

Statistics 1040, Sections 007 & 009, Midterm 1 (200 Points)

Friday, October 5, 2007

Your Name: _____

Question 1: Correlation / Regression I (25 Points)

1. (13 Points) In a study of high-school students, a positive correlation was found between hours spent per week doing homeworks, and scores on standardized achievement tests. The investigators concluded that doing homework helps prepare students for these tests. Does the conclusion *follow from the data*? Answer **yes** or **no**, and **explain briefly**.

2. (12 Points) The freshmen at a large university are required to take a battery of aptitude tests. Students who score high on the mathematics test also tend to score high on the physics test. On both tests, the average score is 60; the SDs are the same too. The scatter diagram is football-shaped. Of the students *who scored about 75 on the mathematics test*:

- (i) just about half scored over 75 on the physics test.
- (ii) more than half scored over 75 on the physics test.
- (iii) less than half scored over 75 on the physics test.

Choose one option and explain.

Question 2: Controlled Experiment / Observational Study (40 Points)

The following is part of an article from Newsweek, April 24, 2006. This question concerns the Women's Health Study, described in the second paragraph.

Take an Aspirin and ...

BY JULIE BURING, SC.D., AND NANCY FERRARI

Aspirin is a wonder drug, plain and simple. At high doses, it quells inflammation; at medium doses, it provides effective pain relief; at low doses, it reduces the blood's ability to clot by inhibiting the action of tiny blood cells called platelets. It makes sense, then, that aspirin might help prevent clot-related cardiovascular events such as heart attack and stroke, even in healthy people. In 1988, the Physicians' Health Study showed exactly that. In healthy men, 325mg of aspirin taken every other day for five years reduced the risk of a first heart attack by 44 percent. That was great news. For men.

It wasn't until March 2005 that the Women's Health Study addressed aspirin's benefits for women. Healthy women — who were at least 45 years old at the start of the study — who participated in the study took either 100mg of aspirin or a placebo every other day for 10 years. Surprisingly, the women taking aspirin experienced no reduction in heart-attack risk. However, aspirin takers were 17 percent less likely to have a stroke. ...

1. (10 Points) Is the study controlled? How do you know?
2. (10 Points) Is the study blind? How do you know?
3. (10 Points) What is a placebo? Why is it used?
4. (10 Points) The article does not say how the women were assigned to the aspirin and placebo groups. What is the best way to do this, and why?

Question 3: Normal Curve (45 Points)

For 167 college students the average handspan size is 20.9 inches, with an SD of 1.9 inches. Fill the blanks in the statements below and **show all the work needed to obtain the answers.**

1. (15 Points) Using the normal curve, approximately what percentage of the students have a handspan of more than 20 inches?

The answer is: _____ %

2. (15 Points) And what percentage of the students have a handspan of more than 24 inches?

The answer is: _____ %

3. (15 Points) If 80 percent of the students have a smaller handspan than the teacher, what is the teacher's handspan?

The answer is: _____ inches

Question 4: Correlation / Regression II (50 Points)

For the 167 college students in Question 3, the relationship between height and handspan size is summarized as follows:

height: average = 68.0 inches SD = 4.0 inches
handspan size: average = 20.9 inches SD = 1.9 inches
 $r = 0.75$

Fill the blanks in the statements below and **show all the work needed to obtain the answers.**

1. (10 Points) Six scatter diagrams are printed on the next page. Which of the scatter diagrams is the correct one for these data? **Circle the correct letter below (No explanation is needed for this part!):**

A B C D E F

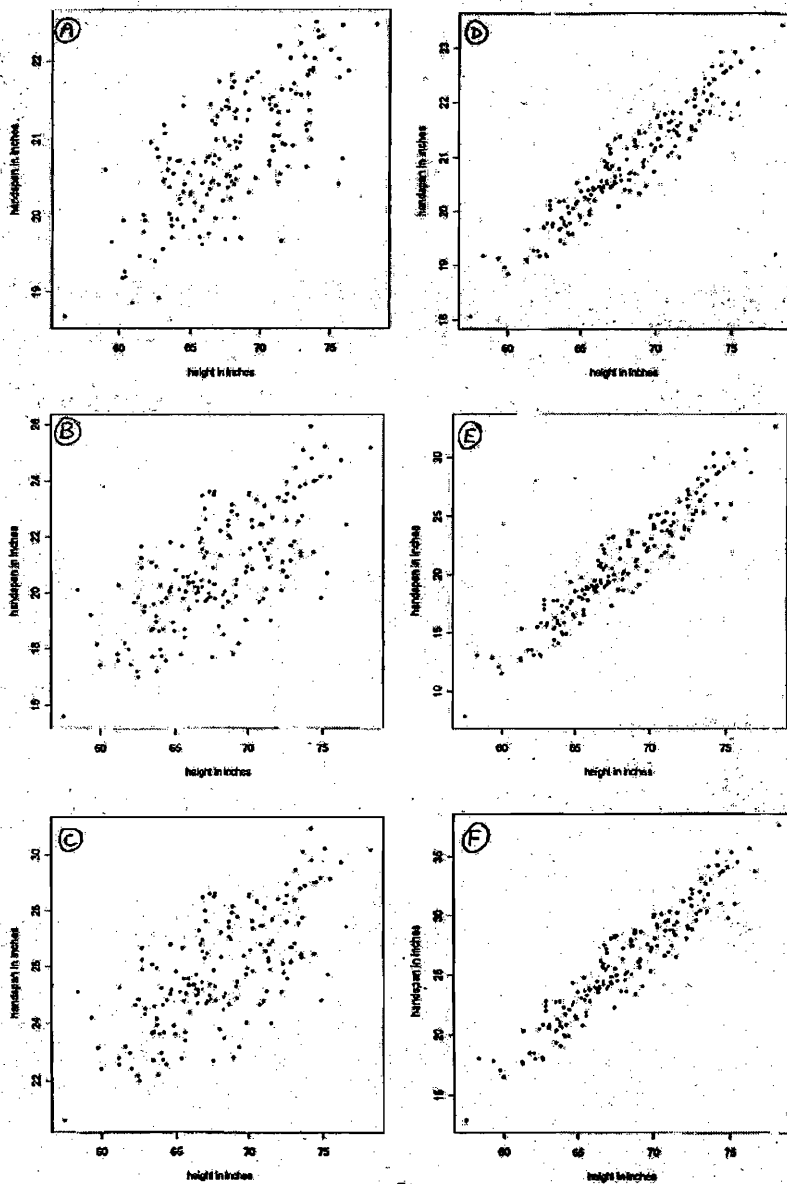
2. (15 Points) Using the summary statistics above, what is the regression estimate for handspan for a student who is 60 inches tall?
The answer is: _____ inches

3. (15 Points) Find the r.m.s. error for your answer in the previous part.
The answer is: _____

4. (10 Points) What would the correlation coefficient be if we changed all the handspan measurements to centimeters? (There are 2.54 centimeters in an inch).
The answer is: _____

Formulas:

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

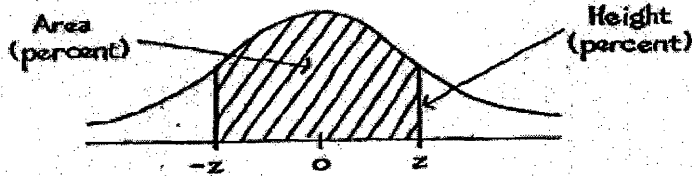


Question 5: Histograms / Change of Scale (40 Points)

An investigator draws a histogram for some height data, using the metric system. She is working in centimeters (cm). The vertical axis shows density, and the top of the vertical axis is 10 percent per cm. Now she wants to convert to millimeters (mm). There are 10 millimeters to the centimeter. On the horizontal axis, she has to change 175 cm to _____ mm, and 200 cm to _____ mm. On the vertical axis, she has to change 10 percent per cm to _____ percent per mm, and 5 percent per cm to _____ percent per mm.

Just fill in the correct answers — no explanation is needed.

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

Statistics 1040, Sections 007 & 009, Midterm 2 (200 Points)

Friday, November 9, 2007

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: EV, SE, and Normal Curve (40 Points)

In a certain town, there are 40,000 registered voters, of whom 15,000 are Democrats. A survey organization is about to take a simple random sample of 1,000 registered voters. **Show your work!**

1. (8 Points) Find the box model.

2. (20 Points) The expected number of Democrats in this sample of 1,000 is _____ with an SE of _____.

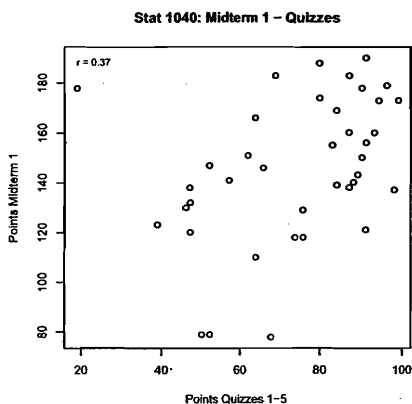
3. (12 Points) The chance that **at least** 500 of the voters in the sample are Democrats is about _____ %.

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: 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 4: Probability and Chance (40 Points)

A bookshelf contains 8 novels, 7 books of poems, 1 dictionary, and 2 copies of Freedman, Pisani, and Purves's "Statistics" textbook. I pick two books at random without replacement. **Answer each of the following questions separately. Show your work!**

1. (5 Points) What is the chance that the first book is a novel or a dictionary?

The chance is _____%

2. (13 Points) If I want to study statistics, what is the chance that I pick at least one copy of the Stats textbook?

The chance is _____%

3. (10 Points) What is the chance that the first two books are both novels?

The chance is _____%

4. (12 Points) What is the chance that I pick one book of poems and a dictionary (in any order)?

The chance is _____%

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:

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

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

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

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

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

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

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

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

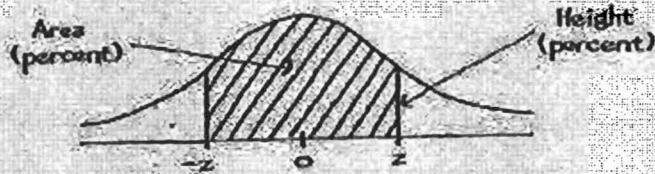
$$\text{SD} = (\text{bigger} - \text{smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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

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

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

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