

# Exam 1

STAT401

September 29, 2006

Name: \_\_\_\_\_

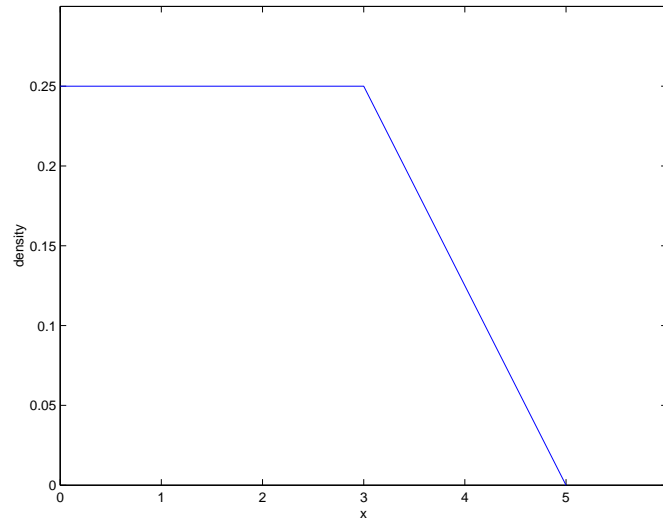
For the following multiple choice questions, choose the one best answer. They are worth 3 points each.

1. An investigator is interested in determining whether physical exercise is useful in alleviating psychological depression in elderly people. The investigator randomly selects individuals from a nursing home then divides the sample into three groups depending on their pre-existing level of depression. Subjects within each group were randomly assigned to exercise or control treatments. After three months of treatment the depression level in each subject was measured. This is an example of a \_\_\_\_\_.
  - (a) stratified random sample
  - (b) simple random sample
  - (c) completely randomized design
  - (d) randomized block design
2. Let  $X_1, X_2, \dots, X_n$  be a random sample of size  $n$  from a non-normal population. Suppose that 10.0 is added to each  $X_i$ . Then
  - (a) The sample mean and sample standard deviation each increase by 10.0
  - (b) Only the sample mean increases by 10.0
  - (c) Only the sample standard deviation increases by 10.0
  - (d) Neither the sample mean nor the sample standard deviation increase by 10.0.
  - (e) The answer is unknown because the population is non-normal.
3. Let  $A$  be the event that Tester wins the election in November and let  $B$  be the event that Burns wins the election in November. 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)  $P(A) + P(B)$
4. Let  $A$  be the event that Tester wins the election in November and let  $B$  be the event that Burns wins the election in November. 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)  $P(A) + P(B)$

5. 100 students took a test on which the mean score was 73 with a standard deviation of 3. A “C” was given to all who scored between a 70 and a 76. Approximately how many C’s were there, assuming scores were normally distributed?
- (a) 68
  - (b) 95
  - (c) 99
  - (d) Can’t be determined with the information given.
6. A distribution of 6 scores has a median of 21. If the highest score increases by 3 points, the median will be:
- (a) 21
  - (b) 21.5
  - (c) 24
  - (d) Cannot be determined without additional information.
  - (e) None of these.
7. 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) the median and  $Q_3$
8. 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.
9. If the events  $A$  and  $B$  are disjoint, then
- (a)  $A$  and  $B$  are independent
  - (b)  $A$  and  $B$  are dependent
  - (c)  $A$  and  $B$  can be either independent or dependent.
10. Chebyshev’s rule states that the proportion of observations that are within 2 standard deviations of the mean is at least
- (a) 0%
  - (b) 68%
  - (c) 75%
  - (d) 95%

11. The plot below displays the density of a continuous variable  $X$ . The probability  $P(X > 2)$  is

- (a) 0
- (b)  $\frac{1}{4}$
- (c)  $\frac{1}{2}$
- (d)  $\frac{3}{4}$



12. A political pundit in November 2000 randomly called different American households and then asked voters “You didn’t vote for George Bush, did 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.

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

13. (6 pts) A biologist sets up an observation post at the southeast corner of Yellowstone Lake, close to the “furthest point from civilization” in the Lower 48. If the probability of observing a grizzly bear on any given day is .21, then what is the probability that a grizzly bear is not observed until the 5th day?

14. (21 pts, 3 pts each for (a)-(g)) A Fish and Wildlife researcher is studying population abundance of trout in a remote creek in Yellowstone National Park. Fish are caught by “electrofishing” - applying an electrical current to the creek temporarily paralyzes the fish and they float to the top of the creek to be caught. Teams of poor, hungry graduate students measure length, weight, and species of each fish caught while also recording creek location, depth, temperature, and flow. The researcher is interested in determining the characteristics of the creek where each species of trout tend to be caught.

(a) Is this an observational study or an experiment? Explain why your answer is correct.

(b) What are the individuals in the study?

(c) Give the explanatory variable(s).

(d) Give the response variable(s).

(e) Give a possible confounding variable.

(f) Give a categorical variable being measured.

(g) Give the sample space corresponding to the measurement of creek depth.

(h) At one section of the creek, the weights of nine trout were measured (in ounces):

17, 21, 9, 8, 29, 20, 6, 12, 15

i. (6 pts) Use the technique given in class to find the five number summary for this data.

ii. (4 pts) Give a boxplot for this data.

iii. (6 pts) Are any outliers present? Use the rule (which depends on the five number summary) discussed in class. Clearly indicate which point(s), if any, are outlier(s), and indicate whether it is a **mild** outlier or an **extreme** outlier.

15. (6 pts)

(a) List the FOUR basic principles of **experimental design**.

(b) Which of these principles of experimental design is most useful for preventing extraneous variables that are unknown to the investigator from becoming confounding variables?

16. (15 pts) In competitive sports, athletes' training sessions are often video recorded so coaches can replay the training sessions for more effective feedback to the athlete. Some people believe the video recording makes athletes nervous and actually decreases their performance. Sixty high-school male competitive tennis players of varying ability have volunteered for this study. Video recording was randomly assigned to 30 of the players; the other half did not experience video recording. The number of successful serves out of 100 is recorded as the measure of performance.

(a) What are the experimental units?

(b) What is the explanatory variable (factor)?

(c) What is the response variable?

(d) Give the sample space of possible outcomes of the response variable.

(e) What specific extraneous variable is being **directly controlled** by the researchers in this study?