

14.4 Correlation Coefficient

Best fit line

GOALS:

1. Associate a regression equation with the correlation coefficient, r .
2. The correlation coefficient, r , provides a measure of how well a straight line fits the data, or how strong a linear relationship exists between the two variables.
3. Compute r from the coefficient of correlation: $r = \sqrt{r^2}$
4. The value of r can range from: $-1 \leq r \leq 1$
5. When $r < 0$, there is a negative correlation between the variables.
6. When $r > 0$, there is a positive correlation between the variables.

Study Ch. 14.4, #118-135, 145, 147

[#109-117, 122, 123, 125]

Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)

[Class Notes](#) [Homework](#)

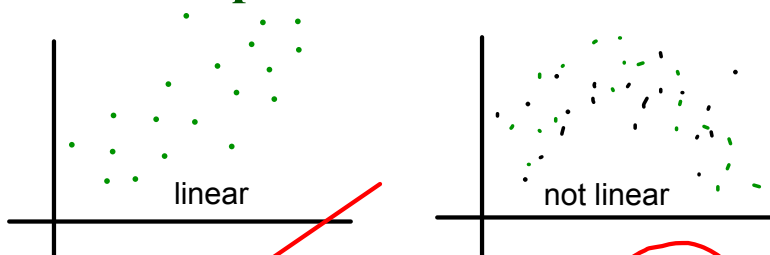


14.4 Correlation Coefficient

Correlation Coefficient, r

Pearson Product Moment Correlation Coefficient, r

Measure of the strength of the linear
relationship between two variables.



Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)

[Class Notes](#) [Homework](#)



14.4 Correlation Coefficient

Correlation Coefficient, r**Pearson Product Moment
Correlation Coefficient, r**

$$r = \frac{\frac{1}{n-1} \sum (x - \bar{x})(y - \bar{y})}{s_x s_y}$$

$$= \frac{\sum xy - (\sum x \sum y) / n}{\sqrt{[\sum x^2 - (\sum x)^2 / n][\sum y^2 - (\sum y)^2 / n]}}$$

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n-1}}$$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)
[Class Notes](#)
[Homework](#)

14.4 Correlation Coefficient

Correlation Coefficient, r

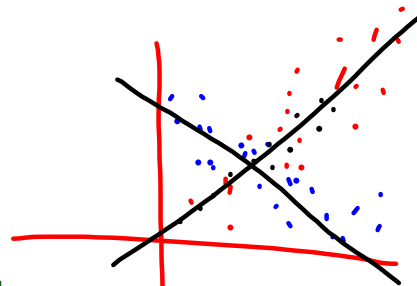
$$-1 \leq r \leq 1$$

**$r > 0$ regression line has positive slope
variables are positively linearly correlated**

**$r < 0$ regression line has negative slope
variables are negatively linearly correlated**

r near +1 or -1 indicates strong linear relationship

r near 0 indicates weak linear relationship



Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)
[Class Notes](#)
[Homework](#)

14.4 Correlation Coefficient

Find on Calculator:

STAT / TESTS / LinRegTTest

xlist: L1

Ylist: L2

2 tailed, left tailed, right tailed

calculate

Output:

$y=ax+b$

2 tailed, rt tailed, lf tailed

t=

p=

df=

a =

b =

s =

$r^2 = .$

r =

Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)

[Class Notes](#)

[Homework](#)

^

14.3 Coefficient of Determination

[geogebra](#)

A random sample of custom homes for sale include the following information: a size of x hundred sq. ft selling at \$ y thousand.

Predict the price of a home that is 3100 ft.

Is the correlation between size and price weak or strong?

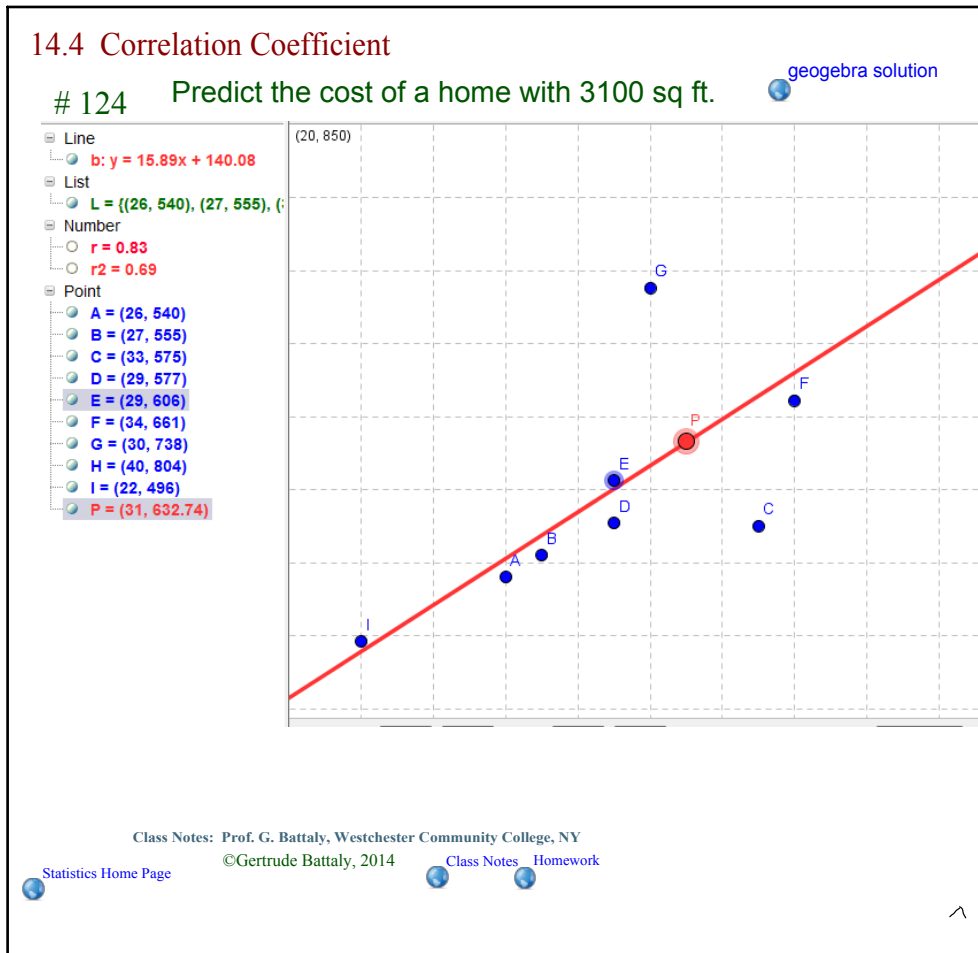
<u>x</u>	<u>y</u>
26	540
27	555
33	575
29	577
29	606
34	661
30	738
40	804
22	496

Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)

[Class Notes](#)

[Homework](#)



14.3 Coefficient of Determination

A random sample of custom homes for sale include the following information: a size of x hundred sq. ft selling at \$ y thousand.
Predict the price of a home that is 3100 ft.
Is the correlation between size and price weak or strong?

geogebra

x	y
26	540
27	555
33	575
29	577
29	606
34	661
30	738
40	804
22	496

$r = 0.829$

r is closer to 1 than to 0. This suggests a **strong positive linear relationship** between size and price

The t-test for the population correlation coefficient $H_0: \rho = 0$, $H_a: \rho \neq 0$ indicates the p value:
 $p = .0058$
 This shows **very strong evidence** that a linear relationship exists and that the regression line is a good predictor.

TEXAS INSTRUMENTS TI-83 Plus

```
LinRegTTest
y=a+bx
b≠0 and ρ≠0
t=3.917864996
p=.0057648354
df=7
↓a=140.0833333
```

TEXAS INSTRUMENTS TI-83 Plus

```
LinRegTTest
y=a+bx
b≠0 and ρ≠0
↑b=15.89351852
s=59.6207536
r²=.6867962161
r=.8287316912
```

Class Notes: Prof. G. Battaly, Westchester Community College, NY
[Statistics Home Page](#) [Class Notes](#) [Homework](#)

14.3 Coefficient of Determination

Hours of Study vs. Test Scores: How are the hours of study and test scores related?

hrs test

x	y
10	92
15	81
12	84
20	74
8	85
16	80
14	84
22	80

$$r = -0.775$$

r is closer to -1 than to 0 .
This **suggests a negative linear relationship** between hours of study and test grades

The t-test for the population correlation coefficient

$$H_0: \rho = 0, \quad H_a: \rho \neq 0$$

indicates the p value:

$$p = .024$$

This shows **strong evidence** that a linear relationship exists and that the regression line is a good predictor .

Use LinRegTTest to get r

```

TEXAS INSTRUMENTS TI-83 Plus
LinRegTTest
y=a+bx
b≠0 and ρ≠0
t=-3.00293955
p=.0239173048
df=6
↓a=94.86698337
  
```

```

TEXAS INSTRUMENTS TI-83 Plus
LinRegTTest
y=a+bx
b≠0 and ρ≠0
↑b=-.8456057007
s=3.538164283
r²=.6004700056
r=-.7748999971
  
```

Class Notes: Prof. G. Battaly, Westchester Community College, NY

[Statistics Home Page](#)

[Class Notes](#) [Homework](#)