

# Linear Regression for Business Statistics

**Interpreting the R-square (a goodness-of-fit measure)**



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

# Linear Regression for Business Statistics

**Interpreting the R-square (a goodness-of-fit measure)**



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

# Linear Regression for Business Statistics

## Interpreting the R-square (a goodness-of-fit measure)

'Overall' variation in Y variable : 'Total' Sum of Squares



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

'Overall' variation in Y variable : 'Total' Sum of Squares

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	4149854856	1.04E+09	19.18472	2.82684E-11
Residual	85	4596596213	54077603		
Total	89	8746451069			



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

**'Overall' variation in Y variable : 'Total' Sum of Squares**

**'Explained' variation in Y variable : 'Regression' Sum of Squares**

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	4149854856	1.04E+09	19.18472	2.82684E-11
Residual	85	4596596213	54077603		
Total	89	8746451069			



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

**'Overall' variation in Y variable : 'Total' Sum of Squares**

**'Explained' variation in Y variable : 'Regression' Sum of Squares**

**'Unexplained' variation in Y variable : 'Residual' Sum of Squares**

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	4149854856	1.04E+09	19.18472	2.82684E-11
Residual	85	4596596213	54077603		
Total	89	8746451069			



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

**'Overall' variation in Y variable : 'Total' Sum of Squares**

**'Explained' variation in Y variable : 'Regression' Sum of Squares**

**'Unexplained' variation in Y variable : 'Residual' Sum of Squares**

$$\text{Total SS} = \text{Regression SS} + \text{Residual SS}$$

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	4149854856	1.04E+09	19.18472	2.82684E-11
Residual	85	4596596213	54077603		
Total	89	8746451069			





## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

*'Overall'* variation in Y variable : 'Total' Sum of Squares

*'Explained'* variation in Y variable : 'Regression' Sum of Squares

*'Unexplained'* variation in Y variable : 'Residual' Sum of Squares

$$\text{Total SS} = \text{Regression SS} + \text{Residual SS}$$

$$\text{R - square} = \frac{\text{Regression SS}}{\text{Total SS}}$$

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	4149854856	1.04E+09	19.18472	2.82684E-11
Residual	85	4596596213	54077603		
Total	89	8746451069			



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- The regression model can explain about 47.45% variation in the House Prices.
- The remaining variation goes unexplained



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- The regression model can explain about 47.45% variation in the House Prices.
- The remaining variation goes unexplained.



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- The regression model can explain about 47.45% variation in the House Prices.
- The remaining variation goes unexplained.
- **R-square is always between 0 and 1.**
- Value closer to 1 is a 'good' model fit.
- Value closer to 0 is a 'poor' model fit.



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- The regression model can explain about 47.45% variation in the House Prices.
- The remaining variation goes unexplained.
- R-square is always between 0 and 1.
- Value closer to 1 is a 'good' model fit.
- Value closer to 0 is a 'poor' model fit.



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

<i>Regression Statistics</i>	
Multiple R	0.6888117
→ R Square ←	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

➤ The regression model can explain about 47.45% variation in the House Prices.

➤ The remaining variation goes unexplained.

➤ R-square is always between 0 and 1.

➤ Value closer to 1 is a 'good' model fit.

➤ Value closer to 0 is a 'poor' model fit.



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

Regression Statistics	
Multiple R	0.6888117
→ R Square	0.47446156 ←
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- The regression model can explain about 47.45% variation in the House Prices.
- The remaining variation goes unexplained.
- R-square is always between 0 and 1.
- Value closer to 1 is a 'good' model fit.
- Value closer to 0 is a 'poor' model fit.



# Linear Regression for Business Statistics

Interpreting the R-square (a goodness-of-fit measure)

A common misconception about R-square...

...a low R-square model is of no use.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

A common misconception about R-square...

...a low R-square model is of no use.

A low R-square implies...

- Perhaps missing some important explanatory variables.
- Predictions would not be that accurate.



## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

A common misconception about R-square...

...a low R-square model is of no use.

A low R-square implies...

- Perhaps missing some important explanatory variables.
- Predictions would not be that accurate.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

A common misconception about R-square...

...a low R-square model is of no use.

A low R-square implies...

- Perhaps missing some important explanatory variables.
- Predictions would not be that accurate.
- *However*, there is value in that the model lets you infer relationship between the X variables and Y variable.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

# Linear Regression for Business Statistics

## Interpreting the R-square (a goodness-of-fit measure)

If focus is on *prediction*

...low R-square is problematic.

If focus is more on *understanding relation* between X and Y variables

...low R-square may not be that problematic.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

If focus is on *prediction*

...low R-square is problematic.

If focus is more on *understanding relation* between X and Y variables

...low R-square may not be that problematic.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

If focus is on *prediction*

...low R-square is problematic.

If focus is more on *understanding relation* between X and Y variables

...low R-square may not be that problematic.



RICE  
JONES GRADUATE  
SCHOOL OF BUSINESS

## Linear Regression for Business Statistics

### Interpreting the R-square (a goodness-of-fit measure)

If focus is on *prediction*

...low R-square is problematic.

If focus is more on *understanding relation* between X and Y variables

...low R-square may not be that problematic.





## Linear Regression for Business Statistics

### The "Adjusted R-square"

<i>Regression Statistics</i>	
Multiple R	0.6888117
R Square	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90



## Linear Regression for Business Statistics

### The "Adjusted R-square"

<i>Regression Statistics</i>	
Multiple R	0.6888117
R Square	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- Mere addition of X variables always increases R-square.



## Linear Regression for Business Statistics

### The "Adjusted R-square"

<i>Regression Statistics</i>	
Multiple R	0.6888117
R Square	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- Mere addition of X variables always increases R-square.
- Adj. R-square adjusts the R-square for the number of X variables in the model.



## Linear Regression for Business Statistics

### The "Adjusted R-square"

Regression Statistics	
Multiple R	0.6888117
R Square	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- Mere addition of X variables always increases R-square.
- Adj. R-square adjusts the R-square for the number of X variables in the model.

Which "R-square" to use?



## Linear Regression for Business Statistics

### The "Adjusted R-square"

Regression Statistics	
Multiple R	0.6888117
R Square	0.47446156
Adjusted R Square	0.44973034
Standard Error	7353.74751
Observations	90

- Mere addition of X variables always increases R-square.
- Adj. R-square adjusts the R-square for the number of X variables in the model.

### Which "R-square" to use?

- Better to use the Adj. R-square.