Social Studies 201

April 13, 2005

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 March 11 and covered in Problem Set 5. There will be five or six questions on the examination and you are expected to answer three of these. Make sure you bring the tables of the normal, t and chi-square distributions and a calculator. 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 CL232, Monday, April 18, 2005.

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 use a confidence level, C, either the one stated in the question or a level you select yourself.

   (a) **Interval estimate for a mean, μ, large sample size**

   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 mean, μ, small sample size**

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

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

using the t-value for $n-1$ degrees of freedom. One assumption built into this method is that the population from which the random sample is drawn is a normally distributed population. Such an assumption may not always be warranted but, if n is small, it is always better to use the t-value rather than the Z-value.

(c) **Interval estimate for a 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.

   (a) **Estimate of mean, 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.

   (b) **Estimate of proportion, p**

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

   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 one stated in the question or a level you select yourself.

(a) **Hypothesis test for a mean, μ, large sample size** As with an interval estimate, a large sample size is when $n \geq 30$.

$$H_0 : \mu = M$$

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

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

(b) **Hypothesis test for a mean, μ, small sample size**

$$H_0 : \mu = M$$

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

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

As with the interval estimate, the assumption is that the population from which the sample is being selected is a normally distributed population. Use the t-value rather than the Z-value whenever $n < 30$.

(c) **Hypothesis test for a proportion (p), large sample size**

$$H_0 : p = P$$

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

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

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.

Last edited April 13, 2005.