15.4 Inferences in Correlation

GOALS:
1. Recognize the assumptions for regression inferences
2. Understand that the computed statistic, the correlation coefficient, r, is an estimate of the population correlation coefficient, ρ
3. Perform a t-test to determine if two variables are linearly correlated: if $\rho \neq 0$, $\rho > 0$, or $\rho < 0$

Study Ch. 15.4, # 99, 101, 109-113, 114

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

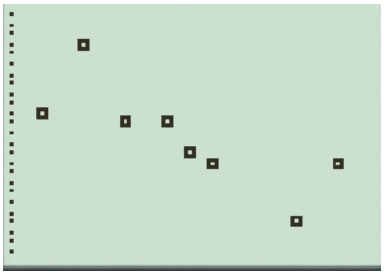
Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Study time and test score:

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

Is there a correlation between study time and test grades?



a=94.867 and b= -0.846
for regression equation,
r= -0.7749, r^2 =0.6005

BUT, What confidence do we have?

Is this **sample r** a good estimate of the population parameter, **ρ** ?

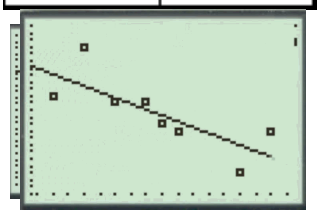
Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Study time and test score:

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



Statistics vs Parameters

$y = a + b\,x$

a=94.867 and b= -0.846

for regression equation,

r= -0.7749, r^2 =0.6005

estimating β_0 β_1 population parameters

$\hat{y} = 94.867 - 0.846\,x$

r= -0.7749 ρ rho

r^2 =0.6005 ρ^2

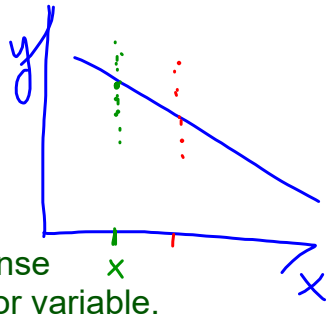
β_1 β_0

15.4 Inferences in Correlation

Assumptions for **Regression Inferences**:

1. **Population regression line exists**:
for each predictor variable, x, there is a population response variable

$y = \beta_0 + \beta_1 x$



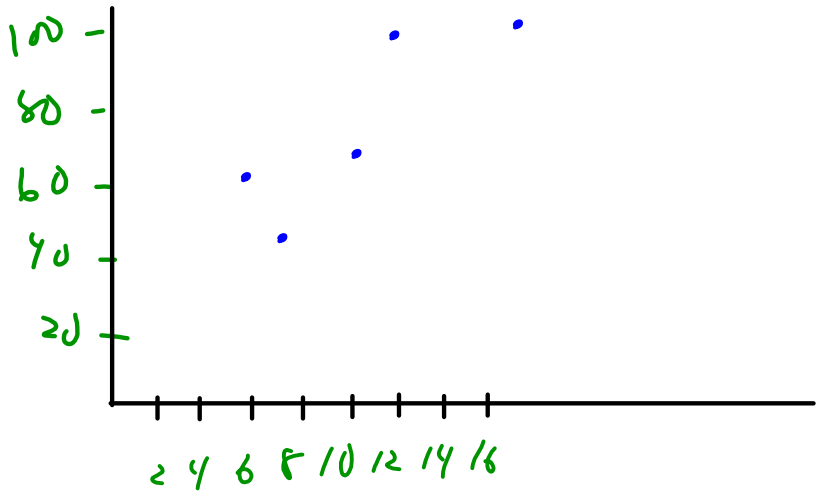
2. **Equal standard deviations**:
the conditional standard deviations of the response variable is the same for all values of the predictor variable.

3. **Normal populations**:
for each value of the predictor variable, the conditional distribution of the response variable is a normal distribution.

4. **Independent observations**:
observations of the response variable are independent of each other.

15.4 Inferences in Correlation

3. **Normal populations**:
for each value of the predictor variable, the conditional distribution of the response variable is a normal distribution.



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

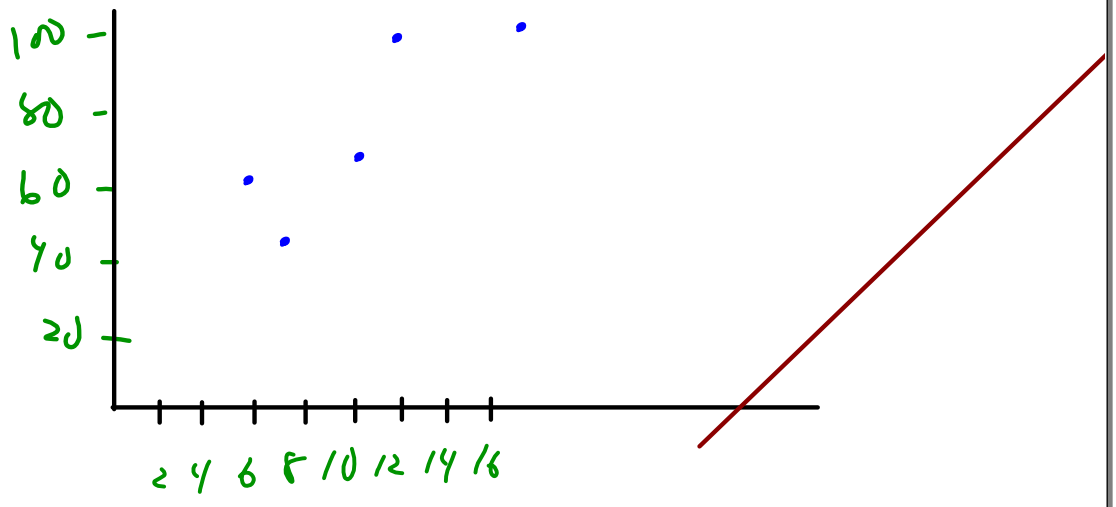
Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

3. **Normal populations**:
for each value of the predictor variable, the conditional distribution of the response variable is a normal distribution.

Points are Data from a **SAMPLE**

$\therefore$ **expect variation**



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

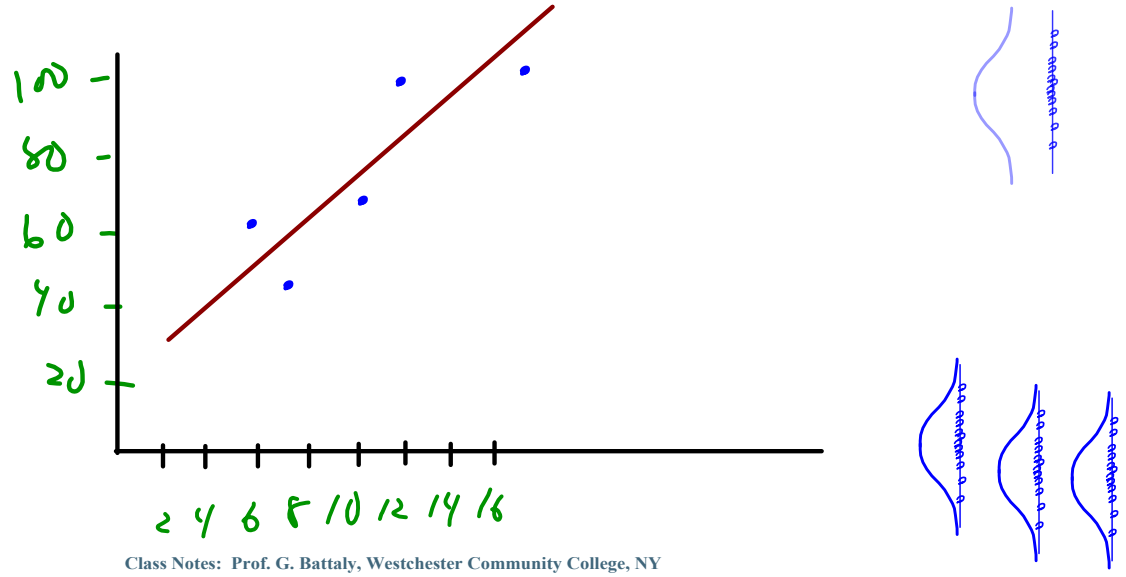
Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

3. **Normal populations**:
for each value of the predictor variable, the conditional distribution of the response variable is a normal distribution.

Points are Data from a **SAMPLE**

∴ **expect variation**



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

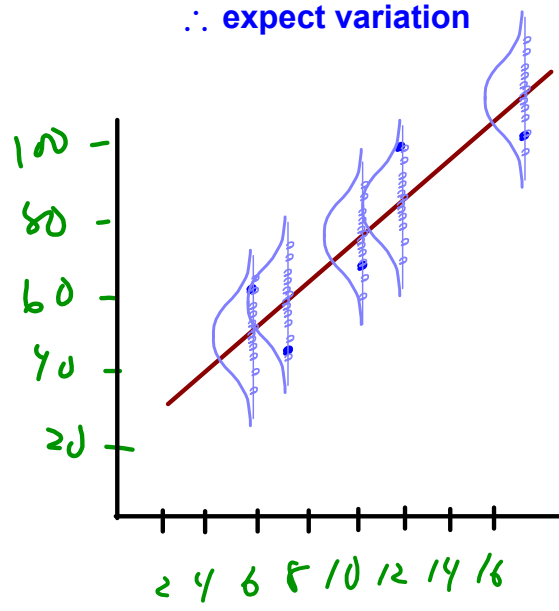
Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

3. **Normal populations**:
for each value of the predictor variable, the conditional distribution of the response variable is a normal distribution.

Points are Data from a **SAMPLE**

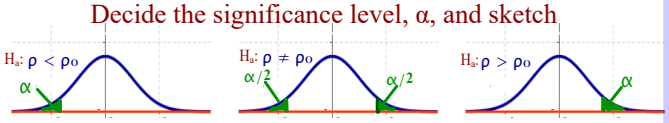
∴ **expect variation**



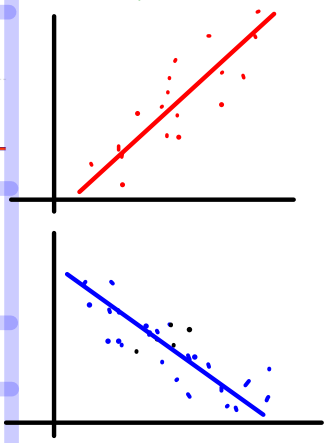
Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Test	Correlation t - Test
1.	State the Null and Alternative Hypotheses: H_0, H_a H_0: $\rho = 0$ H_a: $\rho < 0$ or $\rho \neq 0$ or $\rho > 0$
2.	Decide the significance level, α, and sketch 
3.	Compute the test statistic: t df = n - 2 $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$
4.	Find the P-value
5.	Decision: Rej. H_0 if $P \leq \alpha$
6.	Interpret results

$-1 \leq \rho \leq 1$



$\frac{a}{\frac{1}{b}} = a \cdot b$

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

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Correlation t-Test

(assumptions: population regr. line, =stdev,
normal distribution, independent observations)

STAT / TESTS

LinRegT-Test

Xlist: L1 (list where entered x values)

Ylist: L2 (list where entered y values)

β and ρ: $\neq 0$ < 0 > 0

Calculate

Results include:

test t, p, df, a and b for regression equation,

s, r, r^2

s is standard error of the estimate, the average difference between observed value and the predicted value:

$s_e = \sqrt{\dfrac{SSE}{n-2}}$ ⟷ $\Sigma(y-\hat{y})^2$

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

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Fetal Development:
Age in weeks & Growth - Length of Crown-Rump in mm.

Age (weeks	Length (mm)
10	66
10	66
13	108
13	106
18	161
19	166
19	177
23	228
25	235
28	280

Is there a correlation between age and crown-Rump Length?

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

Statistics Home Page Class Notes Homework

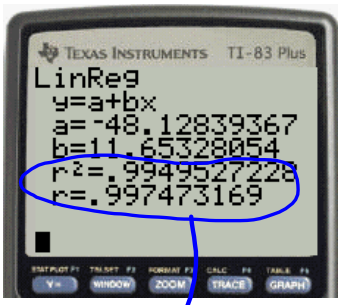
15.4 Inferences in Correlation

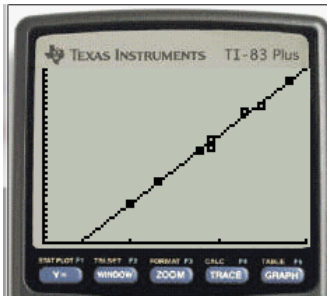
Fetal Development:
Age in weeks & Growth - Length of Crown-Rump in mm.

Is there a correlation between age and crown-Rump Length?

STAT/CALC/#8 LinReg L1, L2, Y1

Age (weeks	Length (mm)
10	66
10	66
13	108
13	106
18	161
19	166
19	177
23	228
25	235
28	280





From Ch 14: Easy to conclude that this is a strong positive linear relationship.

solution w/o confidence

BUT, What confidence do we have?

Is this **sample r** a good estimate of the population parameter, **ρ** ?

Note that r alone is not always easily interpreted. With only 2 points, would get perfect correlation. Larger samples generally result in lower r. If r = 0.4 or - 0.35, should we decide that not linearly correlated? It doesn't seem likely that it is, but maybe it's just highly variable. So --- need a confidence level or significance level.

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

Statistics Home Page Class Notes Homework

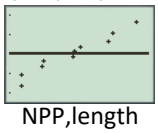
15.4 Inferences in Correlation

Fetal Development:
Age in weeks & Growth - Length of Crown-Rump in mm.

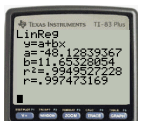
Age (weeks	Length (mm)
10	66
10	66
13	108
13	106
18	161
19	166
19	177
23	228
25	235
28	280

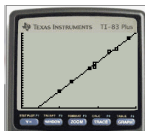
At the 1% significance level, do the data provide sufficient evidence to conclude that age and crown-rump length are linearly correlated?
Assume the popul. regr. line exists, = stdev,
~ n.d., independent observations.

Correlation t-Test
(assumptions: population regr. line, =stdev, normal distribution, independ
STAT / TESTS
LinRegT-Test
Xlist: L1 (list where entered x value
Ylist: L2 (list where entered y value
β and ρ: ≠0 <0 >0
Calculate



NPP,length





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

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

Fetal Development:
Age in weeks & Growth - Length of Crown-Rump in mm.

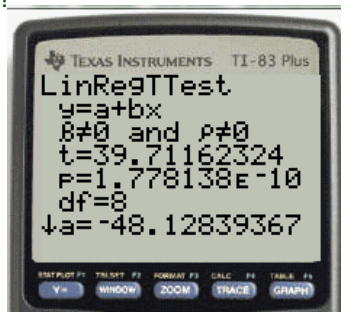
Age (weeks	Length (mm)
10	66
10	66
13	108
13	106
18	161
19	166
19	177
23	228
25	235
28	280

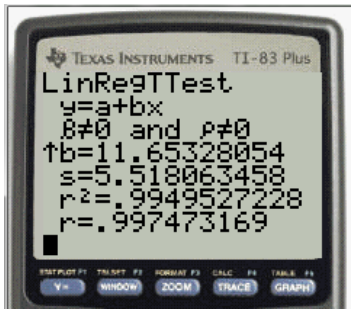
At the 1% significance level, do the data provide sufficient evidence to conclude that age and crown-rump length are linearly correlated?
Assume the popul. regr. line exists, = stdev,
~ n.d., independent observations.

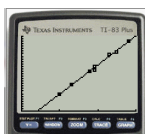
Correlation t-Test
(assumptions: population regr. line, =stdev, normal distribution, independ
STAT / TESTS
LinRegT-Test
Xlist: L1 (list where entered x value
Ylist: L2 (list where entered y value
β and ρ: ≠0 <0 >0
Calculate

1. Assump met. H_0: $\rho = 0$ H_a: $\rho \neq 0$
2. $\alpha = 0.01$ 2-tailed

3. $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{0.9975}{\sqrt{\dfrac{1-0.9950}{8}}} = 39.712$
4. $p = 1.778\,(10^{-10}) = 0.000$
5. $p = 0.0000 < 0.10 = \alpha$ ∴ rej. H_0
6. Conclude: Age and crown-rump length are linearly correlated.







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

Statistics Home Page Class Notes Homework

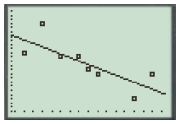
15.4 Inferences in Correlation

From HW: Study time and test score:
Assume regression assumptions are met.

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

At the 1% significance level, do the data provide sufficient evidence to conclude that a negative linear correlation exists between study time and test score for beginning calculus students?

Correlation t-Test
(assumptions: population regr. line, =stdev, normal distribution, independent observations)
STAT / TESTS
LinRegT-Test
Xlist: L1 (list where entered x values)
Ylist: L2 (list where entered y values)
β and ρ: ≠0 **<0** >0
Calculate



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

Statistics Home Page Class Notes Homework

Concl. that study time is NOT neg. correlation to test grade. At $\alpha = 0.01$

15.4 Inferences in Correlation

Study time and test score:
Assume regression assumptions are met.

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

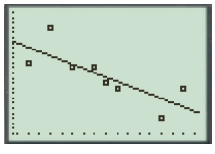
At the 1% significance level, do the data provide sufficient evidence to conclude that a negative linear correlation exists between study time and test score for beginning calculus students?

Correlation t-Test
(assumptions: population regr. line, =stdev, normal distribution, independent observations)
STAT / TESTS
LinRegT-Test
Xlist: L1 (list where entered x values)
Ylist: L2 (list where entered y values)
β and ρ: ≠0 **<0** >0
Calculate

Results include:
test t=-3.00294, p=0.01196, df=6, a=94.867 and b= -0.846 for regression equation,
s= 3.538, **r=-0.7749, r^2 =0.6005**

1. Assump met. H_0: ρ = 0 H_a: ρ < 0
2. α = 0.01 left-tailed
3. $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.7749}{\sqrt{\frac{1-0.6005}{6}}} = -3.003$
4. p = 0.01196
5. p = 0.01196 > 0.01 = α ∴ do NOT rej. H_0
6. Conclude: Not sufficient evidence to conclude that a neg. linear corr exists at 1% sign. level.

p=0.012 > 0.01 = α do NOT reject

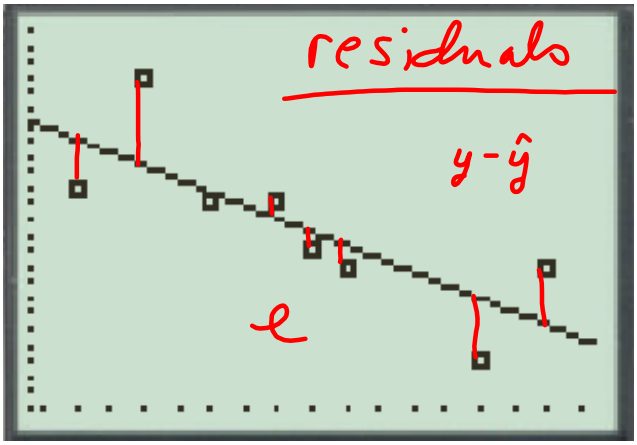


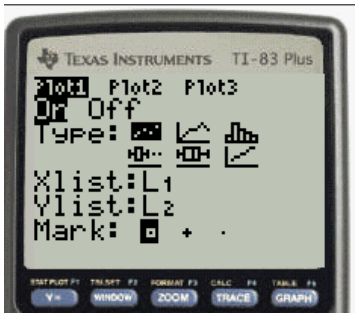
Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

15.4 Inferences in Correlation

How can we decide if the assumptions of regression are met?



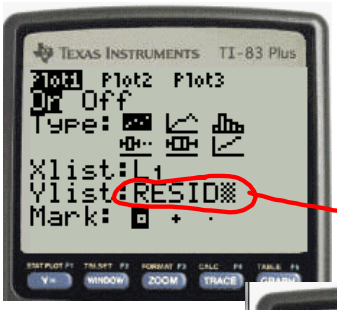


Look at residuals.

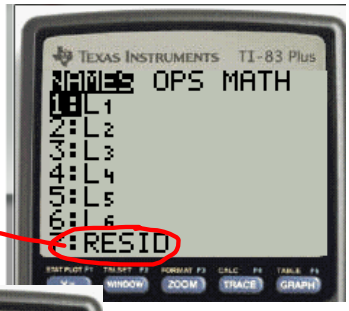
15.4 Inferences in Correlation

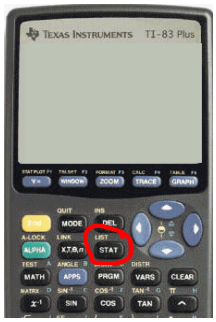
How can we decide if the assumptions of regression are met?

Look at residuals.

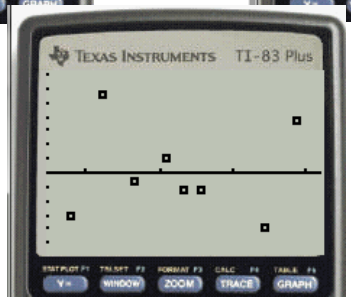


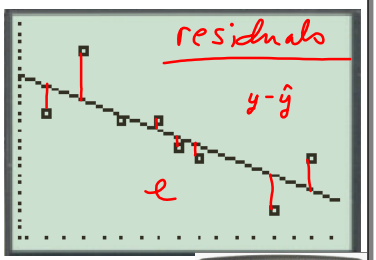
use LIST

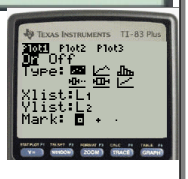




then Zoom/ Stat





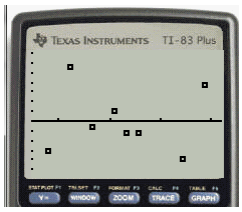


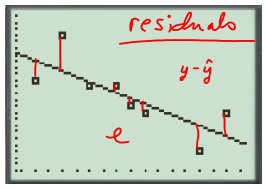
15.4 Inferences in Correlation

How can we decide if the assumptions of regression are met?

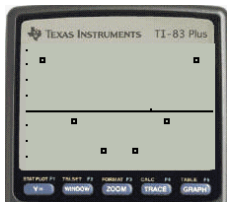
Look at residuals.

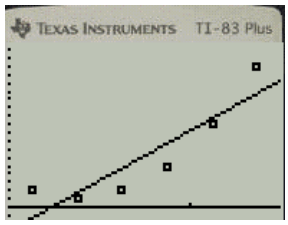
If random above and below axis, then no violations.



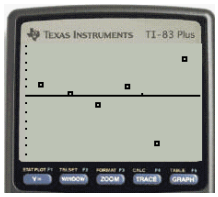


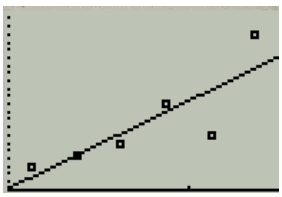
If not random above and below axis, then not linear.





If farther from the axis at one end than the other, may not have = stdev





Residuals: more examples

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

Statistics Home Page Class Notes Homework

The systolic blood pressure was measured for 30 people of different ages.

Age	Systolic BP
39	144
59	140
45	138
47	145
65	162
46	142
67	170
42	124
67	158
56	154
64	162
56	150

Age and Blood Pressure:
Assume regression assumptions are met.

At the 1% significance level, do the data provide sufficient evidence to conclude that a positive linear correlation exists between age and systolic blood pressure?

a) test
b) look at residuals, and interpret, re: assumptions

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

Statistics Home Page Class Notes Homework

The systolic blood pressure was measured for 30 people of different ages.

Age	Systolic BP
39	144
59	140
45	138
47	145
65	162
46	142
67	170
42	124
67	158
56	154
64	162
56	150

Age and Blood Pressure:
Assume regression assumptions are met.
At the 1% significance level, do the data provide sufficient evidence to conclude that a positive linear correlation exists between age and systolic blood pressure?
a) test
b) look at residuals, and interpret, re: assumptions

$H_0: \rho = 0 \quad H_a: \rho > 0$

$\alpha = 0.01$

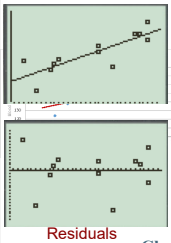
$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.8237}{\sqrt{\frac{1-0.6785}{10}}}$

$t = 4.594$

$p = 0.000495$

$p = 0.000495 < 0.01 = \alpha \therefore$ rej. H_0.

LinRegTTest
y=a+bx
β>0 and ρ>0
t=4.59400907
p=4.9457735E-4
df=10
↓a=93.03043007
↑b=1.030068666
s=7.613306025
r²=.6785074447
r=.8237156334



Residuals

Conclude: The age and BP are positively linearly correlated.

