Stat 2000 International – Sample Midterm 1

Midterm 1 consists of 25 questions: 15 multiple-choice questions (with exactly 1 correct answer) and 10 text-based questions where you have to provide a verbal explanation or calculate one or multiple numerical values. Some of the questions require you to use WebStat or any of the interactivities from within CyberStats.

The exam is worth a total of **200 points**. The number of points for each question is indicated in parentheses at the beginning of each question. You have exactly **60 minutes** to complete the exam. Try to correctly answer as many questions as possible during this time period. **You are allowed to answer questions in any order.** Start with a question that seems the easiest for you. If you cannot answer a question within a short time, move to another question, and come back to the previously unanswered questions toward the end of the exam.

Obviously, you are allowed to correct your answers. However, only your last submitted answer will be graded. If you change a previously correct answer and your last submitted answer is incorrect, you will obtain 0 points for your last submitted answer.

Note:

The actual exam will be fully given within CyberStats. This means you will have to mark your choices of multiple-choice questions and fill in answers to text-based questions within CyberStats. In the actual exam, interactivities will be directly linked to the questions. Make sure to memorize your CyberStats password for the exam.

1. (11 Points) Experiment with the interactivity used in **Questions 17-19** in **Exercises 2** in **Unit A-8: Correlation – Describing Bivariate Data** before answering this question.

Which of these is easiest?

- **a.** Determining the difference between a plot with r = 1 and r = .95
- **b.** Determining the difference between a plot with r = .85 and r = .90
- **c.** Determining the difference between a plot with r = .50 and r = .55
- **d.** Determining the difference between a plot with r = 0 and r = .05

2. (6 Points) A study is done to compare the extent of heart disease in people who drink 1 to 2 alcoholic drinks per day to the extent of heart disease in non-drinkers. The researcher is able to study 200 individuals of each type.

Other factors that might affect the extent of heart disease are smoking habits and exercise habits. The smoking habits of the two groups of people are similar, but those who drank generally exercised less than the non-drinkers.
In this study, the response variable is:

- **a.** exercise
- **b.** heart disease
- **c.** smoking
- **d.** drinking status

3. (6 Points) Suppose you read that a study has found that 38% of college students engage in 'risk-taking behavior'.

Which question should be asked about this study?

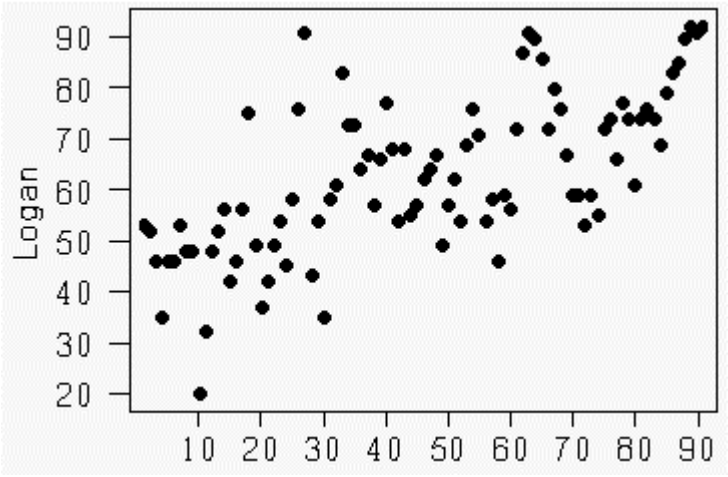
- **a.** How many students were surveyed?
- **b.** How were these students selected to be in the sample?
- **c.** What is the definition of "risk-taking behavior"?
- **d.** All of the above

4. (11 Points) This exercise concerns long-term relative frequency. Use the **WebStat Interactive** associated with **Questions 18-21** in **Exercises 3** in **Unit B-1: Probability Concepts** with the results of a coin tossing experiment simulating 700 tosses. It is not known whether the coin is fair. The data set contains:

- the result of each coin toss (0 for a tail, 1 for a head),
- the average (percentage) of heads up to that point,
- the length of the current streak of heads or tails, and
- the difference between the observed number of heads and the expected number of heads (which is half the number of tosses if the coin is fair).

Determine the shape of a histogram of heads or tails if the coin were fair and were tossed 700 times. Describe and explain the histogram obtained in **WebStat** and compare it with the histogram based on a fair coin.

5. (6 Points) Consider the graph with the daily high temperatures in Logan, Utah from Mar 17, 1996 to June 15, 1996. The correlation coefficient is 0.668, a high correlation.



Which of the following best describes the relationship of temperatures to the time from March 17, 1996 to June 15, 1996?

a. The correlation is very strong. The points follow a general linear trend within the time from Mar 17, 1996 to June 15, 1996.

b. The correlation coefficient of 0.668 suggests a strong relationship, but it is only measuring how close the data are to a straight line, within this time frame. There is a definite pattern, it may not be linear.

c. There is a definite pattern, it may not be linear.

d. All of the above

6. (11 Points) Use WebStat. Load from "**Data > Sample data**" the data set "Left-right feet lengths.dat" and compute the correlation coefficient for the variables "Left" and "Right".

The correlation is closest to

a. zero
b. -.97
c. .97
d. .99

7. (6 Points) The following temperatures were recorded on a February day for 25 cities:

25 31 34 35 36 41 42 42 44 45 45 46 47 48 48 48 49 50 50 52 52 53 53 53 54

The median of these 25 numbers is closest to

a. 25
b. 41
c. 47
d. 51

8. (6 Points) Two events are dependent if when one occurs:

a. The probability of the other is unchanged
b. The other event will always occur
c. The other will never occur
d. The probability of the other is changed

9. (10 Points) Annually, about 25% of the U.S. population gets a cold and between 35 -50% gets the flu, but both illnesses are highly age-related. In 1994, for the 5-24 age group, 32.7 million of 74.8 million got the flu. For people 45 years of age and older, 18.8 million of 81.7 million got the flu.

Calculate the probability of getting the flu for the 45-and-older group.

10. (6 Points) For eye color, brown is dominant and blue is recessive. Suppose both parents have one gene for brown eyes and one for blue. A Punnett square is used to determine the probability of an offspring having brown eyes. This probability is found to be 3/4.

Is this estimate an example of experimental, theoretical or personal probability?

a. This is an experimental probability, since a Punnett square was used to determine the probabilities.
b. This is a theoretical probability, since a Punnett square was used to determine the probabilities.
c. This is a personal probability, since a Punnett square was used to determine the probabilities.
d. This is an experimental, as well as personal probability, since a Punnett square was used to determine the probabilities.

11. (6 Points) A researcher selects a sample from a list of all patients at one of five large hospitals in the following manner. A patient is chosen from the first 25 on the list, then every 25th patient from that point forward is selected.

This is an example of a:

a. simple random sample
b. systematic sample
c. stratified sample
d. cluster sample

12. (6 Points) A screening test for high blood pressure (a diastolic blood pressure of 90mm Hg or higher) produced the following results:

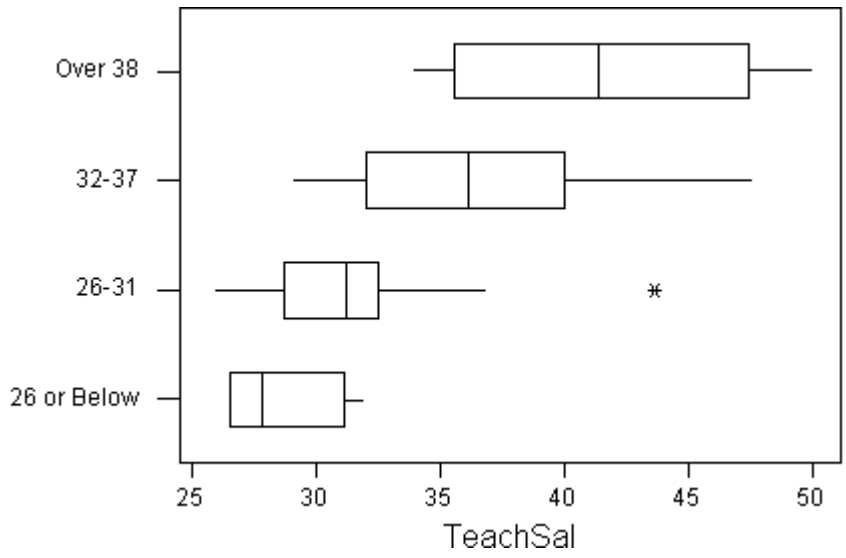
Test Result	Hypertension Present	Hypertension Absent	Total
Positive Test	477	163	640
Negative Test	173	4687	4860
Total	650	4850	5500

What is the probability that a person who tests positive has hypertension?

a. 74.5%
b. 25.4%
c. 9.8%
d. 8.7%

13. (6 Points) We categorize the fifty states plus DC into four groups based on household income, in thousands of dollars. The groups are:

- Income < 26
- 26 <= Income < 31
- 32 <= Income < 37
- 38 <= Income



Which of the following statements is true?

a. About half of the states plus DC in the "over 38" group have teacher salaries between $30,000 and $35,000

b. At least 75% of the states plus DC in the "32 to 37" group have teacher salaries above the median teacher salary of the "26 to 31" group

c. About 50% of the teacher salaries in the "32 - 37" group are below $30,000

d. The largest teacher salary in the "26 or Below" group is smaller than the smallest teacher salary in the "26 to 31" group

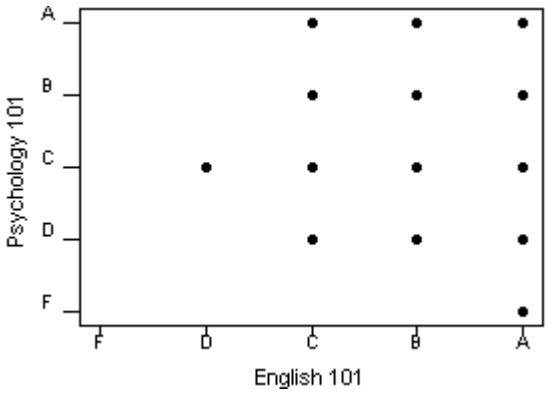
14. (11 Points) The interactive plots associated with **Questions 57-59** in **Exercises 5** in **Unit A-5 Describing Data Graphically** are based on the same class of statistics students. The data are split into groups based on the students' heights in inches. The variable of interest is "MPH," the fastest each person has ever driven a car, in miles per hour.

What is the upper quartile for height?

15. (10 Points) Suppose that 30 students will participate in an experiment in which the effectiveness using a web-based approach to teaching statistics is compared to the effectiveness of a textbook-based approach.

Describe how the researcher could assign participants to the two different approaches.

16. (6 Points) This is a scatterplot showing X = grade in English 101 versus Y = grade in Psychology 101 for 120 U. of I. students.



How many people have a B in English 101 and a C in Psychology 101?

a. 0
b. 1
c. 2
d. There is not enough information to answer the question

17. (10 Points) There is a problem with the situation in this question. Describe the problem.

A list of registered automobile owners is used to select a random sample for a survey about whether people think homeowners should pay a surtax to support public parks.

18. (11 Points) The interactive histogram associated with **Questions 50-56** in **Exercises 5** in **Unit A-5: Describing Data Graphically** exhibits the responses of 136 statistics students when asked, "How many dogs plus cats have you and your family had in your lifetime?"

Do you see skewness? If so, in which direction?

19. (10 Points) The daily high temperatures for the first week in January, 1996, at Berkeley are 67, 67, 57, 57, 57, 61, 60. The mean of these values is 60.8571, and the standard deviation is 4.488.

What are the quartiles of the temperatures?

20. (6 Points) Heights for 100 individuals range from 55 inches to 75 inches. Half the individuals have heights around 65 inches, a quarter have heights near 55 inches, and the rest have heights near 75 inches.

The SD of heights is about:

- **a.** 10 inches or less
- **b.** 20 inches
- **c.** 40 inches
- **d.** 400 inches or more

21. (10 Points) As recently as 1970, most children born with cystic fibrosis (CF), an inherited disease in which a buildup of thick, sticky mucus obstructs lung and pancreatic function, rarely lived beyond their eighth birthday. Today the median life expectancy has risen to 30 years, and new treatments--as well as some currently in development--promise to convert this from one of the most hopeless of ailments to one of the most treatable. CF is caused by a recessive gene. Further, carriers have no symptoms. Among whites 1/25 are carriers, (Recall that each parent provides one of two possible genes: C (no CF gene), or c(CF gene)).

If an unrelated white couple marries, what is the chance they are both carriers?

22. (6 Points) Two dice are thrown. Suppose that the sample space is given by the 36 outcomes { (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), ... }.

Consider the random variable given by the SUM of the dots showing on the uppermost faces of the dice. Call this random variable X. Which of the following statements is FALSE?

- **a.** The set of outcomes where X=2 is a simple event.
- **b.** P(4) = 3/36.
- **c.** The set of outcomes where X=5 is a compound event.
- **d.** P(1) = 1/6.

23. (6 Points) A population consists of 500 Statistics students. A random sample of 50 Statistics students is selected from the population. You are interested in the diastolic blood pressure for men and women.

What is the sampling frame?

- **a.** 50 Statistics students
- **b.** Raw data consisting of diastolic blood pressures
- **c.** Raw data consisting of sex (male/female)
- **d.** 500 students in the population

24. (6 Points) A random sample of individuals in a county has been selected for a survey. You have been hired to conduct the survey and decide to use random digit dialing. The first individual you contact tells you that she has no free time and you should call someone else.

What should you do and why?

a. Call someone else; the next number is in the same county so it does not matter which individual is included

b. Try to get the person to respond anyway; if you substitute someone else the sample will be biased in favor of people with more time

c. Ignore this individual and reduce the sample size by one; people who do not have a lot for free time probably will not respond truthfully anyway

d. Replace this individual with another one chosen randomly from the county; one randomly chosen individual is as good as another

25. (11 Points) The interactivity associated with **Questions 10-13** in **Exercises 2** in **Unit A-7: Scatter Plots** is based on 11 observations with 6 variables. Use it to construct some plots. For the next question, choose among the following choices for the best description of each plot.

- Linear
- Linear, but with an outlier
- Curved
- None of the above

Best description for X = X1 and Y = Y3?