Sociology 592 - Advanced Statistics I
Exam 1 - Answer Key
February 12, 1991

1. (5 pts. each, 20 pts. 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. $P(-1 \le Z \le 1) = 2F(1)$.

False. A true statement is $P(-1 \le Z \le 1) = 2F(1) - 1$; or, $P(-1 \le Z \le 1) = .68$.

b. The null and alternative hypotheses are:

H_0: $\mu = 100$ H_A: $\mu > 100$

In reality, $\mu = 100$. The researcher rejects the null hypothesis. The researcher has committed a Type I error.

True. The researcher has rejected H_0 when H_0 is true.

c. For a binomially distributed variable X, if N = 17 and p = .3, then P(5) = .2081.

True. See Hayes, p. 930, N = 17, p = .30, r = 5. Or, compute

$$\binom{N}{r} p^r q^{N-r} = \frac{N!}{r!(N-r)!} p^r q^{N-r} = \frac{17!}{5!12!} .3^5 .7^{12} = .2081$$

d. If A and B are independent events, then $P(A \mid B) = P(B \mid A)$.

False. A true statement is "If A and B are independent events, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$." Note that there is no reason P(A) has to be the same as P(B).

2. (10 pts. each, 30 pts. 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 pts. extra credit for each additional problem you do correctly.

a. In a recent election, Murphy defeated Jones, 60% to 40%. Of those who voted, 30% were Protestant, 30% were Catholic, and 40% belonged to other Religions. Twenty-four percent of the voters were Protestant supporters of Murphy, while another 15% were Catholics who voted for Jones. Complete the following table:

	Murphy (M)	Jones (J)	Σ
Protestant (P)	$P(M \cap P)$	$P(J \cap P)$	P(P)
Catholic (C)	$P(M \cap C)$	$P(J \cap C)$	P(C)
Other (O)	$P(M \cap O)$	$P(J \cap O)$	P(O)
Σ	P(M)	P(J)	

We are told P(M) = .60, P(J) = .40 ("Murphy defeated Jones, 60% to 40%"). We are also told P(P) = .30, P(C) = .30, P(O) = .40 ("30% were Protestant", etc.) And, we are told $P(M \cap P) = .24$ ("Twenty-four percent of the voters were Protestant supporters of Murphy") and $P(J \cap C) = .15$ ("another 15% were Catholics who voted for Jones"). Note that if 15% were Catholics who voted for Jones, then 15% must be Catholics who voted for Murphy. From there, it is a very simple matter to get the remaining numbers, we simply have to subtract the numbers we know from the marginal totals. This give us

	Murphy (M)	Jones (J)	Σ
Protestant (P)	**$P(M \cap P) = .24$**	**$P(J \cap P) = .06$**	**$P(P) = .30$**
Catholic (C)	**$P(M \cap C) = .15$**	**$P(J \cap C) = .15$**	**$P(C) = .30$**
Other (O)	**$P(M \cap O) = .21$**	**$P(J \cap O) = .19$**	**$P(O) = .40$**
Σ	**$P(M) = .60$**	**$P(J) = .40$**	

b. Recent advances in genetic engineering have made it possible for a women to be 60% successful in choosing the sex of her child. If 11 pregnant women all want girls, what is the probability that a majority of them will be successful?

There are at least 2 approaches. Note that $X \sim N(6.6, 2.64)$. Ergo, $P(X \geq 5.5) = P(Z \geq (5.5 - 6.6)/1.6248 = -.68) = P(Z \leq .68) = F(.68) = .7517$.

Or, let p = probability of having a boy = .40. Then, P(more girls than boys) = P(5 boys or less) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = .0036 + .0266 + .0887 + .1774 + .2365 + .2207 = .7535 (See Hayes, p. 928, N = 11, p = .40)

c. A gambler is willing to bet that, if he rolls 2 dice 24 times, he will get a 12 (i.e. double-sixes) at least once. If you take him up on his bet, who is more likely to win? (HINT: the probability of getting double-sixes on one roll of a pair of dice is 1/36. Because p is so small, the normal approximation works very poorly).

Note that you will win provided all 24 tosses come up 11 or less. The probability any one toss will not come up 12 (which is what you consider success) is 35/36. Hence, if X = number of tosses that come up 11 or less, X is binomially distributed, p = 35/36, q = 1/36, N = 24. For r = 24,

$$\binom{N}{r} p^r q^{N-r} = \frac{N!}{r!(N-r)!} p^r q^{N-r} = \left(\frac{24!}{24!0!}\right)\left(\frac{35}{36}\right)^{24}\left(\frac{1}{36}\right)^{0} = \left(\frac{35}{36}\right)^{24} = .509$$

i.e. the probability you will win is .509. The gambler's chances of winning are 1 - .509 = .491. Your odds are slightly better than his, so you are more likely to win.

d. On a recent 5-question statistics exam, 15% of the students got 1 question right, 20% got 2 questions right, 30% got 3 questions right, 20% got 4 questions right, and the remaining students got perfect scores. Find the mean and variance of the exam scores.

Construct the following table:

x	p(x)	x * p(x)	x^2	x^2 * p(x)
1	.15	.15	1	.15
2	.20	.40	4	.80
3	.30	.90	9	2.70
4	.20	.80	16	3.20
5	.15	.75	25	3.75
Σ		3.00		10.60

So, $\mu = 3.00$ (since $\mu = \Sigma xp(x)$), $\sigma^2 = 10.60 - 3^2 = 1.60$ (since $\sigma^2 = \Sigma x^2 p(x) - \mu^2$).

e. Prove that the following two formulas for the population covariance of X and Y are mathematically equivalent:

$$E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$$

This is a minor variation of the proof we used for the variance.

$$E[(X - \mu_X)(Y - \mu_Y)] = E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y) =$$

$$E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y =$$

$$E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$$

3. (25 points) A political scientist is studying black-white differentials in voting participation. She gathers data from voting precinct #17, which conveniently has 1000 blacks and 1000 whites that are eligible to vote. Her study reveals that 80% of blacks have low incomes and the remainder have high incomes. For whites, 40% have low incomes and the rest have high incomes. Among blacks, 30% of those with low incomes regularly vote in elections, compared to 80% of the high-income blacks. For whites, 20% of those with low incomes regularly vote, compared to 60% of the high-income whites.

a. Finish filling in the numbers for the following table. Remember that, as is already noted in the table, there are a total of 1000 blacks and 1000 whites.

	Black			White		
Vote/Income	Low	High	Σ	Low	High	Σ
Votes						
Does not vote						
Σ			1000			1000

We are told 800 (80%) of blacks have low incomes, which means that 200 (20%) have high incomes. More formally, for blacks, P(Low income) = .8, P(High income) = .2. Of the 800 blacks with low incomes, 30%, or 240, vote, which means that the other 70% (560) do not. More formally, for blacks, P(Vote ∩ Low Income) = P(Low Income) * P(Vote | Low income) = .8 * .3 = .24, P(Not Vote ∩ Low income) = P(Low income) * P(Not vote | low income) = .8 * .7 = .56. For the 200 blacks with high income, 80% (160) vote, which means that 20% (40) do not. More formally, for blacks, P(Vote ∩ High Income) = P(High income) * P(Vote | High income) = .2 * .8 = .16, P(Not Vote ∩ High Income) = P(High income) * P(Not Vote | High income) = .2 * .2 = .04. This means that 400 blacks (240 low income and 160 high income) vote, and the remaining 600 do not. More formally, for blacks, P(Vote) = P(Low income ∩ Vote) + P(High income ∩ Vote) = .24 + .16 = .40, P(Not vote) = 1 - P(Vote) = 1 - .4 = .6.

For whites, we know 400 (40%) are low income and the other 600 (60%) are high income. More formally, for whites, P(Low income) = .4, P(High income) = .6. Of the 400 low-income whites, 20% (80) vote, which means 80% (320) do not. More formally, for whites, P(Vote ∩ Low Income) = P(Low Income) * P(Vote | Low income) = .4 * .2 = .08, P(Not Vote ∩ Low income) = P(Low income) * P(Not vote | low income) = .4 * .8 = .32. Of the 600 high-income whites, 60% (360) vote, which means 40% (240) do not. More formally, for whites, P(Vote ∩ High Income) = P(High income) * P(Vote | High income) = .6 * .6 = .36, P(Not Vote ∩ High Income) = P(High income) * P(Not Vote | High income) = .6 * .4 = .24. Hence, there are 440 white voters (80 low income and 360 high income) and 560 non voters. More formally, for whites, P(Vote) = P(Low income ∩ Vote) + P(High income ∩ Vote) = .08 + .36 = .44, P(Not vote) = 1 - P(Vote) = .56. Thus, the table is

	Black			White		
Vote/Income	Low	High	Σ	Low	High	Σ
Votes	240	160	400	80	360	440
Does not vote	560	40	600	320	240	560
Σ	800	200	1000	400	600	1000

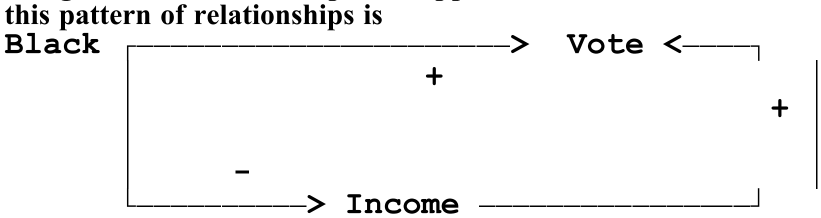
b. What percentage of the voters in this precinct are black?

There are 840 voters altogether (400 black and 440 white). Hence, the proportion of black voters is 400/840 = .476 = 47.6% Or, more formally, P(Black | Vote) = P(Black ∩ Vote)/P(Vote) = .20/.42 = .476 (remember that we are pooling the blacks and the whites here, so there are 2000 cases).

c. As these figures show, blacks are generally poorer than whites, and poor people tend to vote less than do wealthier people. Suppose that blacks had the same income distribution as whites. Suppose further that it continued to be the case that blacks maintained their income-specific voting rates, i.e. 30% of the poor blacks and 80% of the rich blacks voted. What percentage of blacks would then vote?

If there were 400 low-income blacks and 30% of them voted, there would be 120 low-income black voters. And, if there were 600 high-income blacks and 80% of them voted, there would be 480 high-income black voters. Hence, the total number of black voters would be 120 + 480 = 600, or 60%. More formally, using the rules for marginal probability, P(Vote | White income distribution and black voting rates) = P(Low income | White) * P(Voting | Black Low income) + P(High Income | White) * P(Voting | Black High income) = .4 * .3 + .6 * .8 = .60.

Comment: Note that, overall, blacks are less likely to vote than whites (40% as opposed to 44%). However, once you control for income, blacks actually have a higher voting rate. That is, at each income level, blacks are more likely than whites to vote; but because whites are more likely to have high incomes (and high income people are much more likely to vote) whites have a higher voting rate overall. Those who have had my research methods class will immediately recognize this as an example of <u>suppressor</u> effects. A causal model that would be consistent with this pattern of relationships is



4. (25 points) The military is concerned about the reliability of the weapons it is using in Operation Desert Storm. The manufacturer of a particular missile has guaranteed that no more than 20% of its missiles will be defective. Random testing of 100 missiles reveals that 30 are defective. Since the military would hate to unjustly accuse any of its contractors of faulty workmanship, test the manufacturer'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 the manufacturer is correct, what is the probability that a random sample of 100 missiles would contain 30 or more defects?

Step 1. The null and alternative hypotheses are:

H_0: p = .20 (or E(X) = 20)

H_A: p > .20 (or E(X) > 20)

Step 2. The appropriate test statistic is

$$z = \frac{\#\ of\ defects \pm CC - Np_0}{\sqrt{Np_0 q_0}} = \frac{x \pm CC - 20}{\sqrt{16}} = \frac{x \pm CC - 20}{4}$$

To make the correction for continuity, we will add .5 to x if the number of observed defects is less than 20, and we will subtract .5 from x if the number of defects is greater than 20. We will do nothing to x if x = 20.

Alternatively, if you prefer to use the proportion p, the test statistic is

$$z = \frac{p \pm CC/N - p_0}{\sqrt{\frac{p_0 q_0}{N}}} = \frac{p \pm CC/100 - .20}{.04}$$

Step 3. Critical region. For $\alpha = .01$, we will accept H_0 if $z \le 2.33$, otherwise we will reject H_0. This is equivalent to saying we will accept H_0 if $x \le (2.33 * 4 + .5 + 20 = 29.82)$, i.e. we will accept H_0 provided there are 29 or fewer defects.

Step 4. The test statistic equals

$$z = \frac{\#\ of\ defects \pm CC - Np_0}{\sqrt{Np_0 q_0}} = \frac{x \pm CC - 20}{4} = \frac{30 - .5 - 20}{4} = 2.38$$

Or, if you use the proportion p, the test statistic equals

$$z = \frac{p \pm CC/N - p_0}{\sqrt{\frac{p_0 q_0}{N}}} = \frac{p \pm CC/100 - .20}{.04} = \frac{.30 - .5/100 - .20}{.04} = 2.38$$

Without the correction for continuity, the test statistic equals 2.50

Step 5. Decision. Reject H_0. 2.38 > 2.33, so the number of defects falls just outside the acceptance region. Equivalently, the number of defective missiles was greater than 29.82, so reject H_0.

Also, if p = .2, $X \sim N(20, 4^2)$. The probability that 30 or more missiles would be defective if p = .20 = $P(X \ge 29.5) = P(Z \ge (29.5 - 20)/4 = 2.38) = 1 - F(2.38) = 1 - .9913 = .0087$, or about 1 chance in 115. So, either the defense contractor was extremely unlucky with the sample that was drawn, or else the defect rate is higher than .2.