Simple Linear Regression Analysis: an example

Ex:

- A real estate agent wishes to examine the relationship between the selling price of a house and its size (measured in square feet)

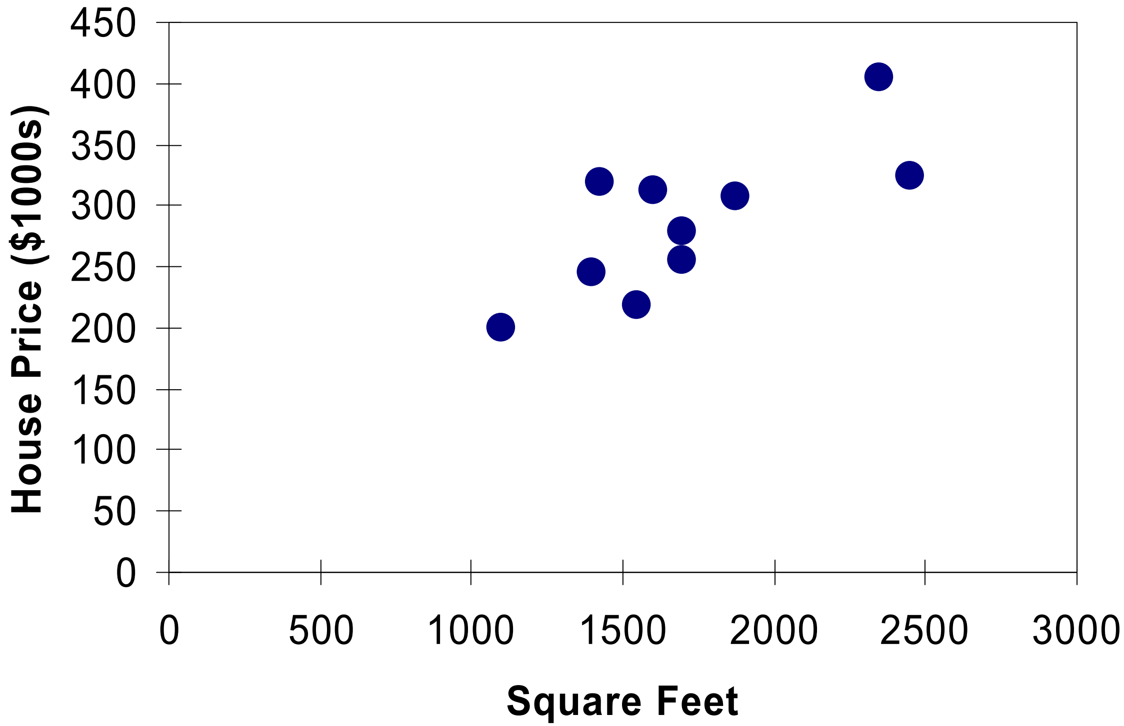
- A random sample of 10 houses is selected
  - Dependent variable (Y) = house price in $1000s
  - Independent variable (X) = square feet

Simple Linear Regression Analysis: an example

House Price in $1000s (Y)	Square Feet (X)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

Simple Linear Regression Analysis: an example

House price model: Scatter Plot



Simple Linear Regression Analysis: an example

	Y	X	$(Y-\bar{Y})$	$(X-\bar{X})$	$(Y-\bar{Y})^2$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
	245	1400	-41.5	-315	1722.25	99225	13072.5
	312	1600	25.5	-115	650.25	13225	-2932.5
	279	1700	-7.5	-15	56.25	225	112.5
	308	1875	21.5	160	462.25	25600	3440
	199	1100	-87.5	-615	7656.25	378225	53812.5
	219	1550	-67.5	-165	4556.25	27225	11137.5
	405	2350	118.5	635	14042.25	403225	75247.5
	324	2450	37.5	735	1406.25	540225	27562.5
	319	1425	32.5	-290	1056.25	84100	-9425
	255	1700	-31.5	-15	992.25	225	472.5
sum	2865	17150	0	0	32600.5	1571500	172500
mean	286.5	1715			3260.05	157150	17250

$$b_1 = \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{172500}{1571500} = 0.109768$$

$$b_0 = \hat{\beta}_0 = \bar{y} - b_1\bar{x} = 286.5 - 0.109768 \cdot 1715 = 98.24833$$

Simple Linear Regression Analysis: an example

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

The regression equation is:

$$\widehat{\text{house price}} = 98.24833 + 0.10977\ (\text{square feet})$$

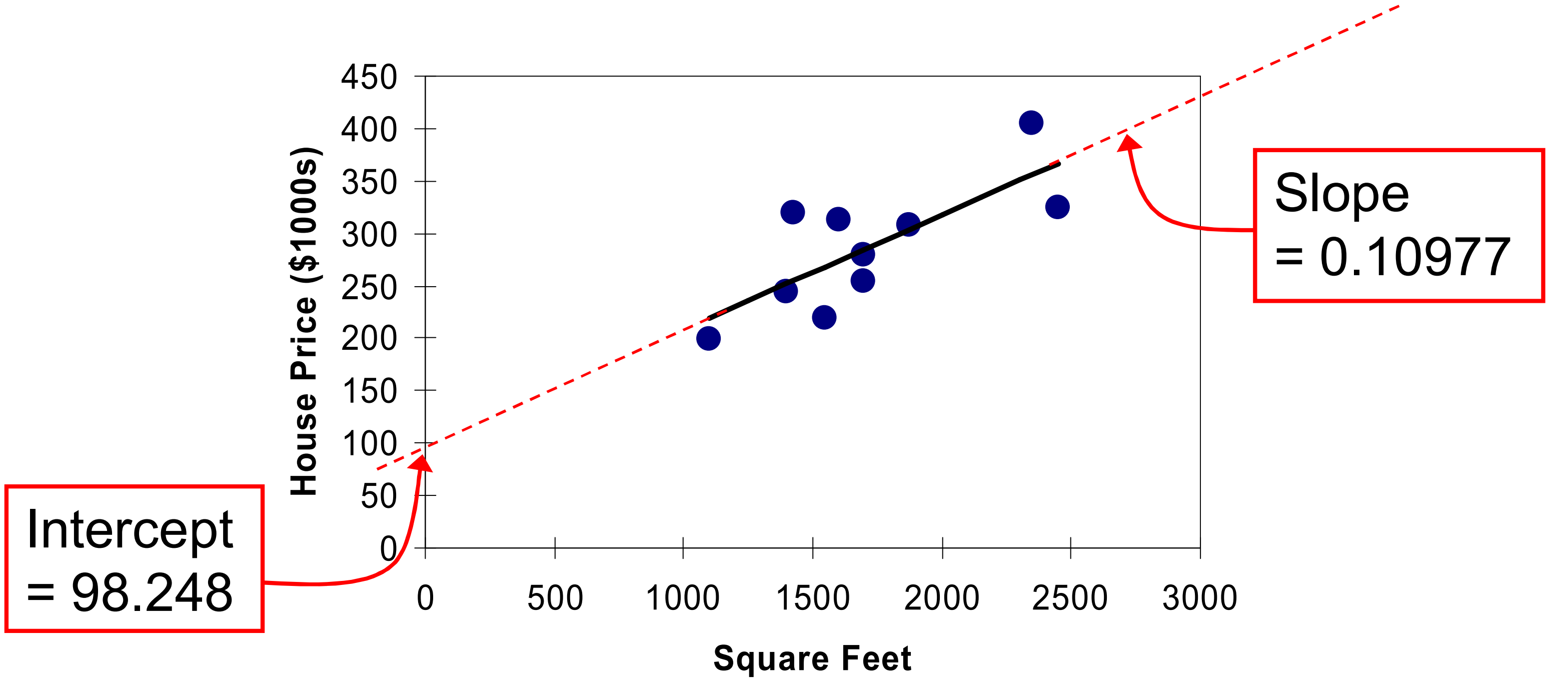
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580

Simple Linear Regression Analysis: an example

House price model: Scatter Plot and Prediction Line



$$\widehat{\text{house price}} = 98.24833 + 0.10977\ (\text{square feet})$$

Simple Linear Regression Analysis: an example

Predict the price for a house with 2000 square feet:

$$\widehat{\text{house price}} = 98.25 + 0.1098\,(\text{sq.ft.})$$

$$= 98.25 + 0.1098(2000)$$

$$= 317.85$$

The predicted price for a house with 2000 square feet is 317.85($1,000s) = $317,850

Simple Linear Regression Analysis: the total variation

- Total variation is made up of two parts:

$$SST = SSR + SSE$$

Total Sum of Squares	Regression Sum of Squares	Error Sum of Squares

$$SST = \sum (Y_i - \bar{Y})^2 \qquad SSR = \sum (\hat{Y}_i - \bar{Y})^2 \qquad SSE = \sum (Y_i - \hat{Y}_i)^2$$

where:

$\bar{Y}$ = Mean value of the dependent variable

Y_i = Observed value of the dependent variable

$\hat{Y}_i$ = Predicted value of Y for the given X_i value

Simple Linear Regression Analysis: the coefficient of determination

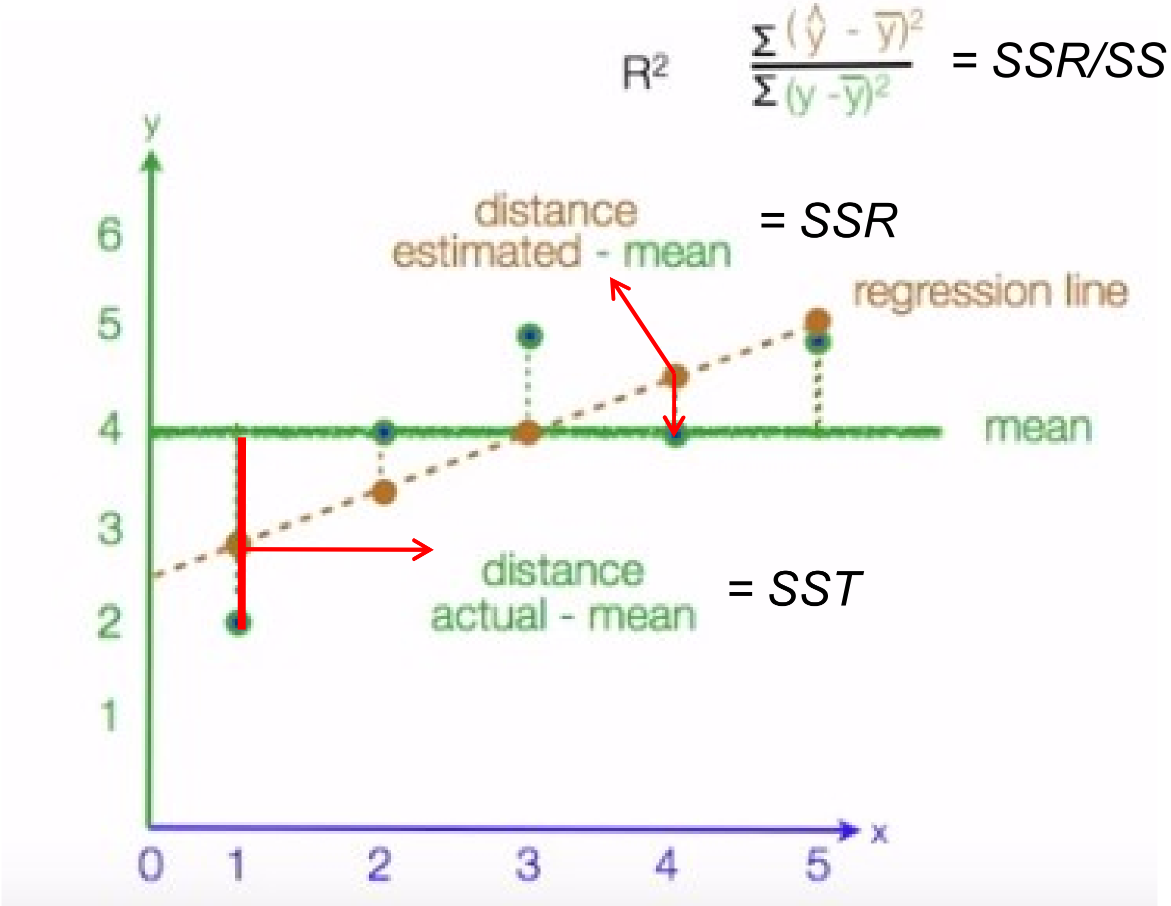
- The coefficient of determination is the portion of the total variation in the dependent variable that is explained by variation in the independent variable
- The coefficient of determination is also called r-squared and is denoted as r^2

$$\boxed{r^2 = \frac{SSR}{SST} = \frac{\text{regression } \textit{sum} \text{ of squares}}{\textit{total} \text{ sum of squares}}}$$

note: $\boxed{0 \leq r^2 \leq 1}$

Simple Linear Regression Analysis: the coefficient of determination

How well the regressed values estimated the real/actual values



= *SSR/SST*

= *SSR*

= *SST*

Simple Linear Regression Analysis: the coefficient of determination

Regression Statistics	
Multiple R	0.76211
R Square	0.58082
Adjusted R Square	0.52842
Standard Error	41.33032
Observations	10

$$r^2 = \frac{SSR}{SST} = \frac{18934.9348}{32600.5000} = 0.58082$$

58.08% of the variation in house prices is explained by variation in square feet

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18934.9348	18934.9348	11.0848	0.01039
Residual	8	13665.5652	1708.1957		
Total	9	32600.5000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	98.24833	58.03348	1.69296	0.12892	-35.57720	232.07386
Square Feet	0.10977	0.03297	3.32938	0.01039	0.03374	0.18580