Stat 1040:

Midterm 1 & 2, Spring 2002

Statistics 1040, Sections 002, 003 & 004, Midterm 1 (200 Points)

February 15, 2002

Your Name: _____

Question 1: Normal Distribution (**40 Points**)

A college that has an excellent track–and–field athletics program runs short on scholarships and cannot further support all of its 100m track athletes. The athletics director wants to make a decision which athletes to support in the future based on their athletic capabilities. Based on the athletes performance over the last few years, it is known that the distribution of running times approximately follows the normal curve with an average of 10.8 sec and a standard deviation of 0.2 sec. Answer the following 2 questions:

1. Which time does an athlete have to run in the end–of–semester competition to belong to the fastest 70% of the runners that will still obtain a scholarship for the next year? (**20 Points**)

2. What are the chances that a randomly selected athlete from this college will set a new world record of 9.7 sec or better in the end–of–semester run? (**20 Points**)

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

In a recent study on SIDS (Sudden Infant Death Syndrome), one hospital collected data on 128 babies who died from SIDS in the last 12 months. They took a random sample of 500 babies (of similar ages) who did not die from SIDS (the "controls"), and they compared the two groups with respect to several variables of interest (e.g. whether the child slept on his or her stomach, birthweight, time of year, whether the mother smoked, whether she breast–fed, socio–economic status, etc.).

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

2. One physician noticed that 63% of the SIDS babies had mothers who smoked during pregnancy, whereas only 26% of the control babies had mothers who smoked during pregnancy. Another physician claimed that low birthweight could be a "confounding factor". Explain what it means for low birthweight to be a "confounding factor". Be specific. (**15 Points**)

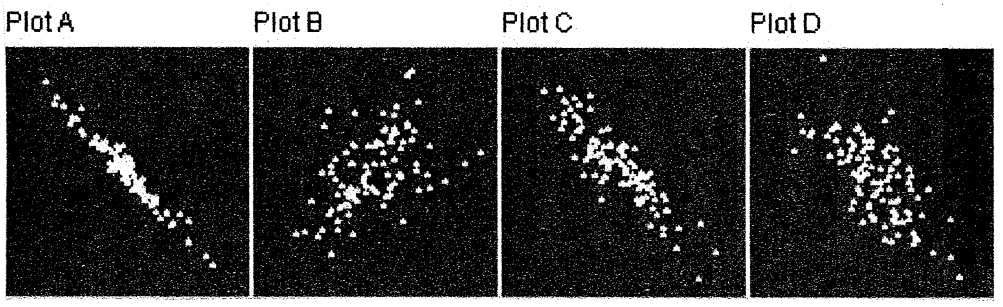
3. If you had access to the data, what would you do to "control for" birthweight? (**15 Points**)

Question 3: Guessing the Correlation Coefficient **(40 Points)**

The correlation coefficients for the data points displayed in these four scatterplots are 4 out of the following 12 values:

$-1.03, \quad -0.97, \quad -0.88, \quad -0.69, \quad -0.46, \quad -0.05, \quad 0.05, \quad 0.46, \quad 0.69, \quad 0.88, \quad 0.97, \quad 1.03$

For each plot below, indicate the corresponding correlation coefficient r:



Correlation for Plot A:

Correlation for Plot B:

Correlation for Plot C:

Correlation for Plot D:

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

In one study, the correlation between the educational level of husbands and wives in a certain town was about 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.

1. Predict the educational level of a woman whose husband has completed 18 years of schooling. **(15 Points)**

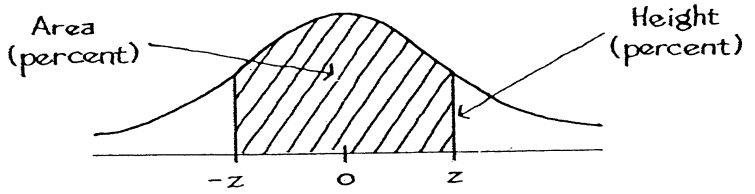
2. Predict the educational level of a man whose wife has completed 15 years of schooling. **(15 Points)**

3. Apparently, well–educated men marry women who are less well–educated than themselves. But the women marry men with even less education. How is this possible? **(10 Points)**

Question 5: Representative Sample (**40 Points**)

In 1998 a researcher took a large representative sample of women in the U.S. She found that the older these women were, the less their daily average meat consumption. True or false and explain: "The data show that as women grow older, their daily average meat consumption drops."

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, Sections 002, 003 & 004, Midterm 2 (200 Points)

March 27, 2002

Your Name: _____

Question 1: Chances and Probabilities **(30 Points)**

I have a bag with 20 balls in it: 10 are red, 8 are blue, and 2 are green.

1. If I draw one ball at random from the bag, what is the chance that I get a red ball or a green ball? **(10 Points)**

2. If I draw two balls at random **without replacement**, what is the chance that I get a red ball, followed by a green ball? **(10 Points)**

3. If I draw three balls at random **with replacement**, what is the chance that I get at least one red ball? **(10 Points)**

Question 2: Regression **(50 Points)**

A student wonders if people of similar heights tend to date each other. To find this out, she measured herself, her dormitory roommates, and all of her female classmates; then she measured the man each woman was currently dating. After making many measurements and analyzing the data, she found out that the women were on average 66 inches tall, with a standard deviation of 2.0 inches; their dates were on average 69 inches tall, with a standard deviation of 2.5 inches. The correlation coefficient between the women's and the men's heights was 0.57.

1. Find the regression equation for predicting the height of a woman's date based on her own height. **(20 Points)**

2. Using your regression equation, predict the height of a date for a woman who is 67 inches tall. **(10 Points)**

3. Find the r.m.s. error for predicting the date's height from the woman's height. **(10 Points)**

4. Does the slope of the regression line (that you found in Part 1) say that men, if they date taller women, will become taller? Why or why not? Explain! **(10 Points)**

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

In Web polls, anyone who views a certain Web page is allowed to vote by clicking on their choice of button. 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. Web based polls such as this are notoriously susceptible to bias. Give **three** possible sources of bias for this poll. **(21 Points)**

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

Question 4: Box Model, EV_{sum}, SE_{sum}, and the Normal Curve **(50 Points)**

Suppose it is known that 10% of all people in Utah have the blood type AB. Suppose we take a random sample of 500 Utahns and want to determine how many of them have the blood type AB.

1. Find the box model for this situation. **(10 Points)**

2. How many people in our sample do you expect to have the blood type AB? **(10 Points)**

3. What is the corresponding standard error? **(10 Points)**

4. What is the chance that fewer than 40 people in our sample have the blood type AB ? **(20 Points)**

Question 5: Law of Averages **(40 Points)**

A die will be rolled some number of times. Which is better: 60 rolls or 600 rolls in the situations listed below? Circle the number and briefly **explain** your answer for each situation:

1. You win \$1 if it shows ⚀ more than 20% of the times. 60 / 600
   (10 Points)

2. You win the dollar if the percentage of ⚀'s is more than 15%. 60 / 600
   (10 Points)

3. You win the dollar if the percentage of ⚀'s is between 15% and 20%. 60 / 600
   (10 Points)

4. You win the dollar if the percentage of ⚀'s is exactly $16\frac{2}{3}\%$. 60 / 600
   (10 Points)

Formulas:

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

$$\text{slope} = r \times \frac{\text{SD}_y}{\text{SD}_x}$$

$$\text{intercept} = \text{avg}_y - \text{slope} \times \text{avg}_x$$

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

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

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

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

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

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

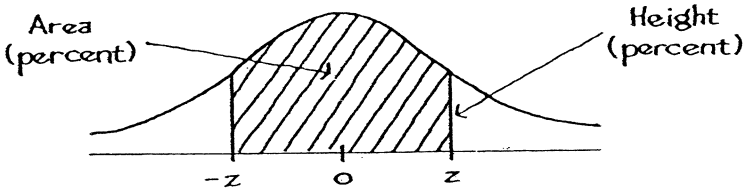
$$\text{SD} = (\text{bigger - smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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

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

$$\text{SD} = \sqrt{\frac{\text{fraction}}{\text{of } \boxed{1}\text{'s}} \times \frac{\text{fraction}}{\text{of } \boxed{0}\text{'s}}}$$

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