POS 416z **Research Methods in Political Science I**
Second hour test **Fall 2002**

Name _____

Please answer each of the following on these test pages in the blanks or spaces provided. Show any significant work (except arithmetic) for part credit. A calculator is recommended. Continue answers on the backs of these test pages if necessary. Statistical tables are attached for your use.

1) Health researchers studied the smoking habits and socioeconomic status (SES) of 356 volunteer subjects.

There were three categories of SES among the subjects — High, Middle, and Low — and there were three categories of smoking habits — current smoker, former smoker, and never smoked. The data are shown at the right.

	Socioeconomic Status			
Smoking	High	Middle	Low	Totals
Current	51	22	43	116
Former	92	21	28	141
Never	68	9	22	99
Totals	211	52	93	356

The researchers planned to use a chi-square analysis of the data in this two-way table to see if there is a statistically significant association between SES and smoking habits in this data.

a) [2 points] What is the null hypothesis in this study?

b) [4 points] How many observations would we expect in the cell "Current, High" under the null hypothesis? [Be sure to be clear *how* you are computing this expected number (give a formula or use words, as you wish), as well as clear about your numerical answer.]

c) [4 points] Let's assume we have computed the expected number of observations in each cell under the null hypothesis and put the values in a second table. Tell how you would compute the *chi-square statistic* from this table of expected values and the table in part (a). (Again, a formula or words. But don't waste time computing it; the result is given in part (d).)

d) [4 points] The chi-square statistic for this data turns out to be 18.51. Use the attached chi-square table to test the null hypothesis. Be sure to tell:

- How many degrees of freedom are involved?
- What do you find out from chi-square table?
- Do you accept or reject the null hypothesis?
- What p-value applies to your conclusion?

2) The web site <http://politicalhumor.about.com> contains an online polling page called the "Poll Vault" (subtitled "funnier than Gallup and just as useless"). In one of their online polls they asked readers "Who makes the better butt of political jokes, George W. Bush, or Bill Clinton? Of the 4098 people responding (so far) to this online poll, 69 percent said "George Bush."

a) [6 points] Find a 95% confidence interval for the proportion of the population this sample represents who think George Bush is the better butt of political jokes.

b) [6 points] Adam, Brenda, and Cody are having an argument about this poll.

- Adam says the sample in the poll is biased, so the real proportion of the US population who feel this way could be as low as, say, 65 percent.
- Brenda says it's not biased, it's really a random sample since anybody could answer the poll if they wanted to, but she still thinks it's possible for the true population proportion to be as low as 65%.
- Cody says they're both wrong, that the true population proportion couldn't be that low because it's outside the 95% confidence interval.

Who is the most right here? Who is the most wrong? Explain. [Write your answer on the back of one of these pages.]

3) Do women score lower on the quantitative part of the Graduate Record Exam than men? To answer this question, researchers obtained the test scores for a random sample of 125 women and 150 men, finding that the mean quantitative GRE score for the men ($\overline{X}$) was 573 and the mean score for the women ($\overline{Y}$) was 556. The researchers knew that the standard deviation of the population of individual GRE scores is about 100.

We can test the null hypothesis that the population means for men and women are equal by considering the *difference* in scores, e.g.,

H_o: $\mu_X - \mu_Y = 0$
H_a: $\mu_X - \mu_Y > 0$ (the alternate hypothesis is that men score higher)

(a) [2 points] $\overline{X}$ and $\overline{Y}$ are normally distributed. How do we know?

(b) [4 points] What's the standard deviation of $\overline{X}$? What's the standard deviation of $\overline{Y}$?

(c) [4 points] If $\overline{X}$ and $\overline{Y}$ are normally distributed, then so is their difference, $\overline{X} - \overline{Y}$ (take my word for it), so we can test the null hypothesis by computing

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_{\overline{X}} - \mu_{\overline{Y}})}{\sigma_{\overline{X}-\overline{Y}}}$$

The denominator is the only part that's unfamiliar, but we might remember that the variance of a sum or difference is the *sum* of the variances: $\sigma^2_{\overline{X}-\overline{Y}} = \sigma^2_{\overline{X}} + \sigma^2_{\overline{Y}}$ [yes, there's a *plus* sign in there], so we can indeed compute the standard deviation in the denominator here.

Compute Z assuming the null hypothesis, $\mu_X - \mu_Y = 0$, is true.

(d) [4 points] Compare your value of Z to the critical Z for a one-sided test at the .05 significance level. Do you reject or accept the null hypothesis? What do you conclude about men's and women's quantitative GRE scores? [Write your answer on the back of one of these pages.]