

17.871, Political Science Lab

Spring 2004

Problem set # 4: Multiple regression, sampling, and hypothesis testing

Handed out: March 18, 2004

Due back: April 6 (April 8 for those scheduled to present on April 6)

1. We are interested in explaining how respondents to public opinion polls judge the morality of President Bill Clinton. We suspect that Republicans will think he was less moral than Democrats. But, we also know/believe that the Republican party increasingly has as its core church-going southern Protestants. Therefore, the dislike of Clinton by Republicans may be more due to the religiosity of Republicans, rather than more traditional partisan differences. To disentangle the separate effects of strength of religious beliefs from partisanship, we resort to multiple regression.

The data are taken from the 2000 American National Election Study.

The following is how the three variables of interest were measured:

clinton_moral: Based on the answer to the question “How well do you believe the following trait describes President Bill Clinton: moral?” 1 = extremely well, 2 = quite well, 3 = not too well, 4 = not well at all.

party_id: party identification. 0 = strong Democrat, 1 = weak Democrat, 2 = Democrat-leaning Independent, 3 = Pure Independent, 4 = Republican-leaning Independent, 5 = weak Republican 6 = strong Republican.

attend: How often the respondent attends religious services. 1 = every week, 2 = almost every week, 3 = once or twice a month, 4 = a few times a year, 5 = never

The following is the variance-covariance matrix between the relevant variables in the regression:

	clinton_moral	party_id	attend
clinton_moral	0.718359		
party_id	0.625115	4.59037	
attend	-0.083485	-0.263852	1.45385

Calculate both the *bivariate* and *multivariate* regression coefficients with **clinton_moral** as the independent variable and **party_id** and **attend** as the dependent variable(s).

2. Students of public opinion are interested in how voters develop perceptions of the policy world. One interesting question is whether people believe crime is getting better or worse and how that relates to the types of information people receive.

People who read newspapers tend to believe crime is not as much of a problem as people who don't read newspapers. People who watch lots of game shows tend to believe crime is more of a problem than people who don't watch game shows.

Consider the following series of regressions. The dependent variable in each case is "crimeworse," which is coded 1 if the respondent believes crime has gotten worse in the past year, 0 if it has stayed the same, and -1 if it has gotten better. The independent variables are "newspaper" (the number of times each week the respondent reads the newspaper) and "wheel," the number of times in the past month the respondent watched "Wheel of Fortune." (The data are taken from the 2000 American National Election Study.)

	(1)	(2)	(3)
newspaper	-0.036 (0.010)	—	-0.037 (0.010)
wheel	—	0.050 (0.017)	0.053 (0.017)
intercept	0.062 (0.045)	-0.059 (0.029)	0.071 (0.045)
n	863	834	834
r ²	.015	.010	.027
rmse	0.83	0.83	0.83

What does this pattern of regression coefficients tell us about the correlation between newspaper reading and watching Wheel of Fortune?

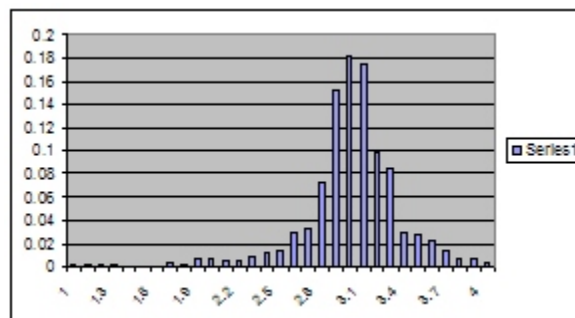
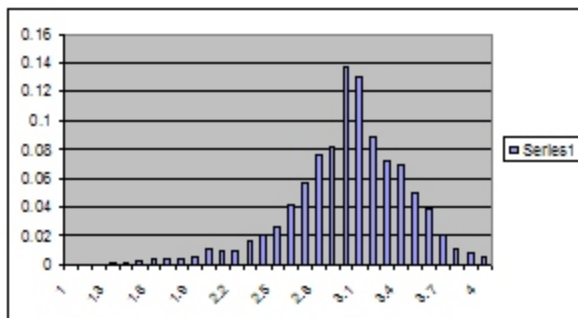
3. True or false: The average of 10 random draws from a distribution with population mean μ is 10 times more variable than the average of 1,000 random draws from the same distribution.
4. If we gave everyone in America an IQ test, we would get an average of 100 with a standard deviation of 16. It would be normally distributed. If we draw a sample of 100 people,

- 4.a what is the probability that the average of the sample will be more than 105?
- 4.b what fraction of the sample should we expect to be between 90 and 110?
- 4.c what fraction of the sample should we expect to be above 140?
5. You are conducting research into how much money MIT undergraduates spent on their spring break trip. You send out a survey that asks people to tell you, in ranges, how much they spent. Here are the results:

\$0	150
\$0-\$100	50
\$100-\$200	40
\$200-\$500	30
\$500-\$1000	20
More than \$1000	5

Provide your estimate of the average cost of a spring break trip and the standard error associated with it.

6. Early in March the Washington Post-ABC News Poll reported that 59% of respondents opposed same-sex marriage. The sample size was 1,202 people. What is the 95% confidence interval around this result?
7. The following two histograms show the distribution of grade point averages of two universities. Both have a mean GPA of 3.0 and standard deviation of 0.7. If you sampled 100 students from University A and 100 students from University B, how would the averages and standard errors of the two samples differ?



8. The 2000 Current Population Survey Special Report on Voting and Registration interviewed 74,174 adults of voting age in November 2000, after the general election. 49,269 of them reported they had voted. Is this number “too high” or “too low,” given the number of eligible voters we know participated in the 2000 presidential election (roughly 52% turnout)?
9. In the survey mentioned above, 22,898 of the 34,580 men and 27,112 of the 39,594 women said they voted. Run the t-test to see if the fraction of men voting equals the fraction of women voting.
10. Find the probability that a normally distributed random variable with mean of 0 and standard deviation of 1 takes on the following values:
 - a. greater than 0
 - b. greater than 1.96
 - c. greater than 2.54
 - d. greater than 1.96 or below -1.96
 - e. between -1.96 and 1.96
11. Imagine a variable with mean of 0 and standard deviation of 1. What is the value of z such that the variable takes on
 - a. a value less than z 99% of the time.
 - b. a value less than z 64% of the time.
 - c. a value greater than z 35% of the time.
12. You are interested in reporting to Dean Robert Redwine about what makes MIT students like the classes they take. You use the data from the Spring 1995 Course Evaluation

Guide to do your research. (Don't worry, the basic story hasn't changed in a decade.). The survey instrument has also basically stayed the same, so if you want to refresh yourself, you can go to the current Course Evaluation Guide web site to see what the questions look like.

Construct a regression equation that explains student satisfaction with their classes, based on at least four independent variables.

Hand in a "do file" that will reproduce the regression that you report.

Type up the table. Write a page that describes the variables you chose, why you chose them, and what you concluded from this exercise.