Student Number:______________________ **Feb 3, 2005**
Midterm Exam **Due Feb 4 at Noon**

Ground rules:

- Breathe!
- DO NOT WRITE YOUR NAME ON THE MIDTERM—write your student number. If you're worried about losing the first page, write your student number on each page.
- You can use your notes or books, a calculator or a spreadsheet, but you may not communicate with other people about this midterm nor the material covered by it.
- In order to receive as much credit as possible, please show all of your work. Showing that you understand the question and know how to set up the solution correctly is more important than arriving at the exact answer.
- Read each question carefully and answer all parts of each question.
- Good luck!
- When you are done, please put in my box in 208 Parrington (the Dean's office) or email it to me. If you put it in the box, please email me to let me know you have done so.

---

1. Locally, the city of Seattle used the Boost program to try to overcome the limitations of I-200 in contracting. The Boost program gave incentives to prime contractors to use small businesses in their work on seven major projects locally (including the new City Hall and Central Library). Since most women's businesses are small, the idea was this was one way to approach the limitations of I-200, which prevents the use of affirmative action in local government contracting (among other things). Prior to I-200, women made up 20% of city contracts. After (even with the BOOST program) women make up on 3.75% of city contracts.

Let's take a look at what we know about the small businesses and woman owned-businesses nationally to see we can come up with a plausible explanation of why this policy did not work. Nationally, women own 30% of businesses in the U.S. 98% of women's businesses have revenues less than $1 million, so most are small businesses. Furthermore, small businesses make up 80% of companies nation-wide.

a. Nationally, what's the probability of being a woman-owned business and being a small business?

Events: W=Women-owned business; S=Small Business

From the paragraph, we know:

$P(W)=.30$

$P(S)=.80$

$P(S|W)=.98$

We want $P(W\cap S)$. Using the multiplicative rule, we know that $P(W\cap S)=P(S|W)P(W)$, so $P(W\cap S)=(.98)(.30)=.294$

So, 29% of all businesses are both owned by a woman and are small.

b. What's the probability of being a women-owned business or being a small business?

We want $P(W\cup S)$. Using the rule of unions, we know that

$P(W\cup S)=P(W)+P(S)-P(W\cap S)=.30+.80-.29=.81$

81% of all businesses in the U.S. are small or owned by a woman.

c. What's the probability that if one owns a small business the owner is a woman?

We want $P(W|S)$. The definition of conditional probability tells us:

$$P(W\mid S)=\frac{P(W\cap S)}{P(S)}=\frac{.294}{.80}=.3675$$

So, given that one has a small business, there is a 38% probability that the business is woman-owned.

d. Are being a small business and being a woman-owned business independent? Make sure to show your work.

If any of these statements are false, then they are not independent.

$P(W|S)=P(W)$? $.37 \neq .30$

$P(S|W)=P(S)$? $.98 \neq .80$

$P(S \cap W)=P(S)P(S)$? $.29 \neq .80^*.30$ $.80^*.30=.24$

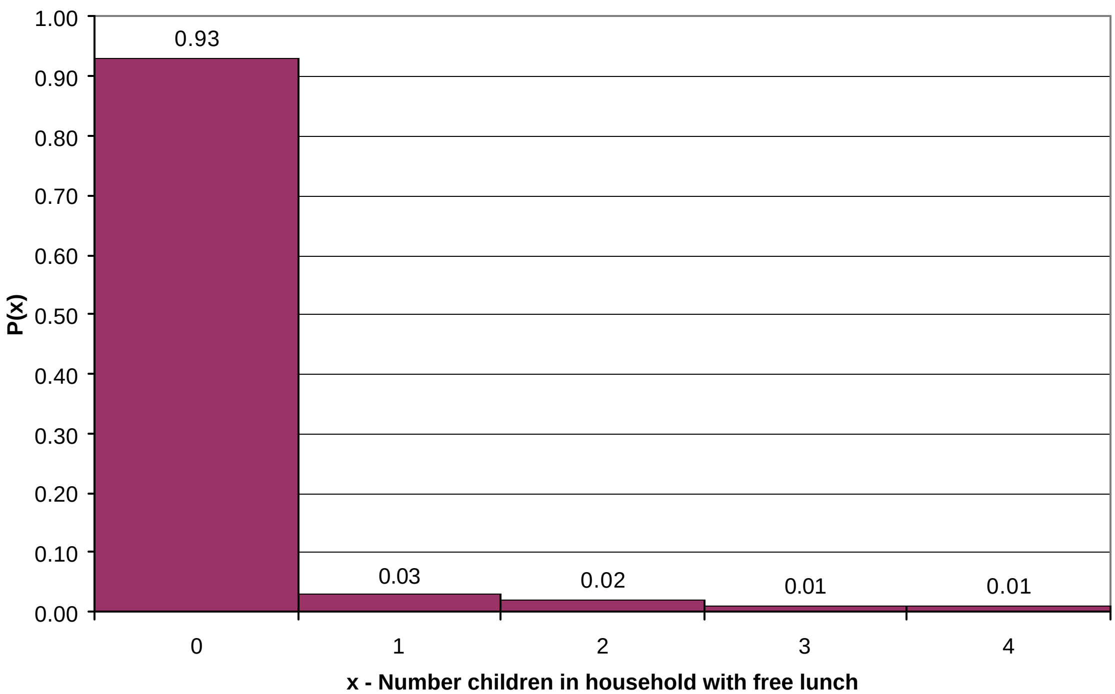
We could have stopped at the first one. Being a woman-owned business and being a small business are not independent—they are dependent. So, being a women-owned business tells you something about the size of the business.

e. Write a paragraph interpreting your results and using what you've learned about small and woman-owned businesses nationally to talk about why the Boost program may not have worked as intended. Write for a non-technical policy audience.

Although most woman-owned businesses are small, not all small businesses are owned by women. While 98% of woman-owned businesses are small, and 80% of businesses are small, only 30% of all business are owned by women. 81% of all businesses are small or owned by a women, but only 29% of businesses are both owned by a woman and are small. Of small businesses, 37% are owned by women. While being a woman-owned business does provide some information about the size of the business, this does not mean that most small businesses are owned by women.

This dynamic may also be true locally and may account for the lack of focus on woman-owned businesses in the Boost program. Since the Boost program focused on small businesses and (if we assume what is true nationally is also true locally) most small businesses are actually owned by men, then it is not a surprise that the program was not more successful in targeting woman-owned businesses.

2. The March 2004 Current Population Survey asked households in the U.S. how many children in their household received free lunch at their schools. The chart below presents the probability distribution for 112,116,531 U.S. households in March 2004. The free program is part of the National School Lunch Program to provide nutritionally balanced low-cost or free lunches to children below the age of 18.



a. What's the probability of randomly selected household having no children in the free lunch program?

From the chart: $P(x=0)=.93$

There is a 93% probability that the household will have no kids in the free lunch program.

b. What's the probability that at least one child in a household is in the free lunch program?

$P(x>0)=P(x=1)+P(x=2)+P(x=3)+P(x=4)=.03+.02+.01+.01=.07$

So, 7% of households have one or more children who receive a free lunch.

c. What's the average number of children who receive a free lunch in a household?

x	p(x)	xp(x)
0	0.93	0
1	0.03	0.03
2	0.02	0.04
3	0.01	0.03
4	0.01	0.04
		Sum=0.14

So, on average, less than one child in a household receives a free lunch.

d. What's the standard deviation of the number of children in a household who receive a free lunch?

x	p(x)	xp(x)	x^2	(x^2)p(x)
0	0.93	0	0	0
1	0.03	0.03	1	0.03
2	0.02	0.04	4	0.08
3	0.01	0.03	9	0.09
4	0.01	0.04	16	0.16
		Sum=0.14		$E(x^2)=0.36$

$\sigma^2 = E(x^2) - E(x)^2 = .36 - (.14)^2 = .36 - .0196 = .3404$

$s = \sqrt{s^2} = \sqrt{.3404} = .58$

So, the standard deviation is .58

e. Based on the information presented, what does the free lunch program cost the federal government? Reimbursements for the free lunch program are $2.24 a child.

Told there are 112,116,531 households. P(x) tells us the percentage of households with a particular number of kids. That's the third column. The fourth column is the number of kids living in those households, summing to 15,696,314.34 kids. 15696315*$2.24= $ 35,159,746.

x	P(x)	number households	number kids
1	0.03	3,363,495.93	3,363,495.93
2	0.02	2,242,330.62	4,484,661.24
3	0.01	1,121,165.31	3,363,495.93
4	0.01	1,121,165.31	4,484,661.24
			15,696,314.34

f. Write a paragraph explaining your results to a non-technical policy maker.

The school free lunch program by no means serves every household in the U.S., as there is a 93% probability that a U.S. household will have no kids in the free lunch program. This could be because the household is too wealthy for their children to participate or because there are no kids in the household. This means that 7% of households have one or more children who receive a free lunch, and that, on average, less than one child in a household receives a free lunch. The is very little variation in the number of children in a household who receive a free lunch. Although the percent of households served seems small, we estimate that 15,696,315 children are served by the program, costing the federal government over $35 million.

3. Serum cholesterol is an important risk factor for coronary disease. Epidemiologists have shown that when you take the natural log (usually written $\log_e$ or ln) of serum cholesterol, it is normally distributed with mean 5.39 and standard deviation 0.23. Suppose the clinically normal range for cholesterol is 150-250 mg%/mL [note, ln(150)=5.01 and ln(250)=5.52].

   a. What proportion of the population has normal levels of cholesterol?

   Want P(150<x<250) or, since the mean and standard deviation are in natural log terms, P(5.01<x<5.52).

   Convert both ends to z scores:

$$z = \frac{x-m}{s} = \frac{5.01-5.39}{.23} = \frac{-.38}{.23} = -1.65 \qquad z = \frac{x-m}{s} = \frac{5.52-5.39}{.23} = \frac{.13}{.23} = .565$$

   P(5.01<x<5.52)=P(-1.65<z<.565)

Use the table to find P(-1.65<z<0). Since the curve is symmetric, this is the same as wanting P(0<z<1.65). From the table, P(0<z<1.65)=.4505.

Use the table to find P(0<z<.565). From the table, P(0<z<.565)=.2130 (I split the difference between .2123 and .2157)

P(5.01<x<5.52)=P(-1.65<z<.565)= P(-1.65<z<0)+ P(0<z<.565)=.4505+.2130=.6635

So, 66% of the population has normal cholesterol levels.

   b. What proportion of people has abnormally low levels of cholesterol?

Want P(x<150) or, since the mean and standard deviation are in natural log terms, P(x<5.01).

Convert to Z score: $z = \frac{x-m}{s} = \frac{5.01-5.39}{.23} = \frac{-.38}{.23} = -1.65$, so we need to use the table to find P(z>-1.65). Since the curve is symmetric, this is the same as wanting P(z>1.65)

From the table, P(0<z<1.65)=4505. P(z>1.65)=P(z>0)-P(0<z<1.65)=.5-.4505=.0495.

So, about 5% of people have abnormally low levels of cholesterol.

c. What proportion of people has abnormally high levels of cholesterol?

Want P(x>250) or, since the mean and standard deviation are in natural log terms, P(x>5.52).

Convert to Z score: $z = \frac{x-m}{s} = \frac{5.52-5.39}{.23} = \frac{.13}{.23} = .565$, so we need to use the table to find P(z>.565).

From the table, P(0<z<.565)=.2130 (I split the difference between .2123 and .2157) To get P(z>0.565)=P(z>0)-P(0<z<.565)=.5-.2130=.287.

So, 29% of people have abnormally high levels of cholesterol.

Some investigators feel that only cholesterol levels over 300 mg%/mL indicate a high risk of heart disease. [Note, ln(300)=5.70]

d. What proportion of the general population do those with levels over 300 mg%/mL represent?

Want P(x>300) or, since the mean and standard deviation are in natural log terms, P(x>5.70).

Convert to Z score: $z = \frac{x-m}{s} = \frac{5.70-5.39}{.23} = \frac{.31}{.23} = 1.35$, so we need to use the table to find P(z>1.35).

From the table, P(0<z<1.35)=.4115.

P(z>1.35)=P(z>0)- P(0<z<1.35)=.5-.4115=.0885.

So, if the high cut-off for cholesterol is 300 mg%/mL, then 9% of people have abnormally high levels of cholesterol.

e. What proportion of those with abnormally high levels of cholesterol do those with levels over 300 mg%/mL represent?

Those with levels over 250 mg%/ML are 29% of the population (abnormally high)

Those with levels over 300 mg%/ML are 9% of the population.

.09/.29=.31. So, 31% of those with abnormally high levels have levels above 300 mg%/ML.

f. Write up and interpret your results in 3a-e for a non-technical audience.

Serum cholesterol is an important risk factor for coronary disease. If the clinically normal range for cholesterol is 150-250 mg%/mL, 66% of the population has normal cholesterol levels, 5% have abnormally low levels of cholesterol, and 29% of people have abnormally high levels of cholesterol. Some investigators, however, feel that only cholesterol levels over 300 mg%/mL indicate a high risk of heart disease. If this is the case, then only 9% of people have abnormally high levels of cholesterol. That is, 31% of those with abnormally high levels have levels above 300 mg%/ML. So, the decision of a high cut-off for cholesterol has large implications for public health, leaving over 2/3s of those who have levels over 250 mg%/mL uncertain if they should be concerned with their cholesterol levels or not.