Handout #7

Title: Foundations of Econometrics | Fall/2015
Course: Econ 367 | Instructor: Dr. I-Ming Chiu

Statistical Inference (chapter seven, eight, and nine)

Notice: the purpose of this handout is not to cover all the details of statistical inference. Instead, I would like to provide the main concepts of statistical inference using some simple examples.

What's statistical inference?
Def: Statistical inference is the process of using data to make inference about the statistical models that generate the data.

e.g. Suppose we believe that number of students that misses each econ 367 lecture can be modeled using Poisson distribution, then we may want to find how large "λ" is from the collected data.

e.g. Suppose we model the exam outcome as a linear function of study time plus some errors; Score = alpha0 + alpha1*time + error. Then we may want to find the estimates of alpha0, alpha1 and some numerical aspects of the error term (e.g. the variance).

We'll be dealing with "parametric" models only; meaning the models we would like to infer to depends on parameters such as ““λ” in the above example.

A. Point estimate (population mean, variance, proportion, and covariance)

You have learned some sample statistics that can be used to represent the population parameters. The following is a short list of "point estimators" that were used earlier.

Sample Mean: $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n} \Rightarrow E(\overline{X}) = \mu_X$

Sample Variance: $s_X^2 = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \Rightarrow E(s_X^2) = \sigma_X^2$

Sample Proportion: $\hat{p} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ (where X_i = 0 or 1) $\Rightarrow E(\hat{p}) = p$

All of these formulas are termed "estimators".

How do we obtain the above "formulas" to estimate population characteristics?
Note: Formula = Estimator
There are three methods for obtaining estimators;
(i) Method of Moment.
(ii) Maximum Likelihood and
(iii) Least Square (for linear regression model)
(iv) Bayesian

Method of Moment

Def: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f(x; \theta_1, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. Then the moment estimators $\hat{\theta}_1; \ldots; \hat{\theta}_m$ are obtained by equating the first m sample moments to the corresponding first m population moments in order to solve for $\theta_1, \ldots, \theta_m$.

K^{th} population moment = $E(X^k)$

K^{th} sample moment = $\dfrac{\sum_{i=1}^{n} X_i^{\,k}}{n}$

e.g. Poisson random variable has a mass function that depends on one parameter, λ, and $E(X) = \lambda$. We equate the following to estimate λ

$\dfrac{\sum_{i=1}^{n} X_i^{\,1}}{n}$ (first sample moment) = $\hat{\lambda}$ (you can remove exponent "1" for X_i)

Maximum Likelihood Method

Likelihood function = $L(X_1, X_2, \ldots, X_n | \theta)$; there can be more than one θ

e.g. Toss a coin ten time and obtain 2 heads. What is the estimate of P?

$$\underset{P}{\text{Maximize}}\ L = \binom{10}{2} P^2 * (1 - P)^8$$

Take the log function on both sides and we would like to

$$\underset{P}{\text{Maximize}}\ \log(L) = 2 * \log P + 8 * \log(1 - P)$$

Logarithm = a monotoic funciton.

We can generalize the above likelihood function as the following

$$\text{Maximize } L = \binom{n}{k} P^{(\sum_{i=1}^{k} X_i)} * (1 - P)^{(n - \sum_{1}^{k} X_i)}$$

Both are exponents

Criteria for selecting Estimators

If the expectation of the estimator equals the true parameter, we say the estimator is unbiased (i.e., $E(\hat{\theta}) = \theta$).

e.g. If X is a Binomial random variable with parameter n and θ. Please show that the sample proportion estimator, $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$, is an unbiased estimator.

e.g. See whether you can show that $E(s_X^2) = \sigma_X^2$

The other two nice properties for an estimator in addition to unbiasedness are **minimum variance** and **efficiency**.

Mean Squared Error (MSE) = $E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + Bias^2$; where bias = $E(\hat{\theta}) - \theta$.

MSE can be used as a criteria to choose the estimator. Suppose both $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators for θ. If $E[(\hat{\theta}_1 - \theta)^2] \leq E[(\hat{\theta}_2 - \theta)^2]$ for all $\theta \in \Theta$, we say $\hat{\theta}_2$ is better or more efficient than $\hat{\theta}_1$.

B. Confidence Intervals (CI)

Standardization of $\overline{X} \Rightarrow Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

In practice σ is usually unknown, therefore, the sample standard deviation s is used to replace σ. However, from equation (3) on page 2 of the previous handout (#6) we learned that the distribution is no longer standard normal, we should write

$$\frac{\overline{X} - \mu}{s / \sqrt{n}} \sim t_{n-1}$$

Use the t distribution[1] table or the corresponding command in Stata, we can write the following

$Pr(-H < t_{n-1} < H^{2}) = 0.95$

$$Pr(-H < \frac{\overline{X} - \mu}{s / \sqrt{n}} < H) = 0.95 \Rightarrow$$

$$Pr(-H^* \frac{s}{\sqrt{n}} < \overline{X} - \mu < H^* \frac{s}{\sqrt{n}}) = 0.95 \Rightarrow$$

$$Pr(H^* \frac{s}{\sqrt{n}} > \mu - \overline{X} > -H^* \frac{s}{\sqrt{n}}{}^{3}) = 0.95 \Rightarrow$$

$$Pr(\overline{X} + H^* \frac{s}{\sqrt{n}} > \mu > \overline{X} - H^* \frac{s}{\sqrt{n}}) = 0.95$$

We can say that there is a 95% chance that the mean μ will lie in the range of $\overline{X} \pm H^* \frac{s}{\sqrt{n}}$.

We can replace H by $t_{1-2\alpha,\, n-1}$, where α is referred to as rejection region in two-tail test used in hypothesis.

e.g. Let's assume that $\overline{X}$ = 68 (student weight), s = 10 and n = 36 from a example of students' weight collected.

$$68 \pm 2.03^* \frac{10}{\sqrt{36}} \cong 68 \pm 3.38$$

95% chance that mean μ will fall in the range (64.62, 71.38).

e.g. In a random sample of 64 law firms, legal charges per hour are found to have a mean of \$40 with a standard deviation 5. Obtain a 99% CI for the average legal charge per hour in the law profession as a whole.

---

[1] If variance is known, we can use the standard normal distribution (Z).

[2] H varies with the size of confidence interval as well as the distribution used (Z or t).

[3] $a < X < b \Rightarrow -a > -X > -b$

C. Hypothesis Testing

Def: A statistical hypothesis is a conjecture about the sampling distribution of a random variable. When a hypothesis completely specifies the distribution, it is called a simple hypothesis; if it is not, it is referred to as a composite hypothesis.

Neyman and Pearson introduce their hypothesis testing approach by specifying the null hypothesis and alternative hypothesis. The former is denoted by H_0 and the latter H_1 or H_A. The null statement is a status quo and our goal is show that whether the evidence from the data is "strong" enough to reject the null hypothesis.

e.g.

H_0: $\mu = 42{,}000$

H_1: $\mu = 45{,}000$

H_0: $\mu \leq 42{,}000$

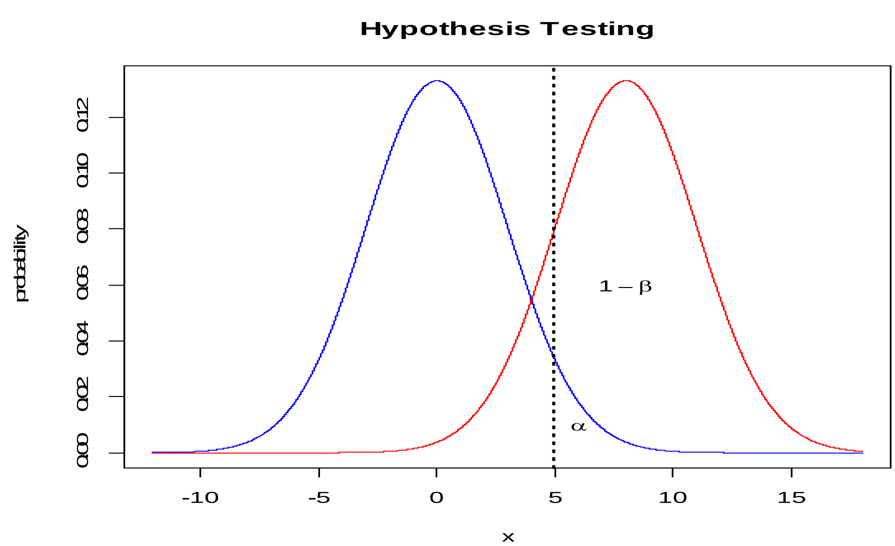
H_1: $\mu > 42{,}000$

Type I and Type II Error

	H_0 is true	H_0 is false
Accept H_0		Type II (β)
Reject H_0	Type I (α)	

Power = $1 - \beta$

Power: the ability to detect a false null hypothesis.
In practice we like our test to have high power given the reject region.



Note: $\alpha = 0.05$ is the rejection region.

One-tail Test

Let's use the previous numerical example from page 5. There we assume that $\overline{X}$ = 68, s = 10 and n = 36. This example is about policy effectiveness. The tax is levied on sweet drinks in order to help children reduce their weight (childhood obesity is a popular research topic in recent years). We are interested in examining whether the average weight is "actually" decreased the sweet drinks tax?

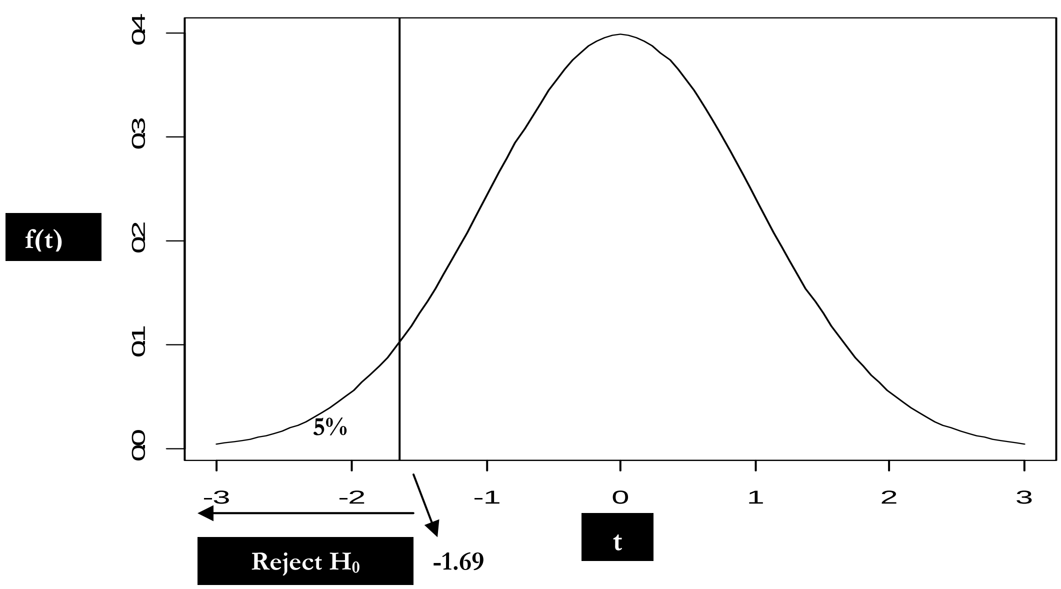
H_0: $\mu = 70$ … Null Hypothesis
H_A: $\mu < 70$ … Alternative Hypothesis

If the null hypothesis is true, then TS (test statistic) has a t distribution and can be represented as follows:

$$TS = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1} \Rightarrow \frac{\overline{X} - 70}{s/\sqrt{n}} = \frac{68 - 70}{10/\sqrt{36}} = -1.2$$ … We can't not reject that $\mu = 70$.

Conclusion: the new tax on sweet food may not be effective to reduce the students' weight.

Let's choose a level of significance that equals 5%[4]; the area under the curve and on the left of the value -1.69. Given the null is true, the chance that TS would fall in the level of significance is only 5%. Meaning, it is unlikely to happen. Therefore, the reasonable judgment is to reject the null hypothesis and accept alternative hypothesis. Note that there is a 5% chance that the null can be true. If this is the case, we may make the type I error[5].

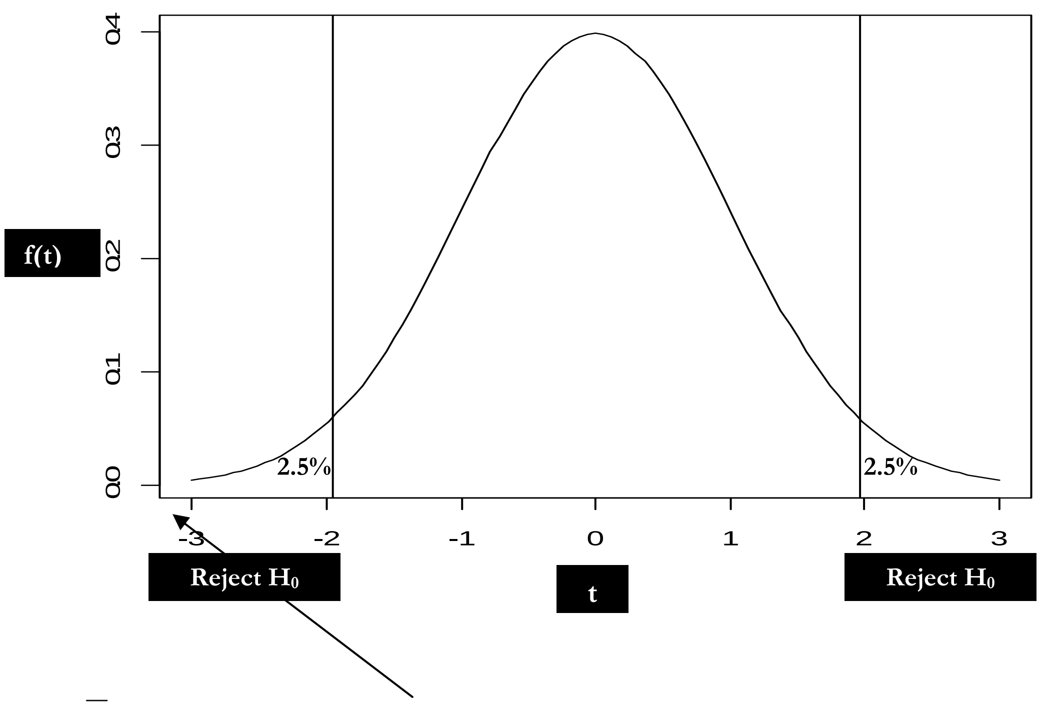


[4] It can also be 1% or 10% depends on researcher's choices.
[5] The Type II error refers to the situation where we accept a false null hypothesis. To reduce Type I error would result in an increase of Type II error. In practice we either reject the null hypothesis or reserve the judgment about it. By doing this we never accept null hypothesis and therefore we **cannot** make Type II error.

Two-tail Test

e.g. The mean lifetime of a random sample of 80 light bulbs produced by a factory is found to be 1460 hours, with a standard deviation s = 110 hours. If μ is the mean lifetime of all the light bulbs produced by the factory, test the hypothesis $\mu = 1500$ against the alternative hypothesis that $\mu \neq 1500$, using the level of significance 0.05.

H_0: $\mu = 1500$ … Null Hypothesis
H_A: $\mu \neq 1500$ … Alternative Hypothesis



$$TS = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{1460 - 1500}{110/\sqrt{80}} = -3.25$$ … We reject that $\mu = 1500$.

***The other method to conduct hypothesis testing is to use P-value. P-value is defined as the lowest significance level at which the null hypothesis can be rejected. Let's use the one-tail test from page 7. Since the TS is -1.2, we can find the P-value is 0.119 or 11.9%. Graphically, it is an area on the left of TS (-1.2) under the t curve.