Sociology 592 - Research Statistics I
Sample Exam 1 Answer Key
September 30, 1994

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 $P(-1.96 \le Z \le 1.96) = .95$, 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 $P(Z \le 0) = .7$ is false. To make it correct, don't just say $P(Z \le 0) <> .7$, instead say $P(Z \le 0) = .5$ or $P(Z \le .525) = .7$.

A. If X and Y are independent, then V(XY) = V(X)V(Y).

False. See expectations rule #16. A true statement is that if X and Y are independent, V(X + Y) = V(X) + V(Y).

B. A waitress claims that she can predict the winner of college football games 60% of the time. Chauvinistic males at the neighborhood bar doubt that any woman could do so well. A random sample of 60 football games is chosen. The waitress picks the winner in 48 of those games. If $\alpha = .05$, the null hypothesis should be rejected.

False. The null and alternative hypotheses are

H_0: $p = .6$
H_A: $p \le .6$

Note that the chauvinistic males would hardly be able to claim victory if the waitress did better than she claims she will. But, that is exactly what happened here: she got 80% right, so the alternative hypothesis that $p \le .6$ is clearly not viable.

C. If a fair coin is tossed 20 times, the probability is .8238 that you will *not* get exactly 10 heads.

True. Note that P(Won't get exactly 10 heads) = 1 - P(exactly 10 heads). As Hayes, Table II shows, when N = 20 and p = .5, the probability of 10 heads is .1762, so 1 - P(exactly 10 heads) = .8238. To confirm that Hays got it right, note that

$$\binom{20}{10} p^{10} q^{20-10} = \frac{20!}{10!(20-10)!} .50^{10} .50^{20-10} = 184756 * .50^{20} = .176197$$

D. N = 16, $\bar{X} = 20$, $s^2 = 4$. The 95% confidence interval is $19.02 \le \mu \le 20.98$.

False. The statement would be true if you substituted σ^2 for s^2. Or, since the true variance is unknown, use a T distribution with 15 d.f. (for $\alpha = .05$, critical value is 2.131) and the confidence interval is

$$\bar{x} \pm (t_{\alpha/2,v} * s/\sqrt{N}), \text{ i.e.}$$

$$20 - (2.131 * 2/4) \le \mu \le 20 + (2.131 * 2/4), \text{ i.e. } 18.9345 \le \mu \le 21.0655$$

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 October 29, 1994. Despite losing injured quarterback Tommie Frazier for the season, the Nebraska Cornhuskers have continued to devastate their opponents and solidify their hold on the #1 ranking in college football. Today, however, they face their greatest challenge: #2 ranked Colorado. While many feel that the Huskers have the most talented and best-coached team in college football history, everyone agrees that Colorado has the luckiest.

After careful analysis, coach Tom Osborne of Nebraska has determined that:

✓ There is a 76% chance that Nebraska will win the game and win the national championship

✓ There is a 6% chance Nebraska will not win the game and yet still win the national championship.

✓ There is an 80% chance that Nebraska will win this game

If Nebraska wins this game, what is the probability that it will win the national championship? If Nebraska does not win this game, what is the probability that it will still go on to win the national championship?

We are told:

P(Win Game ∩ Win National championship) = .76

P(Lose Game ∩ Win national championship) = .06

P(Win Game) = .80

We are asked to find

P(Win National Championship | Win Game) =

<u>P(Win Game ∩ Win National championship)</u> = .76/.80 = .95

P(Win Game)

P(Win National Championship | Lose Game) =

<u>P(Lose Game ∩ Win National championship)</u> = .06/.20 = .30

P(Lose Game)

b. A company believes that its products will appeal to those with incomes between $12,000 and $30,000. If income ~ N($30,000,$8000²), what percentage of the population will be interested in the company's products?

First, convert $12,000 and $30,000 to the corresponding Z scores. If X = 12,000, Z = (12,000 - 30,000)/8,000 = -2.25. When X = 30,000, Z = (30,000 - 30,000)/8,000 = 0. The area between 0 and -2.25 = F(0) - F(-2.25) = F(0) - 1 + F(2.25) = .5 - 1 + .9877755 = .4877755, i.e. almost 49% of the population will be interested in the company's products.

c. Here are the results from a previous cohort's first exam in statistics. Compute the mean and variance of the scores. As the frequencies show, there were 10 students in the class.

Score	Frequency
56.0	1
76.0	1
90.0	1
93.0	4
96.0	1
97.0	1
100.0	1

Let us expand the table as follows:

Score (Xi)	Frequency (fi)	Xi * fi	Xi^2	Xi^2 * fi
56	1	56	3136	3136
76	1	76	5776	5776
90	1	90	8100	8100
93	4	372	8649	34596
96	1	96	9216	9216
97	1	97	9409	9409
100	1	100	10000	10000
Σ	10	887		80,233

Hence, $\mu = 887/10 = 88.7$, $\sigma^2 = 80{,}233/10 - 88.7^2 = 155.61$, $\sigma = 12.474$

d. A polling firm reports that in its sample, 60% of the American public approved of former President Carter's role in the Haiti crisis. The firm further reports that the (approximate) 95% confidence interval for Carter's approval is $.5608 \le p \le .6392$. What was the sample size used in the study? [HINT: Write out the formula for the confidence interval, plug in all the values you can, and then solve for N.]

This falls under confidence intervals, case II, binomial parameter p. The approximate confidence interval is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}},\ i.e.$$

$$.6 - 1.96\sqrt{\frac{.6 * .4}{N}} \le p \le .6 + 1.96\sqrt{\frac{.6 * .4}{N}},\ i.e.,$$

$$.5608 \le p \le .6392$$

This implies

$.6392 = .60 + 1.96\sqrt{\frac{.24}{N}}$	Upper end of the c.i.
$.0392 = 1.96\sqrt{\frac{.24}{N}}$	Subtract .6 from both sides
$.02 = \sqrt{\frac{.24}{N}}$	Divide both sides by 1.96
$.0004 = \frac{.24}{N}$	Square both sides
$\frac{.0004}{.24} = \frac{1}{N}$	Divide both sides by .24
$\frac{.24}{.0004} = N = 600$	Take reciprocals

NOTE: Researchers often need to know in advance how large their sample must be in order to guarantee that the confidence interval is not too large. Hence, calculations similar to the above are done even before any data are collected, so the researcher knows how many cases need to be collected.

e. Prove that if $P(A) > P(B) > 0$, then $P(A \mid B) \geq P(B \mid A)$. [HINT: Write out the formulas for the conditional probabilities.]

P(A) > P(B) ==> 1/P(B) > 1/P(A)	Take reciprocals
==> $P(A \cap B)/P(B) \geq P(A \cap B)/P(A)$	Multiply both sides by $P(A \cap B)$. Note that, when A and B are mutually exclusive events, $P(A \cap B) = 0$ and both sides of the inequality equal 0, otherwise the left side is greater than the right side
==>$P(A \mid B) \geq P(B \mid A)$	Formula for conditional probability

3. (25 points) Jury selection for the trial of the century has begun. Judge Lance Ito is afraid that massive media coverage is going to make it very difficult to find jurors who have not already made up their mind about O. J. Simpson. He also suspects that sports fans will be more likely than others to have formed an opinion. He has therefore commissioned a study to see what relationships, if any, there are between interest in sports, the amount of TV a person watches, and whether or not a person has formed an opinion on the Simpson case. A sample of 200 potential jurors is drawn. One hundred report that they watch only a little TV, while the other 100 report that they watch TV a lot.

For those who claim they only watch TV a little bit:
✓ 52% say they have already formed an opinion on the Simpson case
✓ 70% report that they are *not* sports fans
✓ 80% of the sports fans say they have already formed an opinion on the case

For those who claim they watch TV a lot:
✓ 61% say they have already formed an opinion on the Simpson case
✓ 30% report that they are *not* sports fans
✓ 70% of the sports fans say they have already formed an opinion on the case

a. Complete the following table. Remember that, as is already noted in the table, 100 respondents say they only watch a little TV while 100 report that they watch a lot.

	Only watch a little TV			Watch TV a lot		
Sports interest/opinion	Formed Opinion on O.J.	No opinion on O.J.	Σ	Formed opinion on O.J.	No opinion on O.J.	Σ
Sports fan						
Not a sports fan						
Σ			100			100

For those who watch a little TV, we are told
✓ P(Form opinion) = .52
✓ P(Not sports fan) = .70
✓ P(Form opinion | Sports fan) = .80

Hence, we can easily determine that
P(No opinion) = .48, P(Sports fan) = .30,
$P(\text{Sports fan} \cap \text{Form opinion})$ = P(Sports fan) * P(Form opinion | Sports fan)
= .3 * .8 = .24

Similarly, for those who watch a lot of TV, we are told
✓ P(Form opinion) = .61
✓ P(Not sports fan) = .30
✓ P(Form opinion | Sports fan) = .70

Hence, we can easily determine that

P(No opinion) = .39, P(Sports fan) = .70,

P(Sports fan ∩ Form opinion) = P(Sports fan) * P(Form opinion | Sports fan)

= .7 * .7 = .49

Once you fill in these numbers, the rest of the table falls easily into place:

	Only watch a little TV			Watch TV a lot		
Sports interest/opinion	Formed Opinion on O.J.	No opinion on O.J.	Σ	Formed opinion on O.J.	No opinion on O.J.	Σ
Sports fan	24	6	30	49	21	70
Not a sports fan	28	42	70	12	18	30
Σ	52	48	100	61	39	100

b. Of all those who have formed an opinion on the case, what percent are sports fans?

Note that 113 have formed an opinion (52 who only watch a little TV, 61 who watch a lot). Of these, 73 are sports fans (24 who only watch a little TV, 49 who watch a lot). Hence, 73/113 = 64.6% of those who have formed an opinion are sports fans.

c. As the judge feared, those who watch TV a lot are more likely to have formed an opinion on the case. However, those who watch a lot of TV are also more likely to be sports fans, and sports fans are more likely to have formed an opinion on the case. This makes the judge wonder what role television is having in leading people to form opinions. Suppose that those who watch TV a lot had the same percentage of sports fans as do those who do not watch TV very much. Suppose further that those who watch a lot of TV maintained their sports interest-specific rates of forming opinions on the case, e.g. 70% of the sports fans who watch a lot of TV still formed an opinion on the case. What percentage of those who watch a lot of TV would then have formed an opinion on the Simpson case? Based on this result, what do you think is more important — being a sports fan or watching a lot of TV?

For those who watch TV a little, P(Sports fan) = .3, P(Not a fan) = .7.

For those who watch a lot of TV,

P(Form Opinion | Sports fan) = .7, P(Form Opinion | Not a fan) = .4

Hence, if those who watch a lot of TV had the same mix of fans and non-fans as those who watch a little TV, while maintaining their interest-specific rates of forming opinions, it would be the case that

P(Form opinion) =

$P(\text{Sports Fan})^{\text{Little TV}} * P(\text{Form opinion} \mid \text{Sports fan})^{\text{Lot of TV}}$ +

$P(\text{Not Fan})^{\text{Little TV}} * P(\text{Form Opinion} \mid \text{Not a fan})^{\text{Lot of TV}}$

= .3 * .7 + .7 * .4 = .49

Note that, under these conditions, those who watch a lot of TV would actually be slightly less opinionated than those who currently watch a little TV. This suggests that being a sports fan has more to do with forming an opinion than does how much you watch TV.

4. (25 points) Last Resort Savings and Loan has been stung by criticisms that it does not do as much as it should for low income areas of town. Statistics show that, for all lending institutions countywide, 19% of home mortgage loan applications from low income areas get denied. Last Resort is confident that its denial rate for low income areas is no worse (i.e., no higher) than that. It therefore draws a random sample of 133 home mortgage loan applications it has received from low income areas. It discovers that 37 of these applications were denied.

Test Last Resort's claim at the .01 level of significance. 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?

NOTE: You will receive partial credit if you can at least tell me, if Last Resort is correct, what is the probability that a random sample of 133 applications would contain 37 or more rejections?

Step 1. The null and alternative hypotheses are

$H_{0:}$ E(X) = 25.27 (or p = .19)
H_A: E(X) > 25.27 (or p > .19)

A one tailed alternative is appropriate, since Last Resort claims its rejection rate is no worse (i.e. no higher) than the countywide average.

Step 2. The appropriate test statistic is

$$Z = \frac{\#\ denied\ \pm\ CC\ -\ Np_0}{\sqrt{Np_0q_0}} = \frac{\#\ denied\ \pm\ CC\ -\ 25.27}{4.524234742}$$

To make the correction for continuity, we will add .5 if X < 25.27 and subtract .5 if X > 25.27.

Step 3. We will reject the null hypothesis if $z_c > 2.33$ (remember, we are using the .01 level of signifcance); alternatively, reject the null if X > 2.33 * sqrt(133*.19*.81) + 25.27 + .5, i.e. X > 36.31

Step 4. The computed value of the test statistic is

$$Z = \frac{\#\ denied\ \pm\ CC\ -\ Np_0}{\sqrt{Np_0q_0}} = \frac{37\ -\ .5\ -\ 25.27}{4.524234742} = 2.48$$

Step 5. Reject the null hypothesis. The test statistic of 2.48 is greater than the critical value of 2.33; or equivalently, the observed number of denials, 37, is greater than the critical value of 36.31.

Since F(2.48) = .9934309, there is a 0.66% chance (i.e. about 1 in 150) that the S & L could be correct and the sample still contain this many denials.