

Exam 1 (100 pts)
STAT401 October 4, 2016

Name: KEY

1. Fill in the Blanks. (6 pts) Fill in the correct symbol in the space provided.

- (a) The population mean is μ and is estimated by \bar{x} .
- (b) The sample standard deviation is s , and it estimates σ .
- (c) The population proportion is p , and it is estimated by \hat{p} .

For the following multiple choice questions, choose the one best answer. They are worth 3 points each.

2. To study insomnia among 1115 military veterans, a psychiatrist divides the veterans into 4 disjoint groups by the primary ~~primary~~ psychiatric diagnosis: post traumatic stress syndrome, depression, psychotic, or impulse & addiction. The psychiatrist then randomly chooses two patients from each group. This is an example of a _____.

- (a) stratified random sample
- (b) cluster sample
- (c) simple random sample
- (d) completely randomized design
- (e) randomized block design

3. To determine whether GE employees support a new health-care plan, administrators require that any chosen employee must take part in a poll. A pollster randomly selects thirty employees from the company, and then asks "You don't support the new health-care plan, do you?" The resulting data might suffer from

- (a) Selection bias
- (b) Response bias
- (c) Non-response bias
- (d) There is no bias, a random sample was selected.

4. Half (50%) of the values in a distribution are

- (a) between minimum and maximum
- (b) between Q_1 and Q_3
- (c) between the mean and median
- (d) between the median and Q_3

5. If you are told a data set has a mean of 25 and a variance of 0, what must you conclude?

- (a) Someone has made a mistake.
- (b) There is only one data value in the data set.
- (c) All the values in the data set are 25.
- (d) None of the above.

6. Let A be the event that Donald Trump wins the presidential election in November and let B be the event that Hillary Clinton wins. Assume that $P(A)$ and $P(B)$ are both non-zero. Then $P(A \text{ and } B)$ is

- (a) 0
- (b) 1
- (c) $P(A) \times P(B)$
- (d) $P(A|B)$
- (e) $P(A) + P(B)$

7. Let A be the event that Steve Bullock wins Montana's gubernatorial election in November and let B be the event that Greg Gianforte wins. Assume that $P(A)$ and $P(B)$ are both non-zero. Then $P(A \text{ or } B)$ is

- (a) 0
- (b) 1
- (c) $P(A) \times P(B)$
- (d) $P(A|B)$
- (e) $P(A) + P(B)$

8. If the events A and B are not disjoint, then

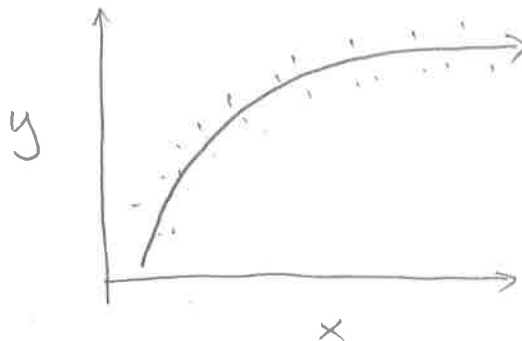
- (a) A and B are independent
- (b) A and B are dependent
- (c) A and B can be either independent or dependent.
- (d) None of the above.

9. (6pts)

(a) Sketch a scatterplot that clearly illustrates a weak negative linear relationship.



(b) Sketch a scatterplot that clearly illustrates a strong positive non-linear relationship.



The remaining questions require short answers. For those that require computations, SHOW YOUR WORK!

10. In June 2006, researchers at Wake Forest University in North Carolina concluded that long term diets containing trans-fats led to alarming patterns of weight gain and insulin resistance in monkeys (*Discover*, December 2006). Two groups of vervet monkeys were randomly assigned to one of two different diets: a common "western style" trans-fat diet or a "Mediterranean" mono-saturated fat diet.

(a) (2 pts) State the population of interest.

Vervet monkeys

(b) (2 pts) Give the name of the experimental design used by the Wake Forest researchers.

Completely randomized design

(c) (2 pts) List the explanatory variable(s).

diet

(d) (2 pts) List the response variable(s).

weight gain, insulin resistance

(e) (4 pts) ~~(4 pts)~~ Explain how a placebo can be used in this study.

Use of a diet with neither trans-fat nor mono-saturated fat in a 3rd group of monkeys.

(f) (4 pts) Give an extraneous variable that can be directly controlled in this experiment. Explain.

Timing of feedings & number of feedings should be the same for both treatment groups.

- (g) (4 pts) A critic believes that the sex of a monkey may confound the effect of diet. Explain how blocking may be used to control for the effect of sex.

Randomly assign male monkeys to both diets.
Randomly assign female monkeys to both diets.

- (h) Over six years, suppose that the twenty-two monkeys in the trans-fat group had the following weight gains (in percentages)

-2.30 0.36 1.41 1.99 2.10 3.67 4.24 4.54 5.58 5.81 6.10
7.30 9.15 9.40 10.06 12.94 13.67 15.24 19.27 20.40 30.5 31.2

- i. (6 pts) Determine the five number summary for these measurements.

$$\bar{x} = \frac{6.1 + 7.3}{2} = \frac{13.4}{2} = 6.7$$

$$Q_1 = 3.67$$

$$Q_3 = 13.67$$

$$\text{min} = -2.3$$

$$\text{max} = 31.2$$

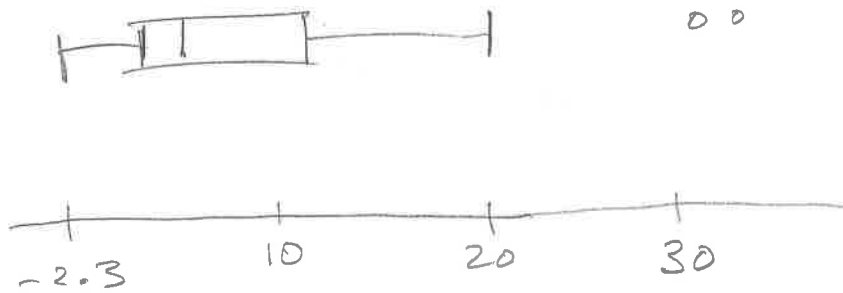
- ii. (2 pts) Determine the interquartile range.

$$IQR = Q_3 - Q_1 = 13.67 - 3.67 = 10.$$

- iii. (4 pts) Use the appropriate rule to identify any outliers.

Using the "1.5 IQR rule", there are NO outliers below Q_1 because $Q_1 - 1.5IQR = 3.67 - 15 = -11.33$. There are 2 outliers above Q_3 , 30.5 & 31.2, because they are larger than $Q_3 + 1.5IQR = 13.67 + 15 = 28.67$.

- iv. (4 pts) Construct a boxplot for this data set. Be sure to use the appropriate graphical representation for any outliers.



- v. (2 pts) Based on the boxplot, describe the shape of the distribution.

Right skewed

- vi. (4 pts) Without calculating the sample mean \bar{x} , indicate whether \bar{x} will be smaller or larger than the sample median. Justify your answer.

$\bar{x} > \tilde{x} = 6.7$ because of the right skew.

- (i) (4 pts) Can results from this study be generalized to the population of interest? Explain.

Only if the experiment was performed on a RS of monkeys. Likely, the experiment was performed on the monkeys kept at Wake Forest. So, no, the results of this study can not be generalized to all monkeys.

- (j) (4 pts) Can results of this study can be used to establish a causal relationship between diet and weight gain? Explain.

Because this was an experiment with random assignment of treatments, then for the monkeys at Wake Forest, the evidence does suggest that diet is causing weight gain.

11. Let X be the number of milliseconds that it takes for the nervous system of the *gryllus bimaculatus* cricket to respond to air vibrations from an attacking predator (like a bird or a wasp). The probabilities associated with the different values of X are given by the following table:

X	3	4	5	6	7 or more
$P(X)$	0.14	0.27	0.28	0.18	0.13

- (a) (2 pts) Give the sample space for X .

$$S = \{3, 4, 5, \dots\}$$

- (b) (4 pts) Use the appropriate rules of probability to explain why the table above gives a valid set of probabilities for all of the outcomes for X .

Because all probabilities are between 0 & 1; and because the probabilities sum to 1.

- (c) (4 pts) Vibrations are applied to $n = 3$ randomly chosen crickets and their nervous systems are recorded. Is this an example of an observational study or an experiment? Explain.

An experiment because a treatment, vibrations, was applied.

- (d) (3 pts) For a randomly chosen cricket, what's the probability that the cricket's nervous system takes 6ms or longer to respond to a vibration?

$$\begin{aligned}
 P(X \geq 6) &= P(6) + P(7 \text{ or more}) = 0.18 + 0.13 \\
 &= 0.31
 \end{aligned}$$

- (e) (4 pts) For $n = 10$ randomly chosen crickets, what's the probability that at least one responds in 3 ms?

$$\begin{aligned}
 P(\text{at least 1}) &= 1 - P(\text{none}) \\
 &= 1 - (.86)^{10} \\
 &= .78
 \end{aligned}$$