10.2 - Regression (and Prediction Intervals)

EXAMPLE 1 Use our data from 10.1 and the above formulas to calculate the linear regression equation. Create a scatter plot and graph the regression line. See the text for other ways of calculating the regression equation.

x	1	2	3	4	5
y	2	4	5	8	8

$$\hat{y} = b_0 + b_1 x \qquad b_1 = r \cdot \frac{s_y}{s_x} \qquad b_0 = \bar{y} - b_1 \bar{x}$$

$\bar{x} = 3$ $\quad \bar{y} = 5.4$

$S_x = 1.5811$ $\quad S_y = 2.6077$

$r = 0.9701$

① $b_1 = 0.9701 \cdot \frac{2.6077}{1.5811}$

$b_1 \approx 1.6$

② $b_0 = 5.4 - 1.6(3)$

$b_0 = 0.6$

$\Rightarrow \hat{y} = 0.6 + 1.6x$ or $\hat{y} = 1.6x + 0.6$

TI 84
LinReg(ax+b)

NORMAL FLOAT AUTO REAL RADIAN MP

LinReg
y=ax+b
a=1.6
b=0.6
r²=0.9411764706
r=0.9701425001

$\hat{y} = 1.6x + 0.6$

Correlation and Regression

	Var1	Var2
1	1	2
2	2	4
3	3	5
4	4	8
5	5	8
6		
7		

Correlation r = 0.9701
r^2 = 0.9412

Regression: y = a + b x

a = 0.6000
b = 1.6000

EXAMPLE 2 **Predicting Temperature From Cricket Chirps**

The following data is the number of cricket chirps in 15 seconds, and the outside temperature at that time:

Cricket Chirps	20	16	20	18	17	16	15	17	15	16	15	17	16	17	14
Temperature (F)	89	72	93	84	81	75	70	82	69	83	80	83	81	84	76

a) Make a scatter plot to determine if there is a linear correlation, and to see if there are any outliers affecting our calculations.

Scatter Plot

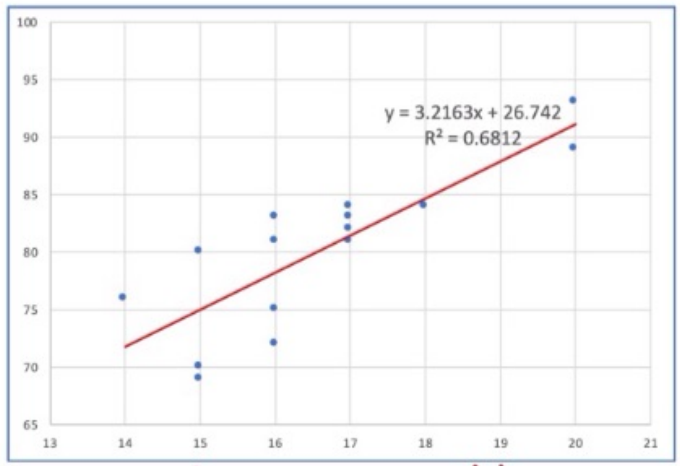
	x	y
1	20	89
2	16	72
3	20	93
4	18	84
5	17	81
6	16	75
7	15	70
8	17	82
9	15	69
10	16	83
11	15	80
12	17	83
13	16	81
14	17	84
15	14	76
16		
17		
18		

	x	y
min	14	69
max	20	93

Axes Values		
min	13	65
max	21	100
ticks	1	5

Set Axes

Add Trendline



No obvious outliers

b) Calculate the correlation coefficient and regression equation.

$r = 0.8254$ $\hat{y} = 3.216x + 26.742$

c) Explain the meaning of r^2 in context of the problem.

$r^2 = 0.6812$

About 68% of the variation in temperature can be explained by the variation in number of chirps.

Note: the number of chirps DOES NOT cause a change in temperature.

d) Use a t-test to determine if the correlation is statistically significant.

Correlation and Regression

	Var1	Var2
1	20	89
2	16	72
3	20	93
4	18	84
5	17	81
6	16	75
7	15	70
8	17	82
9	15	69
10	16	83
11	15	80
12	17	83
13	16	81
14	17	84
15	14	76

Correlation r =	0.8254
r^2 =	0.6812
Regression: y = a + b x	
a =	26.7420
b =	3.2163
degrees of freedom =	13
α =	0.05
critical r_{cr} =	± 0.514
t =	5.2707
P-value =	0.0002

$H_0: \rho = 0$

$H_1: \rho \neq 0$

method I : $P = 0.0002$

method II : $r_c = 0.514$

The linear correlation of $r = 0.8254$ is statistically significant.

e) For what values is the regression equation most reliable for?

The regression equation is most reliable in the range of x the independent variable.

$\Rightarrow 14 \leq x \leq 20$

Making predictions with values outside this interval shouldn't be trusted.

f) What is the best predicted temperature for 19 chirps per 15 seconds?

Using $\hat{y} = 3.216x + 26.742$ with $x = 19$

$\hat{y} = 3.216(19) + 26.742$

$\approx 87.8°$

g) What is the estimated number of chirps one would expect to record at a temperature of 80°F?

Here $\hat{y} = 80$ and solve for x

$80 = 3.216x + 26.742$

$\dfrac{80 - 26.742}{3.216} = x \Rightarrow x \approx 16.6$

We would expect 16 or 17 chirps every 15 seconds.

Correlation and Regression

	Var1	Var2
1	20	89
2	16	72
3	20	93
4	18	84
5	17	81
6	16	75
7	15	70
8	17	82
9	15	69
10	16	83
11	15	80
12	17	83
13	16	81
14	17	84
15	14	76
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		

Correlation r = 0.8254

r^2 = 0.6812

Regression: $y = a + b\,x$

a = 26.7420

b = 3.2163

degrees of freedom = 13

α = 0.1

critical r_{cr} = ± 0.4409

t = 5.2707

P-value = 0.0002

Predictions

standard error of est.: s_e = 3.9359

x = 19

y-hat = 87.8526

90%-Prediction Interval

80.2008 95.5043

90% Prediction Interval

$80.2 < \bar{y} < 95.5$

PREDICTION INTERVAL BY HAND

$$SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x^2(n-1)}}$$

$$PI: \hat{y} \pm t_{\alpha/2} \cdot SE_{\hat{y}} = \hat{y} \pm E$$

$$SE_{\hat{y}} = 3.9359\sqrt{1 + \frac{1}{15} + \frac{(19 - 16.6^2)}{1.7238^2(14)}}$$

$$\approx 4.3208$$

$s = 3.9359$
$n = 15$
$x_0 = 19$
$\bar{x} = 16.6$
$s_x = 1.7238$

$t_{\alpha/2} = 1.7709$

$\hat{y} = 87.85$

$$E = 1.7709(4.3208)$$
$$= 7.6517$$

$$PI = 87.85 \pm 7.6517$$

$$\boxed{(80.2, 95.5)}$$