Econometric Analysis

Review on Estimation and Inference

by

Jin-Lung Lin

Fall, 2007

1 Descriptive Statistics

1. mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. standard deviation: $s_x = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}^{1/2}$
3. skewness = $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{s_x^3(n-1)}$
4. kurtosis = $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{s_x^4(n-1)}$
5. covariance: $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x}(y_i-\bar{y}}{n-1}$
6. correlation: $r_{xy} = \frac{s_{xy}}{s_x s_y}$
7. Correlation matrix: $S = [s_{ij}]$
8. Correlation matrix: $R = [r_{ij}]$

2 Sampling distribution of estimators

1. sample mean:
   If $x_1, x_2, \cdots, x_n$ are a random sample from a population with mean μ and variance σ^2, then $\bar{x}$ is a random variable with mean μ and variance σ^2/n.
2. If $x_1, x_2, \cdots, x_n$ are a random sample from an exponential distribution with $f(x) = \theta e^{-\theta x}$, then the sampling distribution of the sample minimum in a sample of n observations, denoted as $x_{(1)}$ is
   $$f(x_{(1)}) = (n\theta)e^{-n(\theta)x_{(1)}}$$
   with mean $E(x_{(1)}) = 1/(n\theta)$ and variance $Var(x_{(1)}) = 1/(n\theta)^2$.

3 Estimator (or statistic)

Definition 8.1 An estimator is a rule that tells how to calculate the value of an estimate based on the measurements contained in a sample.
Definition 8.2 Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an *unbiased estimator* if $E(\hat{\theta}) = \theta$. Otherwise $\hat{\theta}$ is said to be *biased*.
Definition 8.3 The *bias* of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

4 The Bias and Mean Square Error of Point Estimators

Definition 8.4 The *mean square error* of a point estimator $\hat{\theta}$ is the expected value of $(\hat{\theta} - \theta)^2$:

$$MSE(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right).$$

$$MSE(\hat{\theta}) = V(\hat{\theta}) + (B(\hat{\theta}))^2.$$

5 Some common unbiased point estimators

Table 1: Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Sizes n	Point Estimator $\hat{\theta}$	Expected Value $E(\theta)$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

6 Evaluating the Goodness of a Point Estimator

Definition 8.5 The *error of estimation* ε is the distance between an estimator and its target parameter. That is, $\varepsilon = |\hat{\theta} - \theta|$.

7 Confidence intervals

Let $\hat{\theta}_L$ and $\hat{\theta}_U$ be the lower and upper confidence interval limits respectively, for a parameter θ. Then,

1. *two-sided confidence interval*

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$$

2. *one-sided confidence interval*

$$\begin{aligned} P(\hat{\theta}_L \leq \theta) &= 1 - \alpha \quad \text{or} \\ P(\theta \leq \hat{\theta}_U) &= 1 - \alpha \end{aligned}$$

8 Large-Sample confidence intervals

For large sample, $Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$ has approximately a standard normal distribution.

$$P(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}) = 1 - \alpha$$

9 Selecting the sample size

- Experimental design
- How many samples are needed?
  How accurate is needed?
- $var(\bar{Y}) = \sigma^2/n$

10 Small-sample confidence intervals for μ and $\mu_1 - \mu_2$

For μ

$$\begin{aligned} T &= \frac{\bar{Y} - \mu}{S/\sqrt{n}} \\ P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) &= 1 - \alpha \end{aligned}$$

For $\mu_1 - \mu_2$

$$\begin{aligned} Z &= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) &= 1 - \alpha \end{aligned}$$

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

$$\begin{aligned} Z &= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ S_p^2 &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\ T &= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \end{aligned}$$

Summary of Small-Sample Confidence Intervals for Means of Normal Distribution with Unknown Variance(s)

Parameter	Confidence Interval (ν = degrees of freedom)
μ	$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right), \quad \nu = n-1$
$\mu_1 - \mu_2$	$(\overline{Y_1} - \overline{Y_2}) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $\nu = n_1 + n_2 - 2$ and $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ (requires that the samples are independent and the assumption that $\sigma_1^2 = \sigma_2^2$)

11 Confidence intervals for σ^2

A $100(1-\alpha)\%$ **Confidence interval for** σ^2

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right)$$

12 Relative Efficiency

Definition 9.1 Given two unbiased estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, of a parameter θ, with variances $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$, respectively, then the *efficiency* of $\hat{\theta}_1$ relative to $\hat{\theta}_2$, denoted eff($\hat{\theta}_1$, $\hat{\theta}_2$), is defined to be the ratio

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

13 Consistency

Definition 9.2 The estimator $\hat{\theta}_n$ is said to be a *consistent estimator* of θ if, for any positive number ε,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1,$$

or, equivalently,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

Theorem 9.1 An unbiased estimator $\hat{\theta_n}$ for θ is a consistent estimator of θ if

$$\lim_{n\to\infty} V(\hat{\theta_n}) = 0.$$

Theorem 9.2 Suppose that $\hat{\theta_n}$ converges in probability to θ and that $\hat{\theta_n'}$ converges in probability to θ'.

1. $\hat{\theta_n} + \hat{\theta_n'}$ converges in probability to $\theta + \theta'$.
2. $\hat{\theta_n} \times \hat{\theta_n'}$ converges in probability to $\theta \times \theta'$.
3. $\hat{\theta_n}/\hat{\theta_n'}$ converges in probability to θ/θ', provided that $\theta' \neq 0$.
4. If $\mathrm{g}(\cdot)$ is a real-valued function that is continuous at θ, then $\mathrm{g}(\hat{\theta_n})$ converges in probability to $\mathrm{g}(\theta)$.

Theorem 9.3 Suppose that U_n has a distribution function that converges to a standard normal distribution as $n \to \infty$. If W_n converges in probability to 1, then the distribution function of U_n/W_n converges to a standard normal distribution function.

14 The Rao-Blackwell Theorem and Minimum-Variance Unbiased Estimation

Theorem 9.5 (The Rao-Blackwell Theorem)
Let $\hat{\theta}$ be an unbiased estimator for θ such that $V(\hat{\theta}) < \infty$. If U is a sufficient statistic for θ, denote $\hat{\theta^*} = E(\hat{\theta}|U)$. Then, for all θ,

$$E(\hat{\theta^*}) = \theta \quad \text{and} \quad V(\hat{\theta^*}) \leq V(\hat{\theta}).$$

15 The Method of Maximum Likelihood

Method of Maximum Likelihood
Suppose that the likelihood function depends on k parameters $\theta_1, \theta_2, \ldots, \theta_k$. Choose as estimates those values of the parameters that maximize the likelihood $L(y_1, y_2, \ldots, y_n|\theta_1, \theta_2, \ldots, \theta_k)$.

16 Some large-sample properties of MLEs

Let $\hat{\theta}$ be the MLE of θ, then the MLE of $t(\theta)$ is given by $t(\hat{\theta})$. Under some regularity conditions, $t(\hat{\theta})$ is consistent for $t(\theta)$. For large sample size,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{[\frac{\partial t(\theta)}{\partial \theta}]^2 / nE[-\frac{\partial^2 lnf(Y|\theta)}{\partial \theta^2}]}}$$

has approximately a standard normal distribution. The large sample $100(1-\alpha)\%$ confidence interval for $t(\theta)$ is:

$$\begin{aligned} t(\hat{\theta}) \quad &\pm \quad z_{\alpha/2}\sqrt{[\frac{\partial t(\theta)}{\partial \theta}]^2 / nE[-\frac{\partial^2 lnf(Y|\theta)}{\partial \theta^2}]} \\ \approx t(\hat{\theta}) \quad &\pm \quad z_{\alpha/2}\sqrt{\left([\frac{\partial t(\theta)}{\partial \theta}]^2 / nE[-\frac{\partial^2 lnf(Y|\theta)}{\partial \theta^2}]\right)|_{\theta=\hat{\theta}}} \end{aligned}$$

17 The Elements of a Statistical Test

1. Null hypothesis, H_0
2. Alternative hypothesis, H_a
3. Test statistic
4. Rejection region

Definition 10.1 A type I error is made if H_0 is rejected when H_0 is true. The probability of a type I error is denoted by α. The value of α is called the level of the test. A type II error is made if H_0 is accepted when H_a is true. The probability of type II error is denoted by β.

Table 2: Type I, II errors

Action	Population	
	H_0 true	H_0 not true
Accept H_0		type I error
Reject H_0	type II error	

18 Common large-sample tests

Large-Sample α-Level Hypothesis Tests

$$H_0 : \theta = \theta_0$$

$$H_a = \begin{cases} \theta > & \theta_0 \quad (upper - tail \quad alternative) \\ \theta < & \theta_0 \quad (lower - tail \quad alternative) \\ \theta \neq & \theta_0 \quad (two - tailed \quad alternative) \end{cases}$$

$$Test\ statistic : Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

$$Rejection\ region = \begin{cases} \{z & > & z_\alpha\} & (upper - tail\ \text{RR}\ ) \\ \{z & < & -z_\alpha\} & (lower - tail\ \text{RR}\ ) \\ \{|z| & > & z_{\alpha/2}\} & (two - tailed\ \text{RR}\ ) \end{cases}$$

19 Calculating Type II error probabilities and finding the sample size for the Z test

Sample Size for an Upper-Tail α-Level Test

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

20 Relationships between hypothesis-testing procedures and confidence intervals

Accept the Null hypothesis if the test statistics falls within the confidence interval and reject the null hypothesis otherwise.

21 Another Way to Report the Results of a Statistical Test: Attained Significance Levels or p-values

Definition 10.2 If W is a test statistic, the p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

22 Some comments on hypothesis testing

1. One-sided or two-sided?
2. Selecting α
3. Choosing null hypothesis

23 A Small-Sample Test for μ and $\mu_1 - \mu_2$

Assumption: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$.

$$H_0 : \mu = \mu_0$$

$$H_a : \begin{cases} \mu > & \mu_0 \quad (upper-tail \quad alternative) \\ \mu < & \mu_0 \quad (lower-tail \quad alternative) \\ \mu \neq & \mu_0 \quad (two-tailed \quad alternative) \end{cases}$$

$$Test\ statistic : T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

$$Rejection\ region : \begin{cases} \{t & > & t_\alpha\} & (upper-tail \text{ RR }) \\ \{t & < & -t_\alpha\} & (lower-tail \text{ RR}) \\ \{|t| & > & t_{\alpha/2}\} & (two-tailed \text{ RR }) \end{cases}$$

Small-Sample Tests for Comparing Two Population Means Assumptions: Independent samples from normal distributions with $\sigma_1^2 = \sigma_2^2$.

$$H_0 : \mu_1 - \mu_2 = D_0$$

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0 (upper-tail\, alternative) \\ \mu_1 - \mu_2 < D_0 (lower-tail\, alternative) \\ \mu_1 - \mu_2 \neq D_0 (two-tailed\, alternative) \end{cases}$$

Test statistic : $T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{1/n_1 + 1/n_2}}$, where $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

$$Rejection\ region: \begin{cases} \{t > t_\alpha\} & (upper-tail\ \text{RR}) \\ \{t < -t_\alpha\} & (lower-tail\ \text{RR}) \\ \{|t| > t_{\alpha/2}\} & (two-tailed\ \text{RR}) \end{cases}$$

Here $P(T > t_\alpha) = \alpha$; and degrees of freedom $v = n_1 + n_2 - 2$.

Test of Hypotheses Concerning a Population Variance

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with

$$E(Y_i) = \mu, V(Y_i) = \sigma^2$$

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2 & (upper-tail\quad alternative) \\ \sigma^2 < \sigma_0^2 & (lower-tail\quad alternative) \\ \sigma^2 \neq \sigma_0^2 & (two-tailed\quad alternative) \end{cases}$$

Test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

Rejection region:

$$\begin{cases} \chi^2 > \chi_\alpha^2 & (upper-tail\quad \text{RR}) \\ \chi^2 < \chi_{1-\alpha}^2 & (lower-tail\quad \text{RR}) \\ \chi^2 > \chi_{\alpha/2}^2 \text{ or } \chi^2 > \chi_{1-\alpha/2}^2 & (two-tailed\quad \text{RR}) \end{cases}$$

Notice that χ_α^2 is chosen so that, for v=n-1 degrees of freedom, $P(\chi^2 > \chi_\alpha^2) = \alpha$

24 Test of the Hypothesis $\sigma_1^2 = \sigma_2^2$

Assumptions: Independent samples from normal populations.

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_a : \sigma_1^2 > \sigma_2^2$

Test statistic: $F = \frac{S_1^2}{S_2^2}$

Rejection region: $F > F_\alpha$, where F_α is chosen so that $P(F > F_\alpha) = \alpha$ when F has $v_1 = n_1 - 1$ numerator degrees of freedom and $v_2 = n_2 - 1$ denominator degrees of freedom.

25 Power of Tests and the Neyman-Pearson Lemma

Definition 10.3

Suppose that W is the test statistic and RR is the rejection region for a test of a hypothesis involving

the value of parameter θ. Then the power of the test, denoted by power(θ), is the probability that the test will lead to rejection of H_0 when the actual parameter value is θ. That is,
power(θ)=P(W in RR when the parameter value is θ).
Relationship Between Power and β
If θ_a is a value of θ in the alternative hypothesis H_a, then

$$power(\theta_a) = 1 - \beta(\theta_a)$$

Definition 10.4 If a random sample is taken from a distribution with parameter θ, a hypothesis is said to be a simple hypothesis if that hypothesis uniquely specifies the distribution of the population from which the sample is taken. Any hypothesis that is not a simple hypothesis is called a composite hypothesis.

Theorem 10.1: The Neyman-Pearson Lemma
Suppose that we wish to test the sample null hypothesis $H_0 : \theta = \theta_0$ versus the sample alternative hypothesis $H_a : \theta = \theta_a$, based on a random sample $Y_1, Y_2, \ldots, Y_n$ from a distribution with parameter θ. Let $L(\theta)$ denote the likelihood of the sample when the value of the parameter is θ. Then, for a given α, the test that maximizes the power at θ_a has a rejection region, RR, determined by

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

The value of k is chosen so that the test has the desired value for α. Such a test is a most powerful α level test for H_0 versus H_a.

26 Likelihood Ratio Tests

A Likelihood Ratio Test
Define λ by

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{max_{\Theta \in \Omega_0} L(\Theta)}{max_{\Theta \in \Omega} L(\Theta)}$$

A likelihood ratio test of $H_0 : \Theta \in \Omega_0$ versus $H_a : \Theta \in \Omega_a$ employs λ as a test statistic, and the rejection region is determined by $\lambda \leq k$.

Theorem 10.2 Let $Y_1, Y_2, \ldots, Y_n$ have joint likelihood function L_Θ. Let r_0 denote the number of free parameters that are specified by $H_0 : \Theta \in \Omega_0$, and let r denote the number of free parameters specified by the statement $\Theta \in \Omega$. Then, for large n, $-2ln\lambda$ has approximately a χ^2 distribution with $r_0 - r$ degrees of freedom.