INTERNATIONAL UNIVERSITY OF JAPAN
Public Management and Policy Analysis Program
Graduate School of International Relations

ADC6512 (Winter 2012**)**
Topics in Data Analysis (Panel Data Models Using Stata)

This note summarizes ordinary least squares to refresh your memory and accordingly may NOT be a substitute of textbooks.

1. Ordinary Least Squares

$Y = X\beta + \varepsilon$, where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_N \end{bmatrix}, (X'X)^{-1} = \begin{bmatrix} v_{00} & & & \\ & v_{11} & & \\ & & \dots & \\ & & & v_{kk} \end{bmatrix}$$

$\mathrm{Y = Xb + e}$
$\mathrm{X'Y = X'Xb + X'e}$ [1]
$\mathrm{X'Y = X'Xb}$ since $\mathrm{X'e = 0}$
$b = (X'X)^{-1}X'Y$

2. Gauss-Markov Assumptions

2.1 **Linearity** says that the dependent variable is formulated as a linear function of a set of independent variable and the error (disturbance) term.
2.2 **Exogeneity** says that the expected value of disturbances is zero or disturbances are not correlated with any regressors. $E(\varepsilon_i | X)=0$.
2.3 **Spherical disturbances** says that disturbances have the same variance (**3.a homoskedasticity**) and are not related with one another (**3.b nonautocorrelation**)[2]; $Var(\varepsilon_i|X)=\sigma^2$; $E(\varepsilon\varepsilon'|X)=\sigma^2 I$
2.4 The observations on the independent variable are not stochastic but **fixed** (independent variables in repeated samples without measurement errors.
2.5 **Full rank** assumption says that there is no exact linear relationship among independent variables (no perfect multicollinearity). In order for identifiability of the model parameters, X is an *N* by *k* matrix with rank *k*. Otherwise, | X'X | is zero and $(X'X)^{-1}$ is not defined; $b = (X'X)^{-1}X'Y$ is not solvable.

Caution: Normality says that the stochastic part of the model (disturbances) is normally distributed: $\varepsilon \sim N[0, \sigma^2 I]$.[3] But this assumption is not strictly required for OLS.

[1] Alternatively, $\mathrm{e = Y - Xb}$; $\mathrm{X'e = X'(Y - Xb)}$; $\mathrm{0 = X'Y - X'Xb}$; $\mathrm{X'Y = X'Xb}$
[2] Constant variance is labeled homoscedasticity. The assumptions of observations from their expected values are uncorrelated. Disturbances that meet the twin assumptions of homoscedasticity and nonautocorrelation are called spherical disturbances.
[3] For a given setting of the independent variable x, ε_i, is normally distributed with mean 0 and variance σ^2.

3. Properties of Ordinary Least Squares

This section summarizes key properties of ordinary least squares.

3.1 OLS as Linear Estimator

$\hat{\beta} = b = (X'X)^{-1}X'Y = WY$, where $W = (X'X)^{-1}X'$. It implies that *b* is a linear combination of the dependent variable (random variable) *Y* or a weighted sum of *Y*. Therefore, when sample size (N) is large, the distribution of $\hat{\beta}$ will approach normality.

3.2 OLS as Unbiased Estimator

$$b = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon = \beta + (X'X)^{-1}X'\varepsilon$$
$$E(b) = E(\beta + (X'X)^{-1}X'\varepsilon) = \beta + E[(X'X)^{-1}X'\varepsilon] = \beta + (X'X)^{-1}X'E(\varepsilon) = \beta$$
$$b - \beta = (X'X)^{-1}X'\varepsilon,\ E(b - \beta) = E[(X'X)^{-1}X'\varepsilon] = E(b) - \beta = \beta - \beta = 0$$

Unbiasedness does not, however, mean that b=β. Kennedy (2008:15) states, "[I]t say onlys that, if we could undertake repeated sampling an infinite number of times, we would get the correct estimate 'on the average.' "

3.3 OLS as (Best) Efficient Estimator

$$Var(b) = E[(b - E(b))(b - E(b))'] = E[(b - \beta)(b - \beta)'] = E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] =$$
$$(X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} = \sigma^2 I(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2 I(X'X)^{-1}$$

$\sigma^2 I(X'X)^{-1}$ implies that the variance is constant for all settings of *x*.

3.4 Gauss-Markov Theorem

This theorem says that the least squares estimator $\hat{\beta}$ among all other linear unbiased estimators β has minimum variance (is more efficient). Kennedy (2008: 16) put, the best unbiased estimator is "the unbiased estimator whose sampling distribution has the smallest variance."

3.5 Asymptotic Properties of OLS
- Consistency
- Asymptotic efficiency

4. Reading OLS ANOVA Table

This section summarizes statistics in an ANOVA (Analysis of Variance) table of ordinary least squares.

4.1 Partition of Variance of Y

Total variation of the dependent variable Y is called SST (Total Sum of Squares). SSR (Model Sum of Squares) or SSR (Sum of Squares of Regression) is the variation of Y that is explained by the regression model, while SSE (Error Sum of Squares) is the variation of Y that is NOT explained by the model. SSE is the key information of OLS.

$$\underbrace{\sum (Y_i - \bar{Y})^2}_{Total} = \underbrace{\sum (\hat{Y} - \bar{Y})^2}_{Model} + \underbrace{\sum (Y_i - \hat{Y})^2}_{Residual}$$

SST = SSM + SSE

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y})^2 = e'e$$

$$MSE = \frac{SSE}{N-K} = \frac{1}{N-K}\sum_{i=1}^{n} (Y_i - \hat{Y})^2 = s^2 = \hat{\sigma}^2$$

$$SEE = \sqrt{MSE} = s = \hat{\sigma}$$

- SEE: sum of squared errors or error sum of squares
- MSE: mean squared errors or mean of error sum of squares
- SEE: standard errors of the estimates or square root of mean squared errors (SRMSE)

4.2 Degrees of Freedom

df_{model} = K-1
df_{error} = N-K
df_{total} = N-1 = (K-1) + (N-K)

When adding *j* independent variables, you are losing *j* degrees of freedom (of errors) in your regression model; you are scarifying "parsimony" of your model.

Sources	Sum of Squares	DF	Mean Squares	F
Model	SSM (Model sum of squares)	*K-1*	$MSM = \frac{SSM}{K-1}$	$F = \frac{MSM}{MSE}$
Residual (Error)	SSE (Error sum of squares)	*N-K*	$MSE = \frac{SSE}{N-K}$	
Total	SST	*N-1*		

4.3 Coefficient of Determination

R^2 represents the proportion of the variability in the dependent variable *y* that is accounted for by the independent variables of the model. Technically speaking, R^2 is SSM (or SSR) divided by SST. R^2 will always increase (never decrease) for each independent variable added, even when the new *x* has very little prediction power. Adjusted R^2 gives penalty for adding independent variables.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Adjusted $R^2 = 1 - \frac{(N-1)}{N-K}(1-R^2) = 1 - \frac{e'e/(N-K)}{Y'M^0Y/(N-1)}$,

However, R^2 should not be considered a panacea to evaluate goodness-of-fit of a model. You MUST acknowledge the limitations and risk of misusing R^2 (Kennedy, 2008: 26-28).

4.4 F-test

For overall goodness-of-fit of a model, conduct the F-test for $H_0 : \beta_i = 0$ (i=1…k). H_0 says that all parameters of regressors are not zero, whereas H_a is at leas one parameter is not zero. If k (the number of regressor excluding the intercept) is large, H_0 tends to be too strong!

$$F(K-1, N-K) = \frac{MSR}{MSE} = \frac{SSR/(K-1)}{SSE/(N-K)} = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}$$

4.5 Hypothesis Test for Individual Parameters (T-test)

For individual parameters, you may conduct the one-sample t-test because b is considered as weighted sum of Y.[4] The central limit theorem is applied to this sample mean (weighted sum) with its variance. $\hat{\beta}_k$ has mean β_k and variance $s^2 v_{kk}$, where $s^2 = \hat{\sigma}^2 = MSE$, v_{kk} =kth diagonal element of $(X'X)^{-1}$. See Kennedy (2008: 33) for the structure of $(X'X)^{-1}$.

The H_0 of this t-test is that a particular parameter, say β_k, is zero: $H_0 : \beta_k = 0$ and $H_a : \beta_k \neq 0$.

$$t[N-K] = \frac{\hat{\beta}_k - 0}{\sqrt{Var(\hat{\beta}_k)}}$$

The confidence interval of a parameter is calculated as,

$$\frac{b_k - \beta_k}{\sqrt{s^2 v_{kk}}} \sim t[N-K],\ b_k - t_{critical}\sqrt{s^2 v_{kk}} \leq \beta_k \leq b_k + t_{critical}\sqrt{s^2 v_{kk}}$$

Wald test for general linear restrictions has a hull hypothesis of H_0: $R\beta - q = 0$.

$$F[J, N-K] = \frac{(Rb-q)'\left\{R[s^2(X'X)^{-1}]R'\right\}^{-1}(Rb--q)}{J},$$

where R is J (the number of restrictions) by K matrix, b is the K by 1 vector of parameter estimates, and q is J by 1 matrix for hypothesized values.

For example, $H_0 : \beta_3 = 0$ needs R= [0 0 1 … 0] and q =[0]; $H_0 : \beta_1 + \beta_3 = 2$ needs R= [1 0 1 … 0] and q=[1]; Finally, $H_0 : \beta_3 = 0$ and $\beta_2 + \beta_4 = 2$ needs,

$$R = \begin{bmatrix} 0 & 0 & 1 & ... & ... & 0 \\ 0 & 1 & 0 & 1 & ... & 0 \end{bmatrix} \text{ and } q = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$t^2[N-K] = \frac{(\hat{q}-q)^2}{Var(\hat{q}-q)} = F[1, N-K]$$ for a single restriction of $R\beta=q$. (e.g., $\beta_k = 0$)

$$t[N-K] = \frac{b_i + b_j - k}{\sqrt{Var(b_i) + Var(b_j) + 2Cov(b_i, b_j)}}$$ for a single restriction $\beta_i + \beta_j = k$

In page 19 of Park (2011), for example, 4435.4921=.8827385^2/.00017568 (for output) sqrt(4.80)=-2.1913605=(.8827385-1.62751)/sqrt(.00017568+.11923344+2×(-.00194971))

5. Basics of Statistical Inference

[4] The central limit theorem or normal approximation theorem applies to a weighted sum of random variable Y, parameter estimator of β (Wonnacott and Wonnacott, 1981: 34).

This section discusses basics of statistical inferences for ordinary least squares.

5.1 Parameters versus Statistics

Parameters describe the population, whereas statistics depict samples drawn from the population. Usually there are many statistics that correspond to a parameter. In other word, the mean of a population is unique, while each sample drawn from the population may have a different sample mean. Parameters are what we want to know, whereas statistics are what we already knew from the samples. In general, parameters are denoted by Greek letters, while statistics by alphabet (Table 5.1).

Table 5.1 Comparison of Parameters and Statistics

	Parameter (Population)	Statistic (Sample)
Mean	μ	$\hat{\mu}$ or $\bar{x}$
Variance	σ^2	$\hat{\sigma}^2$ or s^2
Standard deviation	σ	$\hat{\sigma}$ or s
Regression coefficient	$\alpha, \beta, \delta, \gamma$	$\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\gamma}$ or a, b, c, d
Correlation coefficient	ρ	r

5.2 What Is a Valid Hypothesis?

A *hypothesis* is a specific conjecture (statement) about a characteristic of a population. This conjecture may or may not be true; we may not know exactly whether it is true or false forever.

a. *A hypothesis should be specific* enough to be falsifiable; otherwise, the hypothesis cannot be tested successfully. $H_0 : \mu = 0$ (for two-tail test); $H_0 : \mu \geq 0$ or $H_0 : \mu \leq 0$ (for one-tail test) A bad example is “The war on Iraq may or may not produce a positive impact on the U.S. civil society.”
b. A *hypothesis is a conjecture about a population (parameter), not about a sample (statistic).* $H_0 : \mu = 0$ is valid, whereas $H_0 : \bar{x} = 0$ is not. The statistic $\bar{x}$, computed from the sample, is already known!
c. A *valid hypothesis should not be based on the sample to be used to test the hypothesis*. “Hmm… my sample mean is 5, so my null hypothesis is the population mean is 5.” What??? This is a tautology!
d. *An appropriate hypothesis needs to be interesting and challenging (informative)*. A bad example is “Average personal income is zero.”
e. A hypothesis may not be accepted, but be rejected or not rejected (although some scholars use “accept H_0”). We do not know true parameters that we wish to know by drawing samples.

5.3 Null Hypothesis versus Alternative Hypothesis

The *null hypothesis*, denoted H_0, is a specific baseline statement to be tested and often (not always) takes such forms as “no effect” or “no difference.” Why? Simply because it is easy to compute the statistic and interpret the output. The *alternative hypothesis* (or research hypothesis), denoted H_a, is the denial of the null hypothesis and tends to state “significant effect” or “significant difference.” A hypothesis is either *two-tailed* (e.g., $H_0 : \mu = 0$) or *one-*

tailed (e.g., $H_0 : \mu \geq 0$ or $H_0 : \mu \leq 0$).

5.4 Significance Level

a. Type I error is rejecting the null hypothesis (or maintained hypothesis) when it is true. The probability of type I error, conventionally denoted alpha, is the **significance level** or the **size of the test**. The confidence level is expressed as 1-alpha.
b. The significance level, conventionally .05 or .01, plays a role of a criterion to reject or not to reject the null hypothesis on the basis of calculated statistics.
c. Type II error is failing to reject the null hypothesis when it is false. The probability of type II error is called beta. The **power of test**, defined as 1-beta, is the probability that it will correctly lead to rejection of a false null hypothesis.

	Do not reject H_0	Reject H_0
H_0 is true	Correct Decision 1-α : confidence level	Type I Error α : Significance level or the size of the test
H_0 is false	Type II Error β	Correct Decision 1-β : Power of the test

5.5 Procedures of Hypothesis Testing

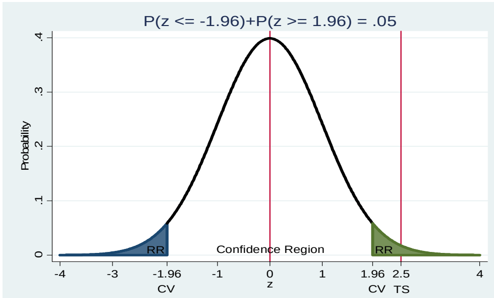
In general, there are five steps for a hypothesis test.

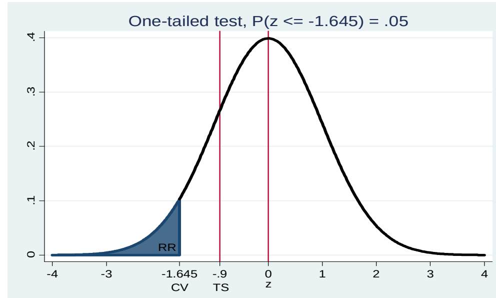
1) State a null and alternative *hypothesis* (one-tailed or two-tailed test)
2) Determine a *test size* (*significance level*) and find the critical value and/or rejection region. You need to pay attention to whether a test is one-tailed or two-tailed to get the right critical value and rejection region.
3) Compute a *test statistic* (test value) and/or its *p-value* or construct the confidence interval. Collect all necessary information (e.g., N, μ, σ, and s) and compute statistics using formula dictated by the model (e.g., z-test and t-test).
4) Reject or do not reject the null hypothesis by comparing the subjective criterion in 2) and the objective test statistic calculated in 3)
5) Draw a conclusion and interpret substantively.

5.6 Subjective Criteria: Critical Value and Rejection Region

Once a test size, a subjective criterion, is determined (the .10, .05, and .01 levels are conventionally used), the critical value (CV) needs to be found. A critical value is a cut point that makes the probability of occurring from the value to the (positive and/or negative) infinity the significance level. In order word, the probability that a random variable x is greater (or smaller) than or equal to the critical value is the significance level or the size of a test. In a two-tailed test, $P(-\infty \leq z \leq -z_{\alpha/2})$ + $P(z_{\alpha/2} \leq z \leq \infty)$=α, where $z_{\alpha/2}$ is the critical value for α significance level. The critical values of z at the .05 and .01 level are 1.96 and 2.58 in a two-tailed test, respectively (left figure below).

The *rejection region (or critical region)* is the area encompassed by the critical value, the (positive and/or negative) infinity, and the probability distribution function curve, and x axis (see the shaded areas above). Why is it called the rejection region (RR)? We reject the null hypothesis if a test statistic falls into this region. This area represents the extent that you are willing to take a risk of rejecting the null hypothesis. This willingness is nothing but the significance level α discussed above.





The significance level determines the critical value and rejection region (critical region) of a hypothesis test. The critical value distinguishes between the rejection region and the confidence region or "probable range." The rejection region visually represents the significance level (test size) on the probability distribution (e.g., the standard normal, *t*, *F*, and Chi-squared distributions). Therefore, one implies the other.

5.7 Objective Criteria: Test Statistic and Its P-value

A test statistic (or test value) is computed from the sample to evaluation the null hypothesis. The information from sample is summarized in this test statistic. *F*, *z*, and *t* values are common examples. For example, ±2.5 and -.9 in the above figures are test statistics.

A *p-value* is the probability that you get even more unlike sample than test statistic. Technically speaking, the p-value is the area underneath the probability distribution curve from the test statistic to the (positive or negative) infinity. For instance, the p-value in the left figure above is the sum of shaded areas from $-\infty$ to -2.5 and from 2.5 to ∞. In the right figure, the p-value is the area from $-\infty$ to -.9 since this is the one-tailed test. A convenient interpretation is that a p-value is amount of risk the researchers should take when rejecting the null hypothesis. A small p-value indicates that it is less risky than expectation (a significance level) to reject to the null hypothesis.

Like the critical value and rejection region, the test statistic and its p-value are both sides of a coin. They both depend on the model and probability distribution used. For example, if you get a *t* value, you need to take a look at the *t* distribution table to get the p-value.

6. Three Approaches for Hypothesis Test

There are three different approaches for hypothesis testing that approaches conduct exactly the same hypothesis test and thus give you the same conclusion. Each approach has both advantages and disadvantages.

6.1 Classical Approach: Test Statistic versus Critical Value

The classical approach asks, "*Is the sample mean one that would likely to occur if the null hypothesis is true?*" This test statistic approach examines how far the sample statistic is away from the hypothesized value (i.e., population mean). This is *point estimation*. If the sample statistic is far away from the hypothesized value, we may suspect that the hypothesized value is not true. If the observance of the sample statistic is likely (or probable), do not reject the null hypothesis; if unlikely, reject the null hypothesis (our conjecture may be wrong).

$$P(\mu_{null} - t_{\alpha/2} \times s_{\bar{x}} < \bar{x} < \mu_{null} + t_{\alpha/2} \times s_{\bar{x}}) = 1 - \alpha$$

The "likely" and "unlikely" mean whether the sample mean is far away from the hypothesized mean. If a (objective) test statistic is smaller than the (subjective) critical value or if the test statistic is not farther away from the hypothesized mean than the critical value, it is more likely to get such a sample mean than you thought if the hypothesized value hypothesis is true. Since you have a probable sample mean, you may not change your conjecture (the null hypothesis); your conjecture appears correct. In the right figure above, the test statistic -.9 is closer to the 0 (mean) than the critical value -1.645. Thus, the sample mean is not unlikely at the .05 significance level.

If a test statistic is larger than the critical value (opposite in the left-hand side), it is less likely than you thought to get such a sample mean. The sample mean is farther away than you thought from the hypothesized mean (the null hypothesis). Therefore, you do have some reason to reject the null hypothesis. Your conjecture may be wrong. In the left figure above, the test statistic -2.5 and 2.5 are farther away from the mean than their critical value counterparts (-1.96 and 1.96). Therefore, the sample mean you got is not likely at the .05 significance level. So you need to reject your conjecture in the null hypothesis.

6.2 P-value Approach: P-value versus the Significance Level α

This approach asks, "*What is the probability that we get a more unlikely (extreme or odd) test statistic if the null hypothesis is true?*" What does "a p-value is smaller than the significance level" mean? If a p-value is smaller than the significance level or falls into the rejection region, you have a smaller risk of being wrong than your subjective criterion (significance level). Accordingly, you become more confident to reject the null hypothesis in favor of the alternative (research) hypothesis because it is less risky than you thought to reject the null hypothesis. Keep in mind that the .05 significance level, for example, means you are willing to take a risk of making a wrong conclusion by that amount. When you reject the null hypothesis, there is still 5 percent chance that your conclusion is wrong (mistakenly rejecting the true null hypothesis).

Put it differently, a small p-value, say .003, indicates that your sample mean is vary far away from the hypothesized mean. You just have only .3 percent chance to observe sample means less than the one you get if the hypothesized mean is true. Obviously, this is an extremely unlikely event. Therefore, it is less risky than you thought to reject the null hypothesis or conclude that the hypothesized mean is wrong.

In the left figure above, the p-value for the test statistic 2.5 is .0124, which is much smaller than the .05 significance level. Therefore you reject the null hypothesis. In the right figure, the p-value for the one-tailed test is .1841, which is greater than .05; do not reject the null hypothesis.

6.3. Confidence Interval Approach: Hypothesized Value versus the Confidence Interval

This approach asks, "*Is the hypothesized value in the null hypothesis the value (parameter) that we would estimate?*" You need to construct the (1- α) percent confidence interval using the critical value and sample statistics (i.e., standard error). This approach uses *interval estimation*. The null hypothesis expects that the hypothesized value exists somewhere in the confidence interval. Unlikely the classical approach, this modern approach focuses on the location of the hypothesized value μ_{null}.

$$P(\bar{x} - t_{\alpha/2} \times s_{\bar{x}} < \mu_{null} < \bar{x} + t_{\alpha/2} \times s_{\bar{x}}) = 1 - \alpha$$

The (1- α) percent confidence interval means that you are (1- α) percent sure that the population mean (not sample mean) exists in the (1- α) percent confidence interval. If the hypothesized value falls into the confidence interval, you may not reject the null hypothesis and conclude that "Yes, my conjecture is right."

In contrast, the presence of the hypothesized value out of the confidence interval indicates you got an odd observation (sample mean) that went beyond your prediction (confidence interval). You may not be confident that your hypothesized value is likely.
Therefore, there must be something wrong in your conjecture (the null hypothesis). In other word, the true parameter does not appear to be the hypothesized value.

Table 6.1. Three Approaches of Hypothesis Testing

Step	Test Statistic Approach	P-Value Approach	Confidence Interval Approach
1	State H_0 and H_a	State H_0 and H_a	State H_0 and H_a
2	Determine test size α and find the critical value	Determine test size α	Determine test size α or 1- α, and a hypothesized value
3	Compute a test statistic	Compute a test statistic and its p-value	Construct the (1- α)100% confidence interval
4	Reject H_0 if TS > CV	Reject H_0 if p-value < α	Reject H_0 if a hypothesized value does not exist in CI
5	Substantive interpretation	Substantive interpretation	Substantive interpretation

* TS (test statistic), CV (critical value), and CI (confidence interval)

6.4 An Example of Hypothesis Test

This example examines if the population mean is 24,000, given a sample mean of 23,450, s=400, μ=24,000, and N=10.

6.4.1 Test statistic approach

i. H_0: μ=24,000, H_a: μ≠24,000. This is a two-tailed test.

ii. α=.05 (rejection region), the critical values are ±2.262, and df is 9=10-1.

iii. $t_{\bar{x}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{23450 - 24000}{400/\sqrt{10}} = -4.35 \sim t(9)$

iv. *Since TS (|-4.35|) ≥ CV (2.262),* reject the null hypothesis at the .05 level. The test statistic is farther away from the mean 0 than the critical value. It is not likely to get such a sample mean if the null hypothesis is true.

v. The average starting salary for nurses is not $24,000. See the left plot below.

6.4.2 P-value approach

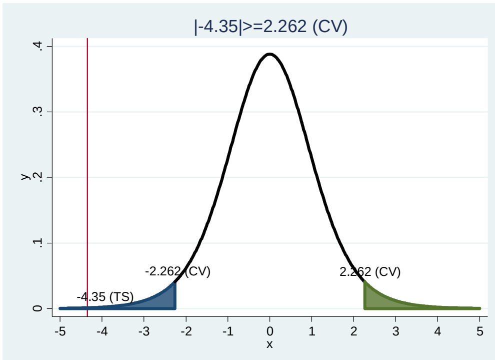
i. H_0: μ=24,000, H_a: μ≠24,000. This is a two-tailed test.

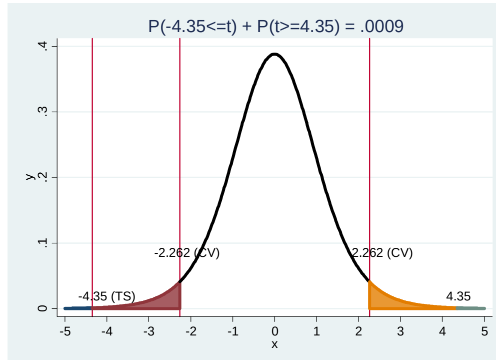
ii. *α=.05 (rejection region)*

iii. $t_{\bar{x}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{23450 - 24000}{400/\sqrt{10}} = -4.35 \sim t(9)$, *p-value* $P(t_{\bar{x}} \le -4.35) = .0009$

iv. *Since p-value (.0009) < α (.05),* reject the null hypothesis at the .05 level. It is extremely less risky to reject the null hypothesis in favor of the alternative hypothesis. You just take only .09 percent of risk when you reject the null hypothesis.

v. The average starting salary for nurses is not \$24,000. See the right plot below.





6.4.3 Confidence interval approach

i. H_0: μ=24,000, H_a: μ≠24,000. This is a two-tailed test.

ii. *α=.05, 95% confidence interval, critical values of ±2.262 with 9 degrees of freedom.*

iii. $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 23450 \pm 2.262 \times \frac{400}{\sqrt{10}} = 23450 \pm 286$, [23,164, 23,736]

iv. *The hypothesized mean 24,000 exists out of the confidence interval between 23,164 and 23,736.* Therefore, we reject the null hypothesis at the .05 level.

v. The average starting salary for nurses is not \$24,000.

7. Asymptotically Equivalent Tests

LR, Wald, and LM tests are asymptotically equivalent in that their statistics follow the chi-square distribution with the degrees of freedom equal to the number of restrictions imposed by the null hypothesis. It implies that in a finite sample three tests may produce substantially different statistics, leading to different conclusions. The choice of the test depends on the purpose of test and computational convenience.

Test	Likelihood Ratio	Wald Test	Lagrange Multiplier
Estimator	Both restricted and unrestricted estimators	Unrestricted estimator $\hat{L}_U$	Restricted estimator $\hat{L}_R$
Distribution	Assumption is needed	Asymptotic Normal, but not heavily depending on distributional assumption	Asymptotic Normal
Hypothesis	If H_0 is true	If H_0 is false	If H_0 is true
Variance	Log Likelihood function	Variance of estimator	Variance of estimator
Constraint		Linear constraint	Nonlinear constraint

7.1 Likelihood Ratio (LR) Test

a. The likelihood ratio test requires both unrestricted and restricted parameter vectors.
b. If the restriction $c(\theta)=0$ is valid, then imposing it should not lead to a large reduction in the log-likelihood function.
c. Under regularity, the large sample distribution of -2lnλ is chi-squared with degrees of freedom equal to the number of restrictions imposed.

d. $-2\ln\left(\frac{\hat{L}_R}{\hat{L}_U}\right) = -2(\ln \hat{L}_R - \ln \hat{L}_U) \sim \chi^2(J)$, where J is the number of restrictions. $\hat{L}_U$ and $\hat{L}_R$ respectively represent the likelihood functions evaluated at the unrestricted and restricted estimators. Restricted estimators cannot be larger than unrestricted estimators since the former is never superior to the latter.

7.2 Wald Test

a. Wald test measures the extent to which the unrestricted estimates fail to satisfy the hypothesized restrictions (Greene, 2011: 569).
b. Wald test only requires computation of the unrestricted model. But it requires computation of covariance matrix appearing in the preceding quadratic form.
c. If the restriction is valid, then $c(\hat{\theta})$ should be close to zero, since the MLE is consistent.
d. Under the null hypothesis of $c(\theta) = q$ in large samples, W has a chi-squared distribution with degrees of freedom equal to the number of restrictions.
e. $W = [c(\hat{\theta}) - q]'(Asy.Var[c(\hat{\theta}) - q])^{-1}[c(\hat{\theta}) - q]$
f. This test is not invariant to the formulation of the restrictions (problematic in nonlinear restriction). Unlike likelihood ratio and Lagrange multiplier tests, Wald test does not rely on a strong distributional assumption.

7.3 Lagrange Multiplier (LM) Test

a. LM or efficient score test is based on the restricted model. If the restriction is valid, the restricted estimator should be near the point that maximizes the log-likelihood.
b. The derivatives of the log-likelihood evaluated at the restricted parameter vector will be approximately zero. The vector of first derivatives of the log-likelihood is the vector of efficient scores. The variance of the first derivative vector is the information matrix.
c. LM test is based on the slope of the log-likelihood at the point where the function is maximized subject to the restriction. Therefore, the slope of the log-likelihood function should be near zero at the restricted estimator, if the restrictions are valid.

Table 1. Assumptions of Ordinary Least Squares

Assumption	Violation/Problem	Impact/Result	Detection	Solution
Linearity	Nonlinear Wrong regressors Changing parameters	Specification error		Transformation (Box-Cox) Change specification
Exogeneity $E(\varepsilon_i\|X)=0$	Limited Dep. Variable	Biased intercept		Logarithmic transformation
Nonspurious Disturbance	Heteroscedasticity	Unbiased estimates Inflated variance	White test Goldfeld-Quandt test Breusch-Pagan test	GLS, FGLS
	Autocorrelation	Biased estimates Incorrect variance	Durbin Watson test Box-Ljung Q test Cochrane-Orcutt	GLS, FGLS IV
Fixed X	Stochastic (Measurement errors in independent variables)		Autoregression, SEM	IV, ILS, SEM
	Reverse regression	Same F, R^2, and t Different SSE and slops		The product of slops equals R^2
	Autoregression with a lagged DV as an IV			
	Simultaneous equation			

	estimation			
Full Ranked	Multicollnearity	Inflated variance of estimates No change in BLUE and R^2	VIF = (1/Tolerance) Eigenvalue Condition number	Drop variables Obtain more observations
Omitting relevant X_i	Multicollinearity	Biased estimates s^2 with smaller variance		Specification
Including irrelevant X_i	Multicollinearity	Unbiased estimates s^2 with larger variance		Specification