Midterm Review
STAT/SOC/CSSS 221:
Statistical Concepts and Methods for the Social Sciences

Professor: Chris Adolph, Political Science and CSSS
Winter Quarter 2012
February 3, 2012

Exam rules

• You may use a calculator or graphing 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.

• Remember: Final grades will be curved, using techniques we will learn later in the quarter, but only to your advantage, not to your detriment.

Concepts to know

population
sample
observation
experiment
interval validity
external validity
confounder
level of measurement
discrete variable
continuous variable
binary variable
ordered variable
nominal variable
additive scale
total scale
histogram
central tendency
mean
median
mode
dispersion
standard deviation
range
quantile
variance
standard deviation
logarithm
log scale
outlier
robustness
expected value
correlation
causation
correlation coefficient
stochastic relationship
deterministic relationship
scatterplot
regression line
regression coefficient
monotonic relationship
contingency table
column percentages
Simpson’s Paradox
What to expect

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

1. Answer questions about 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. You should be comfortable with histograms and boxplots.

2. Interpret a scatterplot or the coefficients of a fitted regression line, as well as explain the difference between a correlation coefficient and a regression coefficient.

3. Interpret a contingency table in column percentage format.

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

1 Scatterplot Example

\[
\begin{align*}
\text{weight}_t &= \beta_0 + \beta_1 \text{height}_t + \varepsilon_t \\
\hat{\text{weight}}_t &= -109.8 + 48.4 \times \text{height}_t \\
\end{align*}
\]

\[ r = 0.43 \]

Figure 1: Weight in pounds and height in feet for a sample of 21 year old males. The best fit line is described by the equation at top right. Source: Centers for Disease Control.

You should be able to:

1. Calculate the fitted weight for specific heights
2. Provide caveats to the above, such as possible confounders
3. Discuss the strength and direction of the relationship, and be ready to identify outliers
4. Distinguish extrapolation and interlopation, and the trustworthiness of each
### Hours of TV watched per day

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1 to 2</th>
<th>3 to 5</th>
<th>6 to 10</th>
<th>&gt; 10</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Happy</td>
<td>513</td>
<td>4718</td>
<td>3387</td>
<td>600</td>
<td>101</td>
<td>9319</td>
</tr>
<tr>
<td>Pretty Happy</td>
<td>702</td>
<td>7857</td>
<td>6731</td>
<td>1435</td>
<td>193</td>
<td>16918</td>
</tr>
<tr>
<td>Not Too Happy</td>
<td>154</td>
<td>1278</td>
<td>1371</td>
<td>479</td>
<td>106</td>
<td>3388</td>
</tr>
<tr>
<td>Sum</td>
<td>1369</td>
<td>13853</td>
<td>11489</td>
<td>2514</td>
<td>400</td>
<td>29625</td>
</tr>
</tbody>
</table>

### Hours of TV watched per day

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1 to 2</th>
<th>3 to 5</th>
<th>6 to 10</th>
<th>&gt; 10</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Happy</td>
<td>37.5%</td>
<td>34.1%</td>
<td>29.5%</td>
<td>23.9%</td>
<td>25.3%</td>
<td>31.5%</td>
</tr>
<tr>
<td>Pretty Happy</td>
<td>51.3</td>
<td>56.7</td>
<td>58.6</td>
<td>57.1</td>
<td>48.3</td>
<td>57.1</td>
</tr>
<tr>
<td>Not Too Happy</td>
<td>11.2</td>
<td>9.2</td>
<td>11.9</td>
<td>19.1</td>
<td>26.5</td>
<td>11.4</td>
</tr>
<tr>
<td>Sum</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 1: Cross-tabulation of self-reported happiness and daily television-watching happiness. The rows show the answer to the question: “Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy.” Source: General Social Survey.

### 2 Contingency Table Example

You should be able to:

1. Identify the elements of the table, particularly the cell and marginal counts
2. Discuss the relationship between the column and row variable using the appropriate table(s)
3. Provide caveats to the above, such as possible confounders