Sociology 592 - Research Statistics I
Exam 1 Answer Key
October 1, 1993

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 $P(Z \le 0) \ne .7$, instead say $P(Z \le 0) = .5$ or $P(Z \le .525) = .7$.

A. In a population of size 20,

$$\sum_{i=1}^{20} X_i = 80, \quad \sum_{i=1}^{20} X_i^2 = 362$$

This population has a variance of 18.1.

FALSE. $\sigma^2 = E(X^2) - E(X)^2 = 362/20 - (80/20)^2 = 18.1 - 16 = 2.1$
Alternatively, the statement would be true if instead $\Sigma X = 0$.

B. IQ ~ $N(100, 20^2)$. To be among the smartest 5% of the population, you have to have an IQ of 139.2 or higher.

FALSE. Note that, to be in the top 5%, you need a z-score of 1.65 (not 1.96). This corresponds to an IQ score of 20*1.65 + 100 = 133.

C. Five women are pregnant. The probability is about .0625, or 1/16, that all of their children will be of the same sex. (Assume that the biological odds of having a boy or a girl are equal.)

TRUE. Note that the children will all be of the same sex if they are either all boys or all girls. P(All boys) = $.5^5$ = 1/32, P(All Girls) = 1/32, hence P(All Girls $\cup$ All Girls) = 1/16 = .0625.

D. The null and alternative hypotheses are:

H_0: $p = .60$
H_A: $p > .60$

A sample of 50 cases yields 20 successes. If $\alpha = .01$, the null hypothesis should not be rejected.

TRUE. Only 40% of the observed cases are successes. Since the alternative hypothesis claims that the number of successes is > .60, there is obviously no evidence to support it. Note that there is absolutely no need to bother computing the value of the test statistic here (and that it wouldn't matter what α equalled.)

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. $\bar{X} = 0,\ N = 25$. Determine the 95% confidence interval when

- $\hat{\sigma} = 5$
- $\sigma = 5$

When σ is not known, use the T distribution. For a T with 25 - 1 = 24 d.f., the critical value is 2.064. Hence, the confidence interval is

$$\bar{X} \pm 2.064 * \frac{\hat{\sigma}}{\sqrt{N}},\ i.e.$$

$$-2.064 \le \mu \le 2.064$$

When σ is known, use the Z distribution. The critical value for z is 1.96. Hence,

$$\bar{X} \pm 1.96 * \frac{\sigma}{\sqrt{N}},\ i.e.$$

$$-1.96 \le \mu \le 1.96$$

b. It is election day, 1996. In summer, Ross Perot's 3rd-party candidacy was considered all but dead; but, thanks to Perot's enormously popular decision to choose Nebraska Senator Robert Kerrey as his running mate, the pollsters all agree that the race is too close to call. Based on the following information, determine who will win: Democrats Bill Clinton and Al Gore, Independents Ross Perot and Bob Kerrey, or Republicans Pat Buchanan and Rush Limbaugh. Be sure to state what percentage of the total electoral vote each ticket receives.

- ✓ 40% of the electoral votes are in the East, 60% are in the West
- ✓ 60% of the Eastern electoral votes go to the Democrats, while the rest are split evenly between the other two tickets
- ✓ Republicans get 5/12 (41.667%) of the Western electoral votes
- ✓ 27% of the total electoral vote consists of Western votes for Perot and Kerrey.
- ✓ Meanwhile, in other news, the Democrats easily maintain their lead in the House of Representatives, Republicans recapture control of the Senate, and the Nebraska Cornhuskers win yet another national football championship.

You are told:

✓ P(East) = .40, P(West) = .60
✓ P(Democrat | East) = .60, P(Republican | East) = .20, P(Indep | East) = .20
✓ P(Republican | West) = 5/12 = 41.67%
✓ P(Indep ∩ West) = .27

From this, you can easily calculate

P (Democrat ∩ East) = P(Democrat | East) * P (East) = .6 * .4 = .24
P(Republican ∩ East) = P(Republican | East) * P (East) = .2 * .4 = .08
P(Independent ∩ East) = P(Independent | East) * P (East) = .2 * .4 = .08
P(Republican ∩ West) = P(Republican | West) * P (West) = 5/12 * .6 = .25
P(Democrat ∩ West) = P(West) - P(Indep ∩ West) - P(Republican ∩ West) =
.60 - .27 - .25 = .08

Hence,
P(Democrat) = P(Democrat $\cap$ East) + P(Democrat $\cap$ West) = .24 + .08 = .32
P(Republican) = P(Republican $\cap$ East) + P(Republican $\cap$ West) = .08 + .25 = .33
P(Independent) = P(Independent $\cap$ East) + P(Independent $\cap$ West) = .08 + .27 = .35

Thus, Perot and Kerry capture a *plurality* of the electoral votes. Note, however, that in the American system, if no candidate receives a *majority* of the electoral (i.e. electoral college) votes, Congress chooses the President and Vice-President. Hence, Clinton is easily re-elected President by the Democratic-controlled House, while Limbaugh is the pick of the Republican Senate. However, the populace, outraged that the vote leaders have been shut out, rises up and overthrows the established government. A constitutional monarchy is established, with Perot as King and Kerrey as Prime Minister. The nation prospers.

c. Five senators are undecided about Bill Clinton's health plan. Clinton needs at least four of their votes. For each Senator, there is a 70% chance they will vote with Clinton, and each senator's vote is independent of the others. What is the probability that Clinton's plan will pass?

Clinton wins if he gets 4 *or* 5 votes. Equivalently, he wins if only 0 or 1 votes go against him. Note that the probability of a Senator not voting for Clinton is .30. Hence, if we Look at Hays, Table II, p. 927, we see that, for N = 5 and p = .30, P(0) = .1681 and P(1) = .3602. Hence, P(0 or 1 votes against) = .1681 + .3602 = .5283.

d. [Slightly hard] A population of families has an unknown mean income μ; the standard deviation of these incomes is known to be $1,000. How large a random sample would be needed to determine the mean income if it is desired that the probability of a sampling error of more than $50 be less than 5 percent? [HINT: Usually you are told N and are asked to find the confidence interval. Here, you have been told the range of the confidence interval and are asked to find N]

The confidence interval is

$$\bar{X} \pm 1.96 * \frac{\sigma}{\sqrt{N}} = \bar{X} \pm \frac{1960}{\sqrt{N}}$$

Hence, we want to choose N such that 1960/sqrt(N) = 50. By simple algebra, we see that sqrt(N) = 1960/50 = 39.2, so N = 39.2^2 = about 1537.

e. [Hard] Prove Expectations rule #15,

V(X + Y) = V(X) + V(Y) + 2 COV (X, Y)

[HINT: Rules 3, 8, and 10 are helpful. Expand squares, and look for ways that terms can be rearranged into known quantities]

Let A = X + Y. Then, $V(A) = E(A^2) - E(A)^2$.

Note that $A^2 = (X + Y)^2 = X^2 + Y^2 + 2XY$, hence $E(A^2) = E(X^2) + E(Y^2) + 2E(XY)$.
Also, E(A) = E(X) + E(Y), Hence $E(A)^2 = E(X)^2 + E(Y)^2 + 2E(X)E(Y)$.
Hence, by rearranging the above terms, we get
$V(A) = E(A^2) - E(A)^2 = E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2E(XY) - 2E(X)E(Y) =$
V(X) + V(Y) + 2COV(X,Y). QED

3. (25 points) There are two banks serving the medium-sized town of Pearson, Indiana. People's Bank dominates in the poorer, older parts of the town, while Continental Bank has the lion's share of the business in the newer, more affluent areas. Last year, each bank received 300 loan applications.

At People's Bank:

✓ 8/15 (53.33%) of the loan applications were approved
✓ Half of the approved loans were from low-income applicants
✓ 2/3 of all loan applicants were low-income

At Continental Bank,

✓ 2/3 (66.67) of all loan applications were approved
✓ 5% of the approved loans were from low-income applicants
✓ 1/6 of all loan applicants were low income.

a. Complete the following table. Recall that, as is already noted in the table, each bank received 300 loan applications.

	People's Bank			Continental Bank		
Income/Loan status	Approved	Denied	Σ	Approved	Denied	Σ
Low income applicants						
High income applicants						
Σ			300			300

For People's, we are told

✓ P(Approval) = 8/15 = 53.33%,
✓ P(Low income | Approval) = .50
✓ P(Low income) = 2/3

Hence, we can easily determine that

Approved = P(Approval) * # Applications = 8/15 * 300 = 160,
P(Low Income ∩ Approval) = P(Low income | Approval) * P(Approval) =
.50 * 8/15 = .2667
Low icome approvals = P(Low Income ∩ Approval) * # applications =
.2667 * 300 = 80
Low income applicants = P(Low income) * # applications = 2/3 * 300 = 200

For Continental, we are told

✓ P(Approval) = 2/3
✓ P(Low income | Approval) = .05
✓ P(Low income) = 1/6

Hence, we can easily determine that

Approved = P(Approval) * # Applications = 2/3 * 300 = 200
P(Low Income ∩ Approval) = P(Low income | Approval) * P(Approval) =
.05 * 2/3 = 1/30
Low icome approvals = P(Low Income ∩ Approval) * # applications =
1/30 * 300 = 10
Low income applicants = P(Low income) * # applications = 1/6 * 300 = 50

All other needed numbers are easily computed by making sure that the row and column totals are correct. Hence, the completed table is

	People's Bank			Continental Bank		
Income/Loan status	Approved	Denied	Σ	Approved	Denied	Σ
Low income applicants	80	120	200	10	40	50
High income applicants	80	20	100	190	60	250
Σ	160	140	300	200	100	300

b. Of all the loans that were approved by both banks, what percentage went to low income applicants?

There are 360 loans made (160 from People's, 200 from Continental). Of these, 90 are low income (80 from Peoples', 10 from Continental). Hence, of all the loans that were made, 90/360, or 25%, went to low income applicants.

c. People's Bank has been criticized for making fewer loans than its competitor. In its defense, People's argues that its higher proportion of low income applicants accounts for its lower approval rate. Suppose that People's had the same mix of low- and high-income applicants that Continental does. Suppose further that People's maintained its income-specific approval rates. What would its overall approval rate be? Based on this evidence, who do you think has the stronger case: People's Bank, or its critics?

At Continental, P(Low income) = 1/6, P(High Income) = 5/6.
At People's, P(Approval | Low Income) = 80/200 = .40,
P(Approval | High Income) = 80/100 = .80

Hence, if People's had the same mix of low- and high-income applicants as Continental does, while maintaining its income-specific approval rates, it would be the case that

P(Approval) = P(Low income) * P(Approval | Low income) +
P(High income) * P(Approval | High income) =
1/6 * .40 + 5/6 * .80 = .733

i.e. People's would approve 73.33% of all loans, or 220 out of 300. Since this is higher than Continental's approval rate, the evidence seems to favor People's over its critics.

4. (25 points) A university is concerned about charges of alcohol abuse on its campus. The university admits that up to 30% of its students drink too much on weekends. Concerned alumnae and parents claim that the figure is much higher than that. To check their suspicions, the university does blood tests on 60 randomly selected students on a Saturday night. Twenty-four (24) students are found to have excessive levels of alcohol in their bloodstream.

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 60 students would contain 24 or more excessive drinkers?

Step 1. The null and alternative hypotheses are

H_0: E(X) = 18 (or p = .30)
H_A:E(X) > 18 (or p > .30)

A one-tailed alternative is appropriate, since the critics believe the drinking rate is higher than what the University admits.

Step 2. The appropriate test statistic is

$$Z = \frac{\#\ intoxicated \pm CC - Np_0}{\sqrt{Np_0 q_0}} = \frac{\#\ intoxicated \pm CC - 18}{4}$$

To make the correction for continuity, we will add .5 if X < 18 and subtract .5 if X > 18.

Step 3. We will reject the null hypothesis if $z_c > 1.65$; alternatively, reject the null if X > 1.65 * sqrt(60*.3*.7) + 18 + .5, i.e. X > 24.36

Step 4. The computed value of the test statistic is

$$Z = \frac{\#\ intoxicated \pm CC - Np_0}{\sqrt{Np_0 q_0}} = \frac{24 - .5 - 18}{\sqrt{12.6}} = 1.55$$

Step 5. Do not reject the null hypothesis. The test statistic of 1.55 is less than the critical value of 1.65; or equivalently, the observed number of intoxicated individuals, 24, is less than the critical value of 24.36 (but just barely!)

Since F(1.55) = .939, there is a 6.1% chance that the university could be correct and the sample still contain this many intoxicated individuals.