Replication  
☒ C. Blocking  
D. Systematic sampling  
E. Double-blind trials

7. A data set has a standard deviation of 2 and a mean of 2. Suppose the maximum value of the data set is changed from 7 to 5. All the other values remain the same, but are less than 5. What is the effect on the standard deviation?

A. The standard deviation will stay the same.  
☒ B. The standard deviation decreases.  
C. The standard deviation increases.  
D. The effect cannot be determined without the actual data.

less spread!

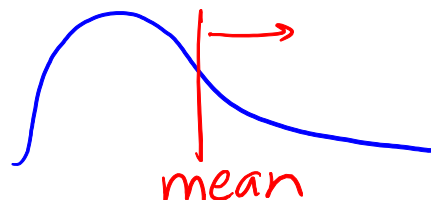
8. A residual plot is used to

A. Determine how strongly the response and explanatory variable are related.  
B. Determine whether there is a causal relationship between the explanatory and response variables.  
☒ C. Determine whether linear regression is an appropriate model for the relationship between the response and explanatory variables.  
D. Determine the direction of the relationship between the explanatory and response variables.  
E. All of the above are true.

A, B, D are scatter plots!

9. In a study of the size of houses in a certain region, the mean size was found to be 1929.5 square feet with a median size of 1200 square feet. Which of the following is most likely true about the shape of the distribution?

A. The distribution is symmetric.  
B. The distribution is skewed to the left.  
☒ C. The distribution is skewed to the right.  
D. The distribution has several low outliers.



A

**Use the following to answer problems 10-12:**

The table below shows results of a poll asking adults whether they were looking forward to the Super Bowl game, looking forward to the commercials or didn't plan to watch.

	Male	Female	Total
Game	280	201	481
Commercials	82	157	239
Won't watch	133	161	294
Total	495	519	1014

10. (3 points) What percent of adults surveyed are male and don't plan to watch the Super Bowl at all?

$$\frac{133}{1014} = 13.11\%$$

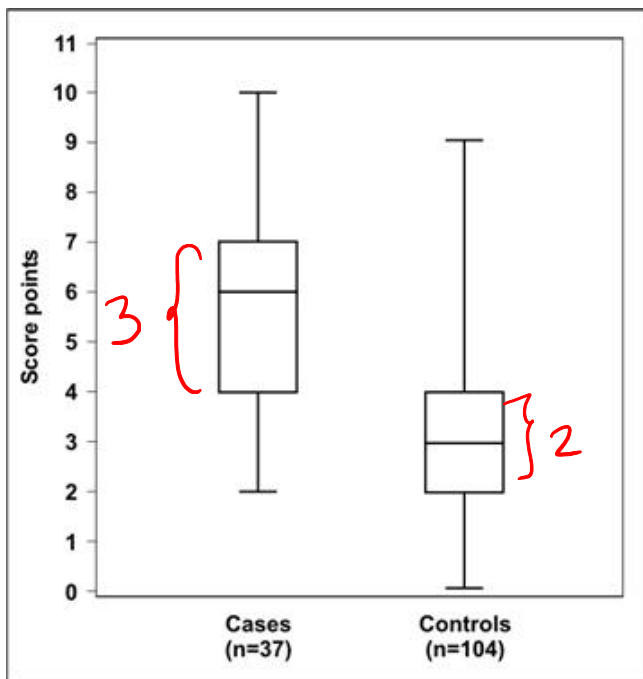
11. (3 points) What percent of adults are looking forward to watching the commercials?

$$\frac{239}{1014} = 23.57\%$$

12. (3 points) Of those looking forward to watching the game, what percent are females?

$$\frac{201}{481} = 41.79\%$$

13. The following boxplots show risk scores for Alveolar Echinococcosis among case and control groups in a Germany study.



Which of the following statements is not true?

- A. 25% of the case group had scores of at least 7.
- B. 75% of the control group had scores under 4.
- C. The IQR for the case group is smaller than that for the control group.
- D. The median for the control group is smaller than that of the case group.

A

**Use the following information to answer problems 14-18:**

A study was conducted in California to investigate the relationship between house size (square feet) and house price (thousands of dollars). The least-squares regression line to predict house price is given below.

$$\hat{y} = 263.5 + 0.174x$$

$$\text{correlation coefficient} = 0.728$$

14. (5 points) Interpret the slope of the given regression line.

For every 1 square foot increase in house size, house price increases by 0.174 thousand dollars (\$174).

15. Interpret the given correlation coefficient.

There is a moderate to strong, positive linear relationship between house size and house price.

16. (5 points) Calculate the residual of a house that is 1200 square feet that recently sold for \$449,500 (449.5 thousand dollars).

$$x = 1200$$

$$y = 449.5$$

$$\textcircled{1} \hat{y} = 263.5 + 0.174(1200)$$

$$= 472.3$$

$$\textcircled{2} e = y - \hat{y}$$

$$-\$22,800$$

$$= 449.5 - 472.5 = -22.8 \text{ thousand dollars}$$

17. Calculate the percent of variation in house price that can be explained by the regression of house price on house size.

$$R^2 = (0.728)^2 \cdot 100\% = 52.9984\%$$

$$\approx \boxed{53.00\%}$$

18. Do you think the interpretation for the y-intercept makes sense in this case? Why or why not?

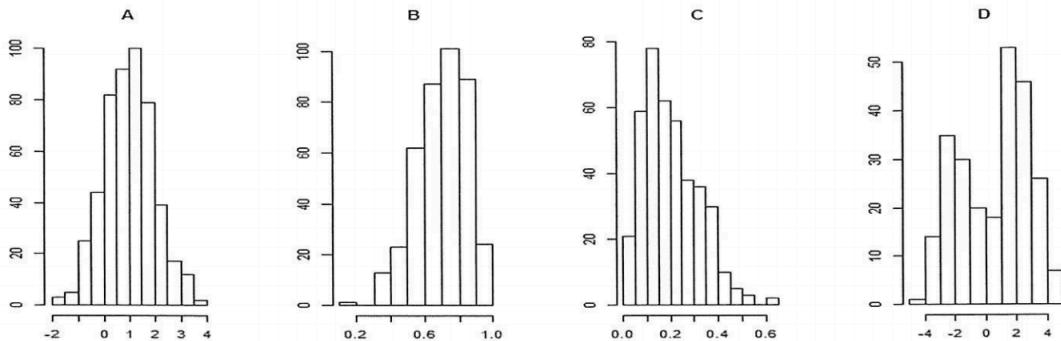
Not really...  
We'd be paying \$263,500 for a house with 0 square feet!

A

19. To analyze the level of employee satisfaction at stores in a certain large city, a grocery store chain randomly selects 10 employees from each of their store locations. What type of sampling did they use?

A. Stratified  
B. Simple random  
C. Systematic  
D. Cluster

20. Match each of the following histograms to their corresponding descriptions.



Symmetric: A  
Skewed to the left: B  
Skewed to the right: C  
Bimodal: D

21. An athlete completed an 800-m race in 140 seconds. The distribution of 800-m race times followed a bell curve with a mean of 150 seconds and a standard deviation of 6. The same athlete also competed in a swim, finished in 10.25 minutes. The distribution of swim times also followed a bell curve, with a mean of 12 minutes and standard deviation of 1.5 minutes. In which event does the athlete have a better standing relative to the other competitors in the event? How do you know? For full points, show all calculations and provide your reasoning in complete sentences.

800-m

$$X = 140$$

$$\mu = 150$$

$$\sigma = 6$$

$$Z = \frac{X - \mu}{\sigma}$$

$$= \frac{140 - 150}{6}$$

$$= -1.67$$

Swim

$$Z = -1.17$$

see explanation on form B key

A

**Use the following information to answer questions 22-23:**

An exercise physiologist is doing a research study on post-menopausal women and bone density. The researcher considers a variety of variables that could have a role in the bone density of this group of women. He looks at a group of 1529 menopausal women and asks them about **whether or not they had taken oral contraceptives** prior to menopause and how many minutes they exercise.

22. What would be the appropriate graphical displays for these two variables, respectively?

- ☒ A. bar chart and histogram  
☒ B. pie chart and bar chart  
☒ C. histogram and boxplot  
☒ D. boxplot and pie chart

23. What is the population in this study?

- GIMME  
☐ A. Women ← too broad  
☒ B. Post-menopausal women  
☒ C. 1529 post-menopausal women ← sample  
☐ D. It cannot be determined from the information given

24. Which of the following studies is an observational study?

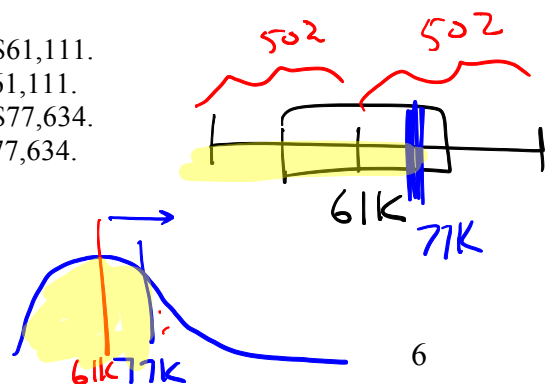
- ☒ A. A group of students is surveyed to determine the percentage will vote in the upcoming election.  
☐ B. A random sample of students is **shown a documentary** on the evolution of reincarnation beliefs through human history. A second random sample is now shown the documentary. The object of the study is to determine whether viewing the documentary affects belief in reincarnation.  
☒ C. To determine whether Vitamin C has any effectiveness at lessening the common cold, one set of cold sufferers is given no Vitamin C, a second set is given 2 grams per day, and a third set is given 5 grams per day.  
☐ D. All of these studies are observational.  
☐ E. None of these studies are observational.

GIMME  
 25. In a random **sample** of 50 college students, it was determined that the average number of times they checked social media sites per day was 4.65. The value of 4.65 is a

- ☐ A. Parameter  
☐ B. Population  
☒ C. Statistic  
☐ D. Sample  
 measure from a sample.

26. The census found that in 2003, households headed by persons between the ages of 45 and 54 had a **median** household income of \$61,111 and a **mean** household income of \$77,634. What does that tell us about household incomes for households headed by persons between the ages of 45 and 54?

- ☐ A. Over half of the households earn more than \$61,111.  
☐ B. Over half of the households earn less than \$61,111.  
☐ C. Over half of the households earn more than \$77,634.  
☒ D. Over half of the households earn less than \$77,634.



A

27. An analyst is looking at the relationship between student height and father's height. He looks at the correlation between student height and father height for all students together and for just the female students. When compared to the correlation for just the female students, would you expect the correlation when using the data for all students to be higher / lower / the same as the correlation for the female student data? Why?

It should be LOWER for all students.  
because adding the male students,  
who are taller and may have  
a different relationship with father  
heights, will add more variability.



### **Goldilocks Survey (Optional)**

#### **Difficulty**

I felt this exam was...

- A. too easy.
- B. too hard.
- C. just right.

#### **Length**

I felt this exam was...

- A. too long.
- B. too short.
- C. just right

What was the most interesting thing you've learned so far in the course?

I find the subjective nature and  
inherent biases the most interesting  
part. Very different than math  
with the ways we may be wrong.