Statistics 1040, Section 006, Midterm 1 (200 Points)

Friday, October 8, 2004

Your Name: _____

Question 1: Controlled Experiment/Observational Study (**60 Points**)

A recent study in Europe looked at a large group of women of childbearing age. The researchers asked each woman how much alcohol they had consumed over the past 12 months. The researchers found that women who drank moderate amount of alcohol were somewhat less likely to have infertillity problems than women who did not drink alcohol at all (November, 2001). The study said it "controlled for age, income, and religion".

1. (**15 Points**) Based on the information above, was this a controlled experiment or an observational study? **Explain briefly.**

2. (**15 Points**) Why did they "control for" age, income, and religion?

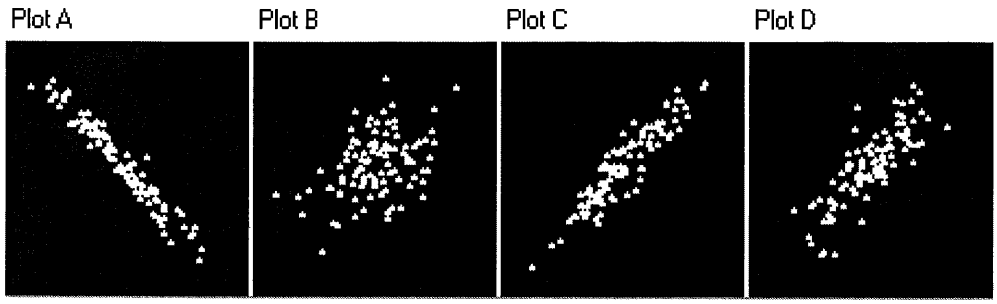
3. (**15 Points**) Is this convincing evidence that infertility would decrease if women with infertility problems started to drink moderate amounts of alcohol? (Note: we are only asking about infertility. There may be other problems introduced by such behavior, but ignore them for answering this question).

4. (**15 Points**) Suggest a possible confounding factor (other than age, income, or religion) and **clearly explain** why you think it might be a confounding factor.

Question 2: Correlation **(30 Points)**

1. **(12 Points)** Match each of the scatterplots with **one** of the correlation coefficients below:

$$-1.00, \quad -0.97, \quad 0.46, \quad 0.82, \quad 0.93, \quad 1.05$$



- Correlation for Plot A: $r =$ _____
- Correlation for Plot B: $r =$ _____
- Correlation for Plot C: $r =$ _____
- Correlation for Plot D: $r =$ _____

There is **no** explanation required for this question.

2. **(18 Points)** Two different investigators are working on a growth study. The first measures the heights of 100 children in inches. The second prefers the metric system, and changes the results to centimeters (multiplying by the conversion factor 2.54 centimeters per inch). A scatter diagram is plotted, showing for each child its height in inches on the horizontal axis, and height in centimeters on the vertical axis.

   (a) **(6 Points)** If no mistakes are made in the conversion, what is the correlation?

   below -1.0, -1.0, close to -1.0, 0.0, close to 1.0, 1.0, above 1.0

   (b) **(6 Points)** What happens to r if mistakes are made in the arithmetic?

   r goes up, r goes down, r stays the same

   (c) **(6 Points)** What happens to r if the second investigator goes out and measures the same children again, using metric equipment?

   r goes up, r goes down, r stays the same

Circle your answers — There is **no** explanation required for this question.

Question 3: Regression **(30 Points)**

Pearson and Lee obtained the following results in a study of about 1,000 families:

average height of husband $\approx$ 68 inches, SD $\approx$ 2.7 inches
average height of wife $\approx$ 63 inches, SD $\approx$ 2.5 inches,
r $\approx$ 0.25

1. **(10 Points)** Predict the height of a wife when the height of her husband is 72 inches. Answer: _____

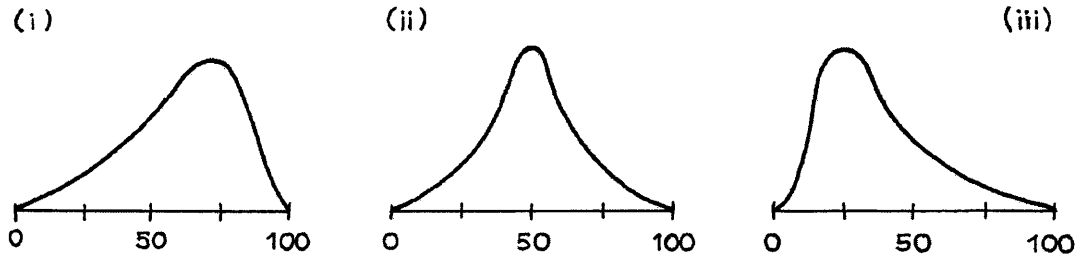
2. **(10 Points)** Predict the height of a wife when the height of her husband is unknown. Answer: _____

3. **(10 Points)** Predict the height of a husband when the height of his wife is 65 inches. Answer: _____

Show your work!

Question 4: Average and Standard Deviation **(40 Points)**

Below are sketches of histograms for three lists.



1. **(10 Points)** In scrambled order, the averages are 40, 50, 60. Match the histogram with the averages:

   Histogram (i): average = _____

   Histogram (ii): average = _____

   Histogram (iii): average = _____

   Just fill in the correct value.

2. **(10 Points)** Match the histogram with the description (circle your answer):

   - The median is less than the average. Histogram (i), (ii), or (iii).
   - The median is about equal to the average. Histogram (i), (ii), or (iii).
   - The median is bigger than the average. Histogram (i), (ii), or (iii).

3. **(10 Points)** Is the SD of histogram (iii) around 5, 15, or 50? Circle your answer.

4. **(10 Points)** The SD for histogram (i) is a lot smaller than that for histogram (iii). **True** or **false**? **Circle your answer and explain**:

Question 5: Normal Distribution **(40 Points)**

The stretched hand spans of women have a distribution that is approximated by a normal curve with a mean of 20 centimeters and a standard deviation equal to 1.5 centimeters.

1. **(15 Points)** For what proportion of women is the stretched hand span greater than 23 centimeters?
   Answer: _____

2. **(10 Points)** Would it be unusual for a woman to have a stretched hand span that's less than 14 centimeters? Support your answer with a calculation.

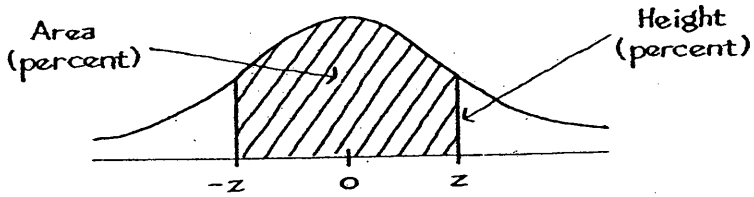
   Answer: **yes, unusual / No, not unusual**

   Calculation/Explanation:

3. **(15 Points)** What stretched hand span is the 25th percentile of the stretched hand spans of women?
   Answer: _____

Show your work!

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 006, Midterm 2 (200 Points)

Friday, November 12, 2004

Your Name: _____

Question 1: Chance/Probability (**50 Points**)

A big box with sports equipment contains 20 balls of which 6 are soccer balls, 10 are basketballs, and 4 are volleyballs. A little boy, barely able to reach into the box, randomly picks 2 balls to play with. **Show your work!**

1. (**10 Points**) What is the chance that the first ball is a volleyball?

   The chance is: _____ %

2. (**10 Points**) What is the chance that at least one of the balls is a soccer ball?

   The chance is: _____ %

3. (**10 Points**) What is the chance that both balls are for the same game?

   The chance is: _____ %

4. (**10 Points**) What is the chance that one ball is a volleyball and the other ball is a basketball?

   The chance is: _____ %

5. (**10 Points**) What is the chance that both balls are volleyballs?

   The chance is: _____ %

Question 2: 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 3: Sampling (**30 Points**)

In Web polls, anyone who views a certain Web page is allowed to vote by clicking on a button that represents their choice. In fact, there is nothing to stop someone voting as many times as they want. The results of one such poll suggest that almost 90% of the US population wants to ban firearm sales. The poll has a very large sample size (over 1 million).

1. (**21 Points**) Web based polls such as this are notoriously susceptible to bias. Give **three** possible sources of bias for this poll.

2. (**9 Points**) Are the sources of bias you listed in Part 1. above a problem even with a very large sample, or does the sample size imply that they can be ignored? **Explain!**

Question 4: Regression (**40 Points**)

For a random sample of 20 car models, the average weight (in pounds) was 3236, with an SD of 523. The average gas mileage (in miles per gallon) was 21.4 with an SD of 4.2. The correlation between weight and gas mileage was –0.87. The scatter diagram was football shaped. **Show your work!**

1. (**10 Points**) Find the equation of the regression line for predicting gas mileage from weight.

2. (**10 Points**) Predict the gas mileage of a car that weighs 3500 pounds.

3. (**10 Points**) Find the r.m.s. error for prdicting gas mileage from weight of cars.

4. (**10 Points**) Would you be surprised if someone told you that one of these cars weighing 3500 pounds got 27 miles per gallon? Explain your answer using the r.m.s. error.

Question 5: Normal Approximation for Probability Histograms (**30 Points**)

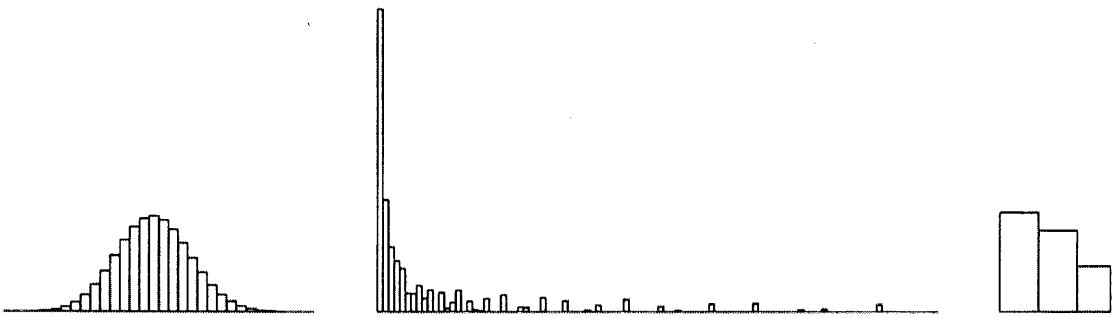
Twenty–five draws are made at random with replacement from the box

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!**

(i) (ii) (iii)



- 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:

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:

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

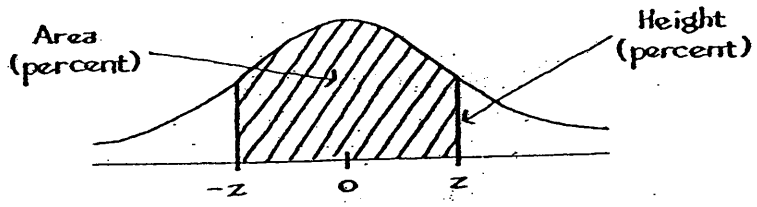
$$\text{SD} = (\text{bigger - smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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

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

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

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