Sociology 592 - Research Statistics I
Exam 1 Answer Key
September 25, 1991

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. The population is 40% Protestant, 30% Catholic, and 30% Other. Twenty-five percent (25%) of the Protestants and 20% of the Catholics smoke, while 10% of the population consists of Others who Smoke. This means that

$$P(\text{Smoke}) = .40 * .25 + .30 * .20 + .30 * .10 = .19$$

SOLUTION. False. Note that you are told the *conditional* probability of smoking for Protestants and Catholics, and the *joint* probability of being both an Other and Smoking, i.e. you are told $P(\text{Smokes} \mid \text{Protestant}) = .25$, $P(\text{Smokes} \mid \text{Catholic}) = .20$, and $P(\text{Smokes} \cap \text{Other}) = .10$. The statement would be correct if you changed it to read "10% of Others smoke" or else changed the formula to

$$P(\text{Smoke}) = .40 * .25 + .30 * .20 + .10 = .26.$$

That is, 10% of the population are Protestant smokers, 6% are Catholic smokers, and 10% are Other smokers, giving us 26% of the population smoking.

B. A fair coin is tossed 12 times. The probability of getting 5 to 7 heads is .6124.

SOLUTION. True. You want to find $P(5 \le X \le 7)$, i.e. $P(5) + P(6) + P(7)$. According to Hayes, p. 928, for N = 12 and p = .5, $P(5) + P(6) + P(7) = .1934 + .2256 + .1934 = .6124$.

C. A defense contractor claims that only 20% of its weapons are defective. Military critics believe that the reliability of the weapons is overstated. A random sample of 100 weapons finds 11 that are defective. If $\alpha = .05$, the null hypothesis should be rejected.

SOLUTION. False. The null and alternative hypotheses are

H_0: $p = .20$
H_A: $p > .20$

Since, in the sample, $\hat{p} = .11$, the alternative hypothesis is clearly implausible. Instead of being less reliable than claimed, it appears that the weapons may actually be better than expected.

D. $P(Z \le -.90) = .8159$

SOLUTION. False. True statements would be

$P(Z \leq .90) = .8159$, $P(Z \geq -.90) = .8159$, or $P(Z \leq -.90) = 1 - .8159 = .1841$.

2. (10 points each, 30 points total) Answer three 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 election day 1992. After once trailing by 40 points in the polls, Democratic Presidential nominee Robert Kerrey of Nebraska and his running mate Jerry Brown of California have turned this into one of the closest elections in history. The electorate has polarized into 6 distinct and non-overlapping groups. Here is a description of each, and how they voted: [HINT: while this question is amusingly written and offers subtle political and social satire, all the stuff you really need to know is in boldface]

Hollywood star gazers (10% of the population): This group votes for whichever candidate will bring the most Hollywood glamour and dazzle to the White House. Traditionally a Republican stronghold, the movie-star looks of Kerrey and Brown have helped keep this group undecided for most of the campaign. For them, the turning point comes 3 days before the election, when it is announced that Kerrey has eloped with actress Debra Winger. **Seventy percent (70%) vote Democratic.**

Willy Horton Haters (10% of the population): This group believes that the nation's top priority, indeed its only priority, is making sure that convict Willy Horton never gets out of jail. As in 1988, **100% vote Republican.**

Football Fanatics (12% of the population): These voters believe that a President is only as good as the College football team behind him. Although Republican George Bush claims Texas and 27 other states as his home, not one of them has a team that can compare with Kerrey's Nebraska Cornhuskers, who in 1992 are storming towards their first national championship in over 20 years. **Seventy-five percent (75%) vote Democratic.**

Impoverished graduate students (4% of the population): This group originally intended to vote for whichever candidate would do the most to raise graduate student stipends. When both candidates promised to eliminate all forms of graduate student support from the federal budget, most graduate students voted on the basis of their second biggest concern: football. **Seventy-five percent (75%) vote Democratic.**

Serious issue voters (30% of the population): This group carefully examines the issues raised by each candidate and votes for the person they feel offers the best and most thoughtful program for America. Unfortunately, extended discussion of Willy Horton, flag burning, Debra Winger, Linda Ronstadt, Governor Moonbeam, Dan Quayle's high school grades, and other similar matters have once again made it impossible for the candidates to go over issues this year. Hence, **Serious issue voters split evenly between the Republicans and the Democrats.**

Not-a-dimes-worth-of-difference-between-them voters (34% of the population): As their name implies, this group believes there is not a dime's worth of difference between either Presidential candidate, hence they toss a nickel to determine their vote. **Fifty percent (50%) vote Democratic.**

Thus, with _____% of the vote, the next President of the United States is _________.

SOLUTION. We want to find P(Democrat) (or else P(Republican)). Use the formula for marginal probability. Let E_1 = Hollywood Star Gazer, E_2 = Willy Horton Haters, E_3 = Football Fanatics, E_4 = Impoverished graduate students, E_5 = Serious issue voters, E_6 = Not-a-dimes-worth-of-difference. We are told

$P(E_1) = .10$, $P(\text{Democrat} \mid E_1) = .70$,

$P(E_2) = .10$, $P(\text{Democrat} \mid E_2) = 0$,

$P(E_3) = .12$, $P(\text{Democrat} \mid E_3) = .75$,

$P(E_4) = .04$, $P(\text{Democrat} \mid E_4) = .75$,

$P(E_5)$ = .30, P(Democrat | E_5) = .50,
$P(E_6)$ = .34, P(Democrat | E_6) = .50.

Hence, according to the formula for marginal probability,

$$P(Democrat) = \sum P(E_i)\ P(Democrat \mid E_i) =$$

$$(.10 * .70) + (.10 * 0) + (.12 * .75) + (.04 * .75) + (.30 * .50) + (.34 * .50) = .51$$

Ergo, with 51% of the vote, the next President of the United States is Robert Kerrey from the great state of Nebraska!

b. (For those who like card problems) There is a card in a hat. It is either the ace of spades or the king of spades, with equal probability. You take another identical ace of spades and throw it into the hat. You then choose a card at random from the hat. You see it is an ace. What are the odds the original card in the hat was an ace? [HINT: The answer to this problem is in your course packet somewhere - but you still have to provide a formal proof!]

SOLUTION. 2/3. Here are the possibilities.
(1) The original card in the hat was an ace. You threw in an ace and then picked the *original* ace.
(2) The original card in the hat was an ace. You threw in an ace and then picked the *second* ace.
(3) The original card was a king; you threw in an ace. You then picked the ace.
In two of three cases, the original card was an ace. QED.

Or, if you prefer to use formulas: Note that, prior to drawing a card, there is a fourth possibility:
(4) The original card was a king; you threw in an ace. You then draw the King.

Each of these 4 possibilities is equally likely. The question asks us to find
P(Original was ace | Ace is drawn). Note that this is equal to
P(Original was an ace ∩ Ace was drawn)/ P(Ace is drawn).

All except the fourth possibility involve drawing an ace, so P(Ace is drawn) = .75. And, in two of the four possible outcomes (1 and 2) the original is an ace and an ace is drawn, hence P(Original was ace ∩ Ace is drawn) = .5. Thus,
P(Original was an ace ∩ Ace was drawn)/ P(Ace is drawn) = .50/.75 = 2/3. QED.

c. If Income ~ N($20,000, $$5000^2$), how low does your income have to be for you to be among the poorest 25% of the population?

SOLUTION. Let X = Income. We want to find x such that
$P(X \le x) = .25$.
This is too hard to solve as it stands - so instead compute
Z = (X - $20,000)/$5,000 (NOTE: Z ~ N(0,1)
and find z for the problem
$P(Z \le z) = .25$.
Note that $P(Z \le z) = F(z) = 1 - F(-z) = .25 \implies F(-z) = .75$. Looking at Hayes, p. 924, we can determine that F(.675) = .75 ==> z = -.675. To get back to the equivalent x score, we compute

x = (-.675 * $5,000) + $20,000 = $16,625.

d. Prove Expectations rule #11 (course packet, p. 26):
If X and Y are independent, COV(X,Y) = 0.
You can use any of the first 10 rules in your proof.

SOLUTION. According to Expectations rule # 9, if X and Y are independent, then

E(XY) = E(X)E(Y).

And, according to rule # 10,

COV(X,Y) = E(XY) - E(X)E(Y).

Ergo, when X and Y are independent,

COV(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0. QED.

e. Here are the results from last semester's first exam in statistics. Compute the mean and variance of the scores. As the frequencies show, there were 13 students in the course.

Score (X_i)	Frequency (f_i)
31	1
65	1
86	1
90	1
94	3
95	2
96	1
98	1
100	1
104	1

SOLUTION. Let us expand the table as follows:

Score (X_i)	Frequency (f_i)	X_i * f_i	X_i^2	X_i^2 * f_i
31	1	31	961	961
65	1	65	4,225	4,225
86	1	86	7,396	7,396
90	1	90	8,100	8,100
94	3	282	8,836	26,508
95	2	190	9,025	18,050
96	1	96	9,216	9,216
98	1	98	9,604	9,604
100	1	100	10,000	10,000
104	1	104	10,816	10,816
Σ	13	1,142		104,876

Hence, μ = 1,142/13 = 87.846, σ^2 = 104,876/13 - 87.846^2 = 350.46, σ = 18.72.

The end of this answer key shows how you could address this problem using SPSS-PC.

3. (25 points) According to the October 1991 issue of Working Mother Magazine,

> During the recession, corporations laid off tens of thousands of workers and many restructured their businesses. But the policies, programs and benefits designed to support working parents not only escaped the ax, they have literally exploded.

A professor decides to study whether the fallout from this explosion has worked its way into academia yet. She draws a random sample of 500 universities, 100 of them private and 400 public. She classifies each university on two variables: (1) "family-friendly" or "not family-friendly" - based on whether the University offers such benefits as on-site child care, paid parental leave, job-sharing, etc.; and (2) "small endowment" vs. "large endowment" - based on the size of the school's endowment relative to the size of the student body.

She finds that 30% of the private schools and 75% of the public schools have small endowments. For Private schools, 1/3 of those with small endowments and 1/2 of those with large endowments are "family friendly." For Public schools, 5/6 of those with small endowments are not family friendly, while 40% of those with large endowments are.

a. Complete the following table. Remember that, as is already noted in the table, there are 100 private universities and 400 public.

	Private			Public		
Family friendly/Endowment	Small	Large	Σ	Small	Large	Σ
"Family Friendly"						
"Not Family Friendly"						
Σ			100			400

SOLUTION. We are told P(Small endowment | Private school) = 30%. Since there are 100 Private schools, this means that 30 have small endowments and the other 70 have large endowments. Further, for Private schools,

P(family friendly | small endowment) = 1/3, which implies that

P(family friendly ∩ small endowment)

= P(small endowment) * P(family friendly | small endowment) = .30 * 1/3 = .10, which means that 10 (.10 * 100) of the small endowment private schools are family friendly while the other 20 are not.

Also for private schools, P(family friendly | large endowment) = 1/2, implying

P(family friendly ∩ large endowment) =

= P(large endowment) * P(family friendly | large endowment) = .70 * 1/2 = .35, which means that 35 (.35 * 100) of the large endowment private schools are family friendly while the other 35 are not.

For Public schools, we are told P(Small endowment | Public school) = .75. Since there are 400 Public schools, this means that 300 have small endowments while the other 100 have large endowments. Further, for Public schools,

P(family friendly | small endowment) = 1/6, which implies that

P(family friendly ∩ small endowment)

= P(small endowment) * P(family friendly | small endowment) = .75 * 1/6 = .125, which means that 50 (.125 * 400) of the small endowment public schools are family friendly while the other 250 are not.

Also for public schools, P(family friendly | large endowment) = .4, implying

P(family friendly ∩ large endowment) =

= P(large endowment) * P(family friendly | large endowment) = .25 * .4 = .10, which means that 40 (.10 * 400) of the large endowment public schools are family friendly while the other 60 are not.

Hence, the completed table is

	Private			Public		
Family friendly/Endowment	Small	Large	Σ	Small	Large	Σ
"Family Friendly"	10	35	45	50	40	90
"Not Family Friendly"	20	35	55	250	60	310
Σ	30	70	100	300	100	400

b. What percentage of the "family friendly" universities are private?

SOLUTION. We are asked P(Private | Family Friendly). As the table shows, there are 135 "family friendly" universities (45 private and 90 public), 45 of which are private. Ergo, 45/135 = 1/3 of the "family friendly" universities are private.

c. As these figures show, Public universities tend to be less family-friendly than Private universities. This may reflect the greater concern Private schools have for their employees. However, Public universities also tend to have smaller endowments than do Private schools, which may hamper their ability to afford benefits. Suppose that public universities were as wealthy as private universities, i.e. 70% (instead of the current 25%) had large endowments. Suppose further that it continued to be the case that Public Universities maintained their endowment-specific family-friendly rates, i.e. 1/6 of the schools with small endowments and 40% of the schools with large endowments were "family friendly". What percentage of Public Universities would then be "family-friendly?"

SOLUTION. Use the Marginal probability rules. We want to find P(family friendly) under the hypothetical situation in which Public Schools are as wealthy as Private schools while maintaining their wealth-specific rates of family friendliness.

For Private schools, P(Small endowment) = .3, P(Large endowment) = .7.
For Public schools, P(family friendly | small endowment) = 1/6,
P(family friendly | large endowment) = .4.

Hence, in this hypothetical situation,

P(Family friendly) = P(small endowment) * P(family friendly | small endowment) + P(large endowment) * P(family friendly | large endowment)
= .3 * 1/6 + .7 * .4 = .33.

Ergo, if Public schools were as wealthy as Private schools, 33% of them would be family friendly, compared to the actual figure of 22.5% (90 out of 400). Note that this would still be less than the family-friendly rate of 45% in the Private schools, suggesting that something besides wealth is also responsible for the Public/Private differences.

4. (25 points) Notre Dame takes great pride in the health and fitness of its students. According to the University, 1/3 of our undergraduates were captains of a high school sports team. A professor, noting how tired and sleepy students seem to be during his classes, suspects that the University must be exaggerating its claims. To check his suspicions, he draws a random sample of 72 students, and finds that 17 of them were captains of a high school sports team. Test the University's claim at the .05 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 the University is correct, what is the probability that a random sample of 72 students would contain 17 or fewer captains?

SOLUTION. Let X = the number of captains found in the sample. Note that X has a binomial distribution, i.e. it is the sum of independent and identically distributed Bernoulli trials. If the null hypothesis is true, $E(X) = Np_0 = 72 * 1/3 = 24$, $\sigma_X^2 = Np_0q_0 = 72 * 1/3 * 2/3 = 16$, $\sigma_X = 4$. Since

$Np_0q_0 > 3$, it is probably safe for us to assume that X is distributed approximately $N(24, 4^2)$. Alternatively, we can work with $\hat{p}$, noting that $\hat{p}$ is distributed approximately $N(1/3, 4^2/72^2)$, i.e. $N(1/3, [1/18]^2)$.

(a) The null and alternative hypotheses are

H_0: E(X) = 24 (or p = 1/3)

H_A: E(X) < 24 (or p < 1/3)

Note that a 1-tailed alternative is clearly called for, since the professor believes students are <u>not</u> in as good of shape as the administration contends.

(b) The appropriate test statistic is

$$Z = \frac{\#\ of\ captains \pm CC - Np_0}{\sqrt{Np_0q_0}} = \frac{\#\ of\ captains \pm CC - 24}{4}$$

To make the correction for continuity, we will add .5 if X < 24 and subtract .5 if X > 24.

If you prefer to use $\hat{p}$,

$$Z = \frac{\hat{p} \pm .5/N - p_0}{\sqrt{\frac{p_0q_0}{N}}} = \frac{\hat{p} \pm .5/72 - \frac{1}{3}}{\frac{1}{18}}$$

(c) Using α = .05, we will reject the null hypothesis if Z < -1.65. Equivalently, we will reject the null hypothesis if the number of captains is less than

Z = -1.65 * 4 + 24 - .5 = 16.9, i.e. we will reject the null if there are 16 or fewer captains.

(d) The computed value of the test statistic is

$$Z = \frac{\#\ of\ captains \pm CC - 24}{4} = \frac{17 + .5 - 24}{4} = -1.625$$

Or, using $\hat{p}$,

$$Z = \frac{\hat{p} \pm \frac{.5}{72} - \frac{1}{3}}{\frac{1}{18}} = \frac{\frac{17}{72} + \frac{.5}{72} - \frac{24}{72}}{\frac{1}{18}} = \frac{\frac{-6.5}{72}}{\frac{1}{18}} = \frac{-117}{72} = -1.625$$

(e) Do NOT reject the null hypothesis. There are just barely enough captains in the sample to keep us from rejecting the University's claim. (Incidentally, note that if you fail to make the correction for continuity, the test statistic = -1.75, which does fall in the critical region. Hence, in this case, failure to make the correction leads to an erroneous conclusion.)

Also, note that F(-1.625) = 1 - F(1.625) = 1 - .9479 = .0521. Hence, the odds are slightly better than 1 in 20 that the University could be correct and the sample contain 17 or fewer captains.

The following shows how problem 2E could be addressed using SPSS-PC. As noted in the comments to the program, SPSS-PC uses the wrong formulas for the variance and standard deviation (since it treats this as a sample rather than a population) but the problem is easily corrected by hand.

SPSS-PC Control Cards:

```
Set lis = 'test1.lis'.
* Exam 1, Problem 2E.  The variances and standard deviations that
* SPSS reports will be slightly off, since SPSS uses the formula for
* the Sample Variance (which uses N - 1 in the denominator) when the
* formula for the Population Variance should be used (since all students
* from the class are included, i.e. there is no sample).  To get the
* correct variance, multiply Spss's variance by 12/13, then take the
* square root of this to get the correct standard deviation.
DATA LIST FREE / Xi Fi.
BEGIN DATA.
 31  1
 65  1
 86  1
 90  1
 94  3
 95  2
 96  1
 98  1
100  1
104  1
END DATA.
WEIGHT BY Fi.
FREQUENCIES /VARIABLES Xi /STATISTICS DEFAULT VARIANCE.
FINISH.
```

Selected output:

```
XI
                                                      Valid     Cum
Value Label                 Value  Frequency  Percent  Percent  Percent

                            31.00         1       7.7      7.7      7.7
                            65.00         1       7.7      7.7     15.4
                            86.00         1       7.7      7.7     23.1
                            90.00         1       7.7      7.7     30.8
                            94.00         3      23.1     23.1     53.8
                            95.00         2      15.4     15.4     69.2
                            96.00         1       7.7      7.7     76.9
                            98.00         1       7.7      7.7     84.6
                           100.00         1       7.7      7.7     92.3
                           104.00         1       7.7      7.7    100.0
                                    -------  -------  -------
                            Total        13    100.0    100.0

Mean         87.846      Std dev      19.484      Variance    379.641
Minimum      31.000      Maximum     104.000

Valid cases      13      Missing cases      0
```

Note that 379.641 * 12/13 = 350.44, which is the correct value for the population variance.