Statistics 1040, Section 009, Midterm 1 (200 Points)

Friday, February 17, 2006

Your Name: _____

Question 1: Correlation (**20 Points**)

Investigators take a sample of DINKS (dual–income families, where husband and wife both work and have no kids). The investigators have data on the husband's income and the wife's income. By definition,

$$\text{family income} = \text{husband's income} + \text{wife's income}.$$

The average family income was around \$50,000, and 10% of the couples had family income in the range \$45,000–\$55,000. Fill in the blanks, using the options given below, and **explain briefly**:

1. (**10 Points**) The correlation between wife's income and family income is _____.

2. (**10 Points**) Among couples whose family income is in the range \$45,000–\$55,000, the correlation between wife's income and husband's income is _____.

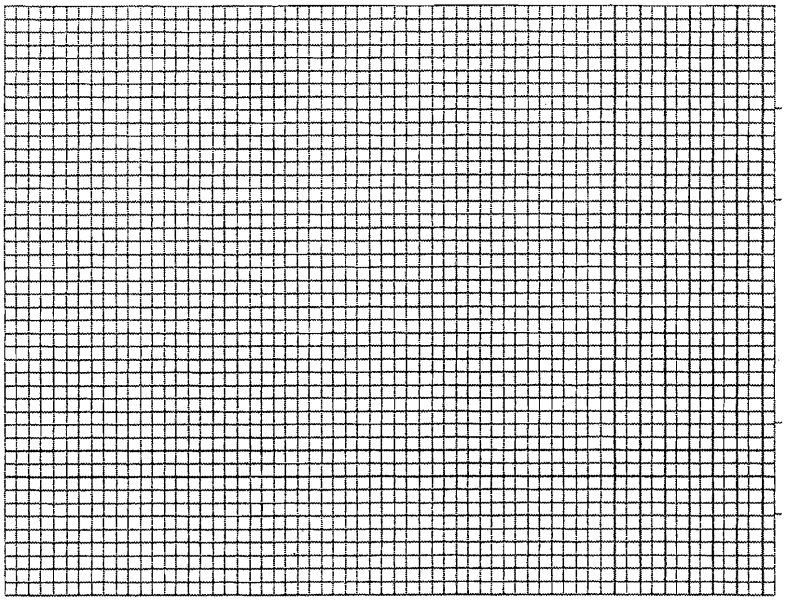
Options: (a) -1 (b) nearly -1 (c) somewhat negative (d) 0
(e) somewhat positive (f) nearly 1 (g) 1 (h) -1.1 (i) 1.1

Question 2: Histograms (**35 Points**)

When the Tribbles invaded the spaceship Enterprise, suppose that crew member Spock decided to take the logical step of seeing what the crew was up against, and he wanted to graphically represent the size of the Tribbles. Suppose that the table below summarizes the heights of the 50 Tribbles he found on the bridge. (Class intervals include the left but not the right endpoints.) [If you don't know what Tribbles are, take a look at `http://www.startrek.com/startrek/view/series/TOS/episode/68744.html` – there are several photos available at this Web site.]

Tribble Height (inches)	Number of Tribbles
3–5	19
5–7	10
7–9	12
9–11	7
11–13	2
	50

1. (**25 Points**) Draw a histogram of these height data, with the vertical axis on the usual **density scale**, both axes labeled, and heights of bars clearly indicated.



2. (**10 Points**) If a Tribble is in the 82^{nd} percentile for height, about how tall is it? (Note: Use the histogram, NOT the normal curve here.)

   Its height is about: _____

Question 3: Normal Curve (**45 Points**)

Spock did some additional height measurements of **all** Tribbles aboard Enterprise (and not only of those found on the bridge) and determined that their overall size closely follows the normal curve, with an average of 6.5 inches and an SD of 2.5 inches.

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

1. (**15 Points**) The percentage of Tribbles that are between 8.0 and 10.0 inches tall is _____.

2. (**15 Points**) The percentage of Tribbles that are less than 4.5 inches tall is _____.

3. (**15 Points**) When using the normal curve, a Tribble that is in the 82^{nd} percentile for height is about _____ inches tall.

Question 4: 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 amount of alcohol were somewhat less likely to have infertillity 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? **Circle your answer and 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 5: 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.

Show your work.

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

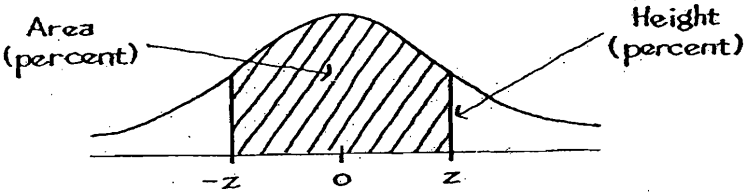
   The answer is: _____ years

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

   The answer is: _____ years

3. (**10 Points**) 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?

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 009, Midterm 2 (200 Points)

Friday, March 31, 2006

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) will be fine. If you do not have to provide an explanation, do not waste your time giving an unneeded explanation.**

Question 1: Sampling **(20 Points)**

(Hypothetical) A survey is carried out by the finance department to determine the distribution of household size in a certain city. They draw a simple random sample of 1,000 households. After several visits, the interviewers find people at home in only 653 of the sample households. Rather than face such a high non–response rate, the department draws a new batch of households, and uses the first 347 completed interviews in the second batch to bring the sample up to its planned strength of 1,000 households. The department counts 3,087 people in these 1,000 households, and estimates the average household size in the city to be about 3.1 persons. Is this estimate likely to be **too low**, **too high**, or **about right**?

Circle your answer and explain clearly!

Question 2: Probability (**40 Points**)

Professor J.S. recently experimented with homegrown chili plants. From 4 seeds that were planted, one seed germinated. Use this chance (25%) as the basis for all calculations in this question and assume that seeds germinate independently from each other.

Show your work!

In a second experiment, J.S. planted 4 more seeds. Determine the following chances:

1. (**15 Points**) The chance that **all 4** of these 4 seeds will germinate is about _____ %.

2. (**15 Points**) The chance that **none** of these 4 seeds will germinate is about _____ %.

3. (**10 Points**) The chance that **at least 1** of these 4 seeds will germinate is about _____ %.

Question 3: Regression (**40 Points**)

A selection of 65 varieties of cereal were tested for calories and sodium (in milligrams) for an one–cup serving. The results can be summarized as follows:

Average sodium = 240 mg; SD = 131 mg;
Average calories = 149 calories; SD = 62 calories; $r = 0.53$.

Show your work!

1. (**10 Points**) Find the equation of the regression line for predicting number of mg sodium in an one–cup serving of cereals from calories.

2. (**10 Points**) Predict the number of mg sodium in an one–cup serving of cereals that has 200 calories per cup.

3. (**10 Points**) Find the r.m.s. error for predicting mg sodium from calories.

4. (**10 Points**) Explain why it would not be a good idea to use the information in the question to estimate the amount of sodium for a cereal with 350 calories per cup.

Question 4: EV, SE, Normal Curve & Sampling (**70 Points**)

Time reported in its March 20, 2006, issue on page 30: *"***74%** *Proportion of female college students and graduates who said women on spring–break trips use drinking as an excuse for behavior like public display of nudity and table dancing."*

Assume that 74% indeed is the true percentage of all female college students and graduates in the US who share this opinion. Suppose an independent researcher wants to do some follow–up study and draws a simple random sample of 400 female college students and graduates in the US.

Show your work!

1. (**10 Points**) Indicate the box model.

2. (**20 Points**) The expected number of these 400 females in the follow–up study who share the opinion given above is _____ with a standard error of about _____.

3. (**20 Points**) The chance that **at least 310** of these 400 females in the follow–up study who share the opinion given above is about _____ %.

4. **(10 Points)** (Hypothetical) Suppose *Time* would have asked the female college students and graduates in their study: *"On your spring–break trips, do you use drinking as an excuse for behavior like public display of nudity and table dancing?"*

   Other things being equal, the percentage of women who would have answered yes to this question would have been **(a) higher than 74%**, **(b) about 74%**, or **(c) much less than 74%**.

   Circle your answer and explain clearly!

5. **(10 Points)** (Hypothetical) Suppose that you were asked by the independent researcher to conduct this follow–up study for him. It is up to you to determine how to draw a sample of 400 females that are representative for all female college students and graduates in the US and obtain their opinion regarding the question originally asked by *Time*. If you have the choice, the best possible way to draw this sample is:

   - (a) Travel to South Padre Island in Texas at the start of Spring Break, go to the beach, ask women (as they arrive on the beach) whether they are college students or graduates, and then ask the first 400 of those for their opinion.
   - (b) To avoid travel, get a list of all current female USU students and female USU graduates, draw a simple random sample of 400 female USU students/graduates, contact them by phone, cell phone, or at home, and ask them for their opinion.
   - (c) Get lists of current female students and female graduates from all colleges (and universities) in the US, draw a simple random sample of 400 female students/graduates, contact them by phone, cell phone, or travel across the country if necessary to reach them at home (this may take weeks!), and ask them for their opinion.

   Circle your answer and explain clearly!

P.S.: If you have never been to South Padre Island or a similar location during Spring Break, take a look at Web pages such as `http://www.spadre.com/springbreak.htm` and think again whether the percentage reported in *Time* might be true or whether it is totally far–fetched...

Question 5: Chance Errors in Sampling (**30 Points**)

Five hundred draws are made at random from the box

$$\boxed{60{,}000 \times \boxed{0} \qquad 20{,}000 \times \boxed{1}}.$$

True or false? Circle your answers. No explanation is needed.

1. (**5 Points**) True / false: The expected value for the percentage of 1's among the draws is exactly 25%.

2. (**5 Points**) True / false: The expected value for the percentage of 1's among the draws is around 25%, give or take 2% or so.

3. (**5 Points**) True / false: The percentage of 1's among the draws will be around 25%, give or take 2% or so.

4. (**5 Points**) True / false: The percentage of 1's among the draws will be exactly 25%.

5. (**5 Points**) True / false: The percentage of 1's in the box is exactly 25%.

6. (**5 Points**) True / false: The percentage of 1's in the box is around 25%, give or take 2% or so.

Formulas:

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

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

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

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

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

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

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

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

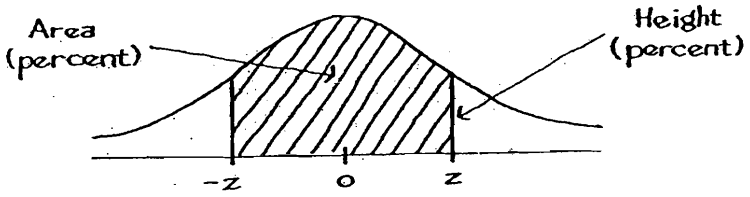
$$\text{SD} = (\text{bigger - smaller}) \times \sqrt{\frac{\text{fraction}}{\text{bigger}} \times \frac{\text{fraction}}{\text{smaller}}}$$

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

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

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

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