

Economics 346
Course Notes

I. What is econometrics? Economic Measurement => numbers to bridge real world/theory gap. Review of Statistics and Opening Course info (cards: name, year, major, prev. courses in math, econ, stat, business, other relevant courses, computer experience, goals for this course) Course misconceptions: Is statistics, must use #s, equations, but will spend time on interpretation rather than proofs, charts to understand, applications. Syllabus stuff. Excel class sessions. Statistics review, but you need to have had statistics for this class. (G&S) Problem sets in Excel or other (not all computer problem sets). Group work ok but must write own answer. Midterm+final exam: drop midterm if final better. Attendance. Phone calls/office hours.

Reference: *Inside Statistics: Against All Odds*, VHS tapes (26 episodes, 2/tape @30 min. each) in Art and Arch. Libe, Douglas 3rd floor.

- A. Why econometrics? How do we know what we know? Take "demand slopes down" relationship. How do we know this is really true? Theory plus tests. Econometrics does tests on data.
1. Statistical application for more than 1 effect. My anecdote vs your anecdote: how do you tell what is really going on? What people say they do is not necessarily how they really act. Nobel Prize winner George Stigler, *Theory of Price*, p. 8: "Our main elements of analysis are people, and people who are influenced by the practices and policies we analyze. Imagine the problems of a chemist if he had to deal with molecules of oxygen, each of which was somewhat interested in whether it was joined in chemical bond to hydrogen. Some would hurry him along; other would cry shrilly for a federal program to drill wells for water instead; and several would blandly assure him that they were molecules of argon. "
 2. Incomplete info, so quantify uncertainty with statistics
 - a. Uncertainty versus randomness: What is randomness? "Random killings" = no predicting versus averages/similarities in large numbers
 - b. Random coin toss: equal chance of H or T, but unpredictable if next toss is H.
 3. Tool to understand world, forecast, find truth.
- B. What do you do with statistics and econometrics?
1. Describe the world: each variable (variables can be qualitative or quantitative; of quantitative, discrete, or continuous)
 - A. Data analysis (gather data, display data, summarize data: mean=central tendency, standard deviation=spread around, z-score to standardize = measure distance from mean per standard

deviation)

- i. Sample mean=average=(sum of data)/n versus median (equal #s of points above and below median value) and mode (most frequently)
- ii. Sample variance: $s^2 = 1/(n-1) \sum (X_i - \mu)^2$ and is measure of spread; Problem: what are units (square miles may make sense, but square dollars?) so use standard deviation = $s = \sqrt{\text{variance}}$

iii. Z-score = $(X_i - \mu)/s$ for each observation

Rule: about 68% of data within 1 standard deviation of mean (normal)

Rule: about 95% of data within 2 standard deviations of mean

Rule: about 99.7% of data within 3 standard deviations of mean

Examples of z-scores, using z tables:

Draw normal curve, shade between -1.2 and 1.2. What is probability? Go to z table, look up 1.2; =77%

Shade 0 to 1: probability = $1/2(68\%) = 34\%$

Shade 0 to 2: Is probability 2 times 0 to 1? No, normal curve is not rectangle; 0 to 2 is $1/2(95\%)=48\%$

Shade -2 to 1: Break down to -2 to 0 + 0 to 1 = $48 + 34 = 82\%$

Shade tails: < -1 and > 1 : Area is "not = 68" = $1 - .68 = 32\%$.

1 tail only, > 1 : $1/2(32) = 16\%$

Parade Magazine example of data description.

- C. Probability (laws of chance), match data generation to known processes: use to test hypotheses about economic theory, forecast, tell amount of certainty (confidence levels, inference, hypothesis testing)
 1. Econometrics uses statistics above and regression relationships among variables to: forecast future economic activity (given past relationships, confidence intervals around forecasts, basic relationships not changing, have important forces covered ==> need theory too), test economic theory
- D. Data Description: 2 or more variables. Relationships between variables.

1. Describe each variable independently
2. Correlation
 - A. Correlation versus causality. Econometrics alone cannot prove causality, assumed in application. Covariance not enough because: depends on units in which x and y happen to be measured. So standardize:
 - B. $r = (\text{Cov}(x,y)/\{\sqrt{\text{Var}(x)\text{Var}(y)}\})$
 - C. Partial correlation coefficient: $r_{xy.z} = (r_{xy} - r_{xz}r_{yz})/(\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)})$ which is the partial correlation for x and y, holding the effect of z constant, and r_{xy} , etc, are the ordinary correlation coefficients for x and y, etc.
 - D. If strong association between 2 variables, then knowing one helps a lot in predicting the other. But when there is a weak association, information about one variable does not help much in guessing the other.
3. Another way to look at correlation coefficient, r.
 - A. Scatter diagram graphing points: GRAPH YOUR DATA; football-shaped cloud of points. How to summarize in a number?
 1. Point of averages (mean of x and mean of y)
 2. But how to do spread: vertically (2 SD's of Y) or horizontally (2 SD's of X)? Could use either or both
 3. Still need strength of association between 2 variables: amount of clustering around a line (Draw correlation near 1, correlation near zero)
 4. So relationship (linear) between 2 variables summarized by: avg of x and SD of x, avg of Y and SD of Y, and correlation coefficient r.
 5. Draw : X, Y axis. How choose: Y is "given" X, so use knowledge; Can be + - 0: graph (-2,1) and (3,2)
 6. Slope = rise/run; intercept
 - C. How to measure spread? Side to side (X standard deviation) or up

and down (Y standard deviation) (Draw \bar{y} , \bar{x} point, and ± 2 SD's for X and Y)

- D. SD line = line that scatter diagram points cluster most closely around. From point of averages, go 1SD(X) across and 1 SD of Y up to get 2nd point. This is enough to draw a line. Slope of this line is **correlation coefficient, r**.
- E. $-1 < r < 1$
- F. Correlation is linear association only, is not the same as causation, and is symmetric X:Y and Y:X
- G. Note, r is pure number, no units; not affected by interchanging 2 variables; adding same number to all values of 1 variable(1,2,2,3 vs 1 1 2 2) vs (2 3 3 4 vs 5 5 6 6); or multiplying all values of 1 variable by same positive number (20 30 30 40 vs 500 500 600 600) all have same correlation (also if leave off labels but draw same points) : all have r about .7
- H. Problems for correlation coefficient:
1. outliers (draw)
 2. nonlinear associations (draw)
 3. Correlations based on rates or averages are usually too big, overstating association. So watch out! Reason: much spread around averages in each series, so if use correlation for averages to estimate correlations for individuals, you eliminate the spread and give a misleading impression of tight clustering. (Draw income:Y and education (X) ; state A has low X,low Y , State B has middle, C high. Scatter football-shaped. Averages almost straight line.
- E. Describe 2 or more variables: Regression. Measures how one variable depends on the other, but is not symmetrical
1. Draw football cloud, correlation Regression line is average Y given X. So increase X 1 SD, say from mean. Then Y increases only r SDs. Why? Take if $r = 0$. Then no association between X and Y. So 1 SD increase in X associated with 0 SD increase in Y, on average. If $r=1$, all points lie on line, SD line, so 1 SD up in X is 1 SD up in Y. Look at $r=-1$. Same story, but line slopes down. In between, more complicated math, but use r as the

factor. This is called the **regression method**. If you graph the average Y's for all the X's, and it makes a straight line, then this will be the regression line. Otherwise, the regression line is a smoothed version of the graph of averages.

2. **Regression fallacy and regression effect:** In re-test situations, bottom group on first test will, on average, show some improvement on second test, and top group will on average fall back. This is *regression effect*. *Regression fallacy* is thinking change is due to something important, and not just the spread around the line of averages.

Example: repeated IQ test: 2 scores differ a little due to chance variability, lucky or unlucky on each test. If first test score very high, that suggests lucky on first test, and second score likely to be lower. (Wouldn't say high score, too bad about the luck!) And vice versa. If error + or - is equally likely, about 4 points on average of 100 IQ with SD of 15, then true 135 equally likely to be 130 and 140 on test. These above mean on frequency diagram=> if take people who scored 140, they could be true below 140 with >0 error or truly above 140 with <0 error. More likely (bell curve) to be positive error.

3. 2 possible regression lines given 2 series: predict Y given X and predict X given Y. They are not the same, and not invertible.

Regression doesn't work for non-linear (parabola, etc.) Linear in coefficients (or parameters): $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$ where Y_i, X_i , can be squared, logged, combination of 2 other variables, etc. but not b's.

F. **Single-Equation Linear Models, Vocabulary (S, Chapter 1)**

Simplest model is $Y = B_0 + B_1X$ (This is ideal model to be approximated by a cloud of data)

Y is **dependent variable**, to be "explained"

X is **independent variable** (later more than 1 X)

Bs are **coefficients** or **parameters**, B_0 is the **constant** or **intercept** term – equation of line – B_0 is value of Y when X is 0. B_1 is **slope** coefficient, showing how much Y changes (in Y units) when X increases 1 unit (in X units)

Linear in the variables means if plot in XY space, get a straight line. Example: $Y = a + bX$ for any a and b. If Y or X are functions like logs, or raised to powers (like X-squared) then the resulting equation will not be a straight line in XY space.

Linear in the parameters (or coefficients) means that the coefficients are in the simplest form (numbers or constants like a and b) and not functions (logs

or combinations of parameters) or powers.

Regressions must be linear in the coefficients or parameters, but need not be linear in the variables.

Model is ideal, may not be observable. If Y is measured with error, then even if exact relationship between X and Y, will get variations. Like weights and measures stones +/- some tolerance. Variation from other sources – difference between Y predicted by equation for a given X_i , (Y_i^{\wedge}) and Y observed (Y_i). The difference, $Y_i - Y_i^{\wedge} = e_i$, the **error** or **residual**. If error is just from measurement variations, expect that central tendency of e_i is 0, and evenly spread above and below.

So $E(Y|X) = B_0 + B_1X$. This is conditional, like difference in correlations SD line and regression line as smoothed connection of averages for limited categories of X.

And expect e_i to be unexplainable by other things in model, so e_i is **stochastic** or **random**.

i. Capture (account for) effects not explained by independent variables such as: omitted (left-out) variables; measurement errors in data (of Y); true relationship has different functional form (shape) than regression; random (unpredictable) events

4. "Truth" can't be observed, but regression approximates by using sample of actual X's and Y's. So use statistical theory focussed on estimated equation, coefficients to organize, understand, to match data to theory.

Now need to refer to observations (annual data, differences in a specific year versus overall) and multiple independent variables (more than 1 X).

$Y_1 = B_0 + B_1X_1 + e_1$ and $Y_2 = B_0 + B_1X_2 + e_2$ up to $Y_n = B_0 + B_1X_n + e_n$; the regression has the same coefficients for every observation; or the regression "holds" for every observation.

If more than 1 X, (**multivariate**) then $Y_1 = B_0 + B_1X_{11} + B_2X_{21} + e_1$ and $Y_2 = B_0 + B_1X_{12} + B_2X_{22} + e_2$ up to $Y_n = B_0 + B_1X_{1n} + B_2X_{2n} + e_n$. B_s are direct effect on Y from a change in one X, **holding other things constant**, including the values of other X variables.

Estimated equation versus model (residual= observed Y minus estimated Y^{\wedge} versus error term = observed Y minus true expected value of Y). (Draw sample in corner of XY "full population" cloud, then get true, estimated lines and different errors.)

II. Ordinary Least Squares (S, chapter 2)

A. Choose beta hats that minimize the summed squared residuals for a sample.

1. Why minimize squared residuals? Easy, Theoretically sound, and estimates have useful properties

2. Easy enough to do by hand, but tedious
3. Squaring versus absolute value of difference from line (larger value to big residuals)
4. Properties of OLS estimators: Goes through means of X and Y (called **centroid**); Sum of residuals is exactly zero; BLUE
5. To get **estimators** (= formula to apply to data to get estimates of the model coefficients), minimize squared residuals = sum of e_i -squared = sum of $(Y_i - B_0 - B_1 X_i)$ -squared. Take derivative with respect to B_0 and B_1 , set derivatives = 0, then get 2 equations. Solve these for B_0 and B_1 (eliminate 1 and substitute) and you get the **normal equations**, or the formula for the coefficients:

$$A. \quad B_1^{\wedge} = \frac{\sum(X_i - \underline{X})(Y_i - \underline{Y})}{\sum(X_i - \underline{X})^2}$$

$$B. \quad B_0^{\wedge} = \underline{Y} - B_1^{\wedge} \underline{X}$$

6. **Standard error of the Estimate (SEE)** is a measure of the fit of the overall equation

$$SEE = \sqrt{\sum(e_i\text{-squared}) / (n-2)}$$

7. **Standard error of the estimated coefficients:** $SE(B_1) = SEE / \sqrt{\sum(X_i - \underline{X})^2}$
8. $SE(B_0)$ also related to SEE = $SEE \sqrt{[\sum X_i^2 / N \sum (X_i - \underline{X})^2]}$
9. $Cov(B_0, B_1) = (-\underline{X} SEE\text{-squared}) / \sum(X_i - \underline{X})^2$

B. Coefficient of determination, R squared, = percentage of the variation of Y around its mean = explained sum of squares ($ESS = \sum(Y_i^{\wedge} - \bar{Y})^2$) divided by total sum of squares ($TSS = \sum(Y_i - \bar{Y})^2$, squared and since $Y_i = Y_i^{\wedge} + e_i$, can be squared Y deviation from mean + squared residuals if 2 times product of Y_i^{\wedge} deviations from \bar{Y} and residuals is zero. Since these are uncorrelated, by assumption, this is so)

1. measures degree of statistical fit
2. Always increases (or at least, never decreases) when add new variable

3. $TSS = ESS + \text{var}(e)$; $\text{var}(e) = \text{sum of squared residuals} = \text{RSS}$
 4. R-squared is $ESS/TSS = \text{percent of changes in } Y \text{ explained by regression line}$
 5. R-squared also $= 1 - \text{RSS}/TSS$
 6. $0 \leq \text{R-square} \leq 1$
 7. Simple correlation coefficient, r is square root of R-squared, with the sign from the direction (sign) of B_1 for a simple regression. This does not hold for a multivariate regression.
- C. Examples of regressions: fertilizer and radioactive waste:
1. Fertilizer: $\hat{Y} = 58.7 + 7.0X$, Graph, predict for $X=3$ (80) and $X=4$ (87)
 2. B_1 is incremental yield for every additional pound of fertilizer
 3. Radioactive waste: regression line $\hat{Y} = 119 + 9X$
 4. Predict for $X=0$ (119) $X=5$ (164), graph
 5. Is radioactivity harmful? This is uncontrolled study, so not proof
- D. Multivariate Regression: B_1 is increase in Y from increase in X_1 , holding all other independent variables in the equation constant.
- E. R^2 is coefficient of determination adjusted for degrees of freedom, so it increases only if improvement in fit from adding new variable is bigger than loss of degree of freedom used up in estimating coefficient of new variable. So use this instead of plain R-squared.
- F. Fit is only part of the quality of a regression equation. Theory, expectations about relationships, etc at least as important as R-squared
- G. How to Use Regressions (S Chapter 3)
1. 6 steps to solve problem with regression.: 1) Review the literature and develop the theoretical model. 2) Specify the model: select the independent variables and functional form 3) Hypothesize about the expected signs of the coefficients 4) Collect the data 5) Estimate and evaluate the equation 6) Document the results
 2. Lagged variables: lag independent variables (dependent special, later) Then interpretation changes: B_2 measures increase in this year's Y from change

in last year's X_2 , holding constant other X variables (if X_2 is lagged variable)

3. Dummy variables: take 1 or zero. Qualitative effects, brand names, etc.
 - A. Intercept dummy: shifts line vertically if =1, X_1 is dummy or binary variable.
 Y = teacher salary, X_1 = MA or no, X_2 = # years experience

If MA, regression equation is: $Y = B_0 + B_1 + B_2X_2$

If no MA, regression equation is $Y = B_0 + B_2X_2$
 - B. Interpretation: If has Masters, adds B_2 \$ to salary
 - C. Can also have slope dummies
 - D. If 2 conditions, yes MA and no MA, then 1 dummy. If 5 brands, then 4 dummies. One fewer dummy than conditions.

III. Classical Model: NLRM (S Chapter 4)

- A. Assumptions
 1. Homogeneous variance of Y 's ($P(Y_1|X_1) = P(Y_2|X_2) = \dots$ and this is sigma-squared
 2. Linearity: $E(Y_i) = B_0 + B_1X_i$, [Note: $E(Y_i) = \mu_i$] Note that this is linear in the parameters, but not necessarily linear in the variables. This also means the model must be correctly specified to be linear (check data plot) and that the error term is additive (+ u_i , not, for example, multiplicative or rising with X .)
 3. Independence: the Y_i are statistically independent
 4. $E(u_i) = 0$ and the variance is constant: ($E(u_i\text{-squared}) = \text{sigma-squared}$. This means no heteroscedasticity.)
 5. The error terms are uncorrelated with each other (no serial correlation)
 6. The X_i are nonstochastic variables whose values are fixed in repeated sampling, no X is a perfect linear function of any other X 's (no multicollinearity) and the X 's are uncorrelated with the error term.
 7. These are the Gauss-Markov assumptions, and given these, the estimators are the best (most efficient) linear unbiased estimates of the coefficients, meaning they have the minimum variance of all linear unbiased estimates. OLS IS BLUE!

- B. Normal Distribution of error (Book's assumption 7). This is usual, but not absolutely necessary. Why make it?
1. Since error term can be thought of as adding all other influences, a composite of minor errors or influences. Data influenced by many small and unrelated random effects are approximately normally distributed. As the number of these rises, the distribution of the error term approaches the normal distribution (by the Central Limit Theorem).
 2. Need normality to use t-statistic and F-statistics.
 3. What is Central Limit Theorem? The mean (or sum) of a number of independent, identically distributed random variables will tend to be normally distributed, regardless of their distribution, if the number of different random variables is large enough.
 4. Better statement of CLT: Take random samples, size n from population with mean μ and standard deviation σ . Then, as n gets large, \bar{X} approaches the normal distribution with mean μ and standard deviation σ divided by the square root of n . Then

$$\Pr(a \leq X \leq b) \cong \Pr\left[\frac{(a-\mu)}{(\sigma/\sqrt{n})} \leq z \leq \frac{(b-\mu)}{(\sigma/\sqrt{n})}\right]$$
 5. Means whatever the shape of the original distribution, if you take averages, the distribution of the averages will tend to normal distribution.
 6. But problems: depends on large sample size and to use it you need σ . So what if n is small and σ unknown?
 - a. Take standard deviation of sample for σ .
 - b. If substitute s for σ in z-score, get sample mean- $\mu/(s \text{ times radical } n)$ which is distributed Student's T (from William Gosset, who used the pseudonym "student". T distribution is more spread out than normal (kurtosis) and amount of spread depends on n : as n rises, the t distribution tends to normal.
- C. Sampling distribution of B^\wedge
1. Unbiased (mean of distribution of samples of B^\wedge is true B , or population mean)
 2. B^\wedge is unbiased estimator if $E(B^\wedge)=B$. If not, is biased estimator. Tradeoff sometimes between bias and precision (lower variance of estimator may mean is biased)
- D. Gauss-Markov Theorem: Given classical assumptions (don't need normality of errors), the OLS estimator of B is the minimum variance estimator from the set of all linear unbiased estimators of B_k for $k=0,1,2,\dots,K$. So OLS is BLUE!
- E. Properties of OLS Estimators: 1) unbiased ($E(B^\wedge) = B$), 2) minimum variance,

3)consistent (as sample size \rightarrow infinity, variance gets smaller, and estimates converge on true parameters, 4) normally distributed (if normal distribution of error term is assumed), so can use statistical tests based on normal distribution. (see Chapter 5).

F. Greenspan and Fed Set Money Supply to get GDP target (example). Using Greenspa.xls data, which is dependent, independent variable in money supply:GDP relationship? Plot in scatter diagram, with Monetary base (billions of \$) on X and GDP (trillion \$) on Y. Is this a straight line?

1. What is regression line? $Y=1.90 + .012X$, $R\text{-square}=.978$. Interpret this regression. If $M=0$ is nonsense, so don't pay attention to intercept.
2. What level should Fed set M at to get GDP of \$9 trillion next year?
3. Try another estimator:
 - a. Break data into 2 groups according to size of M. Calculate means:
 $M1=$
 $M2=$
 - b. Estimator $C = (Y2-Y1)/(M2-M1)$

IV. Basic Statistics and Hypothesis Testing (Chapter 5)

A. Statistical Inference

1. Conclusions about the population drawn from our sample.
2. Form is probability: 95% probability that $B1^{\wedge}$ is between 1 and 5, for example.
3. Or check to see if some statement is likely right or likely wrong: Hypothesis, X determines Y. If $B1$ (slope) coefficient $=0$ in our sample, then this hypothesis is likely wrong. So we check how likely it is that the slope = 0. If the probability that the slope is 0 is .00000001, then it is likely that changes in X are at least a little associated with changes in Y.
4. Use hypothesis testing as process for inference

B. Hypothesis Testing

1. How to choose (specify) the hypothesis to be tested
 - a. Null ($H0$) versus alternative (HA). Null is usually the one the researcher does not believe, and wants to reject. The tests are set up to reject, or to fail to reject. Accept is not allowed, since failing to say something is false is not the same as saying something is true. Not guilty in a murder trial is not the same as innocent.
 - b. May do $H0, HA$ as range: If you are testing demand curves, $H0$ would be that $B1 \geq 0$, since price is believed to be negatively related

- to quantity demanded. H_A is the other side: that $B_1 < 0$. It doesn't matter if the $B_1 = 0$ piece is part of the null or part of the alternative
- c. 1-tailed (as above: that $B <$ or $>$ some number, which doesn't have to be 0) versus 2-tailed: $H_0: B = 0$, $H_A: B \neq 0$. Sometimes negative values, for instance, don't make sense. Then the 1-tailed test is $H_0: B = 0$, $H_A: B > 0$
 - d. Types of errors: Type I: reject the true null hypothesis (convict the innocent) and Type II: fail to reject a false null (let guilty go free). Tradeoff: lower I but raise probability of II, or vice versa.
 - e. Use decision rule to decide criterion for rejecting null. Choose decision rule based on tradeoff of types of errors.
2. What decision rule to use to accept or reject the hypothesis
- a. Divides all possible values of B into accept or reject regions. Regions are probability under frequency diagram for distribution of B . (What is distribution of B , shape of frequency/probability ?)
 - b. If B^\wedge in acceptance region, fail to reject the null. Common probabilities are 90%, 95%, 99% in acceptance region.
 - c. If B^\wedge in rejection region, reject the null. Probabilities in each tail depend on acceptance probability and whether 1 or 2-tailed tests. These probabilities are of Type I errors, that the null hypothesis is really true, but we are rejecting it. If α is 1-probability of acceptance region, then for 1-tailed test, rejection region is α . For 2-tailed test there are 2 regions, one on each end, so each has value $\alpha/2$.
 - d. Border values are called **critical values**. If $B^\wedge > .8$, then reject the null, for example.

C. T-test

1. Use this to test hypotheses about individual coefficients, 1 at a time. Need normally-distributed error term. Appropriate when standard deviation of error is not known, and must be estimated.
2. Like a Z-score in form. If B is the border value implied by the null hypothesis, then the t-statistic is

$$T = (B^\wedge - B) / SE(B^\wedge)$$
 Where B^\wedge is the regression coefficient estimated with OLS, and $SE(B^\wedge)$ is the estimated standard error of B^\wedge .
3. Note that if $B = 0$, the t-statistic is just the ratio of B^\wedge to its standard error. This is the t-statistic printed in standard computer output.
4. Still need to know probability. So look up in t tables (table B1 on back cover of book). T-distribution probabilities approach the normal distribution as degrees of freedom (sample size) rises. Degrees of freedom

are sample size minus number of coefficients you are estimating. For a simple regression (1 X variable, plus the intercept), the available degrees of freedom is $N-2$. So for t-statistic of 2.02, sample size of 30, the probability is 5% for a 2-tailed test and 2.5% for a 1-tailed test.

5. Level of significance (probability of tails, or rejection) = 1 - confidence level = probability of acceptance. (2 is rule of thumb for t-statistic, and approximately relates to 95% confidence level or 5% significance level. This is most common.
6. Confidence interval: $B^{\wedge} \pm t SE(B^{\wedge})$. To make a 95% confidence interval for B^{\wedge} , look up critical t value in table for degrees of freedom in sample. So if $B1^{\wedge} = 10$, $SE(B1^{\wedge}) = 2$, and you have 26 degrees of freedom (sample size was originally 28 for simple regression), then the t critical value (2-sided test) is 2.056. So the 95% confidence interval is $10 \pm 2(2.056) = 10 + 4.112$ to $10 - 4.112 = 5.9 < B1 < 14.1$
7. Q: is zero in this confidence interval? Since the answer is no, we can say that with at least 95% confidence, $B1^{\wedge}$ is not equal to zero. In other words, if a value is not in the range, then we can reject the null hypothesis that $B1 =$ that value, using the t-test.
8. T-test examples: Ann Arbor rents in student housing
9. Problems with the t-test
 - a. Doesn't test theoretical validity, only statistical validity.
 - b. Doesn't test importance, only precision.
 - c. Not appropriate if whole population (because SE of whole population is zero, and dividing by zero is undefined).

D. F-test

1. What are we testing: that $r\text{-square}=0$, or that $B1=B2=B3=\dots=Bk=0$. Joint test that all together (or subset) equal zero, versus individual tests for t-test.
2. $F = (ESS/K)/(RSS/(n-K-1)) = [(\text{sum of squared } Y^{\wedge} - \text{mean of } Y)/K]/[\text{sum of squared residuals}/(n-K-1)]$, where ESS is explained sum of squares and RSS is residual sum of squares, n is number of observations in sample, and K is number of independent variables.
3. Null equation is $Y_i = B_0 + e_i$, which says Y is explained by nothing but its mean plus a random error term.
4. Reject H_0 is $F \geq F_c$, and do not reject H_0 if $F < F_c$, where F_c is the critical F value, determined from Table B-2 or B-3, and using K as the degrees of freedom in the numerator and $n-K-1$ as the degrees of freedom in the denominator.

MIDTERM MIDTERM MIDTERM MIDTERM MIDTERM MIDTERM MIDTERM

A. Specification of a Regression Model: Choosing Independent Variables (Chapter 6)

1. What makes a good model?
 - a. Parsimony: models are always simplifications because reality is so complex that it is not practical as a model. So a good model selects a *few, key* variables and confines minor influences to the error term.
 - b. Identifiable: Given a set of data, parameters must have unique values; this means that only 1 estimate may exist for a given parameter.
 - c. Goodness of fit: high R-bar-squared is better than low. Note that this is not the most important criterion.
 - d. Theoretical consistency: If signs are wrong, even if R-bar-squared is high, model may be bad.
 - e. Predictions: within the sample, R-bar-squared is predictive power. But what about outside the sample, with new data? This is the hardest test for many models.

2. Underfitting the model: Omitting a relevant, important variable from a regression equation causes bias in the estimates of the remaining coefficients to the extent that the omitted variable is correlated with included variables.
 - a. If an important variable is omitted, the error term in your equation is really the true error plus the effect omitted: $u_i = e_i + B_k x_{ki}$, if B_k is the coefficient on x_k , the relevant but omitted variable for observation i .
3. If the left-out variable is correlated with another X variable, say X_1 , then the expected value of B_1^{\wedge} is not B_1 . The intercept is also biased. This bias does not disappear no matter how large the sample. The bias to be expected from leaving a variable out of an equation equals the coefficient of the excluded variable times a function of the simple correlation coefficient between the excluded variable and the included variable in question.
 - b. Even if the X terms are all uncorrelated with the left-out variable, the intercept is biased.
 - c. The variance of the errors is estimated incorrectly. This means that the variance of the B's is also estimated incorrectly.
 - d. So confidence intervals, t-tests, and F-tests can give misleading conclusions.

4. **Once a model is formed based on theory, dropping a variable from the model is not advised.**
5. Including a variable in an equation in which it is actually irrelevant does not cause bias, but it will usually increase the variances of the included variables' estimated

- coefficients, thus lowering their t-values and lowering R-bar-squared.
6. To decide whether to include a variable use these 4 steps:
 - a. Theory: Is there a sound reason for including the variable, and is this variable the best way to achieve that?
 - b. T-test on the variable's coefficient: does the coefficient have the sign predicted by theory, and is it significantly different from zero?
 1. Sometimes you test more than 1 variable at a time (example: if the question is, do brand names matter for demand for orange juice, then would want to test if dummy variables for several brands at once have coefficients different from zero.
 2. This is a **joint test** that, say, $B_3=B_4=B_5=0$ for H_0 , versus that they are not equal to zero, for H_a .
 3. A variation on F-tests is the correct statistic. Run 2 regressions, one including B_3, B_4, B_5 (and the other variables) and one, which we'll call the restricted equation, using subscript R, without B_3, B_4 , and B_5 . Calculate the residual sums of squares for both.
 4. $F = [(RSS_R - RSS)/M] / [RSS/(n-K-1)]$, where RSS_R is the sum of squared residuals from the restricted equation, RSS is the sum of squared residuals from the unrestricted equation (the one with all the variables included), M is the number of restrictions considered (3 in this example), and $n-K-1$ is the number of degrees of freedom in the residuals from the unrestricted equation. N is the sample size, K is the number of X variables, and 1 for the intercept.
 - c. R-bar-squared: Does adding the variable to the equation improve the overall fit of the equation?
 - d. Bias: Do coefficients on other variables change when the variable is added? (If coefficients change after adding the variable, then the equation was biased through omission of a relevant variable.)
 7. Theory, not statistical fit, should be the most important criterion for whether or not to include another variable in a regression equation. Stepwise regression gives biased results and test statistics have distributions different from the standard t-tables.
 8. If all 4 of the criteria in (6) are true, then the variable definitely belongs in the equation. If none of the criteria in (6) are true, then the variable definitely does not belong in the equation.
 9. If some of the criteria are conflicting, then use "judgment".
 10. "Data mining" is to simultaneously try a whole series of possible regression formulations and choose the one with the results that look best to you. You can get the results you want, but they are likely to be worthless.
 11. Scanning: analyze data to develop a testable theory. Not the same as testing a

hypothesis. Note that to test a theory developed by scanning, you need to find a separate set of data. This is especially hard for time series.

12. Sensitivity analysis: running alternate specifications to determine whether results of your regression depend on the particular specification you chose (making them statistical flukes). When interpretations or results are the same for different specifications, they are called "robust".

B. Functional Form (Chapter 7)

1. Constant: Include it but don't rely on it.
 - a. No constant can violate assumption that error terms have zero mean, so leave it in even if it is not significant.
 - b. Don't rely on constant for analysis or inference because:
 - I) contains mean effects of marginal variables not included in regression, so its function as a locator for the equation as a whole interferes with its analysis; and
 - II) the intercepts often lie outside the range of sample data, and very far from means, so estimates are less reliable.
2. Alternative functional forms
 - a. **linear**: used so far. Coefficient interpretation: Change in Y from a 1-unit change in X. Economists like elasticity (ratio of percent changes: % change in Y/% change in X), or $B_k X_k / Y$. This ranges from 0 (when $X_k=0$) to infinity (when $Y=0$)
 - b. **Double-log**: natural log of Y is dependent variable and natural logs of X's are independent variables. Now coefficients are changes in logs, which are % changes, so B_k is % change in Y given a 1% change in X, which is the elasticity. Means constant elasticity. Use for production functions, some cost functions, indifference curves.
 - c. Comes from an **exponential function**: $Y = e^{B_0} X_1^{B_1} X_2^{B_2} e^u$ where e is the natural log base = 2.71828.
 - d. Logarithms: A log is an exponent to which the base is taken to get a number. For example, in base 10, the log of 100 is 2.
 - i. Properties of logs: 1) Change in log (derivative of log) is percentage change. 2) To get log of a product, add logs together: $\text{Log}(XY) = \text{Log}(X) + \text{Log}(Y)$. 3) The log of a variable with an exponent is the exponent times the log of the variable: $\text{Log}(X^2) = 2\text{log}(X)$ 4) Logs of negative numbers and zero are undefined. This means if you have a dummy variable, do not log it.
 - ii. While logs could be any base, the most common are e, the natural log (ln) and base 10.
 - e. **Semi-log**: Some, but not all, variables are in logs. Use when increases in Y happen from a change in X at a decreasing rate. Example: consumption functions.

- f. **Polynomials** : At least 1 of independent variables raised to some power different from 1. Degree of polynomial is highest power. Second degree is quadratic, 3rd degree is cubic, etc. Captures nonlinearities when variable gets increasing importance: Time and Time-squared in growth, age and age-squared in earnings
 - g. **Inverse**: Y is function of reciprocal of at least one independent variable. Captures when y asymptotically approaches a value from above or below, so impact of change in X on Y approaches zero as X increases. Example: Phillips curve.
 - h. **Can't compare R-squared** values of models with different Y's. So you cannot compare directly the R-squared values of log models with linear models. (Different TSS).
3. Problems if functional form is wrong: bad predictions out of sample, large forecasting errors, and incorrect inferences about coefficients within the sample.
 4. Example with GNP data: demand for housing expenditures as function of price and income: what functional form?
 5. Dummy variables: intercept
 - a. Examples: seasonal effects, male/female, brands
 - b. Interpretations: if dummy is case, then Y changes by coefficient's value (in units of Y)
 - c. 1 less dummy than alternatives
 6. Dummy variables: slope, interaction terms, dummy dependent variables.
 - a. Interaction term is additional term in regression that is multiple of other terms. Example is slope dummy variable: allow for extra effect if condition of dummy is met. Example: consumption in wartime, production during strikes.
 - b. Include variable, intercept dummy, and interaction term. Otherwise you get bias in estimate of slope dummy. So equation is $Y=B_0+B_1X_1+B_2D+B_3X_1D+u$
 - c. Piecewise regression allows slopes to change if value of X variable changes, so is special case of interaction dummy. But dummy is 0 if $X < \text{some number}$, 1 is $X \geq \text{some number}$, for example. Dummies can be related to size as well as qualitative characteristics.
 - d. Piecewise regression: $Y=B_0+B_1X_1+B_2D+B_3X_1D+u$
 - e. Allows measures of kinks, such as kinked demand curve.

C. Multicollinearity

1. Perfect vs Imperfect multicollinearity: imperfect multicollinearity means high correlation between independent variables.

2. Consequences of multicollinearity
 - a. Perfect multicollinearity: can't run regression
 - b. Imperfect multicollinearity: high R-squared, low t-statistics, good forecasts
 - c. Multicollinearity means the standard error of the betas (coefficients) will be higher, which means the t-statistics will be lower.
 - d. F tests of the hypotheses will be valid.
 - e. The absolute value of the covariance of the coefficients, B_i and B_j , is very high if X_i and X_j are highly correlated. This means that it is difficult to interpret individual coefficients. The regression can't separate out the effects of only X_i or only X_j . So it divides them the best it can.
 - f. Multicollinearity may improve the forecasts of a model.
 - g. If you have multicollinearity, OLS estimators and SE's will be sensitive to small changes in the data
 - h. The regression coefficients will be sensitive to the exact specification
3. How to tell if you've got it
 - a. Check correlation coefficient (why not covariance?) of each independent variable with others. High values for correlation coefficient of independent variables with other independent variables means multicollinearity
 - b. Low correlations between pairs of independent variables don't mean you are safe from multicollinearity problems, because correlation could be 3-way.
 - c. If you add a new data point and the coefficients change a lot, this is a sign of multicollinearity problems.
 - d. The classic identification: low individual t-statistics but high R-square and F-statistics
 - e. Tests: partial correlations, eigenvalues, etc. are all controversial
4. What to do about multicollinearity
 - a. Ignore: as long as not perfect multicollinearity, OLS estimators are still BLUE, MLE, which means they are unbiased, efficient (but the confidence intervals will be wider), consistent
 - b. Or, eliminate variables (but be careful -- then you can get omitted-variable bias). Remember, if the absolute value of the t-statistic is > 1 , consider carefully before omitting a variable.
 - c. Outside information may help: a priori or cross-section regression + time series. But may have apples/oranges problem: cross-section is not same as time series
 - d. Transform data: first-differences. Levels may be correlated but not changes. This can create other problems, though.
 1. May make residuals serially correlated
 2. You lose one degree of freedom in differencing. This is not a problem if you have a lot of data, but can be a problem for small samples.
 3. Cross-sectional data, and other data that are not ordered cannot be differenced. It makes no sense.

- e. Add data. Increasing the sample size increases precision, even if correlation doesn't go down. This is not always easy or possible.
- f. Other solutions are controversial.

D. Serial Correlation

- 1. Pure vs Impure Serial correlation
- 2. Consequences of Serial Correlation
- 3. Durbin-Watson d-test
- 4. Generalized least squares