New York University
Wilf Family Department of Politics
Fall 2013

Quantitative Research in Political Science I

Professor Patrick Egan

MIDTERM EXAMINATION: NOVEMBER 8, 2013, 10 a.m. - 1 p.m.

This exam is open-book, open-note. You will need either statistical tables like those appearing in the back of your text or access to a statistical software program that can provide similar information. Consultation of Web resources is not permitted.

Be sure to show all your work and to use complete sentences to provide explanations.

Point totals for each question are as follows. They sum to 92 points. **It is recommended that you read all parts to a question–that is, parts (a), (b), (c), etc.–before beginning to answer it.**

- Question 1: 20 points.
- Question 2: 20 points.
- Question 3: 24 points.
- Question 4: 28 points.

1. Any attempt to assess voter turnout with survey data is made more complicated by the challenge of social desirability bias: survey respondents tend to over-report whether they voted. For example, 58.3 percent of voting-age Americans told the Census Bureau that they voted in the 2004 presidential election. But voting statistics show that only 55.5 percent did. (This gap is even greater in years without presidential elections.)

   You are conducting a post-election survey with which you'd like to get the best possible estimate of aggregate voter turnout. You have the choice between two different survey modes. The first is a traditional telephone poll using professional interviewers ("live"). The second is a telephone poll using the interactive voice response ("IVR") technique, in which questions are asked by a recorded voice and participants enter their responses via their telephone keypad. Previous research has found that live and IVR survey modes have different advantages and disadvantages:

   - The *advantage* of IVR is that–since it is more removed from the social context of a conversation between two people–it is less subject to overreporting of the vote than surveys with live interviewers.
   - The *disadvantage* of IVR is that it is subject to more error than surveys with live interviewers.

   Your goal is to estimate p, the proportion of the voting-age population that turned out to vote. We can consider the proportion of respondents reporting they voted in the live and IVR survey modes as two different estimates of p. Call these estimates $\widehat{p}_L = \frac{Y_L}{n_L}$ and $\widehat{p}_I = \frac{Y_I}{n_I}$, respectively.

   (a) The advantages and disadvantages of the live and IVR modes correspond with two properties of estimators we've discussed in class. In a few sentences, describe these two properties, and say how we would expect $\widehat{p}_L$ and $\widehat{p}_I$ to compare regarding these two properties.

   Assume it is the case that:

$$E\left(\widehat{p}_L\right) = ap + b; \;\; VAR\left(\widehat{p}_L\right) = \frac{a^2}{n} p\left(1-p\right) \;\; \text{and}$$
$$E\left(\widehat{p}_I\right) = cp + d; \;\; VAR\left(\widehat{p}_I\right) = \frac{c^2}{n} p\left(1-p\right).$$
$$\text{Further, assume that } a < c, b > d, \text{ and } \frac{b-d}{c-a} > 1.$$

   (b) What is $B\left(\widehat{p}_L\right)$? What is $B\left(\widehat{p}_I\right)$? Show that $B\left(\widehat{p}_L\right) > B\left(\widehat{p}_I\right)$.

   (c) Show (trivially) that $VAR\left(\widehat{p}_I\right) > VAR\left(\widehat{p}_L\right)$.

   (d) We have a situation where one of our potential estimators suffers from greater bias, while the other is subject to more error. In class, we learned of a criterion often used to measure the tradeoff between bias and error. What is this criterion called and what is its formula?

   (e) Now consider a case where $a = .01, b = .1; c = .02, d = .05$ and $p = \frac{1}{2}$. According to the criterion you identified in part (d), which is the better estimator?

2. Consider two independent large-sample means, $\overline{X}$ and $\overline{Y}$, where $\overline{X} < \overline{Y}$. As discussed in class, it is sometimes the case that the confidence intervals constructed about the means μ_X and μ_Y overlap, but the confidence interval constructed about the difference in these means $(\mu_Y - \mu_X)$ does not contain zero.

   (a) Write the inequality that must hold if the CIs for μ_X and μ_Y overlap. The expression should be in terms of $\overline{X}$,$\overline{Y}$, $\sigma_{\overline{X}}$, $\sigma_{\overline{Y}}$.and some $z_{\frac{\alpha}{2}}$,which for simplicity you may refer to as z.

   (b) Write the inequality that must hold if the CI for $\mu_Y - \mu_X$ does not contain zero. This should be in terms of $\overline{X}$,$\overline{Y}$,$\sigma_{\overline{Y}-\overline{X}}$, and z.(the same z as in part (a)).

   (c) Now show that a sufficient condition for the CIs for μ_X and μ_Y to overlap while the CI for $\mu_Y - \mu_X$ does not contain zero is

$$z\sigma_{\overline{Y}-\overline{X}} < \overline{Y} - \overline{X} < z\left(\sigma_{\overline{X}} + \sigma_{\overline{Y}}\right) \ .$$

   (d) To reassure yourself that the inequality in (c) can be satisfied, show that it is always the case that

$$z\sigma_{\overline{Y}-\overline{X}} < z\left(\sigma_{\overline{X}} + \sigma_{\overline{Y}}\right).$$

3. Indicate whether each of the following statements is true or false. Explain each of your answers, using mathematics where necessary.

   (a) Our estimates of population means using large samples rely heavily on assumptions about the shape of the distribution of the population.

   (b) Our estimates of population means using large samples rely heavily on assumptions about the process giving rise to the sample.

   (c) All things being equal, Type I errors are more likely with small samples than with large samples.

   (d) All things being equal, Type II errors are more likely with large samples than with small samples.

   (e)
   $$\frac{\partial \sigma^2_{\overline{Y}}}{\partial \sigma^2_Y} < 0.$$

   (f) In an i.i.d. random sample of size n drawn from the population Y, the observation Y_1 is an unbiased estimate of μ_Y.

   (g) All things being equal, I am more confident in a finding confirming the null $H_0 : \mu_1 - \mu_2 = 0$ as α gets lower.

   (h) The Bernoulli, Binomial, and Poisson distributions bear important similaries to one another.

4. Answer each of the following questions. Show your work.

   (a) You wish to determine whether women are more Democratic than men.

      i. You draw a random sample of 50 men and 50 women, and find that 43 percent of men identify as Democrats, while 56 percent of women do. How sure does this make you that women are more Democratic than men in the general population?

      ii. Now you draw a random sample of n men and n women, and find that 43 percent of men identify as Democrats, while 56 percent of women do. Find the largest n at which you are **not** sure with 95 percent confidence that women are more Democratic than men in the general population.

   (b) You are the leader of an environmental group who hopes that major environmental legislation is passed by Congress and signed into law at some point during the 2013-2014 Congressional session. The number of major environmental laws passed per session can be modeled as the random variable Y with $E(Y) = 1$. What is the chance that at least one piece of major environmental legislation gets passed during the session? Be precise about how you are modeling this process.

   (c) You have discovered some observation Y_i of the random variable Y that is two standard deviations greater than Y's mean, μ_Y.

      i. Find $P(Y \geq \mu_Y + 2\sigma_Y | Y \tilde{} \text{Standard Normal})$.

      ii. Explain why it is non-sensical to model Y as distributed Uniform. HINT: Show that $P(Y \geq \mu_Y + 2\sigma_Y | Y \tilde{} \text{Uniform}) = 0$.