PbAf527 **Midterm Answer sheet**
Winter 2005 **Klawitter**

Question I:
A) The average cigarette price is slightly higher for non-smokers ($1.89) than for smokers ($1.87). There is, however, slightly less variation in the prices faced by non-smokers than by smokers (s=.21 versus s=.23 for smokers). The average index measuring limits on access to cigarettes is also slightly more stringent (14.3 for non-smokers and 13.3 for smokers on a 22 point scale). The price and access differences are very small, but are consistent with a possible beneficial impact of higher price and stricter limits on smoking for teens.

B) To calculate the average price overall, we have to weight the average for each group by the proportion in each group.

$$E(\mathrm{Pr}\,ice) = E(\mathrm{Pr}\,ice \mid S)P(S) + E(\mathrm{Pr}\,ice \mid nonS)P(nonS)$$
$$= (\$1.87)(.28) + (\$1.89)P(1-.28)$$
$$= 1.88$$
$$E(Access) = E(Access \mid S)P(S) + E(Access \mid nonS)P(nonS)$$
$$= (13.3)(.28) + (14.3)P(1-.28)$$
$$= 14.02$$

So overall the price faced by the teens was $1.88 and the access index was just over 14.

C) Here we need to calculate the probability that a student is a smoker given that he or she faces a ban (Pr(S|B))from the pieces we have (the probability that a student who smokes faces a ban and the probability that a student smokes. We can do this with Baye's Theorem.

$$P(S \mid B) = \frac{P(B \mid S) * P(S)}{P(B \mid S) * P(S) + P(B \mid nonS) * P(nonS)}$$
$$= \frac{(.95 * (.28)}{(.95) * (.28) + (.96) * (1-.28)} = .278$$

So, about 28 percent of the people facing a school ban were smokers

D) We can calculate a similar rate of smoker for people who did NOT face a school ban:

$$P(S \mid noB) = \frac{P(noB \mid S) * P(S)}{P(noB \mid S) * P(S) + P(noB \mid nonS) * P(nonS)}$$

$$= \frac{((1-.95)*(.28)}{(1-.95)*(.28)+(1-.96)*(1-.28)} = .327$$

So, about 33 percent of the people not facing a school ban were smokers. So the rate of smoking was lower among those with a school ban (28%) than among those with no ban (33%). This is consistent with the school ban affecting the teen smoking rate. But are there other reasons why these would be related? [For example, students who smoke choosing schools with no smoking ban.]

E) Here we need to find the chances that teens are smokers AND have no rules and the chance that teens are non-smokers AND have no rules, then add them together to get the overall chance of no rules. We'll start by using the multiplicative rule, then add the two mutually exclusive probabilities together.

$$P(S and no R) = (1 - P(R \mid S))P(S)$$
$$= (1-.62)(.28) = .106$$
$$P(nonS and no R) = (1 - P(R \mid nonS))P(nonS)$$
$$= (1-.70)(1-.28) = .216$$

$$P(noR) = P(nonS and noR) + P(S and noR) = .106 + .216 = .322$$

Overall almost a third of students had no rules on their free time, so about two thirds had rules. In a typical group of teens, about 11 percent would smoke and have no rules and about 22 percent would not smoke and have no rules.

F) Here, we just need to add up the rates for the categories of discussions.

For Smokers:

$$P(D > 1 per wk) = P(D 1 per week) + P(D few) + P(D everyday) =$$
$$= 1 - P(D < 1 per wk) = 1 - .18 = .82$$

For non Smokers:

$$P(D > 1 per wk) = 1 - P(D < 1 per wk) = 1 - .15 = .85$$

So, only 82 percent of smokers discussed issues at least once per week, but 85 percent of non-smokers did.

G) Evidence from a 1996 of US teens provides some clues for your work to limit smoking among Washington state teens.

The study found that about a quarter of teens had smoked in the past month (28 percent). Rates were somewhat higher among boys than girls and among whites than among other racial/ethnic groups.

Evidence on the effectiveness of state policy variables is weak—students who had not smoked faced slightly higher cigarette prices ($1.89 per pack versus $1.87 for smokers) and had more limits on access to cigarettes (through minimum age of purchase, packaging and vending machine availability). These differences are small and further multivariate statistical analysis would be required to assess their robustness. There is also weak evidence on the effects of school smoking bans—32 percent of students smoked at schools without bans, but only 28 percent of students smoked at schools with bans on smoking. However, 95 percent of students attended schools that already had bans in place.

Evidence on the efficacy of family interactions is stronger. Students who did not smoke were more likely to have rules on their free time (70 percent versus 62 percent for smokers) and less likely to have parents who smoked (34 percent versus 46 percent for smokers). Although these actions may not drive teen smoking decisions, their association with smoking warrants further investigation.

On the whole, this study suggests that state policies could help curb teen smoking, but the effects would likely be small. It would be worthwhile to investigate how Washington's access and taxation policies compare with other states to assess potential impact.

A more fruitful path might be to target schools to assess school bans on smoking and to gain access to families of teens. The strongest connections to teen smoking came through family interactions (rules, parental smoking, and parent/teen interactions) and I urge the committee to look for further research assessing particular curriculum and health promotion initiatives.

These data are now 8 years old and came from a nationwide study. To better understand their relevancy for our purposes, I suggest looking for more recent data on teen smoking trends and how Washington teen smoking levels compare with the rest of the nation.

Question II:

A) Any student who scores 800 or higher on the SAT gets a score of 800. What percentage of students gets a reported score of 800?

Here we want to calculate the probability that a student scores 800 or above (in upper tail).

$$P(X > 800)$$
$$= P(Z > \frac{(X - m)}{s}) = P(Z > \frac{(800 - 500)}{100})$$
$$= P(Z > 3) = .5 - .4987 = .0013$$

So, less than 1 percent (.13 percent) of students will receive a score of 800 or above.

B) Here we need to have the score that leaves 25 percent of scores in upper tail.

So, we need to the Z* score so that P(Z>Z*)=.25 then convert to the test score. In the table, we get the P(0<Z<Z*) so we need to find the probability that leaves .25 in the upper table and .25 between Z and the mean.

From body of the table, find the probability closest to .25 --.
P(0<Z< **.67**)=.2486

To get X:

$$X = Z * s + m$$
$$= .67(100) + 500 = 567$$

So, the college should limit applications to those with an SAT with 567 or above if they want only those in the 75 percentile or above.

C) Now we need a central region that excludes the top 10 percent and the bottom 20 percent. So we need to find two Z scores corresponding to leaving out 10 percent and 20 percent in the tails, then convert to the test scores.

Want Z* so that Pr(Z>Z*)=.20
=.5 - Pr(0<Z<Z*)=.2

From body of the table, find the probability closest to **.3** (=.5-.2)--.
P(0<Z<. **84**)=.2995 But we want the negative (below the mean).

To get X:

$$X = Z * s + m$$
$$= (-.84)(100) + 500 = 416$$

Similarly, to get the top cut off of 10 percent:

Want Z* so that Pr(Z>Z*)=.10
=.5 - Pr(0<Z<Z*)=.1

From body of the table, find the probability closest to .**4** (=.5-.1)--.
P(0<Z<**1.28**)=.3997

To get X: $X = Z * s + m$

$$= (1.28)(100) + 500 = 628$$

So the college should expect to get applications from students with SAT scores ranging from 416 to 628.

D) Here we need a median SAT score conditional on the score being above the 75th percentile. So, we need to have the score for which the chance of being below is 50%. In this case we need to have the chance that 12.5% of scores would be above (50% of the 25% in this random group from the tail).

From body of the table, find the probability closest to .375 (.5-.125) --.
P(0<Z< **1.15**)=.3749

To get X: $X = Z * s + m$

$$= 1.15(100) + 500 = 615$$

So the median score would be 615, if the college got a random sample of scores above 567 (75th percentile).

What would the mean score be?
To start with, we know that all SAT scores would be above 567, so that the mean score would have to be above 567. From part A, we know that less than 1 percent of scores are over 800, so the average is likely to be much less than that. The probability of any given score is falling rapidly in this part of the distribution (because we are in the tail). Because we have a skewed distribution with a median of 615, we can guess that the mean is more than the median, so above 615. I might guess it would be around 620.

We could use the normal curve formula and generate an exact mean score for this part of the distribution. Or you could try to approximate it by using ranges of scores and their probabilities.