**Stat 2300 International, Fall 2005 – Midterm**

Thursday, October 27, 2005

**Your Name: _____**

**Your CyberStats ID: _____**

The Midterm consists of 35 questions: 20 multiple-choice questions (with exactly 1 correct answer) and 15 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. You have to decide yourself which of these tools you may have to use.

The exam is worth a total of **350 points**. The number of points for each question is indicated in parentheses at the beginning of each question. You have approximately **90 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.

Mark your answers to the multiple choice questions and fill in the spaces for the text-based questions on the **answer sheets**. Do not write your answers on the pages with the questions. However, you can use those pages for personal comments and calculations. Make sure to write your name and your CyberStats ID on the pages with the questions **and** on the answer sheets. You have to turn in the pages with the questions **and** the answer sheets.

For multiple choice questions, mark exactly one of the choices (representing options a, b, c, or d) with a solid dot. Only 1 answer is correct. If you mark more than 1 of the options, this will automatically be an incorrect answer (even if one of the options you have marked is correct). For the text-based questions, indicate the formula you are using, the numerical values you have to fill in, and the final result (e.g., $n*p = 10*0.2 = 20$). Just the (correct) final result will not provide you with the full points for this question. If you use CyberStats to calculate a result, indicate which tool (e.g., binomial calculator) and the result from within CyberStats. Please do not write outside the text boxes.

In case of any problems with CyberStats, please inform your local instructor as quickly as possible and try to answer the questions as far as possible. Return to these questions later on. In case CyberStats is unavailable for an extended time period, your exam score will be adjusted accordingly.

**1.** (8 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

**2.** (8 Points) The following temperatures were recorded on a February day for 25 cities:

25 31 34 35 35 36 36 37 38 39 40 41 41 45 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

**3.** (8 Points) Two events are mutually exclusive 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

**4.** (8 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

**5.** (8 Points) A screening test for high blood pressure (a diastolic blood pressure of 90mm Hg or higher) produced the following results:

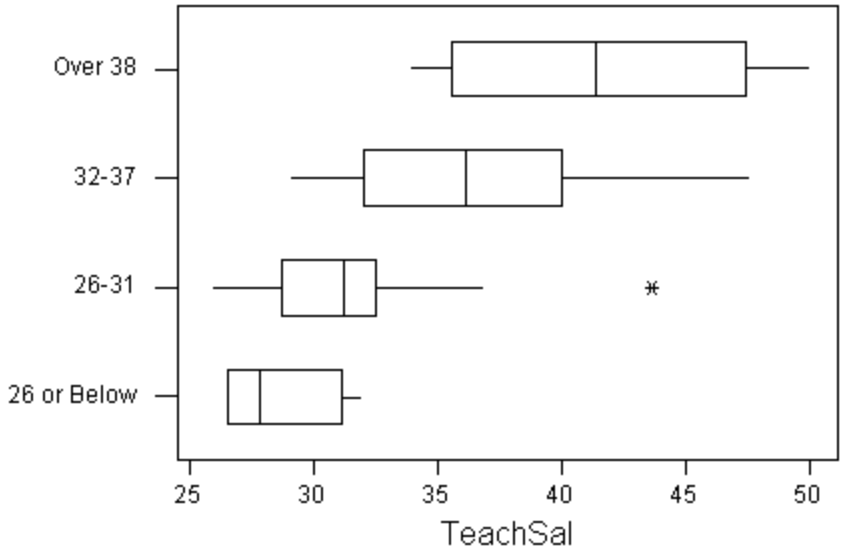
| Test Result | Hypertension Present | Hypertension Absent | Total |
|---|---|---|---|
| Positive Test | 477 | 163 | 640 |
| Negative Test | 173 | 4687 | 4860 |
| Total | 650 | 4850 | 5500 |

What is the probability that a person who tests positive has **no hypertension**?

- **a.** 74.5%
- **b.** 25.5%
- **c.** 9.8%
- **d.** 8.7%

**6.** (8 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 **false**?

**a.** The median income of the states plus DC in the "over 38" group is higher than $40,000 and less than $45,000

**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 75% of the teacher salaries in the "26 to 31" 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 "over 38" group

**7.** (8 Points) Weights for 100 individuals range from 160 pounds to 200 pounds. Half the individuals have weights around 180 pounds, a quarter have weights near 160 pounds, and the rest have weights near 200 pounds.

The SD of weights is about:

**a.** 20 pounds or less

**b.** 40 pounds

**c.** 180 pounds

**d.** 4000 pounds or more

**8.** (8 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 $X$=12 is a simple event.

**b.** $P(10) = 4/36$.

**c.** The set of outcomes where $X$=7 is a compound event.

**d.** $P(2) = 1/36$.

**9.** (8 Points) A life insurance company is determining the mean payout per policy for $80,000 life insurance policies on 50 year old women. These policies have an accidental death benefit that pays beneficiaries twice the policy amount when the death is accidental. The probability is 27/10,000 that a 50 year old woman dies from natural causes during the year she is 50, and the probability is 1/10,000 that she dies from accidental causes during that year.

| $X$ | 0 | $80,000 | $160,000 |
|---|---|---|---|
| Pr($X$) | 9,972/10,000 | 27/10,000 | 1/10,000 |

What is the expected value of the mean amount paid out?

**a.** $0
**b.** $145
**c.** $232
**d.** $80,000

**10.** (8 Points) A machine produces parts. The probability that an individual part is defective is .01. If a random sample of 10 parts contains one or more defective parts, the machine is shut down for repairs.

What is the probability that the machine will be shut down for repairs based on one sample of 10 parts?

**a.** 0.01
**b.** 0.0956
**c.** 0.9044
**d.** 0.99

**11.** (8 Points) 450 manufactured items are randomly selected from a production run and 9 of them are defective.

The best guess for the probability that a randomly chosen item from this production run is defective is:

**a.** 0.02
**b.** 0.04
**c.** 0.20
**d.** 0.83

**12.** (8 Points) A business currently has 5 telephone lines. Suppose the number of lines in use at any one time has the following distribution:

| Phone lines | Probability |
|---|---|
| 0 lines | 0.20 |
| 1 line | 0.30 |
| 2 lines | 0.25 |
| 3 lines | 0.10 |
| 4 lines | 0.08 |
| 5 lines | 0.07 |

What is the probability that **more than 2 lines** are in use?

**a.** 0.50
**b.** 0.25
**c.** 0.18
**d.** 0.15

**13.** (8 Points) Suppose the amount of time that a customer waits to be admitted to a popular restaurant is uniformly distributed. The random variable X measures the waiting time in hours. The probability density function is $f(x) = 0.5$ for $0 < x < 2$, $f(x) = 0$ otherwise.

What is the probability that a customer will have to wait more than 90 minutes (1.5 hours)?

**a.** 1/8
**b.** 1/4
**c.** 3/4
**d.** 7/8

**14.** (8 Points) You are given the following probability distribution where the expected value of $X$ is 6:

| $x$ | 5 | 6 | 7 |
|---|---|---|---|
| Prob($x$) | 0.1 | 0.8 | 0.1 |

What is the standard deviation of $X$?

**a.** 0.100
**b.** 0.200
**c.** 0.447
**d.** 0.553

**15.** (8 Points) A family has three children (G = girl, B = boy).

If we consider the gender of the children, what is the sample space for the gender of the first three children?

- **a.** {Peter, Mary, John}
- **b.** {B, G, B}
- **c.** {BBB, GGG, BBG, GGB, BGB, GBG}
- **d.** {GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB}

**16.** (8 Points) Suppose that we know that two events, *A* and *B*, are independent, and that $P(A) = 0.60$ and $P(B) = 0.25$.

What is the probability of the event "*A* and *B*"?

- **a.** 0.00
- **b.** 0.10
- **c.** 0.15
- **d.** 0.25

**17.** (8 Points) In a partnership, two members decide to make business decisions independently of each other, then compare their conclusions. If they agree, the decision is made to proceed. Partner A makes the right decision 70% of the time and partner B makes the right decision 65% of the time.

What is the probability that **at least one** of them makes the **right** decision?

- **a.** 10.5%
- **b.** 54.5%
- **c.** 45.5%
- **d.** 89.5%

**18.** (8 Points) Three of the following statements are true, while one statement is false.

Which one of the following statements is **false**?

- **a.** The standard deviation of a sampling distribution is an estimate of the general size of the sampling error.
- **b.** A statistic is an unbiased estimator of a parameter if the average of the statistic's sampling distribution equals the parameter.
- **c.** A sample mean is a biased estimator of a population mean.
- **d.** A symmetric sampling distribution occurs if the possible values of the statistic spread out in the same pattern on both sides of the distribution's center.

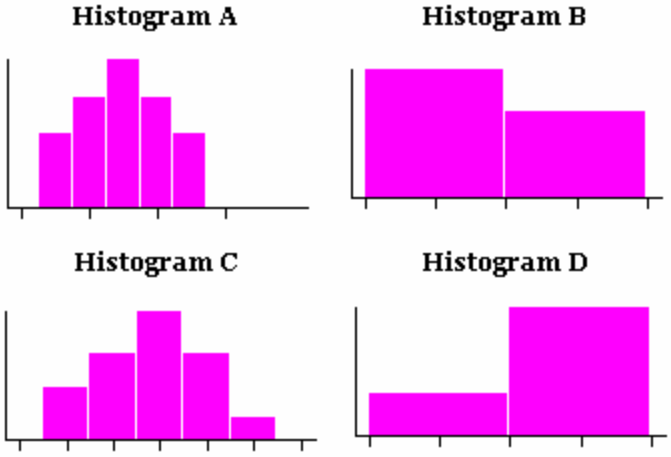
**19.** (8 Points) The average age of all American adults is 36.8 years. The average age of a sample of 25 American adults is 43.2 years.

Which value is a statistic?

- **a.** 36.8 years
- **b.** 43.2 years
- **c.** 25 American adults
- **d.** all of the above

**20.** (8 Points) In a population, 75% of the people are in favor of privatizing a portion of social security, and 25% are opposed.

If the people in favor have values "1", and the people opposed have value "0", then which of the histograms is correct for these values in this population?




- **a.** Histogram A
- **b.** Histogram B
- **c.** Histogram C
- **d.** Histogram D

**21.** (13 Points) The table below contains data on the income level and college graduation status of 79 people.

| | Low income | High income | Total |
|---|---|---|---|
| College grad | 13 | 34 | 47 |
| Not college grad | 26 | 6 | 32 |
| Total | 39 | 40 | 79 |

What is the odds ratio for a person having low income when the person has **not** graduated from college versus when the person has graduated from college?

**22.** (13 Points) Suppose that you suspect that poor performance in a first college level mathematics course is related to whether or not a student has taken a complete high school algebra sequence of courses. You have received the following data:

| | High School Preparation? | | |
|---|---|---|---|
| Grade | Algebra Background | No Algebra Background | Total |
| A, B, C | 43 | 31 | 74 |
| D, F | 2 | 9 | 11 |
| Total | 45 | 40 | 85 |

What is the relative risk of a student **not** receiving a grade of a "A, B, C" when the student has taken a complete high school algebra sequence of courses, compared to when the student has **not** taken a complete high school algebra sequence of courses?

**23.** (13 Points) The mean height of students in a statistics class is 70 inches and the standard deviation is 3 inches.

What is the standardized score for a height of 74 inches?

**24.** (13 Points) Suppose that the pulse rates of a population of students who exercise is normally distributed with a mean equal to 63.9 beats and a standard deviation of 8.

About 95% of the students who exercise have resting pulse rates between about (i) _____ and about (ii) _____.

**25.** (13 Points) The age at which babies, carried to full term, begin walking have a distribution that can be approximated by a normal curve with a mean equal to 10.38 months and a standard deviation equal to 1.40 months.

What proportion of this population begins walking between 9.4 months and 11.4 months?

**26.** (13 Points) A random sample of 25 people was taken from a particular population. The sample mean of the age of these individuals is 26.5 years.

If the population mean is 25.2 years, what is the sampling error?

**27.** (13 Points) The average height of the students in a large university is 68 inches, and the s.d. is 3 inches. Imagine choosing 25 students at random from the university, and finding the mean height of the sample of 25.

What is the standard deviation of the sample mean?

**28.** (13 Points) Assume that 40% of residents of San Francisco are covered by earthquake insurance. The people covered by earthquake insurance have value '1', and those not covered have value '0.' Thus the population mean is 0.40. The population standard deviation is 0.490.

Determine the probability that the sample mean of 400 observations will be between 0.37 and 0.43.

**29.** (13 Points) What is the probability that a chi-squared with 5 degrees of freedom is greater than 7?

**30.** (13 Points) The table below contains data on the income level and college graduation status of 79 people.

| | Low income | High income | Total |
|---|---|---|---|
| College grad | 13 | 34 | 47 |
| Not college grad | 26 | 6 | 32 |
| Total | 39 | 40 | 79 |

If these variables are independent, what is the expected cell count for the upper left cell?

**31.** (12 Points) There is a .10 probability that an individual susceptible to tuberculosis (TB), will contract the disease if he or she comes into contact with a carrier of TB. Suppose that 40 susceptible individuals independently come into contact with a TB carrier.

What is (i) the expected (mean) number who would contact the disease and what is (ii) the corresponding variance?

**32.** (12 Points) Here are summary statistics for the average daily temperatures for a number of U.S. cities in January and July, for 1996:

| | January | July |
|---|---|---|
| Minimum | 12 | 63 |
| Lower quartile | 25 | 70 |
| Median | 31 | 74 |
| Upper quartile | 42 | 78 |
| Maximum | 67 | 85 |

Indicate (i) the interquartile range for January and (ii) the interquartile range for July.

**33.** (12 Points) What is your chance of getting a cold or the flu? Annually, about 25% of the U.S. population gets a cold and between 35 to 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.

What is the probability a younger person (5-24 years of age) will get the flu?

**34.** (12 Points) Take the daily high temperatures for the first week in January, 1996, at Berkeley, but pretend that one was missing and recorded as "9999", so that the data are given as 67, 67, 57, 57, 57, 61, 9999.

Find the average of these seven numbers (that is, pretend you do not realize the 9999 is not a true temperature). Compare this to the correct average of 61. Is it very far off?

**35.** (12 Points) A university realized that about 15% of its students fail the introductory statistics class. Assume that the binomial settings hold.

In a class of 60 students, (i) how many students do we expect to fail? And (ii) what is the probability that at most 12 students from this class fail?