The Coefficient of Determination

Lecture 46
Section 13.9

Robb T. Koether

Hampden-Sydney College

Wed, Apr 17, 2012

Outline

1. The Regression Identity
2. Sums of Squares on the TI-83
3. TI-83 - The Coefficient of Determination
4. Assignment

Outline

1. The Regression Identity
2. Sums of Squares on the TI-83
3. TI-83 - The Coefficient of Determination
4. Assignment

Explaining the Variation in y

- Statisticians use regression models to "explain" y.
- More specifically, through the model they use *variation* in x to explain *variation* in y.

Explaining the Variation in y

- For example, why do some people weigh more than other people?

Explaining the Variation in y

- For example, why do some people weigh more than other people?
- That is, why is there variability in weight?

Explaining the Variation in y

- For example, why do some people weigh more than other people?
- That is, why is there variability in weight?
- One explanation is that some people weigh more than others because they are taller.

Explaining the Variation in y

- For example, why do some people weigh more than other people?
- That is, why is there variability in weight?
- One explanation is that some people weigh more than others because they are taller.
- That is, because there is variability in height.

Explaining the Variation in y

- For example, why do some people weigh more than other people?
- That is, why is there variability in weight?
- One explanation is that some people weigh more than others because they are taller.
- That is, because there is variability in height.
- But that is only a partial explanation.

Explaining the Variation in y

- Statisticians want to quantify how much of the variation in y is explained by the variation in x.
- We measure variation in y by the standard deviation

$$s_y = \sqrt{\frac{\text{SSY}}{n-1}}.$$

- However, the key component is SSY.

The Regression Identity

- There are three different variations that we can measure.
- How the y values in the data deviate from their mean.
- How the y values in the model deviate from their mean.
- How the y values in the data deviate from the y values in the model.

The Regression Identity

- There are three different variations that we can measure.
- That is, compare y to $\overline{y}$.
- How the y values in the model deviate from their mean.
- How the y values in the data deviate from the y values in the model.

The Regression Identity

- There are three different variations that we can measure.
- That is, compare y to $\overline{y}$.
- That is, compare $\hat{y}$ to $\overline{y}$.
- How the y values in the data deviate from the y values in the model.

The Regression Identity

- There are three different variations that we can measure.
- That is, compare y to $\overline{y}$.
- That is, compare $\hat{y}$ to $\overline{y}$.
- That is, compare y to $\hat{y}$.

The Regression Identity

- Variation in the data (Total sum of squares):

$$\text{SST} = \sum (y - \overline{y})^2.$$

- (This is the same as SSY.)
- Variation in the model (Regression sum of squares):

$$\text{SSR} = \sum (\hat{y} - \overline{y})^2.$$

- Residues (Sum of squared Errors):

$$\text{SSE} = \sum (y - \hat{y})^2.$$

The Regression Identity

- SST is the observed variation in y.
- SSR is the predicted variation in y (according to the model).
- SSE is the unexplained variation in y (not explained by the model).

Example - SST, SSR, and SSE

- The following data represent the heights and weights of 10 adult males.

Height (x)	Weight (y)
70	185
65	140
71	180
76	220
68	150
67	170
68	185
72	200
74	210
69	160

Example - SST, SSR, and SSE

- The regression line is

$$\hat{y} = -310 + 7x.$$

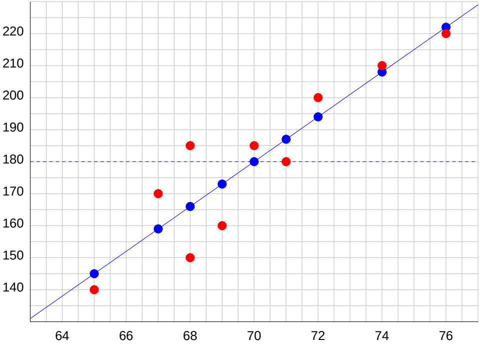
- The model predicts, for example, that if a person is 70 inches tall, he will weigh 180 pounds.
- The model also predicts that a person will weigh an additional 7 pounds for each additional inch of height.

Example - SST, SSR, and SSE

- Compute the predicted weight: $\mathtt{Y}_1\mathtt{(L}_1\mathtt{)} \rightarrow \mathtt{L}_3$.

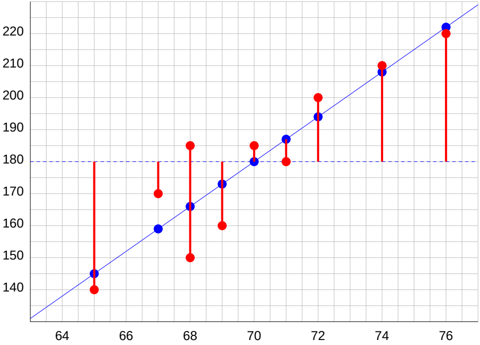
Height (x)	Weight (y)	Pred. Wgt. ($\hat{y}$)
70	185	180
65	140	145
71	180	187
76	220	222
68	150	166
67	170	159
68	185	166
72	200	194
74	210	208
69	160	173

Example - SST, SSR, and SSE



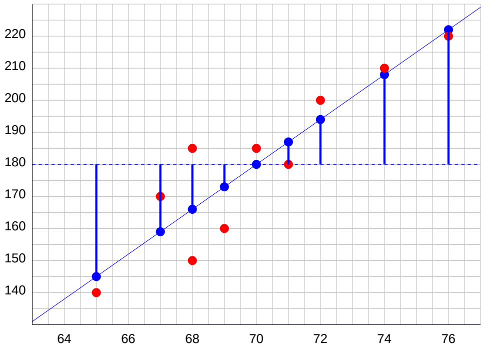
The regression line

Example - SST, SSR, and SSE



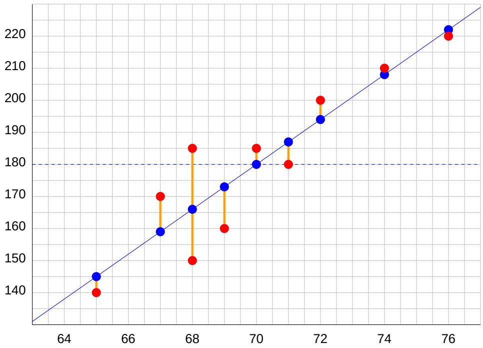
The deviations of y from $\overline{y}$

Example - SST, SSR, and SSE



The deviations of $\hat{y}$ from $\overline{y}$

Example - SST, SSR, and SSE



The deviations of y from $\hat{y}$

Example

- Compute SST.

x	y	$y - \overline{y}$	$(y - \overline{y})^2$
70	185		
65	140		
71	180		
76	220		
68	150		
67	170		
68	185		
72	200		
74	210		
69	160		

Example

- Compute SST: `L2 - ȳ`.

x	y	$y-\overline{y}$	$(y-\overline{y})^2$
70	185	5	
65	140	−40	
71	180	0	
76	220	40	
68	150	−30	
67	170	−10	
68	185	5	
72	200	20	
74	210	30	
69	160	−20	

Example

- Compute SST: `Ans`^{2}.

x	y	$y - \overline{y}$	$(y - \overline{y})^2$
70	185	5	25
65	140	-40	1600
71	180	0	0
76	220	40	1600
68	150	-30	900
67	170	-10	100
68	185	5	25
72	200	20	400
74	210	30	900
69	160	-20	400

Example

- Compute SST: `sum(Ans)`.

x	y	$y - \overline{y}$	$(y - \overline{y})^2$
70	185	5	25
65	140	−40	1600
71	180	0	0
76	220	40	1600
68	150	−30	900
67	170	−10	100
68	185	5	25
72	200	20	400
74	210	30	900
69	160	−20	400
			5950

Example

- Compute SSR.

x	y	$\hat{y}$	$\hat{y} - \overline{y}$	$(\hat{y} - \overline{y})^2$
70	185			
65	140			
71	180			
76	220			
68	150			
67	170			
68	185			
72	200			
74	210			
69	160			

Example

- Compute SSR: $\texttt{Y}_1\ (\texttt{L}_1) \rightarrow \texttt{L}_3$.

x	y	$\hat{y}$	$\hat{y}-\overline{y}$	$(\hat{y}-\overline{y})^2$
70	185	180		
65	140	145		
71	180	187		
76	220	222		
68	150	166		
67	170	159		
68	185	166		
72	200	194		
74	210	208		
69	160	173		

Example

- Compute SSR: L3−y̅.

x	y	$\hat{y}$	$\hat{y} - \overline{y}$	$(\hat{y} - \overline{y})^2$
70	185	180	0	
65	140	145	−35	
71	180	187	7	
76	220	222	42	
68	150	166	−14	
67	170	159	−21	
68	185	166	−14	
72	200	194	14	
74	210	208	28	
69	160	173	−7	

Example

- Compute SSR: `Ans`^{2}.

x	y	$\hat{y}$	$\hat{y} - \overline{y}$	$(\hat{y} - \overline{y})^2$
70	185	180	0	0
65	140	145	−35	1225
71	180	187	7	49
76	220	222	42	1764
68	150	166	−14	196
67	170	159	−21	441
68	185	166	−14	196
72	200	194	14	196
74	210	208	28	784
69	160	173	−7	49

Example

- Compute SSR: `sum(Ans)`.

x	y	$\hat{y}$	$\hat{y} - \overline{y}$	$(\hat{y} - \overline{y})^2$
70	185	180	0	0
65	140	145	−35	1225
71	180	187	7	49
76	220	222	42	1764
68	150	166	−14	196
67	170	159	−21	441
68	185	166	−14	196
72	200	194	14	196
74	210	208	28	784
69	160	173	−7	49
				4900

Example

- Compute SSE.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
70	185			
65	140			
71	180			
76	220			
68	150			
67	170			
68	185			
72	200			
74	210			
69	160			

Example

- Compute SSE: `Y1(L1)` $\rightarrow$ `L3`.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
70	185	180		
65	140	145		
71	180	187		
76	220	222		
68	150	166		
67	170	159		
68	185	166		
72	200	194		
74	210	208		
69	160	173		

Example

- Compute SSE: $\texttt{L}_2 - \texttt{L}_3 \rightarrow \texttt{L}_4$.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
70	185	180	5	
65	140	145	−5	
71	180	187	−7	
76	220	222	−2	
68	150	166	−16	
67	170	159	11	
68	185	166	19	
72	200	194	6	
74	210	208	−7	
69	160	173	−13	

Example

- Compute SSE: `Ans`^{2}.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
70	185	180	5	25
65	140	145	-5	25
71	180	187	-7	49
76	220	222	-2	4
68	150	166	-16	256
67	170	159	11	121
68	185	166	19	361
72	200	194	6	36
74	210	208	-7	49
69	160	173	-13	169

Example

- Compute SSE: `sum(Ans)`.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
70	185	180	5	25
65	140	145	-5	25
71	180	187	-7	49
76	220	222	-2	4
68	150	166	-16	256
67	170	159	11	121
68	185	166	19	361
72	200	194	6	36
74	210	208	-7	49
69	160	173	-13	169
				1050

Example

- We have now found that

$$\text{SSR} = 4900.$$
$$\text{SSE} = 1050.$$
$$\text{SST} = 5950.$$

- We see that

$$\text{SSR} + \text{SSE} = \text{SST}.$$

- This is called the regression identity.

Explaining Variation

- It follows that the total variation in y (SST) can be separated into two parts:
  - SSR = the part "explained" by the model.
  - SSE = the part unexplained by the model.

Explaining Variation

- Furthermore,

$$1 = \frac{\text{SSR}}{\text{SST}} + \frac{\text{SSE}}{\text{SST}}.$$

- So, the fraction

$$\frac{\text{SSR}}{\text{SST}}$$

can be interpreted as the proportion of SST that is explained by the model.

Outline

1. The Regression Identity
2. Sums of Squares on the TI-83
3. TI-83 - The Coefficient of Determination
4. Assignment

TI-83 - Finding SSR, SSE, and SST

TI-83 SSR, SSE, and SST

- Put the *x* values into L_1 and the *y* values into L_2.
- Use `LinReg(a+bx)` L_1, L_2, Y_1.
- Enter $Y_1(L_1) \rightarrow L_3$.
- To get SSR, evaluate $\text{sum}((L_3 - \overline{y})^2)$.
- To get SSE, evaluate $\text{sum}((L_2 - L_3)^2)$.
- To get SST, evaluate $\text{sum}((L_2 - \overline{y})^2)$.

Explaining Variation

- It can be shown that
$$r^2 = \frac{\text{SSR}}{\text{SST}}$$
and, therefore,
$$1 - r^2 = \frac{\text{SSE}}{\text{SST}}.$$
- Therefore, r^2 is the proportion of variation in y that is explained by the model. It is called the coefficient of determination.
- $1 - r^2$ is the proportion that is not explained by the model.

Outline

1. The Regression Identity
2. Sums of Squares on the TI-83
3. TI-83 - The Coefficient of Determination
4. Assignment

TI-83 - Coefficient of Determination

TI-83 Coefficient of Determination

- To calculate r^2 on the TI-83, follow the procedure that produces the regression line and r.
- In the same window, the TI-83 reports the value of r^2.

Coefficient of Determination

Example (Coefficient of Determination)

- In the height-weight example, we found that $r = 0.9075$.
- Then $r^2 = 0.8235$.
- So, 82% of variation in weight can be explained by variation in height.
- The remaining 18% of variation in weight is due to some other factor.

Outline

1. The Regression Identity
2. Sums of Squares on the TI-83
3. TI-83 - The Coefficient of Determination
4. Assignment

Assignment

Homework

- Read Section 13.9, pages 868 - 869.
- Exercises 35(a), 36(f),