

**POL 502**  
**Intermediate Statistics**  
**Lecture 4**  
**February 14<sup>th</sup>, 2005**

MINITAB example – how do we save the  $\hat{Y}_s$ , the predicted values?

How do we get the errors for each observation?

To save the  $\hat{Y}_s$  :

In the regression window we open the “storage” window and click “FITS.”

Then when we run the regression a new column of data appears – the fitted values,  $\hat{Y}_s$ , for that regression.

To get the errors for each observation we go to calc → calculator

Store results in variable \_\_\_\_\_ ← put the first empty column number in here.

In the expression window: Dependent variable – FITS. That is,  $Y - \hat{Y}$ .

We can also see how much of each observation is explained by the regression by creating a new variable called “Explained” as  $\hat{Y} - \bar{Y}$ .

### **Chapter 4 in Studenmund**

#### The Classical Assumptions

If we want the estimates we get from OLS to be accurate, we need several assumptions to hold.

We will break many of them, so we want to understand the consequences and learn how to spot the problems and when possible avoid them.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

there is some error.

the random component that we are not explaining.

That error or the constant or both, includes all the information we are not using.

If I explain winning percentage using only batting, the error term includes the useful information of pitching.

The error tells us a great deal about our model and it should hold certain properties.

These make up most of our regression assumptions.

The Assumptions are:

1. Linearity: The regression model is linear in the coefficients, is correctly specified, and has an additive error term.
2. The error term has a zero population mean.
3. All explanatory variables are uncorrelated with the error term.
4. Observations of the error term are not correlated with each other (no serial correlation).
5. The error term has a constant variance across all X's (no heteroskedasticity).
6. No perfect multicollinearity. The X's are different from each other, and preferably not too closely related.
7. The error term is normally distributed – this is not required but is usually assumed.

Lets go through these 1 at a time.

1. Linearity: The regression model is linear in the coefficients, is correctly specified, and has an additive error term.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

The coefficients all predict a linear relationship.

If we predict non-linear relationships, we can transform the equation.

What are some non-linear relationships?

Diminishing Marginal Utility – a economics term meaning that the value of each additional unit goes down.

for example, the impact on income of the first year of education is greater than the second year which is greater than the third year and so on.

Correctly specified means we are including all the relevant X's  
 $\varepsilon_i$  is not a function of any variable.

2. The error term has a zero population mean.

If the error term is truly random, we should see that in several ways.

$\varepsilon_i$  is distributed randomly around the regression line.

For each  $X_i$ , we are just as likely to miss above as below, on average we will land on the line.

Thus, the distribution of the error term has a mean of Zero which is equivalent to saying its expected value is Zero.

Here we can put in assumption 7 also, that the error term is normally distributed. It is more likely to be closer to the mean than far away.

3. All explanatory variables are uncorrelated with the error term.

If the error term is randomly generated and distributed, it shouldn't be related to the X variables.

If they were related, we might believe X is causing variation in Y that is actually caused by  $\varepsilon_i$ .

This is something that we should test – and it can be tested fairly easily.

In MINITAB, we ask the regression to save our residuals, the errors.

Then we check the correlations between the independent variables and the residuals.

4. Observations of the error term are not correlated with each other (no serial correlation). If the fact that the  $\varepsilon_i$  from one observation is positive increases the probability that the  $\varepsilon_i$  from another observation is positive (or negative) the two are correlated.

This is especially a problem with data that is studied over time.

Each observation is supposed to be randomly selected, but *Time Series* data are sequential and may depend on the previous day to a certain extent.

5. The error term has a constant variance across all X's (no heteroskedasticity).

For any value of X, the variance in  $\varepsilon_i$  is the same.

We call this homoskedasticity.

Homoskedasticity means that our line is fitting just as well at any point on the line.

Not a better fit in one spot versus another spot.

This can be checked by making a graph of the residuals against X.

6. No perfect multicollinearity. The X's are different from each other, and preferably not too closely related.

$$\frac{(\sum yx_1)(\sum x_2^2) - (\sum yx_2)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

becomes 0/0 and regression is impossible.

As  $X_1$  and  $X_2$  get more and more alike, our estimates are problematic.

This can be checked by looking at the correlations between our independent variables.

7. The error term is normally distributed – this is not required but is usually assumed.

## The Normal Distribution – Review

Describe it:

Symmetrical, continuous, bell shaped curve.

Two parameters define the normal distribution, the mean  $\mu$ , and the variance,  $\sigma^2$ .

Can have two normal distributions with different values for the parameters.

The normal is useful to think about our errors, but also to think about the sampling distribution of  $\hat{\beta}$ .

## The Sampling Distribution of $\hat{\beta}$

Remember the sampling distribution of sample means?

What was it?

It was the distribution of means when we took multiple samples.

That was useful for a single variable.

Now we are interested in two or multiple variables.

We want to think about the  $\hat{\beta}$  that we generate as coming from a population of  $\hat{\beta}$ s.

Our  $\hat{\beta}$  came from taking one sample out of the population and seeing the relationship between the variables.

Had we taken another sample, we would have gotten a different  $\hat{\beta}$ .

Think about many many samples and the distribution of  $\hat{\beta}$  – this is what we mean by “Sampling Distribution of  $\hat{\beta}$ ”

If all of the above assumptions hold, we get an “**unbiased estimator**”

- our  $\hat{\beta}$  is an “estimator” in that we use it to estimate the relationship between X and Y.

- an unbiased estimator is one whose sampling distribution has as its expected value, the true value of  $\beta$ .

$$E(\hat{\beta}_k) = \beta_k$$

If we could take an infinite number of samples and get a  $\hat{\beta}$  for each, they would be normally distributed.

We only get one sample and therefore one  $\hat{\beta}$ , so we want to know that it is unbiased.

What else would we like?

Variance of distribution should be as small as possible for an unbiased estimator. This is called EFFICIENCY

We can decrease the variance by increasing our sample size.

As the size of our sample approaches the size of the population, we get closer and closer to the actual value of  $\beta$  and the variance is smaller and smaller. CONSISTENCY

So,

We have this theoretical sampling distribution of  $\hat{\beta}$ s.

We use the estimate we get,  $\hat{\beta}$ , to tell us what we believe the population value is.

We also want to know how much variance there may be in this estimate.

We want to know the variance of the  $\hat{\beta}$ .

Use the Standard Deviation of the Sampling Distribution of  $\hat{\beta}$ .  
This is called the Standard Error.

For one independent variable:

$$SE(\hat{\beta}) = \sqrt{\frac{\sum e_i^2 / (n-2)}{\sum (X_i - \bar{X})^2}}$$

For two independent variables:

$$SE(\hat{\beta}_1) = \sqrt{\frac{\sum e_i^2 / (n-3)}{\sum (X_{1i} - \bar{X}_1)^2 (1-r_{12}^2)}}$$

SE is roughly average miss for  $\hat{\beta}$ .

If all of our assumptions hold our estimator is **BLUE**.  
“Best (minimum variance) Linear Unbiased Estimator.”

Our OLS estimates will have the following properties:

1. They are unbiased.

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

2. They are minimum variance.

The distribution of the coefficient estimates around the true parameter values is as tight as possible.

3. They are consistent. As the sample size increases the estimates converge on the true population parameters.

4. They are normally distributed.

The  $\hat{\beta}_s \sim N(\beta, VAR[\beta])$ .

### Review of Notation Conventions

Population Parameter (True values, but unobserved)		Estimate (Observed from Sample)	
Name	Symbol(s)	Name	Symbol(s)
Regression coefficient	$\beta_k$	Estimated regression coefficient	$\hat{\beta}_k$
Expected value of the estimated coeff.	$E(\hat{\beta}_k)$		
Variance of the error term.	$\sigma^2$ or $VAR(\varepsilon_i)$	Estimated variance of the error term	$s^2$ or $\hat{\sigma}^2$
Standard deviation of the error term	$\sigma$	Standard error of the equation (estimate)	s or SEE
Variance of the estimated coeff.	$\sigma^2(\hat{\beta}_k)$ or $VAR(\hat{\beta}_k)$	Estimated variance of the estimated coeff.	$s^2(\hat{\beta}_k)$ or $\widehat{VAR}(\hat{\beta}_k)$
Standard deviation of the estimated coeff.	$\sigma_{\hat{\beta}_k}$ or $\sigma(\hat{\beta}_k)$	Standard error of the estimated coeff.	$\hat{\sigma}(\hat{\beta}_k)$ or $SE(\hat{\beta}_k)$
Error or “disturbance term.”	$\varepsilon_i$	Residual (estimate of error.)	$e_i$

### Homework

Read the chart on Page 85 10 times. Be sure you understand them.

Exercises 1, 2, 3, 5.