Practice Exam Questions and Solutions for Midterm Exam, Spring 2009 Statistics 301, Professor Wardrop

1. Sarah performs a CRD with a dichotomous response. She obtains the sampling distribution of the test statistic for Fisher's test for her data; it is given below.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.6667	0.0001	0.0001	1.0000
-0.5278	0.0024	0.0025	0.9999
-0.3889	0.0242	0.0267	0.9975
-0.2500	0.1104	0.1371	0.9733
-0.1111	0.2588	0.3959	0.8629
0.0278	0.3220	0.7179	0.6041
0.1667	0.2094	0.9273	0.2821
0.3056	0.0652	0.9925	0.0727
0.4444	0.0075	1.0000	0.0075

   (a) Find the P-value for the first alternative ($p_1 > p_2$) if $x = 0.1667$.

   (b) Find the P-value for the second alternative ($p_1 < p_2$) if $x = -0.2500$.

   (c) Find the P-value for the third alternative ($p_1 \neq p_2$) if $x = -0.1111$.

   (d) Determine **both** the P-value and x that satisfy the following condition: The data are statistically significant but not highly statistically significant for the second alternative ($p_1 < p_2$).

   (e) Determine all combinations of P-values and x's that satisfy the following condition: The data are statistically significant but not highly statistically significant for the third alternative ($p_1 \neq p_2$).

   (f) Suppose that you want to draw the probability histogram of this sampling distribution. How tall is the rectangle centered at $x = 0.0278$.

2. A comparative study, with two treatments, dichotomous response and randomization is performed. The study is not balanced.

   The observed value of the test statistic, x, is a positive number. The exact P-values for all three alternatives are obtained and are below.

   0.3329, 0.6211 and 0.8234.

   Match each P-value with its alternative.

Alternative		
$>$	$<$	$\neq$

3. A comparative study, with two treatments, dichotomous response and randomization is performed. The study is balanced.

   The observed value of the test statistic, x, is a negative number. The exact P-values for all three alternatives are obtained. The three exact P-values are below along with three other numbers. Also note that for the actual value of x, $P(X = x) = 0.1257$.

   0.0863, 0.1215, 0.4336, 0.6921, 0.8672 and 0.9507,

   Identify the three P-value and match each with its alternative.

Alternative		
$>$	$<$	$\neq$

4. Consider an unbalanced study with ten subjects, identified as A, B, C, D, E, G, H, J, K and L. In the actual study,

   - Subjects A, B, C and D are assigned to the first treatment, and the other subjects are assigned to the second treatment.
   - There are exactly four successes, obtained by A, E, G and L.

   This information is needed for parts (a)–(c) below.

   (a) Compute the observed value of the test statistic.

(b) Assume that the Skeptic is correct. Determine the observed value of the test statistic for the assignment that places A, E, H and L on the first treatment, and the remaining subjects on the second treatment.

(c) Assume that the Skeptic is correct about subjects A, B, E and G, but incorrect about subjects C, D, H, J, K and L.

For the assignment that puts B, G, H and J on the first treatment, and the other subjects on the second treatment, determine the response for each of the ten subjects.

5. A comparative study yields the following table of counts.

Treat.	S	F	Total
1	6	4	10
2	6	2	8
Total	12	6	18

Assuming that the Skeptic is correct, determine all possible values of the test statistic.

6. An unbalanced CRD is performed and yields the data below.

Treatment	S	F	Total
1	22	78	100
2	24	56	80
Total	46	134	180

(a) Use the standard normal curve to obtain the approximate P-value for the first alternative, $p_1 > p_2$.

(b) Use the standard normal curve to obtain the approximate P-value for the second alternative, $p_1 < p_2$.

(c) Use the standard normal curve to obtain the approximate P-value for the third alternative, $p_1 \neq p_2$.

7. Amy, Beth and Cathy perform balanced studies; there n's are 30, 60 and 120, respectively. Amazingly, they get identical values for $\hat{p}_1$ and identical values for $\hat{p}_2$ and their $\hat{p}_1 > \hat{p}_2$.

There three P-values for the alternative $>$ are: 0.2189, 0.1000 and 0.3576.

Match each researcher with her P-value.

8. Mike draws a probability histogram for a Fisher's test. You are given the following facts:

- The study is balanced.
- $\delta = 0.02$.
- The height of the rectangle centered at 0.01 is 7.77.

(a) Calculate $P(X = 0.01)$.

(b) Given that the actual value of the test statistic is $x = -0.03$, calculate the exact P-value for the third alternative, $p_1 \neq p_2$.

9. On each of five days next week (Monday thru Friday), Earl will shoot six free throws. Assume that Earl's shots satisfy the assumptions of Bernoulli trials with $p = 0.37$.

(a) Compute the probability that on any particular day Earl makes his first free throw and misses his second and third free throws.

(b) Compute the probability that on any particular day Earl obtains a total of exactly two successes. For future reference, if Earl obtains exactly two successes on any particular day, then we say that the event "Brad" has occurred.

(c) Refer to part (b). Compute the probability that: next week Brad will occur exactly twice, with one of the occurrences on Tuesday. (Note: You are being asked to compute one probability.)

10. Below are tables (1–3) from three different studies. Match each Table to the correct statement below.

(a) This table provides evidence that p increased over the course of the study.

(b) This table provides evidence that p decreased over the course of the study.

(c) This table provides no evidence that p changed over the course of the study.

Table 1

Half	S	F	Total
1st	30	45	75
2nd	40	35	75

Table 2

Half	S	F	Total
1st	60	90	150
2nd	60	90	150

Table 3

Half	S	F	Total
1st	90	110	200
2nd	50	150	200

Table 4

	Current		
Prev.	S	F	Total
S	60	59	119
F	60	80	140
Total	120	139	259

Table 5

	Current		
Prev.	S	F	Total
S	70	50	120
F	49	30	79
Total	119	80	199

Table 6

	Current		
Prev.	S	F	Total
S	18	42	60
F	42	98	140
Total	60	140	200

Above are tables (4, 5 and 6) from three different studies. Note that at least one of these tables is the answer to more than of the following four questions.

(d) Which table is for the study that had an F on its first trial and an S on its last trial?

(e) Which table is for the study in which the subject performed better (S's are good) after an S than after an F?

(f) Which table is for the study that had an F on its first and last trials?

(g) Which table is for the study in which the subject performed the same after an S as after an F?

Solutions

1. (a) The P-value is

$$P(X \geq 0.1667) = 0.2821.$$

(b) The P-value is

$$P(X \leq -0.2500) = 0.1371.$$

(c) The P-value is

$$P(X \leq -0.1111) + P(X \geq 0.1111) =$$

$$0.3959 + P(X \geq 0.1667) =$$

$$0.3959 + 0.2821 = 0.6780.$$

(d) The P-value is in the column headed "$P(X \leq x)$." The only number in this column that satisfies the stated conditions is 0.0267; thus, 0.0267 is the P-value and it corresponds to $x = -0.3889$.

(e) The P-value is the sum of a number from the third column and a number from the fourth column. I will proceed by 'informed trial and error.'

Let's begin in the third column with the entry 0.0267. This area corresponds to $x = -0.3889$. If $x = -0.3889$, then the P-value is $0.0267 + 0.0075 = 0.0342$.

It can be seen rather quickly that for any other entry in the third column the resultant P-value will be smaller than 0.01 or greater than 0.05.

Moving to the fourth column, consider the entry $x = 0.4444$. This gives a P-value of $0.0075 + 0.0025 = 0.0100$, which is highly statistically significant.

It can be seen rather quickly that for any other entry in the fourth column the resultant P-value will be smaller than 0.01 or greater than 0.05.

Thus, the lone correct answer is $x = -0.3889$ which gives a P-value of 0.0342.

(f) There is some round-off error in the values of the x's b/c there are two possible values for δ:

$$0.4444 - 0.3056 = 0.1388, \text{ and}$$

$$0.3056 - 0.1667 = 0.1389.$$

You may use either one w/o having much impact on your answer. Using the former, the height of the rectangle is

$$P(X = 0.0278)/0.1388 =$$

$$0.3220/0.1388 = 2.320.$$

2. Recall that $x > 0$. Define the following quantities.

$$A = P(X \geq x), B = P(X \leq x)$$

$$\text{and } C = P(X \leq -x).$$

The P-values are given below.

Alt.	$>$	$<$	$\neq$
P-value	A	B	$A + C$

Recall from lecture, $A + B$ must be larger than 1.

Consider $A + C$. I will prove by contradiction that it equals 0.6211. First, $A + C$ is larger than A; thus, it cannot be the smallest of the three P-values. If $A + C = 0.8234$, then A and B would be the remaining P-values and they would NOT sum to more than 1.

Given that $A + C = 0.6211$, it follows that A (being smaller) is 0.3329 and B is 0.8234.

3. The fact that the study is balanced implies that the sampling distribution is symmetric, as discussed previously in the class. Recall that $x < 0$. Define the following quantities.

$$D = P(X \geq x), E = P(X \leq x)$$

$$\text{and } F = P(X \geq -x).$$

B/c of symmetry, $E = F$.

The P-values are given below.

Alt.	$>$	$<$	$\neq$
P-value	D	E	$E + F = 2E$

Thus, one of the P-values ($2E$) is twice as large as another (E). Examining the six candidates, we find that $E = 0.4336$ and $2E = 0.8672$. Recall that

$$D + E = 1 + P(X = x) = 1.1257.$$

Thus,

$$D + 0.4336 = 1.1257 \text{ or } D = 0.6921.$$

4. (a) The observed value of the test statistic is

$$x = \hat{p}_1 - \hat{p}_2 = 1/4 - 3/6 =$$

$$0.25 - 0.50 = -0.25.$$

(b) The observed value of the test statistic would be

$$x = \hat{p}_1 - \hat{p}_2 = 3/4 - 1/6 =$$

$$0.75 - 0.17 = 0.58.$$

(c) A and G will remain S's b/c the Skeptic is correct.

B and K will remain F's b/c they have not changed treatment.

C, D, H and J will become S's b/c the Skeptic is incorrect and they have changed treatment.

E and L will remain S's b/c they have not changed treatment.

5. The smallest possible value in the 'a' position in the table is 4, b/c any number smaller than 4 will make 'd' negative. You can check that $a = 4$ gives $x = -0.600$. The largest possible value for 'a' is 10, which gives $x = 0.750$. Next,

$$\delta = \frac{n}{n_1 n_2} = \frac{18}{10(8)} = 0.225.$$

Thus, the possible x's are:

$$-0.600, -0.375, -0.150, 0.075, 0.300,$$

$$0.525 \text{ and } 0.750.$$

6. First, calculate

$$x = 22/100 - 24/80 = 0.22 - 0.30 = -0.08, \text{ and}$$

$$\sigma = \sqrt{\frac{46(134)}{100(80)(179)}} = \sqrt{0.0043045} = 0.0656.$$

Thus, $z = -0.08/0.0656 = -1.22$.

   (a) For the alternative $>$, the approximate P-value equals the area under the snc to the right of $z = -1.22$, which equals 0.8888.
   (b) For the alternative $<$, the approximate P-value equals the area under the snc to the right of $-z = +1.22$, which equals 0.1112.
   (c) For the alternative $\neq$, the approximate P-value equals twice the area under the snc to the right of $|z| = +1.22$, which equals $2(0.1112) = 0.2224$.

7. The largest P-value is Amy's and the smallest is Cathy's. See pages 72–74 of the lecture notes for an explanation of this answer. (View Amy's as the original data; Beth's is for $k = 2$ and Cathy's for $k = 4$.)

8. It helps to draw a picture.

   (a) The probability is given by the area of its rectangle:

   $$0.02(7.77) = 0.1554.$$

   (b) The possible values of the test statistic are:

   $$\ldots - 0.03, -0.01, 0.01, 0.03, \ldots.$$

   By symmetry,

   $$P(X \geq 0.01) = P(X \leq -0.01) = 0.5000.$$

   Thus,

   $$P(X \geq 0.03) = 0.5000 - P(X = 0.01) =$$
   $$0.5000 - 0.1554 = 0.3446.$$

   The P-value sought is

   $$P(X \leq -0.03) + P(X \geq 0.03) = 2P(X \geq 0.03),$$

   by symmetry. Thus, the P-value is

   $$2(0.3446) = 0.6892.$$

9. (a) This is a multiplication rule problem b/c it is a question about a particular sequence.

   $$P(SFF) = pqq = 0.1469.$$

   (b) This is a binomial problem b/c it is about the total number of successes. The probability of exactly two successes is

   $$\frac{6!}{2!4!}(0.37)^2(0.63)^4 =$$
   $$15(0.1369)(0.1575) = 0.3234.$$

   (c) For this part each day is a trial and the probability that a day yields a success is, from part (b), $p = 0.3234$. Then it gets trickier. Let Y denote the total number of successes (Brads) Earl gets on Monday, Wednesday, Thursday and Friday. We want the probability of a Brad on Tuesday **and** $(Y = 1)$. These probabilities are $p = 0.3234$ and

   $$\frac{4!}{1!3!}(0.3234)(0.6766)^3 = 0.4007.$$

   Finally, by the multiplication rule, we multiply these answers to get

   $$0.3234(0.4007) = 0.1296.$$

10. (a) This describes Table 1.
   (b) This describes Table 3.
   (c) This describes Table 2.
   (d) There will be one subtracted from the column total for F and one subtracted from the row total for S. Thus, the row total for F will be one larger than the column total for F; only Table 4 satisfies this condition.
   (e) We need $\hat{p}_1 > \hat{p}_2$; only Table 4 satisfies this condition.
   (f) There will be one subtracted from the column total for F and one subtracted from the row total for F. Thus, the row and column totals for F will be the same number; only Table 6 satisfies this condition.
   (g) We need $\hat{p}_1 = \hat{p}_2$; only Table 6 satisfies this condition.