Quantitative Economics Lecture 5 - Statistical Inference I : Testing A Single Hypothesis

Richard Povey

University of Oxford

1st May 2018

richard.povey@hertford.ox.ac.uk, richard.povey@st-hildas.ox.ac.uk

With thanks to Ian Crawford, Margaret Stevens, Vanessa Berenguer-Rico.

Outline of Lecture

Already Introduced in Lecture 3 on Asymptotics

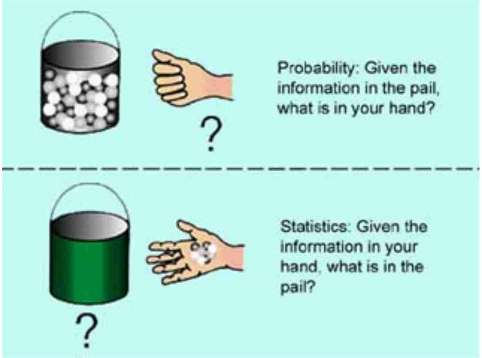
The Law of Large Numbers

The Central Limit Theorem

Statistical Inference

- Data Samples
- Estimators
- Sampling distributions
- Hypothesis Testing
- Confidence Intervals

Probability and Statistics



Populations and Samples

A population

- a complete enumeration of some set of interest
- a mathematical model of a set of interest ("data generating process")

In order to learn about the population, a sample is studied.

Cross-sectional data

- Cross-sectional data sets have one observation per member of the population/sample.

$$\{X_1, X_2, ..., X_n\} = \{X_i\}_{i=1,...,n}$$

It is conventional to assume that cross-sectional observations are mutually independent (i.e. the X_is are i.i.d.).

Populations and Samples

Time series data

- Time-series data are indexed by time.

$$\{X_1, X_2, ..., X_T\} = \{X_t\}_{t=1,...,T}$$

This type of data is characterised by serial dependence (i.e. the X_ts are *not* i.i.d. because the variable at time t is correlated with previous and future values).

Panel data

- Panel data combines elements of cross-section and time-series.

$$\{X_1^1, X_2^1, ..., X_T^1; ...; X_1^n, X_2^n, ..., X_T^n\} = \{X_t^i\}_{t=1,...,T}^{i=1,...,n}$$

It is often assumed that members are mutually independent, but a given member's observations are serially dependent.

- Most examples in QE are either cross-sectional or time series.

Statistical Inference

- We normally have a random sample from the population of interest.
- We are interested in some feature of the population (e.g. the population conditional mean).
- Whatever the population object of interest is (call it θ), it is fixed but unknown.
- Our goal is to draw inferences about this fixed but unknown attribute of the population from a sample.

Statistical Inference

Consider three questions that we might ask:

1. What is the true value of θ?
   More precisely, what is a reasonable guess as to the true value of θ?

2. Is θ equal to some particular value?
   Specifically, is the evidence that θ is not equal to this value so compelling that we can reject it?

3. What are plausible values of θ?
   In particular, is there a range of values that we can confidently claim contains the true value of θ?

Statistical Inference

Analogue Estimators

- These tools allow you to make inferences about the population expectation (e.g. the population mean μ_X) on the basis of sample expectations (e.g, the sample mean $\bar{X}_n$).
- Virtually all of the statistics you will come across in QE are sample expectations (the mean value of some function in a sample).

- Suppose you have a random sample $\{X_1, X_2, ..., X_n\}$ and you want to know what the population mean μ_X is.
- You can never be certain but you can make probabilistic statements using asymptotics and the analogy principle.
- This suggests estimating the population mean with the sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n}[X_i]$. (Note: Unlike in Lecture 3, we suppress $\bar{X}_n$ to $\bar{X}$ or order to keep notation economical.)

Statistical Inference

- We can denote $\bar{X}$ as $\hat{\mu}_X$, where the "hat" indicates a *sample estimator* or a population parameter. From the properties of sample means we know that $E\left[\bar{X}\right] = \mu_X$. We express this by saying that $\hat{\mu}_X$ is an unbiased estimator of μ_X.
- The Law of Large Numbers states that as n increases, $\bar{X}$ converges in probability to μ_X: $\hat{\mu}_X \xrightarrow{p} \mu_X$. We express this by saying that $\hat{\mu}_X$ is a consistent estimator for μ_X.
- In estimator notation, the t-statistic becomes:

$$t_{\hat{\mu}_X} = \frac{\hat{\mu}_X - \mu_X}{\sqrt{\frac{\hat{\sigma}_X^2}{n}}} \xrightarrow{d} N(0,1)$$

- The Central Limit Theorem tells us that the t-statistic has an asymptotic $N(0,1)$ distribution.
- (If the X_is are precisely normally distributed then $t_{\hat{\mu}_X} \overset{d}{=} t(n-1)$, but as n gets large this is well approximated by the standard normal distribution.)

- We don't know the population variance σ_X^2 but the analogy principle suggests estimating σ_X^2 using the sample variance:

$$\hat{\sigma}^2 = \bar{S}^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left[\left(X_i - \bar{X}\right)^2\right]$$

- Proof that $\bar{S}^2$ is an **unbiased estimator** for σ_X^2? This is one of the exercises in the problem sets. *Hint: Rearrange* $E\left[\left(\frac{1}{n-1}\right)\left(X_i - \bar{X}\right)^2\right]$ *to get* $E\left[\left(\frac{1}{n-1}\right)\left(X_i\left(1-\frac{1}{n}\right) - \frac{1}{n}\sum_{j\neq i}^{n}\left[X_j\right]\right)^2\right]$ *and then multiply out and apply the covariance definition and rules from Lecture 3.*

- The Law of Large Numbers implies consistency: $\bar{S}^2 \xrightarrow{p} \sigma_X^2$.

Statistical Inference - Hypothesis Testing

Given our probabilistic description of what we know about the behaviour of the sample mean we can go on to do various things.

1. Hypothesis tests
2. Confidence intervals

We consider testing the hypothesis that the true value of μ is actually μ_0. We set up the null and the alternative hypotheses as

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

By the Law of Large Numbers, the observed sample mean ought to be fairly close to the true population mean.
Hence, if the null hypothesis is true, then sample mean ought to be fairly close to the hypothesized value of the population mean, μ_0.
If we observe a sample mean far from μ_0, then we conclude that $\mu \neq \mu_0$, i.e. we reject H_0.

Using the definition of a standardised mean, then under the null hypothesis $\mu_X = \mu_0$ the t-statistic is a random variable with an asymptotic $N(0,1)$ distribution:

$$t = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\bar{S}^2}{n}}} = \frac{\hat{\mu}_X - \mu_0}{se\left(\hat{\mu}_X\right)} \xrightarrow{d} N(0,1)$$

Thus under the null hypothesis and the Law of Large Numbers then t should be close to zero, we should only see large values of t rarely.

If we do observe a large value for t then either H_0 is true, and we've seen something unlikely.

Or (more likely) H_0 is not true.

Statistical Inference - Hypothesis Testing

The steps involved in hypothesis testing are:

1. State your null hypothesis (H_0) and the alternative (H_a).
2. Calculate your test statistic (t).
3. Derive its asymptotic distribution using the CLT.
4. Compare it to the appropriate critical value (c_α) (depends on your chosen significance level).
5. State whether you reject or fail to reject your null hypothesis.

Important Usually it makes for a clearer and simpler test to select the null hypothesis H_0 as the one under which the distribution of the t-statistic is *known*. So, when you are asked to "test the hypothesis that..." you need to make a judgement whether the hypothesis should be the null or the alternative hypothesis.

Hypothesis tests regarding the population mean μ are usually of one of three types:

1. $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
2. $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$
3. $H_0 : \mu \geq \mu_0$ versus $H_a : \mu < \mu_0$

- Notice that in the case of the last two, the possibility that $\mu = \mu_0$, belongs to the null hypothesis. This is a technical necessity that arises because we compute significance probabilities using the μ in H_0 that is nearest H_a. For such a μ to exist, the boundary between H_0 and H_a must belong to H_0.
- We in fact only know the distribution of $\hat{\mu}$ under H_0 if we rule out a priori $\mu < \mu_0$ or $\mu > \mu_0$ in cases (2) and (3) respectively.

Choosing the Significance Level α (also called the **size** of the test):

- α is the probability of rejecting the null hypothesis when it is in fact *true* - that is, of making a **Type I error**.
- $\alpha = \text{P(Reject } H_0 | H_0 \text{ is true)} = \text{P(Type I error)}$
- A low value of α corresponds to conservative decision-making and scepticism - only reject H_0 if the evidence against is overwhelming.

Statistical Inference - Hypothesis Testing

Choosing the Significance Level α (also called the **size** of the test):

- But we also need to be concerned about **Type II** errors: That is, wrongly accepting the null hypothesis when it is in fact *false*.
- $\beta = \text{P}(\text{Accept } H_0 | H_a \text{ is true}) = \text{P}(\text{Type II error})$
- Usually we cannot precisely evaluate β because the alternative hypothesis is *composite*: $H_a : \mu \neq \mu_0$.
- But if we choose a low α, then β will tend to be high: the test will be less **powerful**:
- $\text{Power} = 1 - \beta = \text{P}(\text{Reject } H_0 | H_a \text{ is true})$

Statistical Inference - Hypothesis Testing

The choice may depend on the relative costs of **Type I** and **Type II** errors. For example:

- H_0 : the global temperature has not changed
- H_a : it has risen
- **Type I error**: conclude that climate change is happening when it is not
- **Type II error**: continue to assume that climate change is not happening when it is
- Both are costly. Choose a high α if you are more concerned about Type II errors.

A different example:

- H_0 : class sizes have no effect on educational outcomes
- H_a : smaller classes improve outcomes
- Policy maker may want a low α if rejecting H_0 means committing extra resources to education.

Example - Average Hourly Earnings

	ahe	*female*	*age*
Mean	16.7711	0.4149	29.7544
Standard Error	0.0980	0.0055	0.0324
Median	14.9039	0	30
Mode	19.2308	0	34
Standard Deviation	8.7587	0.4927	2.8911
Sample Variance	76.7147	0.2428	8.3586
Kurtosis	2.6563	-1.8810	-1.2263
Skewness	1.4108	0.3457	-0.1032
Range	58.9598	1	9
Minimum	2.0979	0	25
Maximum	61.0577	1	34
Count	7986	7986	7986

Sample of 7986 (young) individuals from CPS 2004.

$$\begin{aligned} ahe &= \text{Average Hourly Earnings} \\ female &= \text{1 if female, 0 if male} \end{aligned}$$

If μ is the population mean of ahe, and p is the population proportion of women in the labour force, we can estimate them:

$$\hat{\mu} = 16.7711, \qquad \text{se}(\hat{\mu}) = \frac{8.7587}{\sqrt{7986}} = 0.0980$$

$$\hat{p} = 0.4149, \qquad \text{se}(\hat{p}) = \frac{0.4927}{\sqrt{7986}} = 0.0055$$

Standard Error for Sample Mean of Dichotomous Variable

In the data, *female* = 1 if female, 0 if male. We can think of *female* as a Bernoulli random variable; we want to estimate the population mean:

$$p = \mathrm{P}(\textit{female} = 1)$$

We can do this just as before, with the sample mean: $\hat{p} = 0.4149$

The sample standard deviation and standard error of the estimate are:

$$s = 0.4927 \quad \text{and} \quad \mathrm{se}(\hat{p}) = \frac{0.4927}{\sqrt{7986}} = 0.0055$$

(this is what Excel does)

Books do this slightly differently:

We know that for a Bernoulli random variable Y: $\mathrm{E}(\bar{Y}) = p$ and $\mathrm{Var}(\bar{Y}) = \dfrac{p(1-p)}{n}$

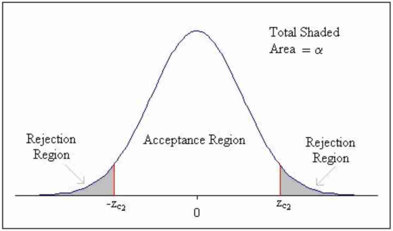
So use the estimate for the mean to work out the standard error:

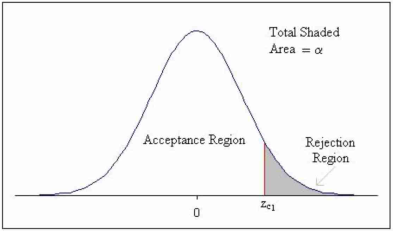
$$\hat{p} = \bar{Y} \quad \text{and} \quad \mathrm{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Gives *almost* the same answer, but slightly smaller because no degrees of freedom correction. Doesn't matter when $n = 7986$

Both $\bar{S} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} [(p_i - \hat{p})^2]}$ and $\sqrt{\hat{p}(1-\hat{p})}$ are **consistent** estimators for $\sigma_Y = \sqrt{p(1-p)}$.

One-Tailed and Two-Tailed Tests





Two-sided Alternative

$H_0 : \mu = \mu_0$
$H_a : \mu \neq \mu_0$

$$t = \frac{\hat{\mu} - \mu_0}{\text{se}(\hat{\mu})}$$

Two-tail test: reject H_0 at significance level α if $|t| > z_{c2}$

One-sided Alternative

$H_0 : \mu = \mu_0$
$H_a : \mu > \mu_0$

$$t = \frac{\hat{\mu} - \mu_0}{\text{se}(\hat{\mu})}$$

One-tail test: reject H_0 at significance level α if $t > z_{c1}$

Note - The diagram illustrates case 2 for a one-tailed test.

Hypothesis Testing Example - Proportion of Females

Two-tailed test

Suppose we are interested in the hypothesis that $p_{female} = 0.5$. We proceed as follows:

1. $H_0 : p_{female} = 0.5$ versus $H_a : p_{female} \neq 0.5$
2. $t = \frac{\hat{p}_{female} - 0.5}{se(\hat{p}_{female})} = \frac{0.4149 - 0.5}{0.0055} = -15.47$
3. Under H_0, $t \xrightarrow{d} N(0, 1)$ by CLT.
4. The critical value is 1.96 (5% significance/95% confidence).
5. Since $|t| > 1.96$, we can reject H_0 and accept H_a.

One-tailed test (case 3)

Suppose we are interested in the hypothesis that $p_{female} < 0.5$. We proceed as follows

1. $H_0 : p_{female} = 0.5$ versus $H_a : p_{female} < 0.5$
2. $t = \frac{\hat{p}_{female} - 0.5}{se(\hat{p}_{female})} = \frac{0.4149 - 0.5}{0.0055} = -15.47$
3. Under H_0, $t \xrightarrow{d} N(0, 1)$ by CLT.
4. The critical value is -1.64 (5% significance/95% confidence).
5. Since $t < -1.64$, we can reject H_0 and accept H_a.

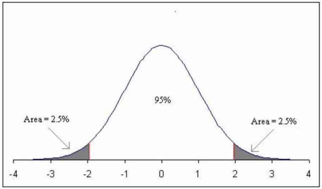
Statistical Inference - Confidence Intervals

Using our ability to switch from the standard normal - described by $N(0,1)$ - and our knowledge of $\hat{\mu}_X$ - described by $N\left(\mu_X, \frac{\bar{S}^2}{n}\right)$ - we can say that :

- we would expect the population mean μ_X to lie within $1.64\sqrt{\frac{\bar{S}^2}{n}}$ of the sample mean $\hat{\mu}_X$ with 90% probability.
- we would expect the population mean μ_X to lie within $1.96\sqrt{\frac{\bar{S}^2}{n}}$ of the sample mean $\hat{\mu}_X$ with 95% probability.
- we would expect the population mean μ_X to lie within $2.58\sqrt{\frac{\bar{S}^2}{n}}$ of the sample mean $\hat{\mu}_X$ with 99% probability.

Confidence Interval Example - Average Hourly Earnings

Recall from Lecture 3 that $\Phi^{-1}\left(1-\frac{0.05}{2}\right)=\Phi^{-1}(0.975)=1.96$.
To calculate the 95% confidence interval we need:
$1.96se(\hat{\mu})=1.96\times 0.0980=0.1921$ and $\hat{\mu}=16.7711$.



$$\mathrm{P}\left(-1.96\leq\frac{\hat{\mu}-\mu}{\mathrm{se}(\hat{\mu})}\leq 1.96\right)=0.95$$

$$\Rightarrow\quad \mathrm{P}\left(\hat{\mu}-1.96\mathrm{se}(\hat{\mu})\leq\mu\leq\hat{\mu}+1.96\mathrm{se}(\hat{\mu})\right)=0.95$$

$$\hat{\mu}-1.96\mathrm{se}(\hat{\mu})=16.5790\qquad \hat{\mu}+1.96\mathrm{se}(\hat{\mu})=16.9632$$

$\Rightarrow$ The 95% confidence interval is $[16.58, 16.96]$.

Similarly for a 99% confidence interval calculate: $\{\hat{\mu}\pm 2.58\mathrm{se}(\hat{\mu})\}\longrightarrow[16.52, 17.02]$.

For 90% confidence interval use $\Phi^{-1}\left(1-\frac{0.1}{2}\right)=\Phi^{-1}(0.95)=1.64$ to calculate:
$\{\hat{\mu}\pm 1.64se(\hat{\mu})\}\rightarrow[16.61, 16.93]$.

Note that the 95% confidence interval is constructed from the estimator for the sample mean $\hat{\mu}$ and its standard error $se\left(\hat{\mu}\right)$, both of which are random variables due to sampling variation. There is a 95% probability that the population mean μ lies within the 95% confidence interval, or in other words, if we calculate many different 95% confidence intervals from many different samples then we will "capture" the population mean in 95% of them. It is, however, *not* meaningful to talk about a particular probability distribution for the population mean μ. This is simply a constant, even though we do not *know* it and have to make inferences about it from our sample.