Social Studies 201

December 4, 2006

Notes about the final examination

The questions for the final examination will centre around interval estimates, samples size, and hypothesis tests, the methods examined since around the second midterm and covered in Problem Sets 4 and 5. There will be five or six questions on the examination and you are expected to answer three of these. Make sure you bring a calculator and the tables of the normal and chi-square distributions. For the examination, you may bring any other written material you wish.

The final examination is from 9:00 a.m to 12:00 noon in CL431, Wednesday, December 13, 2006.

There will be two review sessions, from 12:30 to 2 on Monday, December 11 and from 11 to 1 on Tuesday, December 12. Both review sessions will be held in CL251.

Statistical methods to review

In all of the following, the assumption is that the sample, with sample size n, is a random sample of the population.

1. **Interval estimates**

   Always remember to state the confidence level. This is C per cent, either the level stated in the question or a level you select yourself. In general, a high level, such as 90%, 95%, or 99% should be selected. The 95% confidence level is the most common and has become almost like a default level when constructing confidence intervals.

   (a) **Interval estimate for the mean of a population, μ, large sample size**.

   The first step is to make sure that you clearly define μ, that is, be clear concerning what population mean you are attempting to estimate.

If the sample size is $n \geq 30$, then the interval estimate is

$$\bar{X} \pm Z\frac{s}{\sqrt{n}}$$

Note that s is used as an estimate of the population standard deviation, σ, an approach that is acceptable so long as n is large.

(b) **Interval estimate for a population proportion, p, large sample size**.

If n is larger than 5 divided by the smaller of p or q, then the interval estimate is

$$\hat{p} \pm Z\sqrt{\frac{pq}{n}}$$

For an interval estimate of a proportion, in the denominator of the above expression use $p = q = 0.5$; alternatively, $\hat{p}$ and $\hat{q}$ can be used in the denominator.

2. **Sample size**

E is defined as the accuracy of the estimate or, alternatively, as the error of estimate or the margin of error. The following formulae give the size of a random sample necessary to obtain the accuracy E.

(a) **Estimate of the mean of a population, μ.**

$$n = \left(\frac{Z\sigma}{E}\right)^2$$

Remember to make sure that E and the estimate of σ must be in the same units. Since the population standard deviation, σ is generally unknown, the sample standard deviation, s, is usually used as an estimate of σ. In this case, the formula for obtaining the appropriate sample size is

$$n = \left(\frac{Zs}{E}\right)^2$$

(b) **Estimate of a population proportion,** p.

For the following, E should be expressed as a proportion, not a percentage.

$$n = \left(\frac{Z}{E}\right)^2 pq$$

In order to assure a sufficiently large sample size to obtain accuracy of plus or minus E, use $p = q = 0.5$. If these values are used, the formula for sample size when estimating a proportion is

$$n = 0.25\left(\frac{Z}{E}\right)^2$$

3. **Hypothesis tests**

Always remember to use a significance level, α, either the significance level stated in the question or a level you select yourself. It is most common to select a small significance level, such as 0.01, 0.05, or 0.10. The 'default' confidence level is 95% and the corresponding 'default' significance level is 0.05.

(a) **Hypothesis test for the mean of a population,** μ**, large sample size**.

As with an interval estimate, a large sample size is when $n \geq 30$. M is the hypothesized value of the population mean.

$$H_0 : \mu = M$$

$$H_1 : \mu \neq M \text{ or } \mu > M \text{ or } \mu < M$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

(b) **Hypothesis test for a population proportion (**p**), large sample size**.

If n is larger than 5 divided by the smaller of p or q, then the hypotheses for the test are as follows. P is the hypothesized value

of the proportion of members of a population with a particular characteristic.

$$H_0 : p = P$$

$$H_1 : p \neq P \text{ or } p > P \text{ or } p < P$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Note that the value of p to be used in the denominator of this expression is to be P, the value hypothesized in H_0.

4. **Chi-square test for independence or dependence**

H_0 : No relation between variables – independence or $O = E$

H_1 : Some relation between variables – dependence or $O \neq E$

For the question on chi-square, I will provide the table with observed and expected counts, and the table with the chi-square value, both from the SPSS printout. A large χ^2 value leads to rejection of the null hypothesis and a small χ^2 value means insufficient differences between observed and expected values to reject the null hypothesis. In the table of statistics from the SPSS printout, use the 'Pearson Chi-Square value as the chi-square statistic.

To answer this question on the final examination, state the null and alternative hypotheses, state the significance level, find the chi-square value from the chi-square table, state what the critical region is, and then use the Pearson Chi-Square value to conclude whether you reject or do not reject the null hypothesis. You will also be asked to describe the pattern of the relationship from the table or, if you find there to be no relationship between the two variables, explain why.