Statistics 1040, Section 003, Midterm 1 (200 Points)

Friday, October 3, 2008

Your Name: _____

Question 1: Controlled Experiments / Observational Studies (**35 Points**)

Does Aspartame Cause Cancer? Aspartame is an artificial sweetener found in thousands of products — sodas, chewing gum, dairy products and even many medicines. Some research has suggested that aspartame can cause lymphoma or leukemia in rats.
A recent study by the National Cancer Institute involved 340,045 men and 226,945 women, ages 50 to 69. From surveys they filled out in 1995 and 1996 detailing food and beverage consumption, researchers calculated how much aspartame they consumed. Over the next five years, 2,106 developed cancers such as lymphoma or leukemia. No association was found between aspartame consumption and occurrence of these cancers.

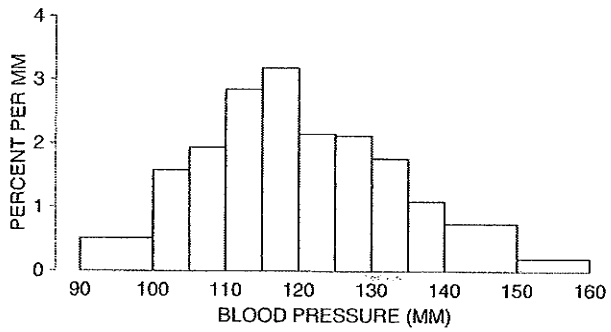
1. (**10 Points**) Was the study a controlled experiment or an observational study? Why? **Circle your answer and explain.**

2. (**15 Points**) Suggest a possible confounding factor for this study and explain *why* your confounding factor might make you doubt their results.

3. (**10 Points**) "It's very reassuring ... It's a large study with a lot of power," said Richard Adamson, a senior science consultant to the American Beverage Association, the leading industry group. Does the large sample size prove that aspartame does not cause cancers such as lymphoma or leukemia? **YES or NO — Circle your answer and explain.**

Question 2: Histograms **(35 Points)**

The figure below is a histogram showing the distribution of blood pressure for all 14,148 women in the Drug Study from Chapter 3, Section 5. Use the histogram to answer the following questions. **Circle your answer.**



1. **(5 Points)** Is the percentage of women with blood pressures above 130 mm around 25%, 50%, or 75%?

2. **(5 Points)** Is the percentage of women with blood pressures between 90 mm and 160 mm around 1%, 50%, or 99%?

3. **(5 Points)** In which interval are there more women: 135–140 mm or 140–150 mm?

4. **(5 Points)** Which interval is more crowded: 135–140 mm or 140–150 mm?

5. **(5 Points)** On the interval 125–130 mm, the height of the histogram is about 2.1% per mm. What percentage of the women had blood pressures in this class interval? About 10.5%, 12.6%, or 21%?

6. **(5 Points)** Which interval has more women: 97–98 mm or 102–103 mm?

7. **(5 Points)** Which is the most crowded millimeter of all? Somewhere between 90 and 100 mm, somewhere between 115 and 120 mm, or somewhere between 140 and 150 mm?

Question 3: Normal Curve (**48 Points**)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For around 210,000 female US citizens who took the General GRE Test in 2005–06, the mean for the quantitative ability portion of the exam was about 520 and the standard deviation was about 135 (`http://www.ets.org/Media/Tests/GRE/pdf/05-06_factors_final.%20pdf.pdf`). We can assume that the histogram follows a normal curve. **Show your work!**

1. (**15 Points**) The percentage of female US citizens who scored **more than 669** on the GRE test is roughly _____ %.

2. (**18 Points**) The percentage of female US citizens who scored **between 351 and 574** is about _____ %.

3. (**15 Points**) In order to be among the top 80% of all female US citizens, a student must have obtained a minimum GRE score of about _____.

Question 4: Regression **(50 Points)**

In one study, the correlation between the educational level of husbands and wives in a certain town was about 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.

Show your work.

1. **(15 Points)** Predict the educational level of a woman whose husband has completed 18 years of schooling.

   The answer is: _____ years

2. **(10 Points)** Calculate the corresponding r.m.s. error.

   The answer is: _____ years

3. **(15 Points)** Predict the educational level of a man whose wife has completed 15 years of schooling.

   The answer is: _____ years

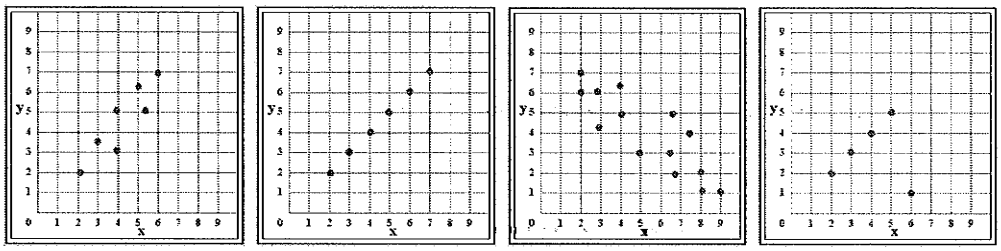
4. **(10 Points)** Apparently, well–educated men marry women who are less well–educated than themselves. But the women marry men with even less education. How is this possible?

Question 5: Guessing the Correlation Coefficient (**32 Points**)

Match the four scatterplots with their correlations from the list:

$-1.03, \quad -1.00, \quad -0.90, \quad -0.85, \quad -0.30, \quad 0.00, \quad 0.30, \quad 0.85, \quad 0.90, \quad 1.00, \quad 1.03$

A B C D



Correlation for Plot A: r = _____

Correlation for Plot B: r = _____

Correlation for Plot C: r = _____

Correlation for Plot D: r = _____

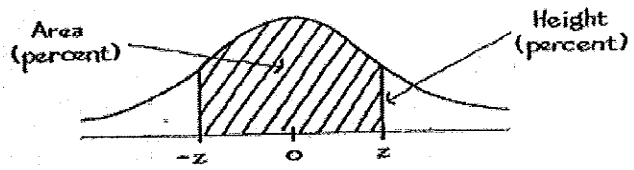
Formulas:

$$\text{avg} = \frac{\text{sum of all numbers}}{\text{how many numbers}}$$

$$\text{SD} = \sqrt{\text{average of [(deviations from avg)}^2\text{]}}$$

$$\text{r.m.s. error} = \sqrt{1 - r^2} \times \text{SD}_y$$

Tables



A NORMAL TABLE

z	*Area*	*z*	*Area*	*z*	*Area*
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

Statistics 1040, Section 003, Midterm 2 (200 Points)

Friday, November 7, 2008

Your Name: _____

Instructions: **Carefully check whether you have to provide an explanation or not. In case you have to provide an explanation, keep it short. Just 1 sentence (or 2 sentences at most) or a short calculation will be fine. If you do not have to provide an explanation, do not waste your time giving an unneeded explanation.**

Question 1: Probability Histograms (**40 Points**)

Four hundred draws will be made at random with replacement from the box
$\boxed{1}\ \boxed{3}\ \boxed{5}\ \boxed{7}\ \boxed{9}$. **Show your work!**

1. (**20 Points**) Estimate the chance that the sum of the draws will be more than 1,500.
   The chance is: _____ %

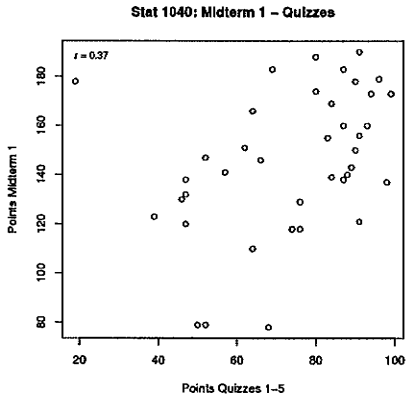
2. (**20 Points**) Estimate the chance that there will be fewer than 90 $\boxed{3}$'s.
   The chance is: _____ %

Question 2: Regression (**50 Points**)

In a recent section of Stat 1040, the following scores for the sum of the first five quizzes and the first midterm were observed:

Quiz 1-5 score:	avg = 73 points;	SD = 19 points;	
Midterm 1 score:	avg = 145 points;	SD = 29 points;	$r = 0.37$.

The scatterplot that shows the data is displayed below and can be assumed to be football–shaped.



Show your work!

1. (**15 Points**) Find the regression equation for predicting the Midterm 1 score from the Quiz 1-5 score.

2. (**8 Points**) Using your regression equation, estimate the Midterm 1 score for a student who had a Quiz 1-5 score of 60 points.

3. (**7 Points**) Find the r.m.s. error for predicting the Midterm 1 score from the Quiz 1-5 score.

4. (**10 Points**) Can we use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points? **YES** or **NO**? **Circle your answer and provide a short explanation.**

5. (**10 Points**) Independently from your previous answer, let us assume that we can use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points. Would you be *surprised* that a student with 19 points in the quizzes got a score of 178 points in Midterm 1? **YES** or **NO**? **Circle your answer and provide a short explanation.**

Question 3: Probability and Chance (**50 Points**)

A class of 28 fourth–graders has 16 boys and 12 girls. This class goes on a field trip. Two children are chosen at random to ride with the teacher.
Answer each of the following questions separately. Show your work!

1. (**10 Points**) What is the chance the first child is a boy?
   The chance is _____%

2. (**10 Points**) What is the chance the second child is a boy?
   The chance is _____%

3. (**10 Points**) What is the chance both children are boys?
   The chance is _____%

4. (**10 Points**) What is the chance neither of the children are boys?
   The chance is _____%

5. (**10 Points**) What is the chance one of the children is a boy and the other is a girl?
   The chance is _____%

Question 4: The Law of Averages **(20 Points)**

Answer the following two questions. **Just circle your correct answer!**

1. **(10 Points)** According to genetic theory, there is very close to an even chance that both children in a two–child family will be of the same sex. Here are two possibilities:

   (a) 15 couples have two children each. In 10 or more of these families, it will turn out that both children are of the same sex.

   (b) 30 couples have two children each. In 20 or more of these families, it will turn out that both children are of the same sex.

   Choose the correct statement among the following statements:

   (i) Possibility (a) is more likely.

   (ii) Possibility (b) is more likely.

   (iii) Possibilities (a) and (b) are equally likely.

2. **(10 Points)** A box contains red and blue marbles; there are more red marbles than blue ones. Marbles are drawn one at a time from the box, at random with replacement. You win a dollar if a red marble is drawn more often than a blue one. There are two choices:

   (a) 100 draws are made from the box.

   (b) 200 draws are made from the box.

   Choose one of the four options below:

   (i) Choice (a) gives a better chance of winning.

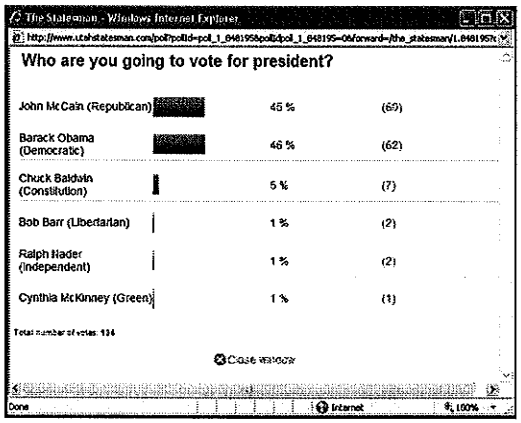
   (ii) Choice (b) gives a better chance of winning.

   (iii) Choices (a) and (b) give the same chance of winning.

   (iv) Can't tell without more information.

Question 5: Chance Errors in Sampling / Accuracy of Percentages (**40 Points**)

The online edition of the Utah Statesman reported the following intermediate standings (on 10/31/08 at 10.15pm) regarding their question "Who are you going to vote for president?":



1. (**20 Points**) For this question part only, assume that the 134 students who replied to this survey represent a simple random sample (SRS) of all USU students. Using the idea of bootstrap, one could estimate that the percentage of Obama voters (a total of 62 students in this sample) in the whole USU student population is about _____ %, give–or–take _____ %. **Show your work!**

2. (**5 Points**) Unfortunately, this survey is not based on a SRS and, therefore, it does not provide a statistically reliable answer (note that many surveys of this type indicate that the survey is not a scientific survey). This particular survey is one example of a _____ survey.

3. (**15 Points**) Provide **three** reasons why a Web survey such as the Utah Statesman survey cited above is liklely to provide invalid results that do **not** relate to the entire population (here: all USU students).

Formulas:

$$\text{r.m.s. error} = \sqrt{1 - r^2} \times \text{SD}_y$$

$$\text{slope} = r \times \frac{\text{SD}_y}{\text{SD}_x} \qquad \text{intercept} = \text{avg}_y - \text{slope} \times \text{avg}_x$$

$$\text{box average} = \frac{\text{sum of all numbers in box}}{\text{how many numbers in box}}$$

$$\text{box SD} = \sqrt{\text{average of [(deviations from box average)}^2]}$$

$$\text{EV}_{sum} = \text{number of draws} \times \text{box average}$$

$$\text{SE}_{sum} = \sqrt{\text{number of draws}} \times \text{box SD}$$

Shortcut formulas for a box that contains only *two* different numbers:

$$\text{average} = \frac{(\text{smaller} \times \text{how many}) + (\text{bigger} \times \text{how many})}{\text{how many tickets in the box}}$$

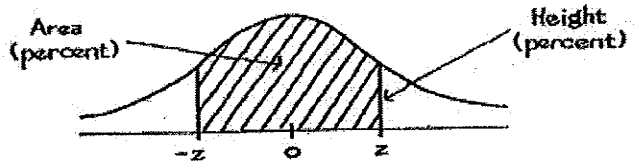
$$\text{SD} = (\text{bigger - smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

Shortcut formulas for a box that contains only $\boxed{0}$'s and $\boxed{1}$'s:

$$\text{average} = \frac{\text{number of } \boxed{1} \text{ 's}}{\text{how many tickets in the box}} \qquad \text{SD} = \sqrt{\frac{\text{fraction}}{\text{of } \boxed{1} \text{ 's}} \times \frac{\text{fraction}}{\text{of } \boxed{0} \text{ 's}}}$$

$$\text{EV}_{\%} = \%\text{ of } \boxed{1} \text{ 's in the box} \qquad \text{SE}_{\%} = \frac{\text{SE}_{\text{sum}}}{\text{number of draws}} \times 100\%$$

Tables



A NORMAL TABLE

z	*Area*	*z*	*Area*	*z*	*Area*
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991