PBAF 527: Quantitative Methods **Assignment 2 Answer Sheet**

Working with Probability

Part I

3.18

a. The sample points would be the possible answers to the question. Thus, the sample points would be: Infant, Child, Medical, Infant & Medical, Child & Medical, Infant & Child, and Infant & Child & Medical.

b. Reasonable probabilities would be equal to relative frequencies. These would be:

P(Infant) = 1,852/30,337 = .061

P(Child) = 17,148/30,337 = .565

P(Medical) = 8,377/30,337 = .276

P(Infant & Medical) = 44/30,337 = .001

P(Child & Medical) = 903/30,337 = .030

P(Infant & Child) = 1,878/30,337 = .063

P(Infant & Child & Medical) = 135/30,337 = .004

c. Define the following event: A = {medical reason}

P(A) = P(Medical) + P(Infant & Medical) + P(Child & Medical) + P(Infant & Child & Medical)

P(A) = .276 + .001 + .030 + .004 = .311

d. Define the following event: B = {infant not used as reason}.

P(B) = P(Child) + P(Medical) + P(Child & Medical)

P(B) = .565 + .276 + .030 = .871

3.32

a. The student is male and a binge drinker would be the event $B \cap A$.

b. The student is not a binge drinker would be the event A^C.

c. The student is male or lives in a coed dorm is the event $B \cup C$.

d. The student is female and not a binge drinker would be the event $B^C \cap A^c$.

3.46

Let A = {Company is trading company} and B = {Company based in Japan}. From the problem, P(B) = 11/20 = .55 and P(A∩B) = 6/20 = .3.

$$P(A\,|\,\mathrm{B}) = \frac{\mathrm{P(A \cap B)}}{P(B)} = \frac{.3}{.55} = .545$$

3.67

Let us define the following events:

S: {School is a subscriber}
N: {School never uses the CNN broadcast}
F: {School uses the CNN broadcast more than 5 times per week}

From the problem, $P(S) = .40$, $P(N|S) = .05$, and $P(F|S) = .20$

a. $P(S \cap N) = P(N|S)P(S) = .05(.40) = .02$
b. $P(S \cap F) = P(F|S)P(S) = .20(.40) = .08$

3.80

First, number the households from 1 to 534,322. Using the random number table, select a starting point that consists of 6 digits. Following either the row or column, select successive 6 digit numbers between 1 and 534,322 until 1000 different 6 digit numbers have been selected. The sample will consist of the 1000 households corresponding to the 1000 different numbers.

Part II

Here is an abbreviated example of Part II.

Here's what the SPSS Crosstabulation looked like:

PERSON*S SEX * PERSON OF HISPANIC ORIGIN Crosstabulation

			PERSON OF HISPANIC ORIGIN		Total
			0.NO	1.YES	
PERSON*S SEX	1.MALE	Count	7860	693	8553
		% within PERSON*S SEX	91.9%	8.1%	100.0%
		% within PERSON OF HISPANIC ORIGIN	47.4%	49.7%	47.6%
		% of Total	43.7%	3.9%	47.6%
	2.FEMALE	Count	8717	700	9417
		% within PERSON*S SEX	92.6%	7.4%	100.0%
		% within PERSON OF HISPANIC ORIGIN	52.6%	50.3%	52.4%
		% of Total	48.5%	3.9%	52.4%
Total		Count	16577	1393	17970
		% within PERSON*S SEX	92.2%	7.8%	100.0%
		% within PERSON OF HISPANIC ORIGIN	100.0%	100.0%	100.0%
		% of Total	92.2%	7.8%	100.0%

This chart shows the percentages of men and women who and whether they are of Hispanic origin among those who responded to the 2004 Washington State Population Survey. Among Hispanics, the gender split is almost equal, with 49.7 percent of respondents being men and 50.3% being women. Of the men who responded, only 8.1 percent were Hispanic. A smaller percentage of women (7.4 percent) were Hispanic. Out of all respondents, only 7.8 percent were Hispanic.

PBAF 527: Quantitative Methods **Assignment 2**
Answer Sheet

Part III

1. The overall poverty rate is the sum of the rate for each group reported (e.g. each region) weighted by its population share. Let Pov stand for the event "poverty", NE for northeast, MW is for Midwest, S is for south, and W is for west. Because the region groups are mutually exclusive and exhaustive we can add the region-specific poverty rates (conditional probabilities) weighted by the proportion of people in each region (to get the "and" events—see book for the multiplicative rule).

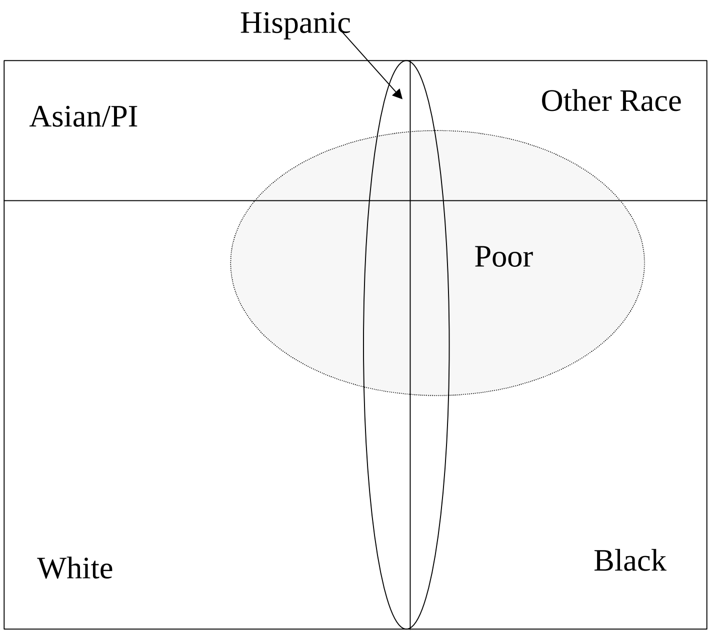
   P(Pov) =P(Pov and NE) + P (Pov and MW) + P(Pov and S) + P (Pov and W)
   $=P(Pov \cap NE) + P(Pov \cap MW) + P(Pov \cap S) + P(Pov \cap W)$
   =P(Pov|NE) x P(NE) + P(Pov|MW) x P(MW) + P(Pov|S) x P(S) + P(Pov|W)x P(W)
   =(.109)(.19) + (.103)(.23) + (.138)(.36) + (.124)(.23)
   = .121

   Thus, the overall poverty rate in 2003 was 12.1 percent. If the poverty rates for any groups or the proportion of the population in groups changes, then the overall probability will change. Similarly, for 1993 using the age groups we could calculate a poverty rate of 15 percent. Let C stand for under 18 (child), A stand for 19-64 (adult), and O stand for 65 and above (older).

   P(Pov) =P(Pov and C) + P (Pov and A) + P(Pov and O)
   $=P(Pov \cap C) + P(Pov \cap A) + P(Pov \cap O)$
   =P(Pov|C) x P(C) + P(Pov|A) x P(A) + P(Pov|O) x P(O)
   =(.227)(.26) + (.123)(.62) + (.122)(.12)
   =.150

   So, the overall poverty rate in 1993 was 15 percent. The overall poverty rate fell from 15 percent in 1993 to 12.1 percent in 2003.

2. If poverty were statistically independent of race, then the poverty rates conditional on race should be equal. [In this data set, the white only, black only, Asian/Pacific Islander only, and Other Race/multiracial categories are mutually exclusive and exhaustive.] If P is poverty, W is white, B is Black, and As is Asian/Pacific Islander, and OR is other races, then by definition statistical independence is



P(Pov|W)=P(Pov|B)=P(Pov|As) =P(Pov|OR)=P(Pov)

But because P(Pov|W) = .105 and the P(Pov|B)= .244 in 2003, there is clear evidence that race provides information about the likelihood of being poor. What kind of data would you want to further explore the cause of race differences in poverty? Could race differences in age distribution, residence, or region explain the race differences in poverty rates?

3. We can use manipulate the conditional probability formula (and use Bayes' Theorem) to get the probability of being a child given poverty status. If C is for child (under 18), then

P(C|Pov)= $\frac{P(Pov \cap C)}{P(Pov)} = \frac{P(C)P(Pov|C)}{P(Pov)}$ [multiplicative rule of probability]

[You could stop here and solve the equation, but I'd like to point out that the equation above equals:

$$= \frac{P(Pov|C)P(C)}{P(Pov|C)P(C) + P(Pov|C^c)P(C^c)}$$

This happens to be Bayes Theorem. The denominators here and just above are equivalent.] [Do you see why?]

In 2003 e know P(Pov|C)= .176 from the chart, P(C)=.25 from the chart as well, and we calculated P(Pov) in problem 1, so

=(.176 x .25)/.121

= .364

So, the probability of a being child given poverty status was about 36 percent in 2003. That is, 36 percent of the poor were children. In the last 30 years, Social Security and Medicare programs resulted in large decreases in the probability of being poor for older people, and this makes children a larger part of the pool of poor people (can you see how this changes the formula?). Increases in single mother households between the 1960s and the 1980s pushed poverty rates for children higher. Also, the poverty rate for working-age adults in more sensitive to economic cycles than are the rates for young or old; the poverty rates for adults increase during recession making children a smaller portion of the total.

4. We want to know the proportion of poor people between 19 and 65 (A) in both 2003 and 1993—the probability of not being under 18 or over 65 given poverty status [P(A|Pov)].

We want, for 2003:

P(A|Pov) = P(A and Pov)/P(Pov)=P(A∩Pov)/P(Pov)

= [P(18-24 and Pov) + P((25-34 and Pov) + P((35-44 and Pov) + P((45-55 and Pov) + P((55-59 and Pov) + P((60-64 and Pov)]/P(Pov)

= [P(18-24|Pov)P(Pov) + P(25-34|Pov)P(Pov) + …]/P(Pov)

= [(.165)(.10) + (.128)(.14) + (.096)(.15) + (.076)(.14) + (.082)(.06) + (.097)(.04)]/.121

=0.56

Alternatively, we know the programs won't be useful for the children and seniors, so let's find the proportion of the poor in those groups then take the complement. We want P(C or O|Pov). Since C and O are mutually exclusive categories, we can add their conditional probabilities (because they have the same denominator):

P(C or O|Pov) =P(C|Pov) + P(O|Pov)

We figured the P(C|Pov) in the last question as .36, similarly

P(O|Pov) = P(Pov and O)/P(Pov)

= P(O)P(Pov|O)/P(Pov) [Multiplicative Rule which gets us to Bayes Theorem again]

=(.102)(.12)/.121

=0.101

So, P(C or O|Pov) = .36 + .10 = .46
Then, since A, Y, and O are mutually exclusive and exhaustive, take the complement:

$$P(A|P) = 1 - .46 = .54$$

For 1993, we can calculate the P(A|Pov) using numbers from the table and the overall poverty rate we calculated before:

$$\begin{aligned} P(A|Pov) &= P(Pov \text{ and } A)/P(Pov) \\ &= P(A)P(Pov|A)/P(Pov) \\ &= (.62)(.123)/.15 \\ &= .508 \end{aligned}$$

Thus, in 1993 just about half and in 2003 only a little over half of the poor were likely to use labor market anti-poverty policies. Note that children will often be living with poor adults, so a program successful for adults would also change the rates for children. Also important is the fact that some programs may not raise incomes above poverty (a fixed threshold) but could increase incomes closer to that rate.