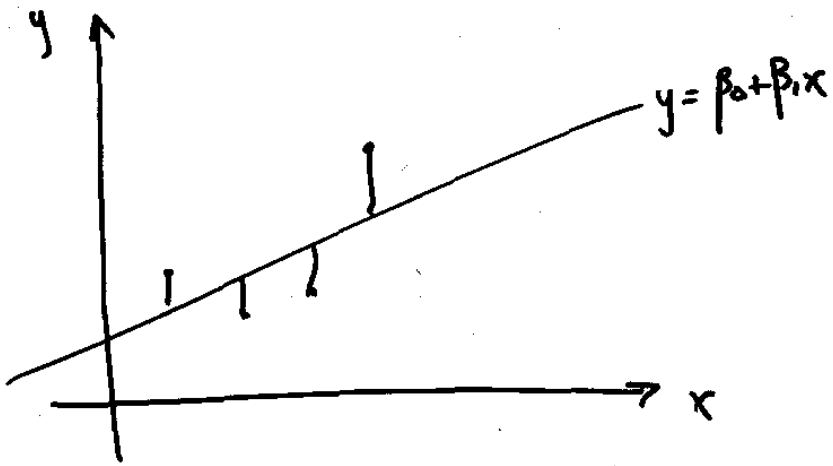
Chapter 11 Regression

552
5-6

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$



Assume β_0 and β_1 are unknown.

Estimate them with b_0 and b_1

The estimated line will be $\hat{y} = b_0 + b_1 x$

The i^{th} residual is $e_i = y_i - \hat{y}_i$

The error sum of squares is

$$SSE = \sum_{i=1}^{n} e_i^2$$

The least squares method finds b_0 and b_1 which minimize SSE.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

$$\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{n} 2(y_i - (b_0 + b_1 x_i))(-1) \stackrel{set}{=} 0$$

$$\sum_{i=1}^{n} y_i - n b_0 - b_1 \sum_{i=1}^{n} x_i = 0 \qquad (3)$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{n} 2\left(y_i - (b_0 + b_1 x_i)\right)(-x_i) \overset{\text{set}}{=} 0$$

$$\sum_{i=1}^{n} x_i y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum x_i y_i - (\bar{y} - b_1 \bar{x}) \sum x_i - b_1 \sum x_i^2 = 0$$

$$\underbrace{\sum x_i y_i - \bar{y} \sum x_i}_{S_{xy}} = b_1 \underbrace{\left(\sum x_i^2 - \bar{x} \sum x_i\right)}_{S_{xx}}$$

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

(4)

$$S_{xx} = \sum x_i^2 - \bar{x} \sum x_i$$

$$= \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$= \sum x_i^2 - n \bar{x}^2$$

$$= \sum (x_i - \bar{x})^2 = (n-1) s_x^2$$

Similarly, $S_{xy} = \sum x_i y_i - \bar{y} \sum x_i$

$$= \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

$$= \sum x_i y_i - n \bar{x} \bar{y}$$

$$= \sum (x_i - \bar{x})(y_i - \bar{y})$$

Now that b_0 and b_1 are known, find the actual minimum value of SSE

$$SSE = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2 \quad (5)$$

$$= \sum_{i=1}^{n} (y_i - (\bar{y} - b_1\bar{x}) - b_1 x_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y} - b_1(x_i - \bar{x}))^2$$

$$= \sum (y_i - \bar{y})^2 + b_1^2 \sum (x_i - \bar{x})^2 - 2b_1 \sum (y_i - \bar{y})(x_i - \bar{x})$$

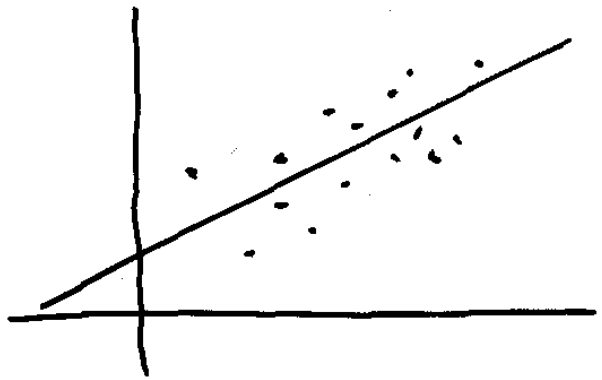
$$= S_{yy} + b_1^2 S_{xx} - 2b_1 S_{xy}$$

$$= S_{yy} + \frac{S_{xy}^2}{S_{xx}^2} S_{xx} - 2\frac{S_{xy}}{S_{xx}} S_{xy}$$

$$= S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

$$SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}} \quad (6)$$

$$S_{yy} = \frac{S_{xy}^2}{S_{xx}} + SSE$$



$$SST = SSR + SSE$$

SST: total; SSR: regression; SSE: error

Source	SS	df	MS	F
REG	SSR	1	$\frac{SSR}{1}$	$\frac{MSR}{MSE}$
ERR	SSE	$n-2$	$\frac{SSE}{n-2}$	
TOT	SST	$n-1$		

(* predictors)

Analysis of Variance (ANOVA) table

What fraction of SST is composed of SSR?

$$\frac{SSR}{SST} = \frac{S_{xy}^2}{S_{xx}} \Big/ S_{yy} = \frac{S_{xy}^2}{S_{xx} S_{yy}} = R^2$$

This is the coefficient of determination

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \text{Correlation coefficient}$$

HW#6 due 5-13

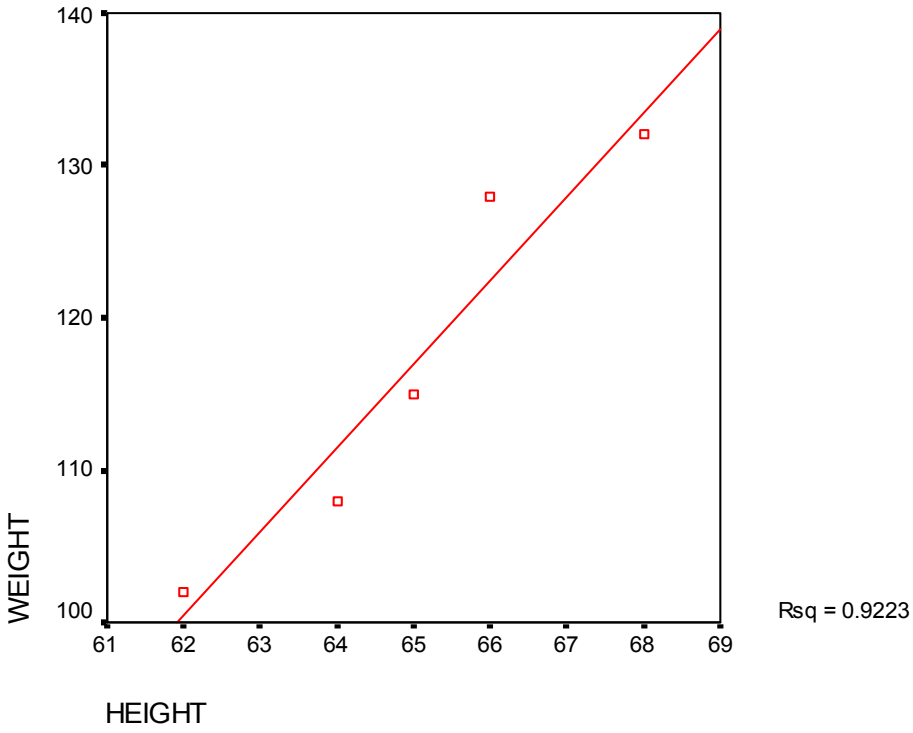
p. 398 #11.9
p. 424 #11.44(a)

} Must use a stat software package

reg2.sav

	height	weight
1	68	132
2	64	108
3	62	102
4	65	115
5	66	128
6	63	.

Graph



Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	HEIGHT[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: WEIGHT

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.960[a]	.922	.896	4.123

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	605.000	1	605.000	35.588	.009[a]
	Residual	51.000	3	17.000		
	Total	656.000	4			

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-240.500	59.955		-4.011	.028
	HEIGHT	5.500	.922	.960	5.966	.009

a. Dependent Variable: WEIGHT

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	100.50	133.50	117.00	12.298	5
Std. Predicted Value	-1.342	1.342	.000	1.000	5
Standard Error of Predicted Value	1.844	3.324	2.523	.737	5
Adjusted Predicted Value	97.71	136.29	116.97	13.931	5
Residual	-3.50	5.50	.00	3.571	5
Std. Residual	-.849	1.334	.000	.866	5
Stud. Residual	-.980	1.540	.004	1.047	5
Deleted Residual	-4.67	7.33	.03	5.444	5
Stud. Deleted Residual	-.971	2.750	.263	1.496	5
Mahal. Distance	.000	1.800	.800	.917	5
Cook's Distance	.037	.395	.259	.154	5
Centered Leverage Value	.000	.450	.200	.229	5

a. Dependent Variable: WEIGHT