

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

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

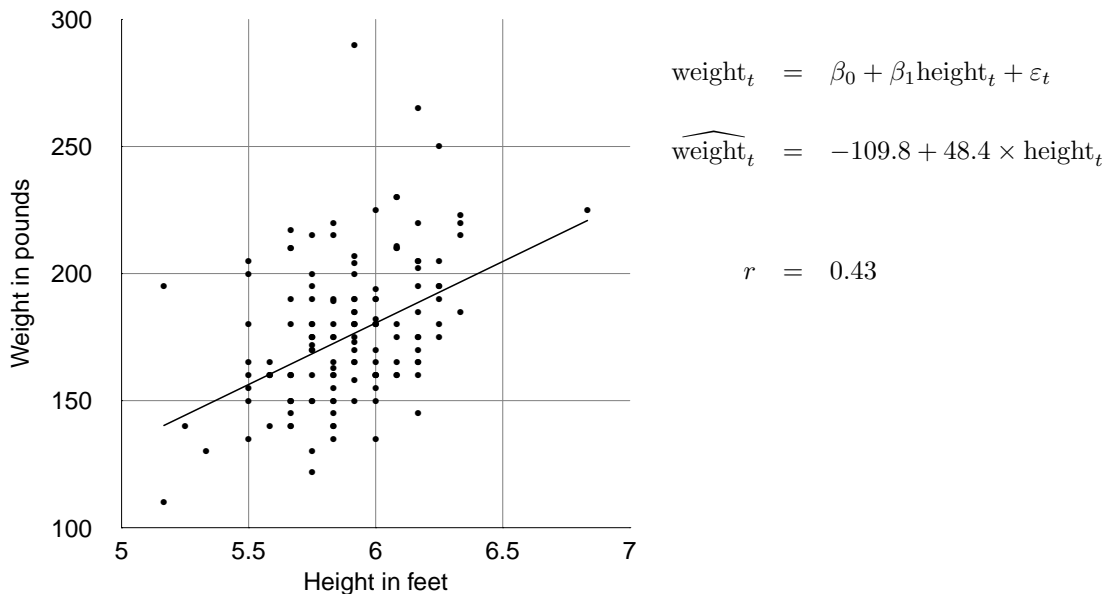


Figure 1: Weight in pounds and height in feet for a sample of 21 year old males. The best fit line is described by the equation at top right. *Source:* Centers for Disease Control.

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 interpolation, and the trustworthiness of each

	Hours of TV watched per day					Sum
	0	1 to 2	3 to 5	6 to 10	> 10	
Very Happy	513	4718	3387	600	101	9319
Pretty Happy	702	7857	6731	1435	193	16918
Not Too Happy	154	1278	1371	479	106	3388
Sum	1369	13853	11489	2514	400	29625

	Hours of TV watched per day					Sum
	0	1 to 2	3 to 5	6 to 10	> 10	
Very Happy	37.5%	34.1%	29.5%	23.9%	25.3%	31.5%
Pretty Happy	51.3	56.7	58.6	57.1	48.3	57.1
Not Too Happy	11.2	9.2	11.9	19.1	26.5	11.4
Sum	100.0	100.0	100.0	100.0	100.0	100.0

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