Midterm Examination Review
CSSS/STAT/SOC 321: Case-Based Social Statistics I

Professor: Chris Adolph, Political Science and CSSS

Teaching Assistant: Aaron Erlich, Political Science

Fall Quarter 2012

October 26, 2012

1 Exam rules

- You may use a calculator, graphing calculator, or phone based calculator, but you may *not* store information in memory or use the internet or any other form of electronic communication. I strongly recommend you bring some sort of calculator to save time on arithmetic.
- You will be given a list of helpful formulas, reproduced below.
- If necessary, grades may be curved upwards, but not downwards.

2 What to expect

Among the questions on the exam, you might see problems which require you to:

1. Summarize the central tendency and variation of a sample. You may need to justify why you are using a mean, median, or mode; a standard deviation or a quantile.
2. Interpret a scatterplot or the coefficients of a fitted regression line.
3. Interpret a contingency table in column percentage format.
4. Solve probability problems like the two examples below.

But there might be other questions drawing on the concepts listed below as well.

3 Concepts to know

population
sample
observation
experiment
interval validity
external validity
confounder
selection bias
measurement bias
level of measurement
discrete variable
continuous variable
binary variable
ordered variable
nominal variable
additive scale
ratio scale
histogram
central tendency
mean
median
mode
dispersion
range
quantile
variance
standard deviation
logarithm
log scale
outlier
robustness
expected value
sampling error
correlation
causation
correlation coefficient
stochastic relationship
deterministic relationship
scatterplot
regression line
regression coefficient
monotonic relationship
contingency table
column percentages
Simpson's Paradox
event
sample space
frequency
probability
independence
marginal probability
joint probability
conditional probability
Bayes' Rule

4 Formulas that will be provided on the exam

Concept	Formula
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard deviation of a sample	$\hat{\sigma}_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2}$
Complement rule	$\Pr(A) = 1 - \Pr(\text{not } A)$
Addition rule (general)	$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
Addition rule (disjoint events)	$\Pr(A \cup B) = \Pr(A) + \Pr(B)$ if $\Pr(A \cap B) = 0$
Conditional probability	$\Pr(A\|B) = \Pr(A \cap B)/\Pr(B)$
Joint probability (independent events)	$\Pr(A \cap B) = \Pr(A)\Pr(B)$ if A and B are independent
Joint probability (generally)	$\Pr(A \cap B) = \Pr(A)\Pr(B\|A)$
Bayes' rule	$\Pr(A\|B) = \Pr(B\|A)\Pr(A)/\Pr(B)$

5 Example Problem: Extra-sensory Perception Experiment

Researchers are recruiting subjects for an experiment on extra-sensory perception (ESP). The experimenter looks at a randomly drawn card with one of four symbols, and asks the subject to guess which of the four symbols is on the card. A normal person, with no ESP, can do no better than guessing at random, and gets it right with probability 1 in 4. But suppose some people actually have ESP, and can guess the right card with probability 2 in 5.

The researchers decide to recruit into their study anyone who takes the test three times, and guesses the correct symbol each time.

Part A. What is the probability a "normal" person will get into the experiment?

Solution: A normal person guesses in each of the three tests, getting it right with probability 1/4 each time. Because success in each trial is independent, we can calculate the joint probability of three successes as:

$$\Pr(\text{Success in trial 1} \cap \text{Success in trial 2} \cap \text{Success in trial 3}) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{64} = 0.015625$$

Part B. What is the probability someone with ESP gets into the experiment?

Solution: Each event is still independent, so we use the same approach, but with the higher probability of success at each trial:

$$\Pr(\text{Success in trial 1} \cap \text{Success in trial 2} \cap \text{Success in trial 3}) = \frac{2}{5} \times \frac{2}{5} \times \frac{2}{5} = \frac{8}{125} = 0.064$$

Part C. Suppose half of the population tested has ESP, and half does not. Given that someone is selected to enter the experiment, what is the probability they have ESP?

Solution: We want to calculate the conditional probability that someone has ESP given that they are in the experiment. There are many valid ways to do this. For example, you could use a tree diagram, or even apply Bayes' Rule. In this example solution, I will show what happens if the recruiters test 2 million people, creating a contingency table which makes it easy to solve the problem.

Step 1. Imagine there are 2 million people. We know the marginal $\Pr(\text{ESP}) = 0.5$. So we can fill in some of the margins of the table:

	ESP	No ESP	
Accepted			
Not accepted			
	1,000,000	1,000,000	2,000,000

Step 2. From parts A. and B., we know the probability that someone with ESP gets accepted, $\Pr(\text{accepted}|\text{ESP}) = 0.064$ and the probability someone without ESP gets accepted, $\Pr(\text{accepted}|\text{no ESP}) = 0.015625$. We can use these to fill out the cells:

	Has ESP	No ESP	
Accepted	64,000	15,625	79,625
Rejected	936,000	984,375	1,920,375
	1,000,000	1,000,000	2,000,000

Step 3. Recall the probability interpretation of the contents of the contingency table:

	ESP	No ESP	
Accepted	$\Pr(\text{Accepted} \cap \text{ESP})$	$\Pr(\text{Accepted} \cap \text{No ESP})$	$\Pr(\text{Accepted})$
Not accepted	$\Pr(\text{Not accepted} \cap \text{ESP})$	$\Pr(\text{Not accepted} \cap \text{No ESP})$	$\Pr(\text{Not accepted})$
	$\Pr(\text{ESP})$	$\Pr(\text{No ESP})$	1

Also recall the definition of conditional probability, $\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$. In our case:

$$\Pr(\text{ESP}|\text{recruited}) = \frac{\Pr(\text{ESP} \cap \text{Accepted})}{\Pr(\text{Accepted})}$$

We can get the appropriate quantities from the table, and divide:

$$\Pr(\text{ESP}|\text{recruited}) = \frac{64{,}000}{79{,}625} = 0.8038$$

6 Example Problem: Will Homer get a parking ticket?

Homer is late for work and can't find a legal parking space. He thinks he might escape a ticket even if he parks illegally. Springfield has only one police officer, Chief Wiggum, and out of the 40 hours he works a week, Chief Wiggum only devotes 1 hour to ticketing illegally parked cars.

But Homer is worried, because he just saw Chief Wiggum driving around campus. Chief Wiggum is easy to spot when he's writing tickets: in fact, Homer spots him 80% of the time during his ticketing hours. When Chief Wiggum is doing other police work, he's much less visible, and Homer only spots him 10% of the time.

Is Homer likely to get away with parking illegally? What is the probability that Chief Wiggum is on parking duty this hour?

Solution:

Notice the question is asking us to find the probability that Chief Wiggum is writing tickets *given* that Homer has spotted him on campus. This is a *conditional probability*.

We want to find $\Pr(\text{ticketing}|\text{spotted})$.

What do we know?

- We know the marginal probability Chief Wiggum will be writing tickets during a given hour, $\Pr(\text{ticketing}) = 1/40$.

- We know the conditional probability Homer will sight Chief Wiggum if he is writing tickets, $\Pr(\text{spotted}|\text{ticketing}) = 4/5$.

- We know the conditional probability Homer will sight Chief Wiggum if he is not writing tickets, $\Pr(\text{spotted}|\text{not ticketing}) = 1/10$.

We know enough to use Bayes' Rule:

$$\begin{aligned}
\Pr(\text{ticketing}|\text{spotted}) &= \frac{\Pr(\text{spotted}|\text{ticketing})\Pr(\text{ticketing})}{\Pr(\text{spotted})} \\
&= \frac{\Pr(\text{spotted}|\text{ticketing})\Pr(\text{ticketing})}{\Pr(\text{spotted}|\text{ticketing})\Pr(\text{ticketing}) + \Pr(\text{spotted}|\text{not ticketing})\Pr(\text{not ticketing})} \\
&= \frac{\frac{4}{5} \times \frac{1}{40}}{\frac{4}{5} \times \frac{1}{40} + \frac{1}{10} \times \frac{39}{40}} \\
&= \frac{0.02}{0.02 + 0.0975} \\
&= 0.170
\end{aligned}$$

The probability that Chief Wiggum is ticketing cars is only 17.0%, even though Homer has spotted him on campus.