Sociology 592 - Research Statistics I
Exam 1 Answer Key
September 30, 1996

Where appropriate, show your work - partial credit may be given. (On the other hand, don't waste a lot of time on excess verbiage.) Do not spend too much time on any one problem. It is legitimate (and probably essential) to refer to results that have previously been proven in class or homework, without re-proving them - for example, you wouldn't need to prove that $P(-1.96 \leq Z \leq 1.96) = .95$, since we have already shown that in class. Likewise, you are free to refer to anything that was demonstrated in the homework or handouts.

1. (5 points each, 20 points total). Indicate whether the following statements are true or false. If you think the statement is false, indicate how the statement could be corrected. For false statements, do not just say that you could substitute "not equals" for equals. For example, the statement $P(Z \leq 0) = .7$ is false. To make it correct, don't just say $P(Z \leq 0) <> .7$, instead say $P(Z \leq 0) = .5$ or $P(Z \leq .525) = .7$.

   a. If A and B are independent events, then $P(A|B) = P(B|A)$.

FALSE. A true statement is $P(A|B) = P(A)$.

   b. $V(X + 7) = 7^2 + V(X) + 2*COV(X, 7)$

FALSE. Recall that $V(X + a) = V(X)$, i.e. adding a constant to all cases does not affect the variance. So $V(X + 7) = V(X)$.

   c. The null and alterative hypotheses are

   H_0: $p = .7$
   H_A: $p > .7$

   In reality, $p = .7$. The researcher rejects the null. A type I error has been committed.

TRUE. The null has been rejected when it shouldn't be, which is a type I error.

   d. $P(Z \geq 1) = .84134474$

FALSE. True statements are $P(Z \leq 1) = .84134474$ or else $P(Z \geq -1) = .84134474$.

2. (10 points each, 30 points total) Answer <u>three</u> of the following. The answers to most of these are fairly straightforward, so do not spend a great deal of time on any one problem. NOTE: I will give up to 5 points extra credit for each additional problem you do correctly.

   a. It is January, 1997. Bill Clinton's second term is only days old, yet already speculation is rampant about who will succeed him. Naturally, much of the attention has come to be focused on one man: Robert Kerrey, the incredibly popular and talented senior senator from Nebraska. Kerrey's fortunes have always been closely linked to that of his favorite football team, the Nebraska Cornhuskers. This past season, the Huskers suffered an early setback but courageously struggled back and claimed the national championship with a convincing win over Florida State in the Sugar Bowl. But, can their luck continue? Kerrey estimates that, in any given season, Nebraska has an 80% chance of winning the national title. What is the probability that Nebraska will win at least three of the next four national championships? How likely is it that the unthinkable will happen and they will not win any championships at all in the next four years?

Note that the question says "at least three", hence we want to find P(3 national championships) + P(4 national championships). We also need the probability of 0 national championships for the last part of the question. Hence, we get

$$\binom{4}{3}.8^3.2^1 = 4*.1024 = .4096$$

$$\binom{4}{4}.8^4.2^0 = 1*.4096 = .4096$$

$$\binom{4}{0}.8^0.2^4 = 1*.0016 = .0016$$

So, the probability of at least three national championships is .8192, the probability of no national championships is .0016

You could also solve the problem by referring to the tables on the binomial distribution. You would say the probability of failure in any given year is point .2 and N = 4. "At least three national championships" is then equivalent to P(0 failures) + P(1 failure), no national championships = P(4 failures).

b. N = 18, $\bar{x}$ = 30, s^2 = 36. Construct the 90% confidence interval. If the null and alternative hypotheses are

H_0: $\mu = 19$
H_A: $\mu < 19$

should you reject or not reject the null? Explain why.

Looking at the T table, we see that for v = N -1 = 17 and 2Q = .10, the critical value is 1.74. Hence the confidence interval is

$$\bar{x} \pm 1.74 * \frac{s}{\sqrt{N}} = 30 \pm \frac{10.44}{\sqrt{18}} = 30 \pm 2.46$$

So, the confidence interval ranges from 27.54 to 32.46.

With regards to the second part of the question, as we'll soon see, you can use confidence intervals for hypothesis testing. HOWEVER, you should not do so when the alternative hypothesis is one tailed. While the sample mean (30) is far from the value specified in the null (19), it is even farther away from the values implied by the alternative (< 19). So, do not reject.

c. Here are the results from a previous cohort's first exam in statistics. Compute the mean and variance of the scores. There were 9 Students in the class.

Score	*Frequency*
68	**1**
77	**1**
86	**2**
90	**1**
94	**1**
98	**1**
101	**2**

Let's expand the table:

	Score	Freq	Freq * Score	Score²	Freq * Score²
	68	1	68	4624	4624
	77	1	77	5929	5929
	86	2	172	7396	14792
	90	1	90	8100	8100
	94	1	94	8836	8836
	98	1	98	9604	9604
	101	2	202	10201	20402
$\sum$			801		72287

So, the mean is 801/9 = 89. The variance is $72287/9 - 89^2 = 110.89$.

d. The University has decided that there are too many weak graduate students here. Therefore, students whose Grade Point Average (GPA) falls in the bottom 25% will be terminated from the University. If GPA ~N(3.4, $.25^2$), how high does your GPA have to be for you to stay in school?

By looking at the Z table, we see that F(.68) = .75, so F(-.68) = .25. So, you need a standardized score of -.68 or better to stay in school. When z = .68, x = 3.4 - (.68 * .25) = 3.23.

e. Following is an edited description of the Powerball Lottery, taken from the "Powerball Main Page" on the World Wide Web. In this Saturday's drawing, the estimated jackpot is $13 million. If you rush out and buy one ticket after the exam today, what is the probability that you will be the jackpot winner? (And if you do win, just remember who it was that gave you the idea!)

"**Game Description**. PowerBall is an on-line lottery game which is a combined large annuitized jackpot lotto game and a cash game. Every Wednesday and Saturday night at 10:59 p .m. Eastern Time, we draw five white balls out of a drum with 45 balls and one red ball out of a drum with 45 red balls. Numbers can be matched in any order. The jackpot (won by matching all five white balls and the red PowerBall) is an annuitized prize paid out over 20 years and will average around $30 million."

Count the total number of possible combinations. First you choose 5 balls from 45 balls, then you choose 1 ball from 45 balls. Hence, the number of combinations is

$$\binom{45}{5} * \binom{45}{1} = 1{,}221{,}759 * 45 = 54{,}979{,}155$$

So, you have about 1 chance in 55 million of winning the Powerball jackpot. Incidentally, note that the average jackpot is "only" about $30 million— the lottery distributes around half the money collected to prize winners and keeps the rest for itself.

3. (25 points) Numerous studies have found that blacks are more likely to have their home mortgage loan applications denied than are whites. However, critics of these studies maintain that higher denial rates for blacks are justified because blacks are more likely to default (i.e. fail to repay) on their loans than are whites.

To see whether higher default rates really do justify racial disparities in lending, a researcher gathered data on 900 loans made to whites and 100 loans made to blacks. For each loan, the researcher recorded the income of the home-owner (dichotomized into "low" or "high") and whether or not the home-owner defaulted on the home. She found that 31% of blacks, but only 25% of whites, defaulted on their loans. She also found that 60% of blacks were low income, compared to 30% of whites. Finally, she found that 4 high income blacks and 90 high income whites defaulted.

a. Complete the following table. Remember, there are 100 blacks and 900 whites.

	Black			**White**		
	Low Inc	High Inc	∑	Low Inc	High Inc	∑
Defaulted						
Did Not Default						
∑			100			900

Here is the completed table. Once you fill in the information given (indicated in *italics*), everything else follows pretty easily.

	Black			White		
	Low Inc	High Inc	∑	Low Inc	High Inc	∑
Defaulted	27	*4*	*31*	135	*90*	*225*
Did Not Default	33	36	69	135	540	675
∑	*60*	40	*100*	*270*	630	*900*

b. What percentage of those who defaulted were black?

31 of the 256 who defaulted were black, or 12.1%.

c. As these figures show, blacks are indeed more likely to default than are whites. However, blacks are also more likely to have low incomes. Suppose that blacks had the same income distribution as did whites, i.e. suppose that only 30% of blacks were low income instead of the current 60%. Suppose further it continued to be the case that blacks maintained their income-specific default rates. What percentage of blacks would then default? If a black and a white of comparable incomes were both to apply for a home mortgage loan, who would be the better "risk" for the bank, i.e. which one would be more likely to pay the loan off?

For blacks, P(Default|Low Income) = 27/60 = .45, P(Default|High Income) = 4/40 = .10. For whites, P(Low income) = .3, P(High Income) = .7. So, if blacks had the same income distribution as whites while maintaining their income-specific default rates, their overall default rate would be (.3 * .45) + (.7 * .10) = .205. That is, the black default rate would be lower than the white rate. Indeed, at each income level, the black default rate is lower than the white default rate; hence when incomes are equal the black is the better credit risk.

4. (25 points) Jury selection for Trial of the Century II is under way. O.J. Simpson's lawyers are concerned that the potential jurors in his civil suit are atypical of the general population — that is, they are more likely than most people to believe that Simpson murdered his wife.

National polls show that 70% of the general population thinks Simpson is guilty. One hundred potential jurors in the Simpson trial are surveyed; of those, 79 report they think Simpson is guilty.

Using the .05 level of significance, test whether the lawyers' fears appear to be justified. Be sure to indicate:

(a) The null and alternative hypotheses - and whether a one-tailed or two-tailed test is called for.

(b) The appropriate test statistic

(c) The critical region

(d) The computed value of the test statistic

(e) Your decision - should the null hypothesis be rejected or not be rejected? Why?

Step 1: The null and alternative hypotheses are

H0: E(X) = 70 (or p = .7)
HA: E(X) > 70 (or p > .7)

Note that the alternative is one tailed. O.J.'s lawyers specifically believe that potential jurors are exceptionally hostile toward their client.

Step 2: The appropriate test statistic is

$$z_c = \frac{Number\ who\ think\ guilty \pm .5 - NP_0}{\sqrt{Np_0 q_0}} = \frac{X \pm .5 - 70}{\sqrt{21}}$$

Step 3: Critical region — reject if the computed statistic is ≥ 1.65, otherwise do not reject. Equivalently, reject if X ≥ (1.65 * √21) + .5 + 70 = 78.06, i.e. reject if 79 or more think O.J. is guilty, otherwise do not reject.

Step 4: The computed value of the test statistic is

$$z_c = \frac{X \pm .5 - 70}{\sqrt{21}} = \frac{79 - .5 - 70}{\sqrt{21}} = \frac{8.5}{\sqrt{21}} = 1.85$$

Step 5: Decision — reject the null. The computed value of the test statistic falls outside the acceptance region. The lawyers' concerns seem to be warranted.