Statistics 1040, Section 008, Midterm 1 (200 Points)

Friday, February 15, 2008

Your Name: _____

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

In the December 10 issue of NEWSWEEK, medical writer Jerry Adler says:

"It's not too soon to start thinking about New Year's resolutions, and here's mine, as a medical writer: I will not report on any amazing new treatments for anything, unless they were tested in large, randomized, placebo–controlled, double–blind clinical trials published in high–quality peer–reviewed medical journals. If that means not telling NEWSWEEK's readers about, say, a new magnetized–water cure for osteoporosis, cancer and autism — well, there are infomercials to fill that gap."

1. (**10 Points**) Explain what it means for a study to be double–blind.

2. (**15 Points**) Give 3 different reasons *why* a medical study should be double–blind.

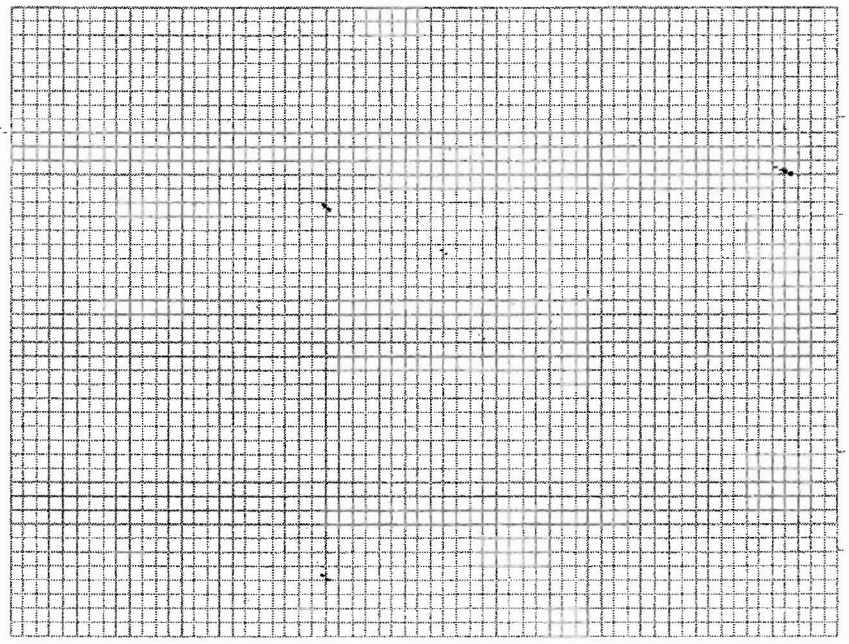
3. (**10 Points**) What is a placebo? Why is it used?

Question 2: Histograms (**45 Points**)

The age distribution of people in the U.S. in 2004 is shown below.

Age	Percent of population
0-5	7
5-15	14
15-20	7
20-25	7
25-30	7
30-35	7
35-45	15
45-55	14
55-65	10
65-75	6
75 and over	6

1. (**20 Points**) Draw a histogram for these data on the graph paper provided. (The class intervals include the left endpoint, not the right; for instance, on the second line of the table, 14% of the people were age 5 years or more but had not yet turned 15. The interval "75 and over" can be ended at 85. Men and women are combined in the data.) Make sure to label the axes.



Use your histogram to answer the following questions on the next page:

2. **(5 Points)** Are there more children age 1, or elders age 71?
   Circle your answer.

3. **(5 Points)** Are there more 21–year–olds, or 61–year–olds?
   Circle your answer.

4. **(5 Points)** Are there more people age 0–4, or 55–59?
   Circle your answer.

5. **(5 Points)** The percentage of people age 35 and over is around 25%, 50%, or 75%?
   Circle your answer.

6. **(5 Points)** To be at the 35^{th} percentile of the age distribution, one has to be about 15 years old, 20 years old, or 25 years old?
   Circle your answer.

Question 3: Normal Curve (**50 Points**)

A grocery store carries a variety of "on the vine" tomatoes with an average weight of 5.0 ounces and an SD of 0.9 ounces. The weights of these tomatoes follow the normal curve. **Show your work!**

1. (**15 Points**) What percentage of them would weigh more than 6.0 ounces?
   The answer is: _____ %

2. (**20 Points**) And what percentage would weigh between 3.7 ounces and 4.7 ounces?
   The answer is: _____ %

3. (**15 Points**) Estimate the 25^{th} percentile of their weights.
   The answer is: _____ ounces

Question 4: Average / SD (**40 Points**)

Part I:

Here is a list of numbers:

0.7 1.6 9.8 3.2 5.4 0.8 7.7 6.3 2.2 4.1
8.1 6.5 3.7 0.6 6.9 9.9 8.8 3.1 5.7 9.1

1. (**10 Points**) Without doing any arithmetic, guess whether the average is around (i) 1, (ii) 5, or (iii) 10. **Circle your answer and explain.**

2. (**10 Points**) Without doing any arithmetic, guess whether the SD is around (i) 1, (ii) 3, or (iii) 6. **Circle your answer and explain.**

Part II:

A study on college students found that the men had an average weight of about 66 kg and an SD of about 9 kg. The women had an average weight of about 55kg and an SD of about 9 kg (Note that 1 kg = 2.2 lb).

1. (**10 Points**) Just roughly, what percentage of the men weighted between 57 kg and 75 kg?
   Answer: _____%
   Fill in your answer and explain.

2. (**10 Points**) If you took the men and women together, would the SD of their weights be (i) smaller than 9 kg, (ii) just about 9 kg, or (iii) bigger than 9 kg?
   Circle your answer and explain.

Question 5: Correlation (**30 Points**)

Investigators take a sample of DINKS (dual–income families, where husband and wife both work and have no kids). The investigators have data on the husband's income and the wife's income. By definition,

$$\text{family income} = \text{husband's income} + \text{wife's income.}$$

The average family income was around \$50,000, and 10% of the couples had family income in the range \$45,000–\$55,000. Fill in the blanks, using the options given below, and **explain briefly**:

1. (**15 Points**) The correlation between wife's income and family income is _____.

2. (**15 Points**) Among couples whose family income is in the range \$45,000–\$55,000, the correlation between wife's income and husband's income is _____.

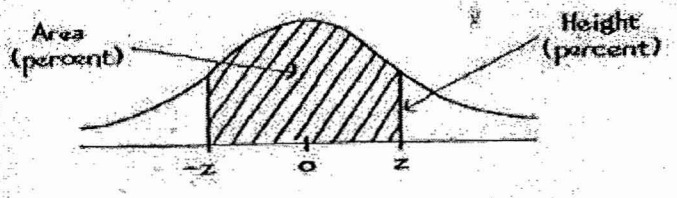
Options (for 1. and 2.): (a) -1 (b) nearly -1 (c) somewhat negative (d) 0 (e) somewhat positive (f) nearly 1 (g) 1 (h) -1.1 (i) 1.1

Formulas:

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

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

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 008, Midterm 2 (200 Points)

Friday, March 28, 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.**

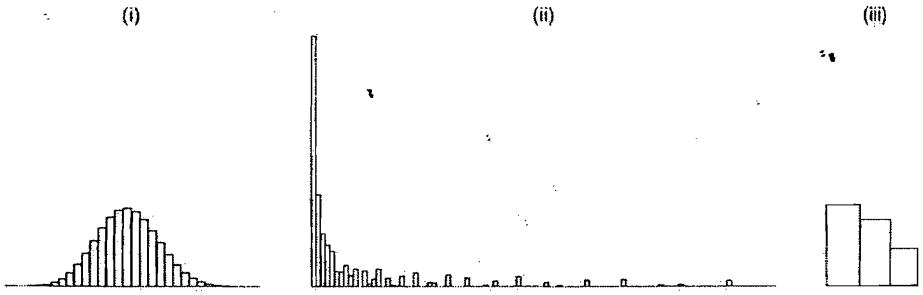
Question 1: Normal Approximation for Probability Histograms (**30 Points**)

Twenty–five draws are made at random with replacement from the box

1	1	2	2	3

.

One of the graphs below is an (empirical) histogram for the numbers drawn. One is the probability histogram for the sum. And one is the probability histogram for the product. Which is which? **Explain!**



- An (empirical) histogram for the numbers drawn is _____.
  Explanation:

- The probability histogram for the sum is _____.
  Explanation:

- The probability histogram for the product is _____.
  Explanation:

Question 2: Probability and Chance (**40 Points**)

For a road trip, a student places the following nine CDs into the glove compartment of his car:

- 5 *modern rock* CDs (Fallout Boy, Hawthorne Heights, The Used, Finger Eleven, Taking Back Sunday),
- 3 *pop* CDs (P!nk, Fergie, Gwen Stefani),
- 1 *American Idol* CD (Jordin Sparks).

On his trip, the student blindly grabs a CD from the glove compartment, listens to it, and places it on the back seat when finished. Then he blindly grabs a second CD from the glove compartment. You should NOT comment on the musical taste of this student, but **answer each of the following questions separately. Show your work!**

1. (**8 Points**) What is the chance that the FIRST CD will be a *pop* CD or the *American Idol* CD? The chance is _____%

2. (**8 Points**) What is the chance that the SECOND CD will be a *pop* CD or the *American Idol* CD? The chance is _____%

3. (**8 Points**) What is the chance that he will listen to Jordin Sparks as one of his two selections? The chance is _____%

4. (**8 Points**) What is the chance that he will listen to **none** of the *pop* CDs? The chance is _____%

5. (**8 Points**) What is the chance that he will listen to **at least one** of the *modern rock* CDs? The chance is _____%

Question 3: EV, SE, and Normal Curve (**50 Points**)

During the 2004 presidential elections, Kerry needed to win the state of Ohio to become the next president. Early on Nov 3, the day after Election Day, Bush had a 51% to 49% lead over Kerry, which related to about 140,000 more votes for Bush in Ohio. However, there were possibly up to 250,000 uncounted provisional ballots at that time. If Kerry could have gotten 140,000 of those, plus 1/2 of the remaining 110,000, plus 1, i.e., a total of 195,001, he would have won Ohio and would have been the next president. However, Kerry eventually conceded to Bush later on Nov 3 (even with many of the provisional ballots still being uncounted) because Kerry's advisors figured out that it was *statistically impossible* for Kerry to win Ohio and thus the election. **Show your work!**

1. (**10 Points**) Assume you are a highly optimistic advisor of Kerry, assuming that he might win up to 70% of the uncounted provisional ballots because a huge majority of these votes come from a population group close to the Democrats. Find the box model.

2. (**15 Points**) The expected number of votes for Kerry from the uncounted provisional ballots is _____ with an SE of _____.

3. (**20 Points**) The chance that **at least** 195,001 of the uncounted provisional ballots are in favor of Kerry is about _____ %.

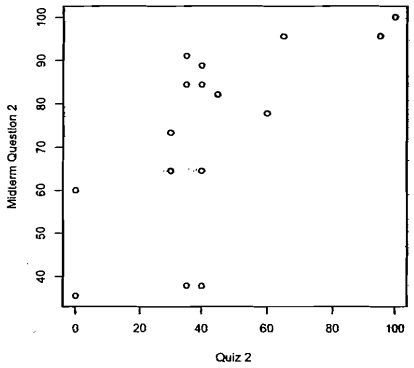
4. (**5 Points**) So, do you agree that it was *statistically impossible* for Kerry to win Ohio and thus the election? **Yes / No**

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

In a particular section of Stat 1040, students had to answer Review Exercise 2 from Chapter 3 of their textbook in Quiz 2. The result was anything but satisfactory, with the median score being an F. Detailed solutions were handed out, together with the graded quizzes. To determine whether students studied the solutions, the instructor basically reused the same question (with some part added) a few weeks later as Question 2 in Midterm 1. For a better comparability, the scores below were adjusted to 100 points.

Midterm Question 2 score:	avg = 73 points;	SD = 21 points;	
Quiz 2 score:	avg = 43 points;	SD = 27 points;	$r = 0.65$.

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 Question 2 score from the Quiz 2 score.

2. (**8 Points**) Using your regression equation, estimate the Midterm Question 2 score for a student who had a Quiz 2 score of 80 points.

3. **(7 Points)** Find the r.m.s. error for predicting the Midterm Question 2 score from the Quiz 2 score.

4. **(10 Points)** Would you be *surprised* if a student with 85 points in Quiz 2 would have obtained a Midterm Question 2 score of 40 points? **YES** or **NO**? **Circle your answer and provide a short explanation.**

5. **(10 Points)** As mentioned above, all scores were adjusted as if graded out of 100 points. However, the Quiz 2 scores were originally graded out of 20 points, that means, each individual Quiz 2 score was multiplied by 5 for this question. Therefore, we had an original average score of _____ points and an original SD of _____ points when grading out of 20 points.

Question 5: Chance Errors in Sampling **(30 Points)**

Five hundred draws are made at random from the box

$$\boxed{60{,}000 \times \boxed{0} \qquad 20{,}000 \times \boxed{1}}.$$

True or false? Circle your answers. No explanation is needed.

1. **(5 Points)** True / false: The expected value for the percentage of 1's among the draws is exactly 25%.

2. **(5 Points)** True / false: The expected value for the percentage of 1's among the draws is around 25%, give or take 2% or so.

3. **(5 Points)** True / false: The percentage of 1's among the draws will be around 25%, give or take 2% or so.

4. **(5 Points)** True / false: The percentage of 1's among the draws will be exactly 25%.

5. **(5 Points)** True / false: The percentage of 1's in the box is exactly 25%.

6. **(5 Points)** True / false: The percentage of 1's in the box is around 25%, give or take 2% or so.

Formulas:

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

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

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

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

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

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

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

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

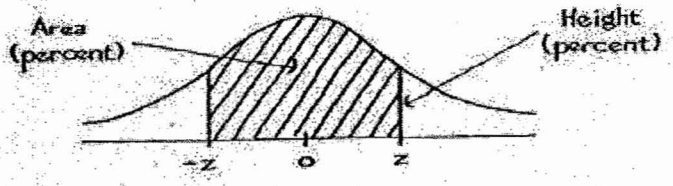
$$\text{SD} = (\text{bigger - smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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