Statistics 410, Fall 2011

Solutions to homework 5, due 10/26/11

6.1 (a) $\alpha = .5$. $\beta = 0$.
(b) $\alpha = .0625$, $\beta = 0$.
(c) $\alpha = .5^{25}$, $\beta = 0$.
(d) $\alpha = .5^{100}$, $\beta = 0$.

6.2 (a) α is the probability of no aces in 24 rolls if the die is fair. This is $\frac{5}{6}^{24} = .0126$.
(b) $\beta = 0$. That is, if the die is altered, we are guaranteed to reject H_0 since no aces will come up in our 24 rolls.

6.5 α will be the probability of $K = 11$ or 12 if the treatment is ineffective. This is

$$\binom{12}{12}(.5)^{12} + \binom{12}{11}(.5)^{12} = .0032.$$

When $\pi = .8$, β will be the probability of retaining the null hypothesis. That is the probability that $K \leq 10$, which is

$$1 - P(K = 12) - P(K = 11) = 1 - \binom{12}{12}.8^{12} - \binom{12}{11}.8^{11}.2 = 1 - .069 - .206 = .725.$$

Similarly, when $\pi = .95$, β is

$$1 - P(K = 12) - P(K = 11) = 1 - \binom{12}{12}.95^{12} - \binom{12}{11}.95^{11}.05 = 1 - .5404 - .3413 = .1183.$$

6.6 (a) α will be the probability of $K \geq 16$ if $\pi = .5$, so

$$\sum_{k=16}^{20} \binom{20}{k}.5^{20} = .0059.$$

(b) With $\pi = .8$, β will be the probability that $K < 16$, which is

$$1 - \sum_{k=16}^{20} \binom{20}{k}.8^{k}.2^{20-k} = 1 - .6295 = .3705.$$

(c) We do the same calculation as above with $\pi = .95$.

$$1 - \sum_{k=16}^{20} \binom{20}{k}.95^{k}.05^{20-k} = .0026.$$

6.11 (a) H_0 is the hypothesis that men are equally likely to overreport and underreport their height.
(b) H_a is the hypothesis that men are more likely to overreport their height.
(c) The significance level is .1.
(d) The test statistic is that of the 11 men who report inaccurately, 10 overreported their height.

(e) The p-value is the probability under H_0 that 10 or 11 men overreported their height. This is

$$.5^{11} + 10(.5)^{11} = .0054$$

(f) The p-value is less than the significance level so we reject H_0 and conclude that men tend to overreport their height.

6.12 (a) H_0 is that the average daily high temperature was 18 degrees.

(b) H_a is that the average daily high temperature was below 18 degrees.

(c) $\alpha = .05$.

(d) Our statistic is that out of the 29 days when the high temperature wasn't 18 degrees, it fell below on 21 of them.

(e) The P-value is the probability of getting 21 or more of the 29 days below 18 degrees if the high temperature is just as likely to be above or below 18 degrees. This is

$$\sum_{k=21}^{29} \binom{29}{k} .5^{29} = ..012.$$

(f) Our p-value is below α, so we would reject H_0 if this made any sense.

This is not an appropriate example for a sign test for two reasons: One is that we aren't taking a sample, in fact we have complete information about thei high temperatures in May (we have all 31 days). The other is that just looking at the days above and the days below ignores how far above and below those days are (but taking averages does *not* ignore that).

6.14 (a) Our null hypothesis is that ambidextrous people have a .5 probability of being men (or women).

(b) H_a is that ambidextrous people are more likely to be men than women.

(c) Our significance level is $\alpha = .01$.

(d) Our test statistic is that out of our sample of 26 ambidextrous people, 19 are men.

(e) Our p value is

$$\sum_{k=19}^{26} \binom{26}{k} (.5)^{26} = .0145.$$

(f) Our p-value is above α so we retain H_0. We don't have enough evidence to conclude more men than women are ambidextrous at the .01 significance level.

6.15 (a) H_0 is that a color blind person is equally likely to be a man or a woman.

(b) H_a is that a color blind person is more likely to be a man.

(c) $\alpha = .01$.

(d) Our test statistic is that out of our sample of 54 color blind people, 50 are men.

(e) Our p-value is

$$\sum_{k=50}^{54} \binom{54}{k} (.5)^{54} = 1.9 \times 10^{-11}.$$

The p-value is (much) less than α, so we reject H_0 and conclude that color blind people are more likely to be men.

6.16 203 people feel Americans are as willing to work as hard in the past.

(a) H_0 is that this doesn't depend on age. Since our group has twice as many people over 30 as under 30, this corresponds to a probability of $\pi = 1/3$ that a person thinking Americans are as willing to work as hard is in the under 30 group.

(b) Our alternative hypothesis is that $\pi < 1/3$. (That is that fewer adults under 30 think Americans are willing to work as hard as in the past.)

(c) $\alpha = .05$.

(d) The test statistic is that 58 of our group of 203 who think Americans are as willing to work as hard as in the past are in the under 30 group.

(e) The p-value is the probability that 58 or fewer of our group of 203 are in the under 30 group, assuming $\pi = 1/3$. This is

$$\sum_{k=0}^{58} \binom{203}{k} \frac{1}{3}^{k} \frac{2}{3}^{203-k} = .085.$$

(f) The p-value is greater than α so we retain H_0 and don't detect a difference relative to age.

6.19 (a) H_0 is that subjects with headaches are just as likely to be in the treatment group as the control group.

(b) H_a is that they are more likely to be in the treatment group (that is, that the mediation increases the likelihood of headaches).

(c) $\alpha = .5$

(d) Our test statistic is that 9 out of the 12 headache sufferers are in the treatment group.

(e) p-value is the probability (assuming H_0) that 9 or more headache sufferers would be in the treatment group. This is

$$\sum_{k=9}^{12} \binom{12}{k} (.5)^{12} = .073.$$

(f) Our p-value is greater than α so we retain H_0 and find insufficient evidence to conclude the medication causes headache.

6.20 (a) H_0 is that the medicine has no affect on headaches, so that a headache sufferer has probability $\pi = .9$ of being in the treatment group (since that is 90% of the subjects).

(b) H_0 is that $\pi > .9$.
(c) $\alpha = .05$.
(d) Our test statistic is that 60 of our sample of 64 headache sufferers are in the medicated group
(e) Our p-value is

$$\sum_{k=60}^{64} \binom{64}{k} (.9)^k (.1)^{64-k} = .22.$$

(f) $.22 > .05$, so we retain H_0 and there is insufficient evidence to conclude the medication causes headaches (at the 5% significance level).

6.26 (a) We need to calculate $P(K > 8)$ if H_0 holds. This will be

$$(.5)^{10} + 10 \cdot (.5)^{10} = .0107.$$

(b) This will be $P(K \leq 8)$ if $\pi = .9$. That's

$$1 - \sum_{k=9}^{10} \binom{10}{k} .9^k .1^{10-k} = .2639.$$

6.33 If the jurors were chosen randomly, the probability of choosing a woman would be $\pi = 102/350 = .2914$. We test for the possiblity that $\pi < 102/350$.
(a) H_0 is that $\pi = 102/350$.
(b) H_a is that $\pi < 102/350$.
(c) The significance level is $\alpha = .01$.
(d) The test statistic is that out of the 100 choices, 9 were women.
(e) The p-value is the probability of 9 or fewer women given 100 choices, assuming H_0.

$$\sum_{i=0}^{9} \binom{100}{i} (102/350)^i (248/350)^{100-i} = 9 \times 10^{-7}.$$

(f) This is less than .01, so we reject H_0 and conclude the judge was biased against choosing women (or there are some other variables we aren't measuring which account for those choices).

6.34 If the new drug has no effect, we would expect our failures to occur proportionally to the number of people in the two groups. So we would expect 2/3 of our failures to be in the new drug group. In other words, given a failure, it should have $\pi = 2/3$ chance of being in the new drug group.
(a) H_0 is $\pi = 2/3$.
(b) H_a is $\pi < 2/3$. This is the situation that the new drug is better than the standard treatment, so we expect fewer failure in the new drug group under this hypothesis.
(c) The significance level is $\alpha = .01$.

(d) The test statistic is that out of our 11 failures, 4 are in the new drug group.
(e) The p-value is the probability of 4 or fewer failures in the new drug group assuming H_0. This is

$$\sum_{i=0}^{4} \binom{11}{i} (2/3)^i (1/3)^{11-i} = .039.$$

(f) .039 is greater than .01 so we retain H_0 and conclude there is not enough evidence to show this new drug does better than the standard treatment at the .01 significance level.

7.2

$$f(x) = \begin{cases} 0 & x < 5, x \geq 10 \\ 1/5 & 5 \leq x \leq 10. \end{cases}$$

7.3 (a) $F(x) = (3/2)(x^2 - x^3/3)$ for $0 \leq x \leq 1$, and 0 for $x < 0$, 1 for $x > 1$.
(b) $F(1/4) = (3/2)(1/16 - 1/192) = 3/32 - 1/128 = .0859375$.
(c) $1 - F(.25) = .914063$.
(d) $F(.75) = .632813$.
(e) $F(.75) - F(.25) = .546875$.

7.4 (a)

$$F(x) = \begin{cases} 0 & x \leq -1 \\ .5 - x^2/2 & -1 \leq x \leq 0 \\ .5 + x^2/2 & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

(b) $F(-.5) = .375$.
(c) $P(X \geq 0) = .5$
(d) $F(.75) = .5 + 9/32 = .78125$.
(e) $.78125 = .375 = .40625$.

7.5 (a)

$$F(X) = \begin{cases} 0 & x < -1 \\ \frac{x^2}{2} + x + \frac{1}{2} & -1 \leq x \leq 0 \\ -\frac{x^2}{2} + x + \frac{1}{2} & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

(b) $F(-1/2) = 1/8$.
(c) $P(X \geq 0) = 1 - F(0) = .5$.
(d) $F(.5) = 7/8$.
(e) $P(-.5 < X \leq .5) = F(.5) - F(-.5) = 3/4$.

7.6 (a) $F(x) = 1 - e^{-x}$ for $x \geq 0$ and 0 for negative x.
(b) $F(0) = 0$.
(c) $F(1) = 1 - 1/e$.
(d) $P(X \geq 1) = 1 - F(1) = 1/e$.
(e) $P(1 < X \leq 2) = F(2) - F(1) = 1 - 1/e^2 - (1 - 1/e) = 1/e - 1/e^2$.

7.11 We need to find t_1 so $F(t_1) = .25$ and t_2 so $F(t_2) = .75$.

Using the formula for $F(t)$ at the top of p. 199, we see

$$.25 = F(t_1) = 1 - e^{\frac{-t_1}{80}} \text{ so } .75 = e^{\frac{-t_1}{80}} \text{ so } \ln(.75) = -t_1/80$$

This give $t_1 = (-80)\ln(3/4) = (-80)[\ln(3) - \ln(4)] = 80[\ln(4) - \ln(3)] = 80\ln(4/3)$.

To find t_2 we do something similar to get $\ln(.25) = -t_2/80$ and so $t_2 = 80\ln(4)$.

Thus $t_2 - t_1$ is $80\ln(3)$.

7.12 As above, we are going to want to find y_1 with $F(y_1) = .25$ and y_2 with $F(y_2) = .75$.

$$.25 = -3 + \frac{4}{3}y_1 - \frac{1}{9}y_1^2$$

yields $y_1^2 - 12y_1 + 29.25 = 0.$, giving $y_1 = 3.4$. (There is another solution to this quadratic equation, but F of that solution is 1.)

Similarly,

$$.75 = -3 + \frac{4}{3}y_2 - \frac{1}{9}y_2^2$$

gives $y_2^2 - 12y_2 + 33.75 = 0$, so $y_2 = 4.5$ (again, there is another solution, but F of that other solution is 1).

So the interquartile range is $y_2 - y_1 = 4.5 - 3.4 = 1.1$.