AP Statistics Free Response Review

1. Long-term records from the Serengeti National Park in Tanzania show interesting ecological relationships. When wildebeest are more abundant, they graze the grass more heavily, so there are fewer fires and more trees grow. Lions feed more successfully when there are more trees, so the lion population increases. Here are data on one part of this cycle, wildebeest abundance (in thousands of animals) and the percent of the grass area burned in the same year:

Wildebeest (1000s)	Percent burned	Wildebeest (1000s)	Percent burned
396	56	622	60
476	50	600	56
698	25	902	45
1049	16	1440	21
1178	7	1147	32
1200	5	1173	31
1302	7	1178	24
360	88	1253	24
444	88	1249	53
524	75		

a) To what extent do these data support the claim that more wildebeest reduce the percent of grasslands that are burned?

b) How rapidly does burned area decrease as the number of wildebeest increase?

2. Biological measurements on the same species often follow a normal distribution quite closely. The weights of seeds of a variety of winged bean are approximately normal with mean 525 mg and standard deviation 100mg.
a) What percent of seeds weigh more than 500 mg?

b) If we discard the lightest 10% of these seeds, what is the smallest weight among the remaining seeds?

3. U.S. car sales change month by month because of car prices, the national economy, and other factors that affect the sales data. How about gas price? Does it also affect car sales in some manner? To answer this question, we analyze U.S. car sale data from January 1000 to June 2001, a period when gas prices were generally increasing, to see if we can find some relation between gas price and car sales.

a) As gas price goes up, people think more about gas mileage when they are deciding what car to buy. Figure 1.1 shows side-by-side boxplots of gas mileage for each size class of car—subcompact, compact, midsize, and large. Write a few sentences comparing the distributions of gas mileage for the four car classes.

A reasonable theory is that as gas prices increase, people will buy smaller cars. Figure 1.2 displays a scatterplot of the average price of regular unleaded gasoline and the percent of cars sold that month with gas mileage greater than 26.75 miles per gallon (mpg). The least-squares regression line has been added to the plot, and some computer regression output appears at the top of the graph.

b) Interpret the slope and y-intercept of the LSRL in the context of the problem.

c) During on e of the months in this study, the average gas price was $1.50. Calculate the residual for that month.

d) What is the value of the coefficient of determination and interpret what that value means in context of the problem.

4. Is "tartar control" Crest toothpaste more effective at preventing tartar buildup on teeth than regular Crest toothpaste? Researchers recruit 210 volunteers (120 men and 90 women) who first get a free tooth cleaning. Then the researchers randomly assign 105 to the "tartar control" toothpaste group and 105 to the regular toothpaste group. All 210 people are given unmarked tubes of the appropriate toothpaste and instructed on how often to brush their teeth each day. They are not told which group they are in, and the two kinds of toothpaste look the same. After 6 months, they are rated on tartar buildup by dentists who don't know who was in each group.

a) Is this study and experiment or observational study? Justify your answer.

b) Did the researchers use any form of blinding in this experiment? Justify your answer.

c) Why did the researchers give all of the volunteers a free tooth cleaning?

d) The researchers used a completely randomized design in this study. Describe a design that uses blocking to improve on the original design of the study. Give one reason why the block design might be preferable.

e) Identify a possible lurking variable that might be confounded with how the toothpaste controls tartar. Explain how the variable you chose could lead to confounding.

5. Here's the opening of a press release from June 2004: "Starbucks Corp. on Monday said it would roll out a line of blended coffee drinks intended to tap into the growing popularity of reduced-calorie and reduced-fat menu choices for Americans." You wonder if Starbucks customers like the new "Mocha Frappuccino Light" as well as the regular version of this coffee.

a) Describe a matched-pairs design using 40 regular Starbucks customers that would help answer this question. Be sure to discuss blinding.

b) Explain how you would use the partial table of random digits below to do the randomization that your design requires. Then use your method to assign treatments to the first 3 subjects. Show your work clearly on your paper.

07511   88915   16853   84569   79367   32337   03316   81486   69487   60513   09297   00412   71238   27649   39950

c) Would it have been better to use a completely randomized design instead of a match pairs design? Justify your answer.

6. A large university wants to gather student opinion about parking for students on campus. It isn't practical to contact all enrolled students.

a) Give an example of a way to choose a sample of students that is poor practice because it depends on voluntary response.

b) Give an example of a bad way to choose a sample that doesn't use voluntary response.

c) Now suggest a method of obtaining a sample that should lead to more reliable results.

7. In the language of government statistics, you are "in the labor force" if you are available for work and either working or actively seeking work. The unemployment rate is the proportion of the labor force (not of the entire population) who are unemployed. Here are data from the Current Population Survey for the civilian population aged 25 years and older in 2004. The table entries are counts in thousands of people.

Highest education	Total Pop.	In labor force	Employed
Didn't finish high school	27,669	12,470	11,408
High school but no college	59,860	37,834	35,857
Less than bachelor's degree	47,556	34,439	32,977
College graduate	51,582	40,390	39,293

a) Find the unemployment rate for people with each level of education. How does the unemployment rate change with education?

b) What is the probability that a randomly chosen person 25 years of age or older is in the labor force?

c) If you know that a randomly chosen person 25 years of age or older is a college graduate, what is the probability that he or she is in the labor force?

d) Are the events "in the labor force" and "college graduate" independent? Justify your answer.

8. Choose a young person (aged 19 to 25) at random and ask, "In the past seven days, how many days did you watch television?" Call the response X for short. Here is the probability model for the response:

Days X	0	1	2	3	4	5	6	7
Probability	0.04	0.03	0.06	0.08	0.09	0.08	0.05	???

a) What is the probability that a young person watches television for seven days?

b) Calculate the mean of the random variable X. Interpret this value in context.

c) Suppose that you asked 100 randomly selected young people to respond to the question and found that the mean of their responses was 4.96. Would this result surprise you? Justify your answer.

9. All human blood can be "ABO-typed" as one of O, A, B, or AB, but the distribution of the types varies a bit among groups of people. Here is the distribution of blood types for a randomly chosen person in the United States:

Type:	O	A	B	AB
probability:	0.45	0.4	0.11	0.04

a) What is the probability that in a randomly chosen couple the wife has type B blood and the husband has type A?

b) What is the probability that one of the randomly chosen couple has type A and the other has type B?

c) What is the probability that at least one of the randomly chosen couple has type O blood?

11. In the 4 x 100 medley relay event, four swimmers swim 100 yards, each using a different stroke. A college team preparing for the conference championship looks at the times their swimmers have posted and creates a model based on the following assumptions:
- The swimmers' performances are independent.
- Each swimmer's times follow a normal model.
- The means and standard deviations of the times (in seconds) are as shown:

Swimmer	Mean	SD
1 (backstroke)	50.72	0.24
2 (breaststroke)	55.51	0.22
3 (butterfly)	49.43	0.25
4 (freestyle)	44.91	0.21

a) What are the mean and standard deviation for the relay team's total time in this event?

b) The team's best time so far this season was 3:19.48. (That's 199.48 seconds.) Do you think the team is likely to swim faster than this at the conference championship? Justify your answer.