

## Lecture 10: Multiple Regression II

- $R^2$  as a Measure of Model Fit – Chapter 4.3
- Sums of Squares and Effect Tests – Chapter 4.4
  - Sums of Squares SST, SSE and SSR
  - ANOVA Table
  - $R^2$
  - F-statistic and the F-test for ALL coefficients together
  - Effect Test Table and Tests for Individual Coefficients
- **Example:** Gasoline Mileage for New Cars [cont]

## SST and SSE

- In multiple regression, this works the same general way as in ordinary regression.
- The Total Sum of Squares has the same formula:

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

- The Sum of Squares for Regression has the same sort of formula:

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

– The only difference here is that the formula for the prediction of the conditional mean involves all the x-variables –

$$\hat{Y}_i \triangleq b_0 + b_{1i}x_{1i} + \dots + b_{Ki}x_{Ki}.$$

## Sum of Squares for Regression

- The decomposition of the sums of squares also works just like in ordinary regression.
- The Sum of Squares for Regression is called SSR in *Dielman*, and Sum of Squares for the Model in JMP.
- The basic decomposition is

$$SST=SSR+SSE.$$

- Hence, from SST and SSE you can find SSR as

$$SSR=SST - SSE$$

- These sums of squares appear in the “Analysis of Variance Table” (aka ANOVA table)

## ANOVA Table

- Here is the JMP ANOVA table for our data

Analysis of Variance						
Source		DF	Sum of Squares	Mean Square		F Ratio
Model	$K$	4	SSR 2291.2	SSR/DFR 573	MSR/MSE 210	
Error	$n-K-1$	217	SSE 589.9	SSE/DFE 2.72		Prob > F
C. Total	$n-1$	221	SST 2881.1			<.0001

Note the **D**egrees of **F**reedom:

$$DFR = K, DFE = n-K-1, DFT_{\text{otal}} = n-1.$$

Note the **M**ean **S**quares:

$$\text{“Mean Square for Model”} = MSR = SSR/DFR$$

$$s_e^2 = MSE = SSE/DFE.$$

Finally **NOTE**:

$$F = MSR/MSE.$$

## The “Coefficient of Determination”, $R^2$

- Relative to the terms in the ANOVA table this also has the same definition and interpretation as in ordinary regression:

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

- **THUS,  $R^2$  is the proportion of squared variability accounted for by the linear regression using all the x-variables** relative to squared variability of the best fit that does not use any of the x-variables.
- One difference between the situation here and in ordinary regression: In ordinary regression,  $R^2$  is the square of a sample regression coefficient – There’s no such interpretation here!

- $R^2$  appears in the JMP Summary of Fit Table:

### Summary of Fit

RSquare	$R^2$	0.795
RSquare Adj		0.791
Root Mean Square Error	$s_e$	1.65
Mean of Response	$\bar{Y}$	19.40
Observations	$n$	222

- See *Dielman* for definition of “Adjusted  $R^2$ ”. We don’t find this concept particularly useful, and so will ignore it.
- Always  $0 \leq R^2 \leq 1$ .
  - $R^2 = 0$  means the linear regression equation is useless as a means of predicting the  $Y_i$ .
  - $R^2 = 1$  means it provides a perfect fit to the  $Y_i$ .
- For our data,  $R^2 \approx 80\%$ . This is a pretty high value.

## F-Statistic and F-Test

- The F-statistic provides a means of **testing** that the linear model is of no value.
- The null hypothesis corresponding to this test is

$$H_0 : \mu_{Y_i|x_1, \dots, x_K} = \beta_0, \text{ a constant for every } x_1, \dots, x_k.$$

- The alternative is that our linear model is at least a teeny bit useful:

$$H_a : \mu_{Y_i|x_1, \dots, x_K} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki}$$

with *at least one* of the  $\beta_j \neq 0$ .

- Larger values of  $R^2$  suggest that the null is false, and the alternative is more plausible.
- R. A. Fisher ( $\approx$  1924) invented the test of this  $H_0$  vs  $H_a$ .

## Fisher's F-Test

- R. A. Fisher decided to use F to test  $H_0$  vs  $H_a$ . The definition of F is

$$F = \frac{MSR}{MSE}. \quad \{\text{Note that } F \geq 0.\}$$

- Larger values of F suggest that  $H_a$  is more reasonable than  $H_0$ .
- Hence the F-Test goes as follows:

Get F from the ANOVA table, and at level  $\alpha$

**Reject  $H_0$**  in favor of  $H_a$  when  $F > F(\alpha; DFR, DFE)$

where  $F(\alpha; DFR, DFE)$  comes from a table stored in JMP.

- Things to notice
  - The critical value  $F$  depends on  $\alpha$ , DFR and DFE.
  - The test rejects only when F is large, not when it's small.

## More Facts About the F-Test

- F and  $R^2$  are directly related. As one increases, so does the other.
- As  $F \nearrow$  from 0 to  $\infty$  the value of  $R^2 \nearrow$  from 0 to 1.
- Hence, rejecting for  $F \geq F(\cdot)$  is same as rejecting for  $R^2 \geq R(\cdot)$ .
- If  $H_0$  is true the values of F will be moderately close to 1; hence you'll find in the table that the critical values for  $F$  are somewhat larger than 1.
- **For our data**  $MSR = 570$  and  $MSE = 2.72$ . Hence
$$F = 570/2.72 = 210$$
- This is a huge value; the critical value from Table B. 4 is about
$$F(.05; 4, 217) \approx 2.4.$$
- Hence we **Reject  $H_0$**  at level  $\alpha = 0.05$  or any other reasonable level. (P-value  $< 0.0001$ .)

## Interim Summary

- The standard F-test provides a means of rejecting  $H_0$  and thus validating the **Alternative Hypothesis** ( $H_a$ ) that at least one of the model coefficients,  $\beta_j$ , is not zero.
- It is built in a simple way from the Sums of Squares and Degrees of Freedom in the ANOVA table.
- $R^2$  is also built from the same Sums of Squares and is directly related to F.
- The F-test rejects when F is too large, and this is the same as rejecting when  $R^2$  is too large. (Here, “Too large” is not really so big.)
- $R^2$  measures the “strength” of the linear relationship in the model, as compared to the situation with  $\beta_1 = \dots = \beta_K = 0$ .

See addendum for detailed derivation and description of the relation between  $R^2$  and F.

# Test About Individual Coefficients

## The “Effect Tests” Table

### and its Relation to the “Parameter Estimates” Table

- In Lecture 9 we saw how to use entries in JMP’s Parameter Estimates table to construct CIs for the coefficients  $\beta_j$ , and also how to use these entries to test hypotheses about them.
- Now we’ll see where these tests actually come from in terms of sums of squares like those in the ANOVA table.
- This can help explain why a given coefficient such as  $\beta_j$  should be interpreted as the effect of  $x_j$  after controlling for the effect of all the other variables.
- It will also lead us to be able to test other hypotheses of interest in multiple regression types of settings.

## Example

### Testing the effect of the variable “Seats”

- We want to understand how Seats effects MPG\_City after controlling for the three other variables in our study. (These are **HP**, **Wt(1000lb)** and **Length**.)
- To do so we begin by seeing how well these three other variables can do to predict MPG.
- We find this out by analyzing a multiple regression model with *only these three variables as predictors*.
- Here is the basic table for that analysis

ANOVA Table for Regression of **MPG\_City**  
on the three variables **HP**, **Wt**, and **Length**

Source	DF	Analysis of Variance		Mean Square	F Ratio
		Sum of Squares			
Model	<b>3</b>	<b>SSM<sub>R</sub></b>	2274.6	758	272.5
Error	<b>218</b>	<b>SSE<sub>R</sub></b>	606.5	2.78	Prob > F
C. Total	221	<b>SST</b>	2881.1		<.0001

- The subscript “**R**” on **SSM** and **SSE** indicates that these Sums of Squares are for this “**R**educed” model having only 3 predictors.
- Note that we don’t need a subscript on SST, because that’s the same as in the original ANOVA table for the full model with 4 predictors.
- As a reminder, here’s the table for the full model:

ANOVA Table for Full Regression Model of **MPG\_City**  
on all 4 variables **HP**, **Wt**, **Length** and **Seats** ( $= x_3$ ).

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	$SSR_F$ 2291.2	573	210
Error	217	$SSE_F$ 589.9	$MSE_F$ 2.72	Prob > F
C. Total	221	$SST$ 2881.1		<.0001

- Compare the SSE terms in the two tables.
- $SSE_F \leq SSE_R$  .
- **This inequality will always be true.** {Why?}
- We use the difference as a numerical measure of the effect of **Seats** after controlling for **HP**, **Wt**, and **Length**.
- **THUS**

$$SS(\text{Seats} | \text{HP, Wt, Length}) = SSE_R - SSE_F .$$

$$= 606.5 - 589.9 = 16.6$$

Test for  $H_0 : \beta_3 = 0$  vs  $H_a : \beta_3 \neq 0$  in the 4-factor Model

- This is a test of whether there is an effect of **Seats** ( $x_3$ ) after controlling for the other three factors **HP**, **Wt**, and **Length**.
- It looks at how much the inclusion of **Seats** has reduced SSE below the value of  $SSE_R$ , and compares this to  $MSE_F$ .
- Thus, the statistic is

$$F_{\text{Seats}} = \frac{SS(\text{Seats} | \text{HP}, \text{Wt}, \text{Length})}{MSE_F}$$
$$= \frac{16.6}{2.72} = 6.12$$

- This should be compared to the F-Table with 1 DF in the numerator and  $DFE = n - 1 - K = 217$  DF in the denominator.
- From Tables B.4 and B.5 we reject at  $\alpha=.05$ , & not at  $\alpha=.01$ .

The entries in the tables are approx 3.95 for alpha = 0.05 and approx 6.75 for alpha = 0.01.

## Effect Test Table

- The important part of the above information appears in JMP's Effect Test Table.

Source	DF	Effect Tests		
		Sum of Squares	F Ratio	Prob > F
HP	1	82.2	30.2	<.0001
Wt (1000lb)	1	499.7	184	<.0001
Seats	1	<b>16.6</b>	<b>6.12</b>	<b>0.0141</b>
Length	1	5.25	1.93	0.1660

- You can see that the P-Value is 0.0141, and this corresponds to our finding that we reject at  $\alpha=.05$ , & not at  $\alpha=.01$ .
- Reminder: This is a Test for  $H_0 : \beta_3 = 0$  vs  $H_a : \beta_3 \neq 0$  after controlling for the effect of HP, Wt, and Length.

## Relation to the Test Based on the Parameter Estimates Table

- In Lecture 9 we used entries in the Parameter Estimates table to test the same null hypothesis.
- That's not a problem, because **the two tests we've constructed are exactly equivalent**. To see this, look again at the Parameter Estimates table:

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	$b_0 = 31.50$	1.77	17.84	<.0001
HP	$b_1 = -0.0154$	0.00280	-5.50	<.0001
Wt (1000)	$b_2 = -3.77$	0.278	-13.56	<.0001
Seats	$b_3 = 0.337$	0.136	<b>2.47</b>	<b>0.0141</b>
Length	$b_4 = 0.0175$	0.0126	1.39	0.1660

- **Note that the P-values in this table and in the Effects Test table are exactly the same**

- It's not an accident that they're the same:
- They're the same because

$$6.12 = 2.47^2$$

and because the square of a t-statistic is an F-statistic with 1 DF in the numerator.

(To check the claim about F and t you can compare the entries in the first column of Table B.4 (for  $\alpha = .05$ ) with the squares of the entries in the column for  $\alpha = .025 = .05/2$  in the t-table, B.3. For example the entry in the t-table for 8DF is 2.306 and the entry in the F-table is 5.32; and – sure enough -  $2.306^2 = 5.32$ .)

- Consequently, you can use either table to get the test.
- We can also use these tables to test whether each of the other variables has a non-zero effect, after controlling for the remaining variables.
- The answer (at any reasonable  $\alpha$ ) is that HP and Wt do have significant effects after controlling for the remaining 3 variables; but for Length we cannot reject the null hypothesis of no effect since the P-value is  $P = 0.166$ .

## Notes

- You might wonder what is the source of the value for the Std Error in the Parameter Estimates table. That number actually comes about as an indirect result of the previous calculations using Sums of Squares.
- The preceding discussion has shown how to find  $F_{\text{Seats}}$  from Sums of Squares in the least squares analysis. Then we've seen that the t-ratio is the square root of the partial F-value  $F_{\text{Seats}}$ . From this fact we got  $t = 2.47$  in the case of Seats. Finally, by its definition the t-ratio for Seats is

$$2.47 = t = \frac{b_3}{SE_{b_3}}.$$

- Solving for  $SE_{b_3}$  yields the desired value,

$$SE_{b_3} = \frac{b_3}{2.47} = \frac{.337}{2.47} = .136.$$

- This SE can then be used to construct CIs for the coefficient of Seats after controlling for the other three variables, as we've done in Lecture 9.

## Relation between F and $R^2$

Recall,

$$F = \frac{SSR/DFR}{SSE/DFE} \quad \text{and} \quad R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

From this we can get

$$\frac{DFR}{DFE} F = \frac{SSR}{SSE} \quad \text{and} \quad 1 - R^2 = \frac{SSE}{SST}.$$

The two equations involving  $R^2$  can be combined by taking their ratio to get

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

This can then be combined with the second expression for F to yield

$$F = \frac{DFE}{DFR} \cdot \frac{R^2}{1 - R^2}.$$

This shows that F and  $R^2$  are directly related and that F increases from 0 to  $\infty$  as  $R^2$  increases from 0 to 1.