Introduction to Biostatistics BC 203
Midterm Exam
Tuesday, October 24, 2003

(Closed Book)

There are 33 `questions', each worth 3 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 stem and leaf plot displays the average miles per gallon for 19 new car models.

Stem	Leaf	
0	1	1
1	9,3,7,0,8,7,9	0, 3, 7, $\boxed{7}$, 8, 9, 9
2	2,7,4,1,1,3,2	1, 1, 2, 2, 3, 4, 7
3	3,3,1,0	0, 1, 3, 3
4		

SORT data FIRST

(a) What is the median miles per gallon in this sample of new cars?

$\frac{n+1}{2}$ th obsn $= 10^{th}$ 21 mpg is Median

Circled above

(b) What is the 25% percentile in this sample of new cars?

$\frac{n+1}{4}$ th obsn $= 5^{th}$ 17 mpg is 25%

Boxed above

(c) What is the 75% percentile in this sample of new cars?

$3\left(\frac{n+1}{4}\right)^{th}$ obsn $= 15^{th}$ 27 mpg is 75%

Triangle above

(d) What is the Inter-quartile range in this sample?

$IQR = 27 - 17 = 10$ 10 mpg

(75% - 25%)

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

Yes obsn 1 is an outlier b/c it is below lower Fence

lower Fence $17 - 1.5(10) = 2$

upper Fence $27 + 1.5(10) = 42$

2. The number of reported stolen hats (crimes) in New York City and Phoenix are given in the following table, which is broken down by hair color, along with the population size (in thousands):

New York City					Phoenix				
Hair	**Crimes**	**Pop.**	Rate	Pop Freq	**Hair**	**Crimes**	**Pop.**	Rate	Pop Freq
Blond	16	1	16	.02	**Blond**	40	5	8	.04
Brown	6	3	2	.06	**Brown**	20	20	1	.16
Black	12	6	2	.12	**Black**	35	35	1	.28
Red	52	13	4	.26	**Red**	84	42	2	.336
Grey	112	7	16	.14	**Grey**	64	8	8	.0640
None	1600	20	80	.40	**None**	600	15	40	.12
Total	1798	50	35.96		**Total**	843	125	6.744	

(a) What is the crude crime rate in NYC (in thousands)?

$\frac{1798}{50} = 35.96$

35.96

* notice that the strata-specific rate are all proportional by a factor of 2. also notice that the crude rate are NOT proportional by the same factor b/c they are using different populations.

(b) What is the crude crime rate in Phoenix (in thousands)?

$\frac{843}{125} = 6.744$

6.744

Problem 2 (continued)

(c) What is the ratio of directly adjusted crime rates (NYC to Phoenix), if Phoenix is used as the standard population?

Phoenix adjusted to itself is just the same crude rate 6.744.

NYC adjusted to phoenix is $A\text{Rate} = 16(.04) + 2(.16) + 2(.28) + 4(.336) + 16(.064) + 80(.12)$

$= 13.488$

Ratio of rates is $\frac{13.488}{6.744} = \boxed{2}$

(Check **one** box)

(d) Suppose you were able to pick any standard population and then you re-calculated the ratio of directly adjusted rates as you did in part (c) above. Then it is true that

- ☐ You could only make the ratio as large as you wish
- ☐ You could only make the ratio as small as you wish
- ☒ You can not change the ratio
- ☐ You can make the ratio as large or as small as you wish

↳ b/c the strata specific rates are proportional.

(Check **one** box)

(e) Suppose you were able to pick any standard population and then you calculated the same ratio using the indirectly adjusted rates. Then it is true that

- ☐ You could only make the ratio as large as you wish
- ☐ You could only make the ratio as small as you wish
- ☐ You can not change the ratio
- ☒ You can make the ratio as large or as small as you wish

↳ b/c the strat specific pop frequencies vary with some larger in NYC (i.e. Grey, None) and some larger in phoenix (e.g. Dbrd, Red)

* Note this is always true! b/c the pop freq must total to 100%, it can never be the case that all strat are larger or smaller in one group.

3. The following table shows data on 518 primary diagnosis stratified by age.

Age	Diabetes	Asthma	Arthritis	Cardiac	total
30-39	27	56	20	30	133
40-49	32	32	25	24	113
50-59	30	14	43	43	130
Over 60	29	7	65	41	142
total	118	109	153	138	518

(a) What is the probability of being diabetic and under 40 in this population?

$$P(D \text{ and under } 40) = \frac{27}{518} = \boxed{0.0521}$$

A = arthritic
O = over 60

(b) What is the probability of being Arthritic or over 60, but not both?

$$P(A \text{ OR } O, \text{But not Both}) = P[(A \text{ and } O^c) \text{ OR } (A^c \text{ and } O)]$$

$$P(A)+P(O)-2P(A \text{ and } O)$$

$$\frac{153+142-2(65)}{518} = \frac{165}{518}$$

From tables $= \frac{(20+25+43+29+7+41)}{518} = \frac{88+77}{518} = \boxed{\frac{165}{518} = 0.3185}$

	A	A^c	
O^c	88	288	376
O	65	77	142
	153	365	518

A = Asthmatic

(c) For individuals who are between 40 and 49, what is the probability of being Asthmatic?

$$P(A \mid 40\text{-}49) = \frac{P(A \text{ and } 40\text{-}49)}{P(40\text{-}49)} = \frac{32/518}{113/518} = \frac{32}{113} = 0.2832$$

D = diabetic
50+ = over 50

(d) What is the probability of being over 50 given that you are not Diabetic?

$$P(50+ \mid D^c) = \frac{P(50^+ \text{ and } D^c)}{P(D^c)} = \boxed{\frac{213}{400} = 0.5325}$$

	D	D^c	
50^+	59	213	272
50^-	59	187	246
	118	400	518

$$= \frac{(14+7+43+65+43+41)/518}{(109+153+138)/518}$$

$$= \frac{213}{400}$$

4. Suppose I take one Vioxx pill a day and the probability of an adverse event from Vioxx is 0.1. Assume that the effects of Vioxx are not cumulative, that is, the effect of Vioxx is independent from day to day.

(a) What is the probability that I will have at least one adverse event in a single week?

$Y = \#$ of AEs in one week $\quad P(AE) = .1 = \theta$

$Y \sim BIN(7, 0.1)$

$$P(Y \geq 1) = 1 - P(Y=0) = 1 - \binom{7}{0}\theta^0(1-\theta)^{7-0}$$

$$= 1 - (.9)^7$$

$$= 1 - .4783$$

$$= \boxed{0.5217}$$

(b) What is the probability that I have an adverse event on Monday and Friday of that week?

Let S = successful AE.

monday friday

$$P(SFFFSFF) = P(S)^2 \cdot P(F)^5$$

$$= (.1)^2 \cdot (.9)^5 = \boxed{0.0059}$$

Problem 4 (continued)

(c) What is the probability that I have two weeks in a row where I have adverse events on Monday and Friday?

$$P(SFFFSFFSFFFSFF) = (.1)^4(.9)^{10} = \boxed{(.0059)^2}$$

OR

$$[P(SFFFSFF)]^2 = [(0.1)^2(.9)^5]^2 \quad \text{b/c weeks are indp}$$

$$= \boxed{[.0059]^2}$$

$$= 3.48 \times 10^{-5}$$

(d) How many adverse events would I expect to have over a period of one year?

X = # of AEs in one year

$X \sim BIN(365, .1)$

$$E[X] = 365(.1) = \boxed{36.5 \text{ per year}}$$

(f) What is the standard deviation of the number of adverse events in one year?

$X \sim BIN(365, .1)$

$$SD(X) = \sqrt{Var(X)} = \sqrt{n\theta(1-\theta)}$$

$$= \sqrt{365(.1)(.9)}$$

$$= \boxed{5.73}$$

5. Scholastic Aptitude Test (SAT) scores are standardized into percentiles to make them comparable across different versions of the test. In Rhode Island, college-bound high school seniors have a mean SAT verbal score of 502 (with a standard deviation of 108, i.e., σ =108) and a mean SAT math score of 504 (σ =111). Assume that verbal and math scores are independent and normally distributed.

Scores $\begin{cases} V \sim N(502, 108^2) \\ M \sim N(504, 111^2) \end{cases}$

(a) What is the probability that a randomly selected senior will score between 340 and 664 on the verbal portion of the exam?

$$P(340 < V < 664)$$

$$= P\left(\frac{340-502}{108} < \frac{V-\mu}{\sigma} < \frac{664-502}{108}\right)$$

$$= P(-1.5 < Z < 1.5)$$

$$= 1 - 2(0.067) = \boxed{0.8660}$$

(b) What is the probability that a randomly selected senior will score over 600 on the verbal portion of the exam?

$$P(V > 600) = P\left(\frac{V-\mu}{\sigma} > \frac{600-502}{108}\right)$$

$$= P(Z > 0.9074) = 0.166$$

Note Rounding makes a difference

$$P(Z > 0.91) = 0.181$$

Full credit for either!

(c) Find the top 10th percentile of SAT verbal scores for seniors.

$$P(V > ?) = .1$$

$$P\left(\frac{V-\mu}{\sigma} > \frac{?-502}{108}\right) = .1$$

$$P\left(Z > \frac{?-502}{108}\right) = .1$$ But we know $P(Z > 1.28) = .1$

So $$1.28 = \frac{?-502}{108} \Rightarrow ? = 502 + 1.28(108)$$

$$? = \boxed{640.24 \text{ or } 640}$$

Problem 5 (continued)

(d) Suppose your friend is excellent in English and always scores 600 or more on the verbal SAT. What is the probability that your friend will score over 700 on the verbal SAT?

$$P(V > 700 \mid V > 600) = \frac{P(V>700 \text{ and } V>600)}{P(V>600)} = \frac{P(V>700)}{P(V>600)}$$

$P(V>600)$ from part (b)

$P(V>700) = P(Z > \frac{700-502}{108})$

$= P(Z > 1.83) = 0.034$

$$= \frac{0.034}{0.166} = \boxed{0.2048}$$

(f) What is the expected combined math and verbal score for a high school senior?

$$E[V+M] = E[V] + E[M] = 502 + 504 = \boxed{1006}$$

b/c $V \sim N(502, 108^2)$

$M \sim N(504, 111^2)$

(g) What is the standard deviation of the combined math and verbal score for a high school senior?

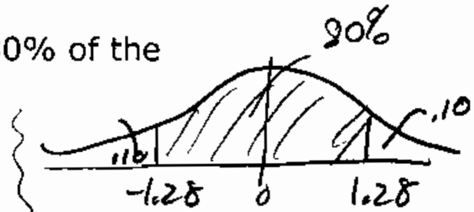
$$SD(V+M) = \sqrt{Var(V+M)} = \sqrt{Var(V) + Var(M) + 2Cov(V,M)}$$

Zero b/c V is indep of M

$$= \sqrt{108^2 + 111^2}$$

$$= \boxed{154.87}$$

(h) Calculate the interval that captures the middle 80% of the combined test score distribution for seniors.



$$P(?_{lower} < V+M < ?_{higher}) = .80$$

$$P\left(\frac{?_l - E[V+M]}{\sqrt{Var(V+M)}} < \frac{V+M-E[V+M]}{\sqrt{Var(V+M)}} < \frac{?_H - E[V+M]}{\sqrt{Var(V+M)}}\right)$$

$P(-1.28 < Z < 1.28) = .8$

$$P\left(\frac{?_l - 1006}{154.87} < Z < \frac{?_H - 1006}{154.87}\right) = .80$$

$$P(-1.28 < Z < 1.28) = .80$$

So $?_l = 1006 - 1.28(154.87) = 807.76$

$?_H = 1006 + 1.28(154.87) = 1204.23$

$\boxed{\text{Interval is } [807, 1204]}$

Problem 5 (continued)

Now suppose that a randomly selected senior takes the SATs 5 times in an attempt to raise his scores.

V_1, V_2, V_3, V_4, V_5 $\quad \bar{V}_5$ is the sample mean.

(i) What is the variance of the average verbal score?

$$VaR(\bar{V}_5) = \frac{VaR(V)}{5} = \frac{108^2}{5} = \boxed{2,332.8}$$

(j) What is the probability that the average verbal score will be greater than 550?

$$P(\bar{V}_5 > 550) = P\left(\frac{\bar{V}_5 - E[\bar{V}_5]}{\sqrt{VaR(\bar{V}_5)}} > \frac{550 - E[\bar{V}_5]}{\sqrt{VaR(\bar{V}_5)}} \right)$$

$E[\bar{V}_5] = E[V] = 502$

$VaR(\bar{V}_5)$ From (i)

$$= P\left(Z > \frac{550-502}{\sqrt{2332.8}}\right)$$

$$= P(Z > 0.9938) = \boxed{0.161}$$

(k) What is the probability that he scores over 550 on each verbal exam?

Let S = Scores over 550 on verbal Exam

then $P(SSSSS) = [P(S)]^5$ b/c Exams are indep

$= [0.33]^5$

$= \boxed{0.0039}$

and $P(S) = P(V > 550)$

$= P\left(Z > \frac{550-502}{108}\right)$

$= P(Z > 0.444)$

$= 0.33$

6. Suppose a voter poll is taken in three states. In state A, 50% of voters support the liberal candidate, in state B, 60% of the voters support the liberal candidate, and in state C, 35% of the voters support the liberal candidate. Of the total population of the three states, 40% live in state A, 25% live in state B, and 35% live in state C.

L = support liberal
A = state A
B = state B
C = state C

$P(L|A) = .5 \quad P(L|B) = .6 \quad P(L|C) = .35$

$P(A) = .4 \quad P(B) = .25 \quad P(C) = .35$

(a) How often do the voters support a liberal candidate?

$P(L) = ?$

$$P(L) = P(L|A)\cdot P(A) + P(L|B)\cdot P(B) + P(L|C)\,P(C)$$

$$= .5(.4) + .6(.25) + .35(.35)$$

$$= \boxed{0.4725}$$

*Law of total probability (weighted average of state-specific rates!)

(b) What is the probability that a voter lives in state B if we know that they supported the liberal candidate?

$P(B|L) = ?$

$$P(B|L) = \frac{P(L|B)\,P(B)}{P(L|B)\cdot P(B) + P(L|A)\cdot P(A) + P(L|C)\,P(C)}$$

$$= \frac{P(L|B)\cdot P(B)}{P(L)}$$

notice! b/c $P(B|L) = \dfrac{P(B \text{ and } L)}{P(L)}$

$$= \frac{.6(.25)}{.4725}$$

$$= \boxed{0.3175}$$

or use a table!

	A	B	C	
L	20	15	12.25	
$\bar{L}$	20	10	22.75	
	40	25	35	100

known from Q

$40(.5) = 20$

$25(.6) = 15$

$35(.35) = 12.25$

$P(L) = \dfrac{47.25}{100}$

made this up!

known from question

$$P(B|L) = \frac{P(B \text{ and } L)}{P(L)} = \frac{15/100}{47.25/100} = \frac{15}{47.25} = \boxed{0.3175}$$

Problem 6 (continued)

(c) What is the probability that two independent voters from state B do not support a liberal candidate?

$$P(\bar{L} \text{ and } \bar{L} | B) = [P(\bar{L} | B)]^2$$
$$= [1 - P(L|B)]^2$$
$$= (1 - .6)^2$$
$$= .4^2 = \boxed{.16}$$

(d) How often do two independent voters from the same state both fail to support a liberal candidate?

$$P(\bar{L} \text{ and } \bar{L} | \text{same state}) = ?$$

note this is different From
$$P(\bar{L} \text{ and } \bar{L}) = (P(\bar{L}))^2$$
$$= (1 - P(L))^2$$
$$= (.5275)^2$$
$$= 0.2783$$

$$P(\bar{L} \text{ and } \bar{L} | \text{same state}) =$$
$$P(\bar{L} \text{ and } \bar{L} | A)P(A) + P(\bar{L} \text{ and } \bar{L} | B)P(B) + P(\bar{L} \text{ and } \bar{L} | C)P(C)$$
$$= [1 - P(L|A)]^2 P(A) + [1 - P(L|B)]^2 P(B) + [1 - P(L|C)]^2 P(C)$$

b/c this includes people From two different states!

$$= (.5)^2 .4 + (.4)^2 .25 + (.65)^2 .35$$
$$= \boxed{0.2879}$$

NOT equal

note: mathematically we can write
$$P(\bar{L} \text{ and } \bar{L} | \text{same state}) = P(\bar{L} \text{ and } \bar{L} | A \text{ or } B \text{ or } C)$$
w/ A, B, C being mutually exclusive