

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

Friday, October 7, 2005

Your Name: _____

Question 1: Controlled Experiment/Observational Study (25 Points)

The New Zealand Herald reported on September 26, 2005:

“Left-handed women are more than twice as likely as right-handers to suffer from breast cancer before reaching menopause, Dutch scientists said on Monday. [...] Cuno Uiterwaal, an assistant professor [...] and his colleagues studied 12,000 healthy, middle-aged women born between 1932–1941 who were part of a breast screening programme. The scientists determined their hand preference and followed up their medical history to see which women developed breast cancer. [...] Left-handed women are more than twice as likely to develop pre-menopausal breast cancer as non-left handed women. [...] Other risk factors such as family history of breast cancer, number of pregnancies, smoking habits, and social and economic status were considered. [...]”

1. (5 Points) Is this an example of a **controlled experiment** or an **observational study**? **Circle your answer and explain.**
2. (10 Points) Why does the article emphasize “Other risk factors such as family history of breast cancer, number of pregnancies, smoking habits, and social and economic status were considered”? **Explain clearly.**
3. (10 Points) Does this result necessarily imply that if all left-handed women immediately switch over to right-handedness, they will reduce, on average, their risk to obtain breast cancer to the same rate as it is reported for right-handed women? **Explain clearly.**

Question 2: Normal Distribution (45 Points)

According to the U.S. Department of Agriculture, one 3-ounce serving of trimmed sirloin beef contains, on average, 7.4 grams of fat. Assume that the amount of fat for such servings closely follows the normal curve, with a standard deviation of 0.4 gram.

Fill the blanks in the statements below and **show all the work needed to obtain the answer:**

1. (15 Points) The proportion of servings that contain between 6.9 and 7.1 grams of fat is _____.

2. (15 Points) The proportion of servings that contain more than 8.3 grams of fat is _____.

3. (15 Points) Only 1% of the servings will contain less than _____ grams of fat.

Question 3: Change of Scale (40 Points)

From the subjects in a health survey, the following data were collected:

average height = 68 inches SD = 2.5 inches

average blood pressure = 120 mm SD = 15 mm

correlation $r = -0.2$

You want to provide a summary of these results to a friend in Europe and report heights in centimeters (instead of inches) to make it easier for your friend to interpret the results. Recall that 1 inch = 2.54 centimeters.

1. (10 Points) The average height (in centimeters) is: _____

2. (10 Points) The SD [for the height] (in centimeters) is: _____

3. (10 Points) The correlation r is: _____

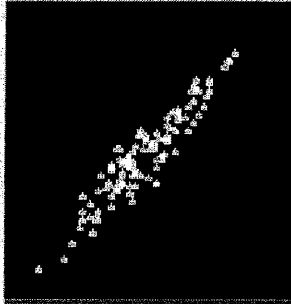
4. (10 Points) Circle the correct answer: r is measured in
 - centimeters
 - inches
 - mm
 - inches · mm
 - centimeters · mm
 - inches · centimeters
 - none of these – r is a unitless number

Question 4: Guessing the Correlation Coefficient (30 Points)

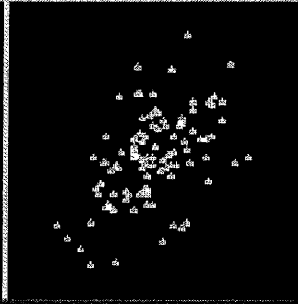
Match the four scatterplots with their correlations from the list:

-1.03, -0.99, -0.89, -0.50, -0.05, 0.50, 0.93, 1.03

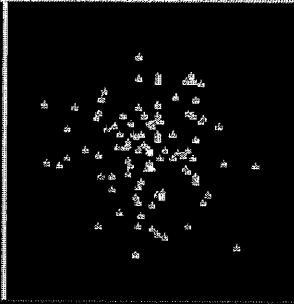
Plot A



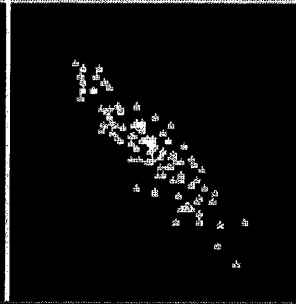
Plot B



Plot C



Plot D



Correlation for Plot A: $r =$ _____

Correlation for Plot B: $r =$ _____

Correlation for Plot C: $r =$ _____

Correlation for Plot D: $r =$ _____

Question 5: Regression (60 Points)

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

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

Show all the work needed to obtain the answer.

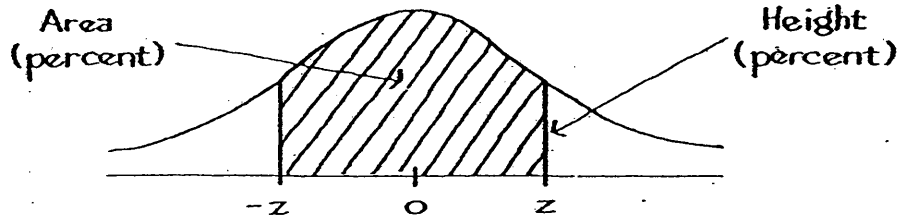
1. **(15 Points)** Predict the height of a wife when the height of her husband is 72 inches.

2. **(15 Points)** Predict the height of a wife when the height of her husband is 68 inches.

3. **(15 Points)** Predict the height of a wife when the height of her husband is unknown.

4. **(15 Points)** Predict the height of a husband when the height of his wife is 68 inches.

A NORMAL TABLE



<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>
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 11, 2005

Your Name: _____

Question 1: Regression (50 Points)

From the subjects (all men) in a health survey, the following data were collected:

Average height = 68 inches SD = 2.5 inches

Average blood pressure = 120 mm SD = 15 mm

Correlation = -0.2.

The scatter diagram is football-shaped.

Show your work!

1. (15 Points) Find the regression equation for predicting blood pressure from height.
2. (10 Points) Using your regression equation, estimate the blood pressure of a man who is 73 inches tall.
3. (15 Points) Find the r.m.s. error for predicting the blood pressure from the height.
4. (10 Points) The correlation coefficient tells us that, on average, taller men have **higher / lower** blood pressures than shorter men, and that the relationship between blood pressure and height is quite **strong / weak**. (Circle the correct word in each pair of choices.)

Question 2: Chance Errors in Sampling (40 Points)

A group of 50,000 tax forms has an average gross income of \$37,000, with an SD of \$20,000. Furthermore, 20% of the forms have a gross income over \$50,000. A group of 900 forms is chosen at random for audit. To estimate the chance that between 19% and 21% of the forms chosen for audit have gross incomes over \$50,000, a box model is needed.

1. (5 Points) Should the number of tickets in the box be **900** or **50,000**?

Circle your answer.

2. (5 Points) Each ticket in the box shows
a zero or a one or **a gross income**

Circle your answer.

3. (5 Points) **True or false:** the SD of the box is \$20,000.

Circle your answer.

4. (5 Points) **True or false:** the number of draws is 900.

Circle your answer.

5. (12 Points) Find the chance (approximately) that between 19% and 21% of the forms chosen for audit have gross incomes over \$50,000. **Show your work!**

The chance is: _____ %

6. (8 Points) With the information given, can you find the chance (approximately) that between 9% and 11% of the forms chosen for audit have gross incomes over \$75,000? **Either find the chance, or explain why you need more information.**

The chance is: _____ %

Question 3: Chance/Probability (50 Points)

Six children attend a party. There are 7 party favors: 3 pink favors and 4 blue favors. The children are each given a party favor at random (which they keep — so this is without replacement).

Answer each of the parts independently from the other parts. **Show your work!**

1. (10 Points) What is the chance that the first child gets a pink favor?

The chance is: _____ %

2. (10 Points) What is the chance that the second child gets a pink favor?

The chance is: _____ %

3. (10 Points) If I see that the first 2 children received pink favors, what is the chance that the third child also gets a pink favor?

The chance is: _____ %

4. (10 Points) What is the chance that the left-over favor is blue?

The chance is: _____ %

5. (10 Points) What is the chance that the first two children get favors of different colors?

The chance is: _____ %

Question 4: Normal Approximation for Probability Histograms (30 Points)

One hundred draws are made at random with replacement from a box with ninety-nine tickets marked "0" and one ticket marked "1". True or false? Circle your answer and explain (with box model calculations where necessary).

1. (15 Points) The sum will be around 1, give or take 1 or so.

True / False

Explanation:

2. (15 Points) There is about a 68% chance that the sum will be in the range 0 to 2.

True / False

Explanation:

Question 5: Sampling (30 Points)

Part 1: (20 Points)

For each of the following, decide whether this describes a simple random sample (SRS).
Just circle your answer.

- (5 Points) A student newspaper asked readers to respond to the question "Do you think that there should be more student activities on the weekends?" An overwhelming 95 percent said "yes". The article reporting the results concluded that 95 percent of all students feels this way.

This is a SRS: **yes** / **no**

- (5 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 a SRS: **yes** / **no**

- (5 Points) Fifteen state parks are to be selected from 1000 state parks in such a way that each has an equal chance of being selected. A random number generator on a computer is used to select 15 integers between 1 and 1000. Based upon those integers, the state parks are selected from a numbered list.

This is a SRS: **yes** / **no**

- (5 Points) A researcher chooses a random sample of households, then interviews every member of the selected households.

This is a SRS: **yes** / **no**

Part 2: (10 Points)

A population consists of 100 individuals who have been numbered from 1 to 100 for the purpose of taking a simple random sample of ten individuals. Which of the following sets of ten is most likely to be chosen as the sample?

Just circle the correct answer:

- (a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- (b) 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
- (c) 3, 17, 24, 39, 41, 47, 66, 73, 87, 96
- (d) These are all equally likely.

Formulas:

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

$$\text{slope} = r \times \frac{SD_y}{SD_x} \quad \text{intercept} = \text{avg}_y - \text{slope} \times \text{avg}_x$$

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

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

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

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

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

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

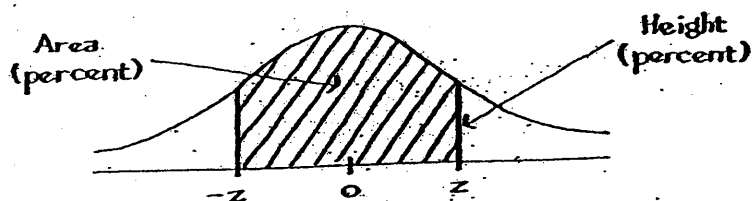
$$\text{SD} = (\text{bigger} - \text{smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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

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

$$EV_{\%} = \% \text{ of } \boxed{1} \text{'s in the box} \quad SE_{\%} = \frac{SE_{sum}}{\text{number of draws}} \times 100\%$$

Tables



A NORMAL TABLE

<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>
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