Stat 1040:

Midterm 1 & 2, Fall 2003

Statistics 1040, Section 004, Midterm 1 (200 Points)

October 3, 2003

Your Name: _____

Question 1: Normal Distribution (**50 Points**)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. A longterm average for the verbal ability portion of this exam of all college seniors and graduates who take this exam is 494, with a standard deviation of 115. Assuming that the histogram of all GRE scores follows the normal curve, answer the following questions. **Show your work.**

1. (**25 Points**) If you received a score of 650 on the GRE exam, what percentile of the distribution would you be in?

   Answer: _____ th percentile

2. (**25 Points**) A graduate school program in English will admit only students with GRE verbal ability scores in the top 30%. What is the lowest GRE score you must have to be accepted in this graduate program?

   Answer: _____

Question 2: Correlation and Regression (**60 Points**)

A study of the IQs of husbands and wives obtained the following results:
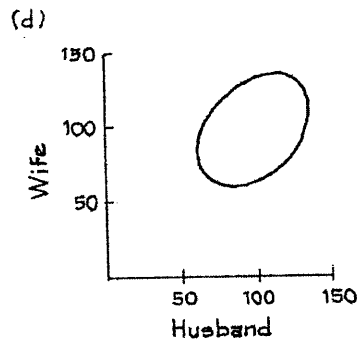
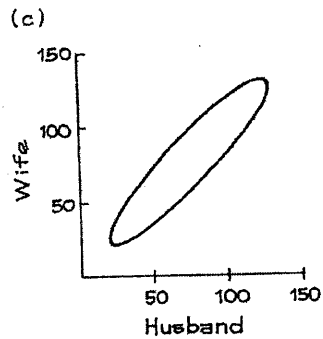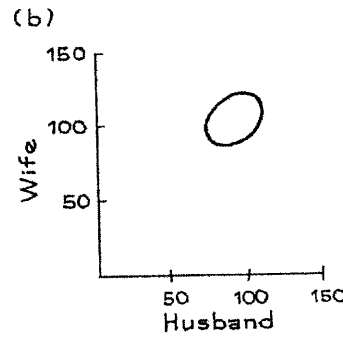for husbands, average IQ = 100, SD = 15
for wives, average IQ = 100, SD = 15
$r = 0.6$

1. (**15 Points**) One of the following is a (summarized) scatter diagram for the data. Which one?
   Circle your answer and **explain briefly** why you reject the others.
   (a)   (b)   (c)   (d)

(a)

(b)

(c)

(d)

2
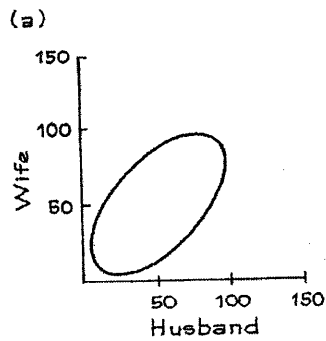
2. **(15 Points)** Predict the IQ of the wife whose husband has an IQ of 130.

3. **(15 Points)** Predict the IQ of the husband whose wife has an IQ of 118.

4. **(15 Points)** Apparently, intelligent men tend to marry women who are less intelligent than themselves. On the other hand, women tend to marry men who are even less intelligent! How is this possible?

Question 3: Controlled Experiment/Observational Study (**60 Points**)

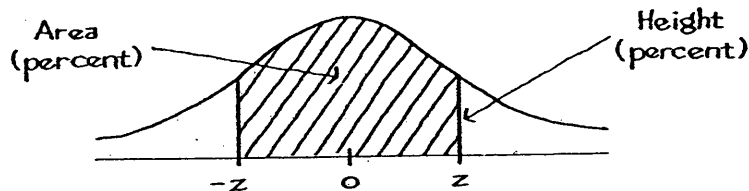A recent study in Europe looked at a large group of women of childbearing age. The researchers asked each woman how much alcohol they had consumed over the past 12 months. The researchers found that women who drank moderate amounts of alcohol were somewhat less likely to have infertility problems than 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 4: Average and Standard Deviation (**30 Points**)

True or false, and **explain briefly** (if false, explain what happens instead):

- (**5 Points**) If you add 7 to each entry on a list, that adds 7 to the average.
  True / False?

- (**5 Points**) If you add 7 to each entry on a list, that adds 7 to the SD.
  True / False?

- (**5 Points**) If you double each entry on a list, that doubles the average.
  True / False?

- (**5 Points**) If you double each entry on a list, that doubles the SD.
  True / False?

- (**5 Points**) If you change the sign of each entry on a list, that changes the sign of the average. **True / False?**

- (**5 Points**) If you change the sign of each entry on a list, that changes the sign of the SD. **True / False?**

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 004, Midterm 2 (200 Points)

November 7, 2003

Your Name: _____

Question 1: Chances and Probabilities (**40 Points**)

A shelf contains 9 books: 5 novels, 3 books of poems, and a dictionary. **Show your work!**

1. (**8 Points**) If one book is picked at random from the shelf, what is the chance that it is a novel?

2. (**8 Points**) If one book is picked at random from the shelf, what is the chance that it is a novel or a dictionary?

3. (**8 Points**) If two books are picked at random from the shelf, what is the chance that they are both novels?

4. (**8 Points**) If two books are picked at random from the shelf, what is the chance that neither of them is a novel?

5. (**8 Points**) If two books are picked at random from the shelf, what is the chance that at least one of them is a novel?

1

Question 2: Sample Size and SE% (**10 Points**)

A professor at a university wants to know what percentage of students visits the university Web page on a regular basis. This professor randomly selects and interviews 100 students (Study 1). Unknown to this professor, another professor at the same university has assigned the same study as a course project — however, only 25 randomly selected students will be interviewed during the course project (Study 2). Even without knowing the exact box and EV%, we can make a statement about the SE%'s for the two studies. Just circle the correct answer(s). (You don't have to provide any further explanation here.)

1. The SE% of Study 1 will be 4 times as big as the SE% of Study 2.

2. The SE% of Study 1 will be 2 times as big as the SE% of Study 2.

3. The SE% of Study 1 will be $\sqrt{2}$ times as big as the SE% of Study 2.

4. The SE% of Study 1 will be about the same as the SE% of Study 2.

5. The SE% of Study 2 will be 4 times as big as the SE% of Study 1.

6. The SE% of Study 2 will be 2 times as big as the SE% of Study 1.

7. The SE% of Study 2 will be $\sqrt{2}$ times as big as the SE% of Study 1.

Question 3: Regression (60 Points)

The average height of married American women in their early twenties is about 64.5 inches and the standard deviation is about 2.5 inches. The average height of married American men of the same age is about 68.5 inches, with a standard deviation of about 2.7 inches. The correlation between husbands heights and wives heights is about $r = 0.5$. **Show your work!**

1. **(20 Points)** Find the regression equation for predicting the height of a husband based on the wife's height in young couples.

2. **(10 Points)** Using your regression equation, predict the height of a husband of a woman who is 67 inches tall. The predicted height of the husband is _____.

3. **(15 Points)** Find the r.m.s. error for predicting the husband's height from the wife's height. The r.m.s. error is _____.

4. **(15 Points)** Would you be surprised to find out that a women who is 67 inches tall has a husband who is 77 inches tall? Why or why not? Explain!

3

Question 4: Box Model, Sums, Percentages, and the Normal Curve (**70 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. Answer the questions below. Where necessary, find the box model, box average and box SD. **Show your work!**

1. (**35 Points**) Find the chance that between 19% and 21% of the forms chosen for audit have gross incomes over $50,000. The chance is _____%.

2. (**35 Points**) Find the chance that the total gross income of the audited forms is over $33,000,000. The chance is _____%.

Question 5: Probability Histograms (20 Points)

1. (**10 Points**) Twenty-five draws are made at random with replacement from each of the boxes below:

A) $\boxed{\boxed{0}\ \boxed{1}}$ B) $\boxed{9 \times \boxed{0} \quad 1 \times \boxed{1}}$ C) $\boxed{24 \times \boxed{0} \quad 1 \times \boxed{1}}$

The probability histograms for the sums are shown below, in scrambled order. Match the histogram with the boxes. **Briefly explain your choices.**
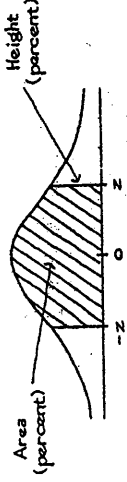
(i)   (ii)   (iii)
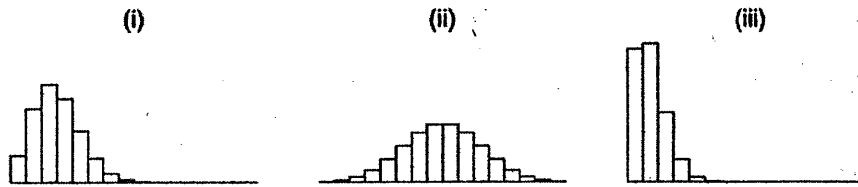
(i) goes with box _____

(ii) goes with box _____

(iii) goes with box _____

2. (**10 Points**) Shown below are probability histograms for the sum of (a) 100, (b) 400, and (c) 900 draws from the box $\boxed{99 \times \boxed{0} \quad 1 \times \boxed{1}}$. Which histogram is which? **Explain briefly.**

(i)   (ii)   (iii)
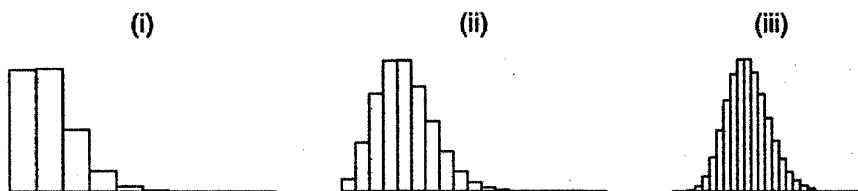
(i) goes with sum _____

(ii) goes with sum _____

(iii) goes with sum _____

5

Tables

Area (percent)
Height (percent)

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

Formulas:

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

$$\text{slope} = r \times \frac{SD_y}{SD_x} \qquad \text{intercept} = avg_y - \text{slope} \times avg_x$$

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

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

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

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

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

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

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

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

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

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

6