Stat 1040:

Midterm 1 & 2, Fall 2002

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

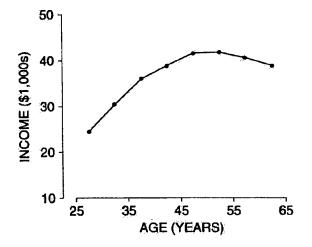
October 4, 2002

Your	Name:	

Question 1: Representative Sample (25 Points)

The graph below summarizes the results from a study based on a representative sample of men age 25–64 in 1993, who were working full time that year; the graph shows average income for each age group.

True or false, and explain: the data show that on average, if a man keeps working, his income will increase until age 50 or so, then start decreasing. If false, how do you account for the pattern in the data?



Source: March 1993 Current Population Survey; CD-ROM supplied by the Bureau of the Census through the U.C. Survey Research Center.

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: Controlled Experiment/Observational Study (30 Points)

As part of a study on exercise and health, a group of 1,000 people was followed for 5 years. At the beginning of the study, the researchers asked each person whether they exercised regularly or not. At the end of the study, the researchers measured several health—related variables, and in doing so, they noticed that the death rate for the exercise group was much lower than for the no–exercise group.

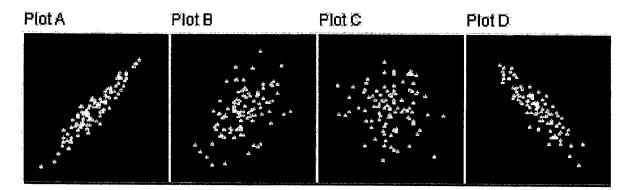
1. (10 Points) Is this an example of a controlled experiment or an observational study? Explain.

2. (20 Points) Does this result necessarily imply that if people who do not exercise start to exercise regularly, they will live longer, on average, than if they do not? Explain clearly.

Question 4: Guessing the Correlation Coefficient (40 Points)

Match the four scatterplots with their correlations from the list:

$$-1.03, \quad -0.99, \quad -0.89, \quad -0.50, \quad -0.05, \quad 0.50, \quad 0.93, \quad 1.03$$



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 ≈ 2.7 inches, average height of wife ≈ 63 inches, SD ≈ 2.5 inches, $r \approx 0.25$. Show all the work needed to obtain the answer.

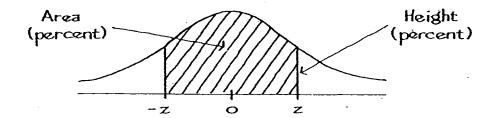
1. (15 Points) Predict the height of a wife when the height of her husband is 72 inches.

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



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	
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		1.80	92.81	3.30	
0.35		1.85	93.57	3.35	
0.40		1.90		3.40	
0.45	34.73	1.95		3.45	
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	
0.60	45.15	2.10	96.43	3.60	
0.65	48.43	2.15		3.65	
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	5 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	
1.10	72.87	2.60	99.07	4.10	
1.15	74.99	2.65	99.20	4.15	5 99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	5 99.9979
1.30	80.64	2.80	99.49	4.30	-
1.35	82.30	2.85	99.56	4.35	
1.40	83.85	2.90	99.63	4.40	
1.45	85.29	2.95	99.68	4.45	5 99.9991

Statistics 1040, Section 006, Midterm 2 (200 Points)

November 8, 2002

Your Name:
Question 1: The Expected Value and Standard Error (40 Points) A hundred draws are made at random with replacement from the box
1. (5 Points) If the sum of the draws is 321, what is their average?
2. (5 Points) If the average of the draws is 3.78, what is the sum?
3. (30 Points) Estimate the chance that the average of the draws is between 3 and 4.

Question 2: Probability and Chance (40 Points)

In a box of 12 chocolates, three are mint, two are orange, five are caramel, and two are cherry. I choose two chocolates at random (without replacement!). Answer each of the following questions separately.

1. (5 Points) What is the chance that the first chocolate is mint?

2. (10 Points) What is the chance that the first two chocolates are both mint?

3. (10 Points) What is the chance that the first chocolate is mint and the second is orange?

4. (15 Points) If I only like caramel, what is the chance that I choose at least one chocolate I like?

Question 3: Regression (50 Points)

In a study between the amount of rainfall and the quantity of air pollution removed, the following data were collected:

Daily rainfall (mm): Average = 5, SD = 1.5

Air pollution removed (in $\mu g/m^3$ particles removed): Average = 122, SD = 9.5 Correlation = -0.98.

1. (15 Points) Find the regression equation for predicting the air pollution removal from the amount of daily rainfall.

2. (10 Points) Using your regression equation, predict the air pollution removal when the daily rainfall is $4.8 \ mm$.

3. (10 Points) Find the r.m.s. error for predicting the air pollution removal from the amount of daily rainfall.

4. (15 Points) Would you be surprised if you read that the daily rainfall of 4 mm had removed the amount of air pollution of 150 $\mu g/m^3$? Why or why not? Explain your reasoning, using the r.m.s. error.

Question 4: Sampling (30 Points)

Suppose you have a telephone directory for your local college from which you randomly select 100 names. To find out how students feel about a new pub on campus, you call the 100 numbers and interview the person who answers the phone.

1. (21 Points) Give three possible sources of bias for this survey.

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

Question 5: Normal Approximation for Probability Histograms (40 Points)

A coin is tossed 100 times. True or false, and explain.

(Answer and explain each of the following questions separately.)

1. (10 Points) The expected value for the number of heads is 50. True/False?

2. (10 Points) The expected value for the number of heads is 50, give or take 5 or so. True/False?

3. (10 Points) The number of heads will be 50. True/False?

4. (10 Points) The number of heads will be around 50, give or take 5 or so. True/False?

Formulas:

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

slope =
$$r \times \frac{SD_y}{SD_x}$$

$$intercept = avg_y - slope \times avg_x$$

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

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

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

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

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

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

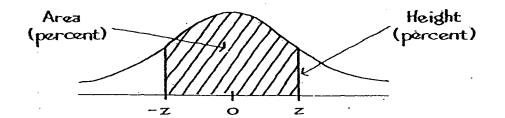
$$SD = (bigger - smaller) \times \sqrt{\begin{array}{c} fraction \\ bigger \end{array}} \times \begin{array}{c} fraction \\ smaller \end{array}$$

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

$$average = \frac{number of [1]'s}{how many tickets in the box}$$

$$SD = \sqrt{\begin{array}{c} fraction \\ of \boxed{1} \text{ 's } \times \begin{array}{c} fraction \\ of \boxed{0} \text{ 's} \end{array}}$$

A NORMAL TABLE



Z	Area	Z	Area	Z	Area
00.0	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
08.0	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. 9 9	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