Introduction to Biostatistics BC 203
Midterm Exam
Tuesday, October 24, 2006

(Closed Book)

There are 30 `questions', each worth 3.33 points.

Write out your solutions and circle your final answer. Show all your work on these pages. If you need more space write on the back of the page.

You have one and one half hours to complete this exam.

Name: Solutions

1. The following collection of numbers represents the ages of 13 junior congressmen. Make the calculations needed for a box plot, and enter your results in the appropriate blanks.

35, 35.5, 36, 37.5, 38, 39, 39, 41, 41, 41, 43, 45.5, 47.5

(a) median = 39 $\frac{n+1}{2}^{th}$ obs = 7^{th} obs

(b) upper hinge = 41 $3\frac{(n+1)}{4}^{th}$ obs = 10^{th} obs

(c) lower hinge = 37.5 $\frac{n+1}{4}^{th}$ obs = 4^{th} obs

IQR = 41 - 37.5 = 3.5

(d) Length of lower whisker = 2.5 (37.5 - 35)

Upper Fence 41 + 1.5(3.5) = 46.25

lower Fence 37.5 - 1.5(3.5) = 32.25

(e) Length of upper whisker = 4.5 (45.5 - 41) * whisker goes to obs closest ~~to~~ Fence from Hinge

(f) Are there any outliers in this data set? If so, identify them.

47.5 (beyond Fence)

2. The table below displays the number of tax returns filed and audited by the IRS in 1997.

	1997 Tax Returns					
Income	**Filed**	**Audited**	Rate			
0-25K	59,211,700	1,076,945	0.0182			
25-50K	27,263,000	259,794	0.0095			
50-100K	17,019,200	196,582	0.0116			
100k+	4,540,800	129,320	0.0285			
Total	108,034,700	1,662,641	0.0154			

(a) What was the IRS Audit rate in 1997?

$$\frac{1,662,641}{108,034,700} = 0.0154$$

(b) If the income distribution in 1997 was uniform (i.e., 25% for each income category), what would have been the IRS audit rate in 1997?

$$\Sigma w_i R_i = .25(.0182) + .25(.0095) + .25(.0116) + .25(.0285) = \boxed{.0170}$$

$$= \frac{\Sigma R_i}{4} = \bar{R}_i \text{ (average rate)}$$

b/c weights are all =.

Problem 2 (continued)

(Check **one** box)

(c) The rate asked for in part (b) is

- ☑ a directly adjusted rate .
- ☐ a indirectly adjusted rate.
- ☐ a crude rate.
- ☐ not actually a rate.

(Check **one** box)

(d) The income distribution for American Caucasians in 1997 was shifted toward higher income levels. So it must be true that the audit rate for American Caucasians in 1997 was

- ☐ higher than the overall audit rate in 1997.
- ☐ lower than the overall audit rate in 1997.
- ☐ equal to the overall audit rate in 1997.
- ☑ either higher or lower than the overall Audit rate in 1997, depending on how the distribution was shifted.

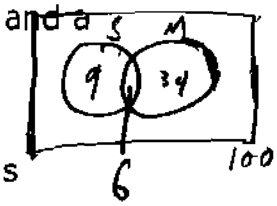
(Check **one** box)

(e) Suppose that the audit rate for incomes less than $25K was 0.0078. Given that the distribution of income levels among American Caucasians is shifted toward higher income levels, it must be true that the audit rate for American Caucasians in 1997 was

- ☑ higher than the overall audit rate in 1997.
- ☐ lower than the overall audit rate in 1997.
- ☐ equal to the overall audit rate in 1997.
- ☐ either higher or lower than the overall Audit rate in 1997, depending on how the distribution was shifted.

3. Suppose 15% of traffic accidents involve a SUV and 40% of traffic accidents involve a motorcycle. Accidents involving both a SUV and a Motorcycle occur 6% of the time.

$P(S) = .15 \quad P(M) = .4 \quad P(S \cap M) = .06$



(a) What is the probability that neither a SUV nor a motorcycle is involved in an accident?

$P(\overline{S \text{ and } M}) = 1 - P(S \text{ and } M) = 1 - (.15 + .4 - .06) = .51$

$\boxed{51\%}$

(b) What is the probability that either a SUV or a Motorcycle is involved, but not both?

$P(SM^c) + P(MS^c)$

$P(SM^c \text{ and } MS^c) = .15 + .4 - .06 - .06 \quad OR = (.15 - .06) + (.4 - .06)$

$= .43 \qquad = .09 + .34$

$\boxed{= .43}$

(c) What is the probability of an SUV accident given a Motorcycle is involved?

$P(S|M) = \dfrac{P(S \text{ and } M)}{P(M)} = \dfrac{.06}{.4} = \boxed{.15}$

(d) Are SUV and Motorcycle accidents independent? (Justify your answer.)

$P(S|M) = P(M)$ So yes S and M are indep.

4. In the United States, 3% of babies born have some type of genetic anomaly. At Women and Infants' Hospital (in Rhode Island) 33 babies are born each day. Assume all births are independent.

(a) What is the probability that at least one baby will be born with a genetic anomaly today?

$Y \sim Bin(33, .03)$

$P(Y \geq 1) = 1 - P(Y=0) = 1 - \binom{33}{0} .03^0 (1-.03)^{33}$

$= 1 - 0.3660$

$= \boxed{0.6340}$

(b) What is the probability that exactly one baby will be born with a genetic anomaly each day for 3 consecutive days?

$P(3 \text{ days w/ 1 baby}) = P(Y=1)^3$ b/c indep

$= \left[\binom{33}{1} .03^1 (.97)^{32}\right]^3 = [.3735]^3$

$\boxed{= .0521}$

(c) What is the probability that exactly 3 babies will be born with a genetic anomaly in the next three days?

$X \sim Bin(99, .03)$ b/c Births are indep.

$P(X=3) = \binom{99}{3} .03^3 (.97)^{96}$

$= \boxed{0.2275}$

Problem 4 (continued)

(d) How many babies with genetic abnormalities are expected over the next three days?

$$E[X] = n\theta = 99(.03) = \boxed{2.97}$$

(f) What is the standard deviation of the proportion of babies with genetic abnormalities in the next three days?

$$SD\left(\frac{X}{n}\right) = \sqrt{Var\left(\frac{X}{n}\right)} = \sqrt{\frac{1}{n^2} Var(X)}$$

$$= \sqrt{\frac{1}{n^2} n\theta(1-\theta)}$$

$$= \sqrt{\frac{\theta(1-\theta)}{n}}$$

$$= \sqrt{\frac{.03(.97)}{99}}$$

$$= 0.0171$$

5. The distribution of heights for American women is normally distributed with a mean of 65.5 inches and a standard deviation of 2.5 inches. The distribution of heights for American men is also normally distributed with a mean of 69 inches and a standard deviation of 3.1 inches. Suppose that 55% of Americans are male.

(a) What is the probability that a randomly selected woman is at least 62 inches tall?

$X \sim N(65.5, 2.5^2)$

$P(X > 62) = P\left(Z > \frac{62-65.5}{2.5}\right)$

$= P(Z > -1.4) = 1 - 0.081 = 0.919$

(b) Find the top 15.2th percentile of American woman's heights.

$P(Z > ?) = .152$ but $P(Z > 1.03) = .152$

$P\left(\frac{X-\mu}{\sigma} > ?\right) = .152$ so $\# = 1.03(2.5) + 65.5$

$P(X > \underbrace{?\sigma + \mu}_{\#}) = .152$ $= \boxed{68.075}$

(c) In a sample of 10 women, what is probability that their average height will be between 64 and 66 inches tall?

$X_1, \dots, X_{10} \sim N(65.5, 2.5^2)$

$P(64 < \bar{X}_{10} < 66) = P\left(\frac{64-\mu}{\sigma/\sqrt{n}} < Z < \frac{66-\mu}{\sigma/\sqrt{n}}\right)$

$= P\left(\frac{64-65.5}{2.5/\sqrt{10}} < Z < \frac{66-65.5}{2.5/\sqrt{10}}\right)$

$= P(-1.89 < Z < 0.6325)$

$= 1 - 0.029 - .264$

$= \boxed{.707}$

Problem 5 (continued)

(d) What is the probability that a randomly selected individual is between 62 and 68 inches tall?

I is an individual gender unknown.

$$P(62 < I < 68) = P(62 < I < 68 \mid I = F)P(F) + P(62 < I < 68 \mid I = M)P(M)$$

$$= P\left(\frac{62-65.5}{2.5} < z < \frac{68-65.5}{2.5}\right)P(F) + P\left(\frac{62-69}{3.1} < z < \frac{68-69}{3.1}\right)P(M)$$

$$= P(-1.4 < z < 1)(.45) + P(-2.25 < z < -0.322)(.55)$$

$$= (1 - .081 - 0.159)(.45) + (.374 - 0.012)(.55)$$

$$= .76(.45) + 0.362(.55) = \boxed{0.5411}$$

(e) An individual is randomly selected from the population and found to be between 62 and 68 inches tall. What is the probability that this individual is female?

$$P(F \mid 62 < w < 68) = \frac{P(F \text{ and } 62 < w < 68)}{P(62 < w < 68)}$$

$$= \frac{P(62 < w < 68 \mid w = F)\,P(F)}{0.5411}$$

Same as Bayes theorem.

$$= \frac{0.76(.45)}{0.5411} = 0.632$$

$$\boxed{63.2\%}$$

Problem 5 (continued)

A European colleague asks you to translate your results to the metric system, specifically into centimeters (cm). One centimeter equals 2.54 inches. For example, 62 inches equals 157.5 cm.

(f) What is the distribution of women's height in cm (give the family, mean and variance)?

$X \sim N(65.5, 2.5^2)$

$Y = 2.54X$

$E[Y] = 2.54E(X) = 2.54(65.5) = 166.37$

$Var(Y) = 2.54^2 Var(X) = 2.54^2 2.5^2 = 40.32$

$\boxed{Y \sim N(166.37, 6.35^2)}$

$\sqrt{40.32} = 6.35$

(g) What is the probability that a randomly selected woman is at least 157.5 cm tall?

$P(Y > 157.5) = P(Z > \frac{157.5 - 166.37}{6.35})$

$= P(Z > -1.4)$

$= \boxed{0.919}$

notice $P(Y > 157.5) = P(X > 62)$!

6. On Monday morning, 20 people with signs of the flu arrive at a Doctor's office and two of those people have the avian flu. Ten patients can be seen in the morning and ten can be seen in the afternoon. However patients are seen in the order that they sign in, so their order is effectively random.

(a) What is the probability that both people with the avian flu are seen in the morning? $X \sim \text{Hyper}(20, 2, 10)$

$$P(X=2) = \frac{\binom{2}{2}\binom{18}{8}}{\binom{20}{10}} = \boxed{0.2368}$$

(b) What is the expected number of people with Avian flu seen in the morning?

$$E[X] = 2\,\frac{10}{20} = \boxed{1}$$

On Tuesday morning, another 20 people with signs of the flu arrive at the Doctor's office. But this time four people have the avian flu. Assume Monday's and Tuesday's patients are independent.

(c) What is the expected number of patients with avian flu seen on both mornings?

$$Y \sim \text{Hyper}(20, 10, 4)$$

$$E[Y] = \frac{10}{20}\,4 = 2$$

$$E[X+Y] = E[X] + E[Y] = 1 + 2 = \boxed{3}.$$