11. **Exact normal** distribution theory. We assumed the random draws are from the **normal** (Gaussian) distribution. Hence, if $y_t \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\hat{\mu}$ is also normal.

    (Why? Because sums of normals are normal). The estimator (sample mean) has sample variation. (Why? Because different realizations have different time-series means.)

    The **sample mean** is an **unbiased** estimator,

    $$E(\hat{\mu}) = \mu$$

    with **variance**

    $$\text{Var}(\hat{\mu}) = \frac{1}{T^2}(T\sigma^2) = \frac{\sigma^2}{T} \to 0 \ \text{ as } \ T \to \infty$$

    We can say, the **sample mean** is **normally distributed** with mean μ and variance σ^2/T. We write it as,

    $$\hat{\mu} \sim N(\mu, \sigma^2/T)$$

    As $T \to \infty$ sample mean becomes more precise (the **randomness vanishes**). The distribution becomes **degenerate**.

    Another way to write this is in the standardized **nondegenerate or asymptotic** form,

    $$\sqrt{T}\frac{(\hat{\mu} - \mu)}{\sigma} \sim N(0, 1) \tag{6}$$

    Test the hypothesis that the true mean is μ with the t-ratio (z-statistic).

    First, we need to **estimate** σ^2 (to simplify, we ignore degrees of freedom correction.)

    $$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\mu})^2$$

    which itself is a **random variable** and has a $\chi^2(T-1)$ distribution. (Why? because it's a function of the y_t which are random variables. Also, the square of a normal is chi-squared.)

    Hence, when we substitute $\hat{\sigma}^2$ into (6), the distribution is not normal (we have a ratio of a normal random variable to the square-root of a chi-square variable). William Gosset did the math and showed this ratio to have the **t-distribution** with $T-1$ degrees of freedom.

    $$z = \sqrt{T}\frac{(\hat{\mu} - \mu)}{\hat{\sigma}} \sim t_{T-1}$$

    Call

    $$\frac{\hat{\sigma}}{\sqrt{T}} = se(\hat{\mu})$$

the **standard error** of $\hat{\mu}$. So we can also write,

$$z = \frac{(\hat{\mu} - \mu)}{se(\hat{\mu})} \sim t_{T-1}$$

These are **exact** results. That means for any T, we can look up the results in a table. The t-distribution applies exactly for any sample size T.

The story behind the **student t distribution**: In the 1890s, **William Gosset** was studying chemical properties of barley with small samples, for the Guinness company (yes, that Guinness). He showed his results to the great statistician Karl Pearson at University College London, who mentored him.

Gossett published his work in the journal *Biometrika*, using the pseudonym Student, because he would have gotten in trouble at Guinness if he used his real name.

Hypothesis testing method (statistical inference)

(a) Assume the null hypothesis is true. (e.g., $\beta = 0$)

(b) Determine the sampling distribution of your test statistic under the null hypothesis. (e.g., the t-statistic, follows a student-t for small samples, and $N(0, 1)$ for larger samples).

(c) Ask if the observed test statistic, computed using data, could reasonably be drawn from the null distribution.

If answer is yes, data are consistent with the null. You cannot reject the null hypothesis

If answer is no, then you can reject the null. The classical hypothesis testing methodology is due to R.A. Fisher.

Figure 3: t-test Review: Two-sided Test

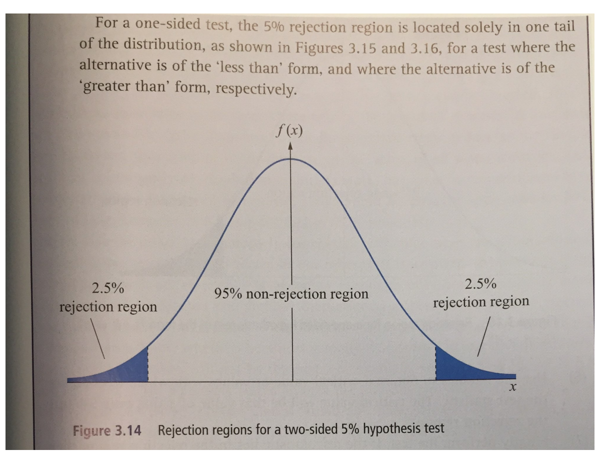
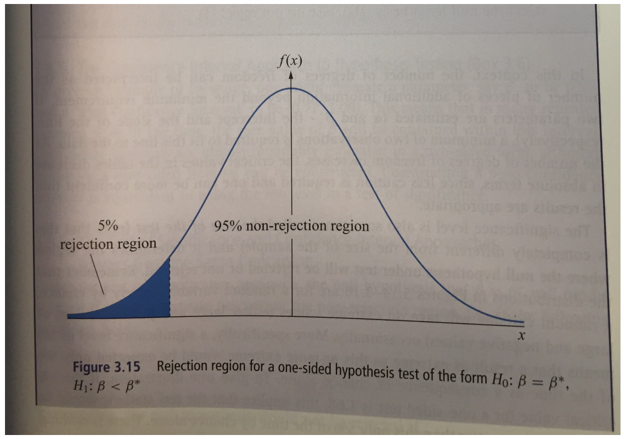


Figure 4: t-test Review: One-sided Test



12. Convergence (**large sample or asymptotic theory**). With time-series, we don't like to assume the underlying shocks are normally distributed. We don't like to be so specific as to name any one particular distribution. In fact, a pretty strong assumption that we often make is simply that the shock sequence is i.i.d. (no distribution named) with mean μ and variance σ^2.

    If you don't assume a distribution, then how can you do a t-test?

    - **Law of large numbers**. Under certain **regularity conditions**, as $T \rightarrow \infty$

    $$\hat{\mu} \overset{a.s.}{\rightarrow} \mu$$

    - **Central limit theorem**. Under certain regularity conditions,

    $$\frac{(\hat{\mu} - \mu)}{se(\hat{\mu})} \overset{D}{\rightarrow} N(0,1)$$

    The second result is **amazing**, right? Philosophically, it's saying something about the universality of the normal distribution.

    What these two results say is, under certain regularity conditions (e.g., the time-series is stationary and ergodic), if we had a large sample ($T = \infty$), the z-statistic has the standard normal distribution. The way econometricians say it is "z is asymptotically normal."

    In applications, for finite T, we use these asymptotic results even though we never have $T = \infty$, and hope that the excact but unknown distribution of z is well approximated by the asymptotically normal distribution.

2 Regression Review

Text: Chapter 3.

2.1 Regression In Population

1. Think about regressing y_t on x_t. Both variables are time-series, both are stationary and ergodic.
$$y_t = \alpha + \beta x_t + \epsilon_t \tag{7}$$
The error term ϵ_t is i.i.d. The systematic part of **regression** is also referred to as **projection**. The ϵ_t is projection **error**.

2. Here is how I want you to think of regression. The right side is conditional expectation
$$E\left(y_t|x_t\right) = \alpha + \beta x_t$$
because $E\left(\epsilon_t|x_t\right) = 0$. The conditional expectation is the best linear predictor (forecast) of y_t, conditional on the information x_t.

3. Next, take the mean,
$$\begin{aligned} E\left(y_t\right) &= \alpha + \beta E\left(x_t\right) \\ \alpha &= \mu_y - \beta\mu_x \end{aligned}$$
Substitute to eliminate the constant. Right now, we're not interested in the constant (we will be later in the course when we study alpha).
$$y_t - \mu_y = \beta\left(x_t - \mu_x\right) + \epsilon_t$$
This is regression stated in deviations from the mean form. β is a function of moments from the joint (or bivariate) distribution between x_t and $y_{t.}$

Multiply both sides by $x_t - \mu_x$,then take expectations, solve for β
$$\begin{aligned} \left(y_t - \mu_y\right)\left(x_t - \mu_x\right) &= \beta\left(x_t - \mu_x\right)^2 + \epsilon_t\left(x_t - \mu_x\right) \\ \underbrace{E\left[\left(y_t - \mu_y\right)\left(x_t - \mu_x\right)\right]}_{\mathrm{Cov}(x_t, y_t)} &= \beta\underbrace{E\left(x_t - \mu_x\right)^2}_{\mathrm{Var}(x_t)} + \underbrace{E\left[\epsilon_t\left(x_t - \mu_x\right)\right]}_{0} \end{aligned}$$
$$\beta = \frac{\mathrm{Cov}\left(x_t, y_t\right)}{\mathrm{Var}\left(x_t\right)} = \frac{\mathrm{Cov}\left(x_t, y_t\right)}{\mathrm{sd}\left(x_t\right)\mathrm{sd}\left(y_t\right)}\frac{\mathrm{sd}\left(y_t\right)}{\mathrm{sd}\left(x_t\right)} = \rho\left(x_{t,} y_{t,}\right)\frac{\sigma\left(y_t\right)}{\sigma\left(x_t\right)}$$
This what we are estimating when we run regression. In finance, **covariance** is **risk.**

4. In financial econometrics, y_t is an asset return and x_t is a risk-factor. We use regression to compute the **betas**, which is **exposure** of this asset to the **risk factor**.

   We are less concerned about instrumental variables and establishing cause and effect.

   We are more concerned about understanding reduced form correlations and in understanding the statistical **dependence** across time series, and **dependence** of observations across time.

Want to know why it's called regression? Darwin's cousin, **Galton**, was studying heights. x_i is average height of parents, y_i is height of kid. He's running the regression

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Height distribution doesn't change across generations:

$$E(y_i) = E(x_i) = \mu$$

$$\text{Var}(y_i) = \text{Var}(x_i) = \sigma^2$$

which implies

$$\beta = \rho\left(x_{i,} y_i\right) \frac{\sigma\left(y_i\right)}{\sigma\left(x_i\right)} = \rho$$

$$\alpha = \mu_y - \beta\mu_x = \mu(1-\beta) = \mu(1-\rho)$$

Hence,

$$E(y_i|x_i) = \alpha + \beta x_i = (1-\rho)\mu + \rho x_i$$

Expected height of child is weighted average of parent and population height. Child not expected to be has tall as parents. Hence, **regression to the mean**.

2.2 Least squares estimation of β

1. Just like we did with the sample mean and sample variance, compute the sample counterparts to the population moments. The least squares estimator of β is,

$$\begin{aligned}
\hat{\beta} &= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\mu}_y\right)\left(x_t - \hat{\mu}_x\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2} \\
&= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\beta x_t + \epsilon_t - \beta\hat{\mu}_x - \hat{\mu}_\epsilon\right)\left(x_t - \hat{\mu}_x\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2} \\
&= \beta + \frac{\sum_{t=1}^{T}\epsilon_t\left(x_t - \hat{\mu}_x\right)}{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2}
\end{aligned}$$

   where we get the last equation because $\frac{1}{T}\sum\mu_\epsilon\left(x_t - \hat{\mu}_x\right) = \hat{\mu}_\epsilon\frac{1}{T}\sum\left(x_t - \hat{\mu}_x\right) = 0$.

In the last line, $\hat{\beta}$ is a linear combination of the regression errors ϵ_t, which are random variables. Therefore $\hat{\beta}$ is a random variable and has a distribution.

2.3 Normal regression theory with nonstochastic regressors. (Inference)

Here, mimic what we did with the sample mean.

1. Find the mean of the estimator
2. Find the variance of the estimator
3. Use the above to get the standard error of the estimator
4. Form the t-ratio

- In your basic econometrics course, you assumed the $x's$ are exogenous constants (not random variables). We interpret them as something under control of the experimenter. The expermment is something like x is food and y is some sort of test outcome like cholesterol. ϵ is random variation unrelated to diet. We assume the ϵ_t are i.i.d. normal with mean 0 and variance σ_ϵ^2.

  First, $\hat{\beta}$ is unbiased. Take expectations,

$$E\left(\hat{\beta}\right) = \beta + \frac{1}{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2} \underbrace{E\left(\sum_{t=1}^{T} \epsilon_t \left(x_t - \hat{\mu}_x\right)\right)}_{\text{0 b/c x's treated as constant}} = \beta \tag{8}$$

  Compute variance

$$\begin{aligned} E\left(\hat{\beta} - \beta\right)^2 &= \left(\frac{1}{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2}\right)^2 E\left(\sum_{t=1}^{T} \epsilon_t \left(x_t - \hat{\mu}_x\right)\right)^2 \\ &= \left(\frac{1}{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2}\right)^2 \left(\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2\right)\sigma_\epsilon^2 \\ &= \frac{\sigma_\epsilon^2}{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2} = V\left(\hat{\beta}\right) \end{aligned}$$

  Hence, if we assume normality (of the ϵ_t), we get normality of the estimator,

$$\hat{\beta} - \beta \sim N\left(0, V\left(\hat{\beta}\right)\right)$$

- To do the t-test, we need to estimate σ_ϵ^2. The regression residual is $\hat{\epsilon}_t$,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2$$

The standard error of $\hat{\beta}$ is

$$se\left(\hat{\beta}\right) = \sqrt{\hat{V}\left(\hat{\beta}\right)} = \frac{\hat{\sigma}_\epsilon}{\sqrt{\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2}}$$

and the z statistic has the t_{T-2} distribution (lose 2 degrees of freedom due to constant and slope).

$$z = \frac{\hat{\beta} - \beta}{se\left(\hat{\beta}\right)} \sim t_{T-2}$$

2.4 Time-series regression theory with stochastic regressors

1. Two assumptions we cannot make when dealing with time series in financial econometrics are (1) the regressors are non-stochastic and (2) the regression errors are normal.

2. When we relax these assumptions, we lose the exact distributional features of least squares (i.e., the ratio of the least squares estimator to the standard error is no longer student t with $T-2$ degrees of freedom.

3. If we assume y_t and x_t are weakly stationary and ergodic time-series,

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\left(x_t - \hat{\mu}_x\right)^2 \xrightarrow{a.s.} Q \\ \hat{\sigma}_\epsilon^2 &= \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2 \xrightarrow{a.s.} \sigma^2 \end{aligned}$$

where Q is a constant, then using the same **large-sample asymptotic** logic we above to show it is possible to show (actually with a lot of tedius and high-level math) using the law of large numbers and the central limit theorem,

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \rightarrow N\left(0, \sigma_\epsilon^2 \frac{1}{TQ}\right)$$

which in practical terms means

$$\frac{\sqrt{T}\left(\hat{\beta}-\beta\right)}{\frac{\hat{\sigma}_{\epsilon}}{\sqrt{\frac{1}{T}\sum(x_t-\hat{\mu}_x)^2}}} \rightarrow N\left(0,1\right) \tag{9}$$

where

$$\hat{\sigma}_{\epsilon}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2$$

Let

$$se\left(\hat{\beta}\right) = \frac{\hat{\sigma}_{\epsilon}}{\sqrt{\sum_{t=1}^{T}\left(x_t-\hat{\mu}_x\right)^2}}$$

Cancel out the $\sqrt{T}$ in (9) and hope that the t-ratio (or the z-statistic)

$$z = \frac{\left(\hat{\beta}-\beta\right)}{\text{se}\left(\hat{\beta}\right)} \sim N\left(0,1\right)$$

is a good approximation to the unknowable exact distribution.

4. In this course, we are interested in t-ratios. We are not interested in F-statistics. Nobody's opinion ever was changed by having a significant F-stat and insignifcant t-ratios.

5. But we are also interested in R^2, the measure of goodness of fit.

$$R^2 = \frac{SSR}{SST} = \frac{\sum\tilde{y}_t^2}{\sum\tilde{x}_t^2} = 1-\frac{\sum\hat{\epsilon}_t^2}{\sum\tilde{x}_t^2} = 1-\frac{SSE}{SST}$$

2.5 Frisch-Waugh theorem

What does 'controlling for z mean in multiple regression?

- Suppose we looking at the regression (ignore constants)

$$y_t = \beta x_t + \delta z_t + \epsilon_t \tag{10}$$

  β is the coefficient of interest, but endogeneity of x_t makes us want to 'control' for z_t.

- Regress y_t on z_t and x_t on z_t, save the residuals y_t^o and x_t^o

$$y_t = az_t + y_t^o$$

$$x_t = bz_t + x_t^o$$

- Now regress y_t^o on x_t^o

$$y_t^o = \beta x_t^o + \epsilon_t \tag{11}$$

  It is not a typo that the slope is β in both (10) and (11). They are identical. This is called the Frisch-Waugh theorem.

  When we control for z_t, it means we are looking at the relationship between y_t and x_t after removing any influence z_t has on both y_t and x_t.