HG
Nov. 08

# ECON 4130

# Supplementary Exercises 5 - 7

## Exercise 5

For this exercise you need to have read lecture notes to Rice chapter 8 - especially section 0, 1, and 3.

**a.** Let $\theta$ be an unknown parameter in an econometric model, $\hat{\theta}$ an asymptotically normally distributed estimator based on *n* observations, and with a consistently estimated standard error

$$\text{se}(\hat{\theta}) = \frac{\hat{b}}{\sqrt{n}}$$

such that

$$U_n = \frac{\hat{\theta} - \theta}{\text{se}(\hat{\theta})} = \sqrt{n}\frac{\hat{\theta} - \theta}{\hat{b}} \overset{\text{approximately}}{\sim} N(0,1) \quad \text{for any possible } \theta .$$

Based on $U_n$ we have an approximate $1-\alpha$ confidence interval (CI) for $\theta$ given by

(1) $\hat{\theta} \pm z_{\alpha/2}\text{se}(\hat{\theta})$

We want to test the two-sided hypothesis, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where $\theta_0$ is a known hypothetical value. Let $z_{\alpha/2}$ be the upper $\alpha/2$ - quantile in *N*(0, 1). An approximate $\alpha$ -level test is given by

(2) Reject $H_0$ if $W_n < -z_{\alpha/2}$ or $W_n > z_{\alpha/2}$ (i.e., if $|W_n| > z_{\alpha/2}$)

where $W_n = \dfrac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}$ is the test statistic used.

Show that the test criterion (2) is equivalent to the following test criterion based on the CI in (1):

(3) Reject $H_0$ if $\theta_0$ lies outside the CI, $\hat{\theta} \pm z_{\alpha/2}\text{se}(\hat{\theta})$ .

[**Note.** Hence a CI of the form (1) can always be used to test two-sided hypotheses about $\theta$. Not only that: If we reject $H_0$ by the test in (2), we may conclude (with strong evidence) not only that $\theta \neq \theta_0$, but also on which side of $\theta_0$ the unknown $\theta$ lies. For example, if $\theta_0$ lies outside to the left of the CI, we may not only conclude that $\theta \neq \theta_0$, but also (with equally strong evidence) that $\theta > \theta_0$. This is simply because the CI itself shows that the true $\theta$ then (with strong evidence) lies to the right of $\theta_0$.]

**b.** Let $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ be a vector of three unknown parameters in an econometric model.

Suppose that some estimating principle has produced an approximately normally distributed estimator $\hat{\beta}' = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ such that

$$\hat{\beta} \overset{\text{approximately}}{\sim} N(\beta, \Sigma)$$

where $\Sigma$ is (a consistent estimate of) the covariance matrix given by

$$\Sigma = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 4 & 1 \\ 3 & 1 & 16 \end{pmatrix}$$

(i) Suppose further that the estimates of the $\beta_j$'s are given in the table. Fill in the standard errors in the table, and calculate the three correlation coefficients, $\text{corre}(\hat{\beta}_1, \hat{\beta}_2)$, $\text{corre}(\hat{\beta}_1, \hat{\beta}_3)$, and $\text{corre}(\hat{\beta}_2, \hat{\beta}_3)$.

**Table 1**

| Coefficient | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|---|---|---|---|
| Estimate | 10 | 15 | 16 |
| Standard error | | | |

(ii) Calculate an approximate 95% CI for $\beta_1$, and perform a test (with level of significance 5%) of $H_0 : \beta_1 = 5$ versus $H_1 : \beta_1 \neq 5$.

**c.** Introduce the parameters

$$\theta_1 = \beta_1 - \beta_2$$
$$\theta_2 = \beta_1 - \beta_3$$
$$\theta_3 = \frac{\beta_2 + \beta_3}{2}$$
$$\theta_4 = \beta_1 - \theta_3$$

Produce a table as **Table 1** with estimates (based on $\hat{\beta}$) and standard errors for the three parameters $\theta_1, \theta_2,$ and, $\theta_4$.

**d.** Test the following three two-sided hypotheses (level 5%) by using CI's

(i) $H_0 : \beta_1 = \beta_2$ vs $H_1 : \beta_1 \neq \beta_2$ (i.e., $H_0 : \theta_1 = 0$ vs $H_1 : \theta_1 \neq 0$)
(ii) $H_0 : \beta_1 = \beta_3$ vs $H_1 : \beta_1 \neq \beta_3$
(iii) $H_0 : \beta_1 = \theta_3$ vs $H_1 : \beta_1 \neq \theta_3$

In each case where the test leads to rejection of $H_0$, state the direction of the alternative in the conclusion (for example, if the test of (i) leads to rejection, state one of "$\beta_1 < \beta_2$" or "$\beta_1 > \beta_2$" as your conclusion instead of just "$\beta_1 \neq \beta_2$").

## Exercise 6

Let *X* be the number of traffic accidents occurring during *t* months in a region. Assume that *X* is poisson distributed with parameter $\lambda t$ (i.e., $X \sim \text{pois}(\lambda t)$).

**a.** Explain why the parameter $\lambda$ can be interpreted as a theoretical incidence rate, i.e., the expected number of accidents per month.

**b.** We cannot observe $\lambda$ directly, but we can observe *X* instead. Show that the estimator $\hat{\lambda} = X/t$

i) is unbiased for all t,
ii) is consistent as $t \to \infty$ (use Chebyshev's inequality).

**c.** Using the fact that *X* is approximately normally distributed when $\lambda t$ is large ($\geq 10$ is usually considered sufficient), develop an approximate $1-\alpha$ confidence interval (CI) for $\lambda$ based on *X*.
[**Hint:** Show first, using Slutsky's lemma, that

$$\sqrt{t}\,\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}}} \;\underset{t\to\infty}{\overset{D}{\to}}\; Z \sim N(0,1)$$

Note also that $\sqrt{t}\,\dfrac{\hat{\lambda}-\lambda}{\sqrt{\lambda}} = \dfrac{X-\lambda t}{\sqrt{\lambda t}}$ ]

## Exercise 7

We are interested in the monthly incidence rate of traffic accidents in Norway. From Statistical Office Norway (SSB), we obtain the number of traffic accidents registered in the period 2003 – 2005, as given in table 1,

**Table 1**

| Year | No. of traffic accidents |
|---|---|
| 2003 | 8266 |
| 2004 | 8425 |
| 2005 | 8078 |
| Sum | 24769 |

We want a 95% CI for the monthly incidence rate based on these results. Let $X_1, X_2, X_3$ denote the rv's behind the three observations in table 1. Our first approach is to calculate a "t-interval" for the incidence rate, called $\lambda$ , based on the following model

**Model 1** $X_1, X_2, X_3$ are *iid* with $X_i \sim N(\mu, \sigma^2)$ where $\mu = 12\lambda$

[**Hint:** When $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim N(\mu, \sigma^2)$, we remember from the basic statistic course that an (exact) $1-\alpha$ CI for $\mu$ (the so called "t-interval") is

$$\bar{X} \pm t_{1-\alpha/2,\; n-1} \cdot \frac{S}{\sqrt{n}}, \qquad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \text{ and}$$

where $t_{1-\alpha/2,\; n-1}$ is the $1-\alpha/2$ percentile in the t-distribution with $n-1$ degrees of freedom. ]

**a.** Calculate the 95% CI for $\mu$ based on model 1 and transform the interval to a corresponding CI for $\lambda$. Explain why the interval for $\lambda$ must have the same degree of confidence as the one for $\mu$. Discuss briefly whether the assumptions in model 1 appear reasonable or not.

**b.** An alternative approach is to assume that the total number of registered accidents in the period 2003 - 2005, $X = X_1 + X_2 + X_3$, is poisson distributed, i.e.,

**Model 2** $X \sim \text{pois}(t\lambda)$ with $t = 36$

Calculate an approximate 95% CI for $\lambda$ based on model 2, and compare with the CI in **a**.

**c.** One reasonable criticism that can be raised against the model 2, is the apparently unrealistic assumption of *constant incidence rate* that underlies the poisson model, which, among other things, implies that all the months of the year have the same incidence rate, $\lambda$. For example, table 2, that gives the number of accidents for January and June, appears to support this criticism.

**Table 2**

| | Number of accidents | |
|---|---|---|
| Year | January | June |
| 2003 | 576 | 805 |
| 2004 | 616 | 847 |
| 2005 | 588 | 853 |

Luckily, the poisson model offers an easy way to accommodate this criticism. To see this, first prove the following result [**Hint:** Use the mgf for the poisson distribution]:

**Property 1** Let $Y_1, Y_2, \ldots, Y_k$ be independent and poisson distributed with $Y_j \sim \text{pois}(\mu_j)$ for $j = 1, 2, \ldots, k$. Then $Y = Y_1 + Y_2 + \cdots + Y_k \sim \text{pois}(\mu)$ where $\mu = \mu_1 + \mu_2 + \cdots + \mu_k$.

**d.** In order to accommodate the criticism, we suggest the following model: Let $Y_{ij}$ be the number of accidents in month *j* ( $j = 1, 2, \ldots, 12$ ) in year *i* ( $i$ = 1,2 3). Assume

**Model 3** The $Y_{ij}$'s are independent and poisson distributed with $Y_{ij} \sim \text{pois}(\lambda_j)$ for $j = 1, 2, \ldots, 12$ and $i = 1, 2, 3$.

Show that $X = \sum_{i=1}^{3}\sum_{j=1}^{12} Y_{ij} \sim \text{pois}(36\bar{\lambda})$ where $\bar{\lambda} = \frac{1}{12}\sum_{j=1}^{12}\lambda_j$ is the average monthly incidence rate.

**e.** Show that the estimator in exercise 6, $\hat{\lambda} = \frac{X}{t}$, where $t = 12r$ is the number of months and *r* the corresponding number of years, is unbiased for $\bar{\lambda}$ and with variance, $\text{var}(\hat{\lambda}) = \frac{\bar{\lambda}}{t}$. This shows that $\hat{\lambda}$ also is consistent for $\bar{\lambda}$ as $t \to \infty$ (why?).

**f.** Explain why the CI in **b** is still valid, but now as an approximate 95% CI for the new parameter, $\bar{\lambda}$.