

# Exam 1 Solutions

STAT401      September 29, 2006

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

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

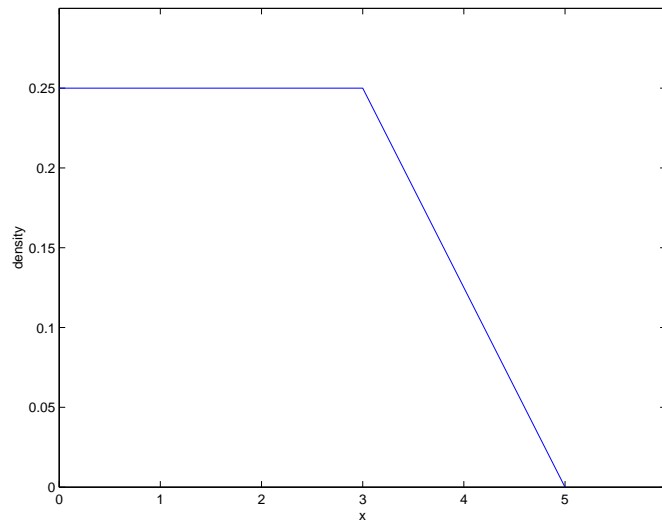
1. An investigator is interested in determining whether physical exercise is useful in alleviating psychological depression in elderly people. The investigator randomly selects individuals from a nursing home then divides the sample into three groups depending on their pre-existing level of depression. 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 \_\_\_\_\_.  
(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  
(b) Only the sample mean increases by 10.0
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
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  
(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
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
7. Half (50%) of the values in a distribution are  
(b) between  $Q_1$  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?  
(c) All the values in the data set are 25.
9. If the events  $A$  and  $B$  are disjoint, then  
(b)  $A$  and  $B$  are dependent

10. Chebyshev's rule states that the proportion of observations that are within 2 standard deviations of the mean is at least

(c) 75%

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

(c)  $\frac{1}{2}$



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

(b) Response bias

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?

By independence and complementarity, the probability is  $.21(.79^4) \approx .08$ . Thus, the probability of not seeing a bear until the 5th day is .08.

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.

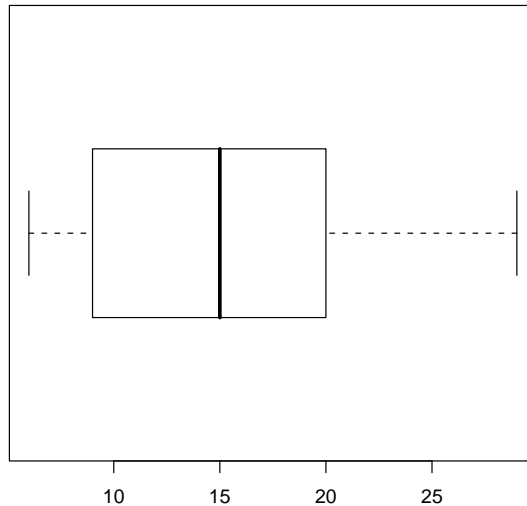
(a) This is an observational study since no treatment is being applied to the trout. The electrofishing is NOT a treatment!

(b) The individuals in the study are the trout.

- (c) The explanatory variable is the species of trout.
- (d) The response variables are creek location, depth, temperature and flow.
- (e) Possible confounding variables are the type of vegetation along the side of the creek, altitude, and predation.
- (f) Trout species is the only categorical variable being measured.
- (g) The sample space corresponding to the measurement of creek depth is [0feet, 100feet]. I chose 100 feet as the upper end point of the interval since creeks don't generally get any deeper than 10 feet, never mind 100 feet.
- (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) Using the technique given in class, the five number summary for this data is (6, 8.5, 15, 20.5, 29).
- ii. (4 pts) Give a boxplot for this data.



- iii. (6 pts) The interquartile range for this data is  $IQR = 20.5 - 8.5 = 12$ . The “1.5IQR Rule” states that mild outliers are between  $1.5IQR=18$  and  $3IQR=36$  from the nearest quartile and extreme outliers are farther than 3 IQR's from the nearest quartile. Thus, there are no outliers.

15. (6 pts)

- (a) The FOUR basic principles of **experimental design** are: Direct Control, Random Assignment, Replication and Blocking.
- (b) Random Assignment 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) The experimental units are the male high school tennis players.
  - (b) The explanatory variable is “video recording” with two levels: yes and no.
  - (c) The response variable is the number of successful serves out of 100
  - (d) The sample space of possible outcomes of the response variable is  $S = \{0, 1, 2, \dots, 100\}$ .
  - (e) Gender is being **directly controlled** by the researchers in this study since only male high-school tennis players are in the experiment.