ECNS561, Fall 2011

Take-home Graded Practice Opportunity 2

Due date: September 30, 2011, 5:00 p.m.

1. Is it the case that if $\lim_{n\to\infty} \widehat{\theta}_n = \theta$ implies $\text{plim}_{n\to\infty}\, \widehat{\theta}_n = \theta$? Prove.

   Recall that if $\lim_{n\to\infty} \widehat{\theta}_n = \theta$, then for a large n and some small value ε the following is true: $|\widehat{\theta}_n - \theta| < \varepsilon$. Therefore, $Prob[|\widehat{\theta}_n - \theta| > \varepsilon] = 0$.

2. Suppose that a sequence of random variables Y_n follows the probability density function:

$$Y_n = \begin{cases} 0 & \text{with probability } (n-1)/n \\ n^2 & \text{with probability } 1/n \end{cases}$$

   Show that $\text{plim}_{n\to\infty}\, Y_n = 0$, but that the $\lim_{n\to\infty} E[Y_n] = \infty$.

   We know that the Y_n can take on either a value of 0 or n^2. So the only case that we need to truly consider in the limit is when Y_n takes on the value n^2. Therefore, we consider only the case:

$$\lim_{n\to\infty} \text{Prob}[|Y_n| > \varepsilon] = \frac{1}{n} \;\to\; 0$$

   To determine the asymptotic expected value, we determine the following:

$$\lim_{n\to\infty} E[Y_n] = \lim_{n\to\infty} \left\{ 0 \cdot \frac{(n-1)}{n} + n^2 \cdot \frac{1}{n} \right\} = \lim_{n\to\infty} \{n\} \to \infty$$

3. Suppose that $\overline{Y}$ and $S^2_{\overline{Y}}$ are the mean and sample variance associated with a random sample $\{Y_1, Y_2, \ldots, Y_n\}$, which is drawn from a population with some finite variance, σ^2_Y.

   (a) What is the expected value of $S^2_{\overline{Y}}$?
   The expected value of $S^2_{\overline{Y}}$ is simply the population variance, σ^2_Y.

   (b) Prove that $E[S_{\overline{Y}}] \leq \sigma_Y$. *Hint:* For a concave function $h(\cdot)$, Jensen's inequality states that $E[h(Z)] \geq h(E[Z])$.
   Provided the hint, we can state that the $E[S^2_{\overline{Y}}] \geq (E[S_{\overline{Y}}])^2$. Using the result in part (a), this simplifies to $\sigma^2_Y \geq (E[S_{\overline{Y}}])^2$. Then, taking the square root of both sides yields that $E[S_{\overline{Y}}] \leq \sigma_Y$.

4. Consider the table 1, which describes wheat prices in four regions and the protein levels associated with each price. In general, protein levels indicate the quality of wheat. the table represents random samples from each of the four regions.

   Complete the following:

   (a) First, investigate the data. What might they tell you about the supply and demand dynamics in each region? What do they tell you about the effect of quality? Justify your answers.

   Generally, it appears that higher quality wheat, as expected, has a price premium relative to lower quality wheat. That is, there is likely a positive correlation between quality (protein) and prices. Also, region 4 has the lowest prices for all protein levels, indicating that either there is a higher supply of all protein types or lower demand for these types. The converse can be said for region three, where prices are highest. Again, either supply is low or demand is high for wheat in the region.

   (b) Estimate the mean population wheat price across all protein levels.

$$\overline{P}_{wheat} = \frac{1}{n} \sum P_{wheat} = \$5.87$$

   (c) Estimate the sampling variance and sampling standard deviation? Interpret these results – what inferences can we make about the population?

$$S_P^2 = \frac{1}{n-1} \sum (P_{wheat} - \overline{P}_{wheat})^2 = 0.00485$$

$$S_P = \sqrt{S_P^2} = 0.070$$

   Relative to the sampling mean, the sampling standard deviation is very small. This indicates that wheat prices are relatively tightly disbursed around the central tendency.

   (d) Compute a 95% confidence interval around the population mean of the wheat price. Then compute a 95% confidence interval around the population mean of 15% protein wheat price. What inferences can you make when comparing the two confidence intervals?

   The critical value from a t-distribution with 14 degrees of freedom is 2.145. Therefore, the confidence interval around the population mean of all wheat prices is:

$$\left[\overline{P}_{wheat} - 2.145 \cdot \frac{S}{\sqrt{n}}, \ \overline{P}_{wheat} + 2.145 \cdot \frac{S}{\sqrt{n}}\right] = [\$5.84, \$5.91]$$

To determine the confidence interval around only the 15% protein wheat, we need to first calculate the mean and standard error, and then determine the critical value from the t-distribution with 3 degrees of freedom.

$$\overline{P}_{wheat,15\%} = \$5.97$$

$$S_{P,15\%} = 0.024$$

$$SE_{P,15\%} = 0.012$$

$$t_{crit} = 3.182$$

$$[\$5.93,\ \$6.00]$$

The confidence interval for the 15% protein wheat is substantially above the confidence interval for the average wheat price. The difference is so substantial, that the upper bound of the overall wheat price does not overlap with the lower bound of the 15% protein wheat price. This implies that the two means are statistically different from each other.

(e) Set up, test, and interpret each of the following hypotheses:

i. At a 95% confidence level, the price of wheat is less than \$5.85/bu.

$$H_0: \ P_{wheat} = \$5.85 \qquad\qquad H_a: \ P_{wheat} < \$5.85$$

$$t_{stat} = \frac{5.87 - 5.85}{0.0175} = 1.14$$

$$t_{crit} = 1.753$$

Because $|t_{stat}| < t_{crit}$*, we fail to reject the null hypothesis in favor of the alternative hypothesis.*

ii. At a 99% confidence level, the average price of 12% and 13% protein wheat is greater than \$5.80/bu.

$$H_0: \ P_{wheat,12-13\%} = \$5.80 \qquad\qquad H_a: \ P_{wheat,12-13\%} > \$5.80$$

$$t_{stat} = \frac{5.82 - 5.80}{0.014} = 1.42$$

$$t_{crit} = 2.998$$

Because $|t_{stat}| < t_{crit}$, we fail to reject the null hypothesis in favor of the alternative hypothesis.

iii. At a 90% confidence level, the average price of 14% and 15% protein wheat is not equal to the average price of 12% and 13% protein wheat.

$$H_0: \; P_{wheat,12-13\%} = P_{wheat,14-15\%}$$
$$H_a: \; P_{wheat,12-13\%} \neq P_{wheat,14-15\%}$$

$$t_{stat} = \frac{5.82 - 5.93}{0.016} = -6.87$$

$$t_{crit} = 1.895$$

Because $|t_{stat}| > t_{crit}$, we reject the null hypothesis in favor of the alternative hypothesis.

(f) For each of the tests in part (e), compute p-values. Interpret.

i. *To determine the p-value, you need to calculate the probability of the t_{stat} value (1.14, from part e) from the t-distribution with 15 degrees of freedom. Because we are performing a test in the lower tail of the t-distribution, we use the formula:*

$$\text{p-value} = 1 - t(1.14, 15) = 0.864$$

We are able to reject the null hypothesis only at a 86.4% significance level. Cannot reject the null at the 5% significance level.

ii. *To determine the p-value, you need to calculate the probability of the t_{stat} value (1.42, from part e) from the t-distribution with 7 degrees of freedom. Because we are performing a test in the upper tail of the t-distribution, we use the formula:*

$$\text{p-value} = t(1.42, 7) = 0.099$$

We are able to reject the null hypothesis only at a 9.9% significance level. Cannot reject the null at the 1% significance level.

iii. *To determine the p-value, you need to calculate the probability of the t_{stat} value (6.87, from part e) from the t-distribution with 7 degrees of freedom. Because we are performing a test in both tails of the t-distribution, we use the formula:*

$$\text{p-value} = 2 \times (t(6.87, 7)) = 0.0002$$

We are able to reject the null hypothesis at a 0.02% significance level.

Table 1: Table of wheat prices and protein levels in four regions

Region	**Wheat price per bushel**	**Protein level**
1	\$ 5.80	12%
1	\$ 5.83	13%
1	\$ 5.89	14%
1	\$ 5.98	15%
2	\$ 5.79	12%
2	\$ 5.84	13%
2	\$ 5.88	14%
2	\$ 5.95	15%
3	\$ 5.84	12%
3	\$ 5.89	13%
3	\$ 5.92	14%
3	\$ 5.99	15%
4	\$ 5.76	12%
4	\$ 5.80	13%
4	\$ 5.87	14%
4	\$ 5.94	15%

5. Consider the following three vectors:

$$\boldsymbol{A} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{(3\times1)} \qquad \boldsymbol{B} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}_{(2\times1)} \qquad \boldsymbol{C} = \begin{bmatrix} 1 & 4 \end{bmatrix}_{(1\times2)}$$

Calculate:

(a) $\boldsymbol{A} \cdot \boldsymbol{B}'$

$$\begin{bmatrix} 2 & 2 \\ 4 & 4 \\ 6 & 6 \end{bmatrix}$$

(b) $\boldsymbol{C}' + \boldsymbol{B}$

$$\begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

(c) $\boldsymbol{D} = \boldsymbol{B}' \cdot \boldsymbol{B}$

$$\boldsymbol{D} = \begin{bmatrix} 8 \end{bmatrix}$$

(d) $\boldsymbol{D}^{-1}$

$$\begin{bmatrix} 0.125 \end{bmatrix}$$

(e) $\boldsymbol{B}'\boldsymbol{C}'$

$$\begin{bmatrix} 10 \end{bmatrix}$$

(f) $\boldsymbol{D}^{-1} \cdot \boldsymbol{B}'\boldsymbol{C}'$

$$\begin{bmatrix} 1.25 \end{bmatrix}$$

6. Applied practice. Please turn in your code and a summary of your output. If you are going to do the assignment in anything other than SAS, please also email me your code. You don't have to print the entire output window; just provide the relevant statistics asked for in the problems.

   Part of this week's applied practice will have you use simulations to prove several properties and theorems that we discussed in class. The other part will have you perform hypothesis tests.

   Simulation Analysis

   (a) Simulate random samples from the distribution $Y \sim N(2, 2^2)$. Show that $\lim_{n\to\infty} E[\overline{Y}_n] = \mu_Y$ (asymptotic unbiasedness) and plim $(\overline{Y}_n) = \mu$ (law of large numbers). *Hint*: the SAS example I showed in class has code for simulating from a normal distribution.

   (b) Prove graphically the central limit theorem using simulated values from a uniform distribution $Y \sim U[0, 4]$. *Hint*: To simulate a uniform distribution in SAS, use the function `rand(''Uniform'')*b`, where `b` is the upper bound of the uniform distribution (lower bound, `a`, is assumed to be 0).

   Hypothesis Tests

   1 Using the data presented in problem 4, perform the hypothesis tests in part (e). In SAS, you can use the `proc ttest` procedure, but note that the SAS default is to perform a one-sided hypothesis test. Use the "sides" option to alter the default setting. Interpret the results, including the p-values.

   See SAS code under the Homework link on the course website.