§10.1 Correlation
§10.2 Regression

Project Due Thursday -
2-SampTTest (Interval) Graph, P-value, Conclusion
2 Prop Z Test (Interval) Graph, p-value, Conclusion

Final Thurs. 12/20 at 7 AM
Practice Final Due 12/13 thurs.

§10.1 Correlation is a measure of How close a set of Bivariate Data fits a Line.

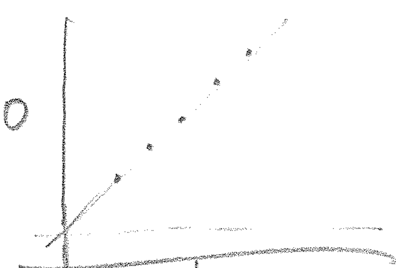
$$r = \sqrt{\frac{(\text{observed date} - \text{predicted})^2}{\text{Predicted}^2}} = \sqrt{\frac{(y-\hat{y})^2}{(\bar{y}-\hat{y})^2}}$$

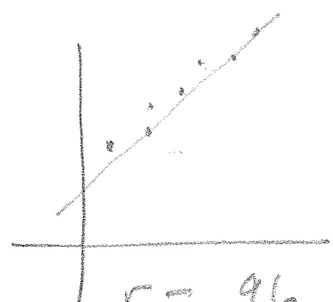
↑
From Calculator

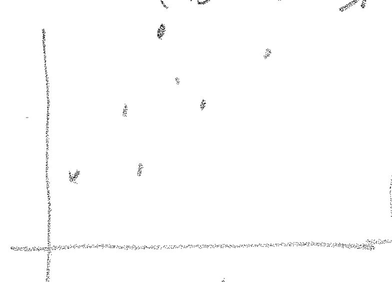
$-1 \le r \le 1$

(x,y) = point ← Bi Variate → Two Variable

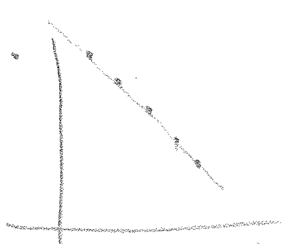
m ≥ 0


r = 1


r = .96


r = .56


r = 0


r = −1


r = −.95

m < 0

Population Parameters	Sample Statistics
μ	$\bar{x}$
p	$\hat{p}$
σ	s
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$
rho $- \rho$	r
β_0, β_1	b_0, b_1

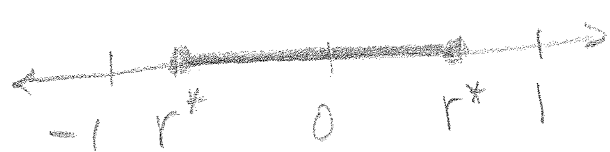
$H_0: \rho = 0$ → There is Not a significant Linear correlation
Do Not use Regression Line $\hat{y} = ax+b$
use $\bar{y}$ as best predicted Value

$H_1: \rho \neq 0$ → There is a significant Linear relationship
We can use $\hat{y} = ax+b$

If $|r| > r^*$

Reject H_0
Accept H_1

If $|r| < r^*$

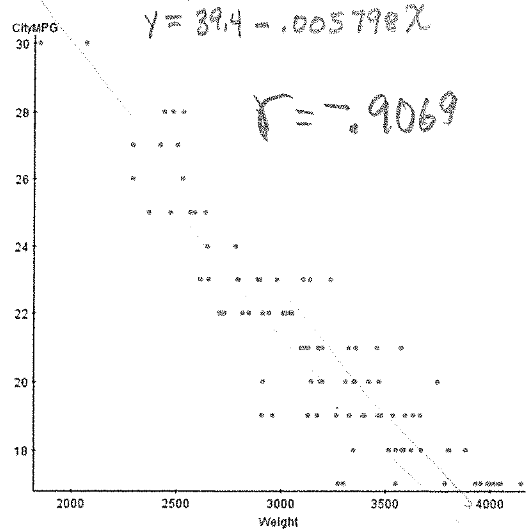


fail to Reject H_0
Do Not use Regression Line.

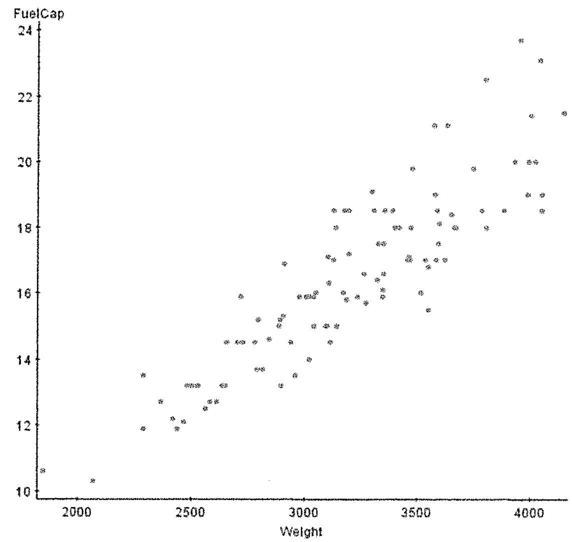
MPG VS. Weight
Strong Negative Correlation
CityMPG = 39.411663 - 0.00579757 Weight   Good prediction
R (correlation coefficient) = -0.9069 $= r$

$y = 39.4 - .005798x$

$r = -.9069$

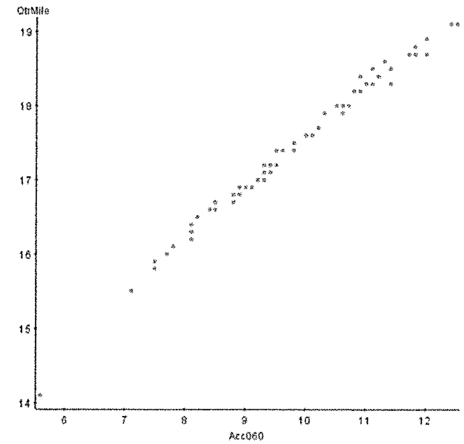


Weight Vs. Fuel Capacity
Strong positive correlation
R (correlation coefficient) = 0.894



Time to accellerate 0 to 60 and Quarter Mile Time

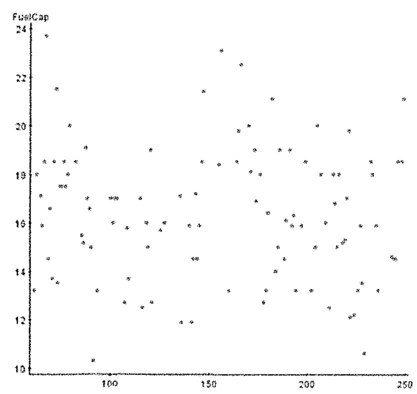
R (correlation coefficient) = 0.9942
Use regression equation
Very strong correlation regression equation will make reliable predictions



No correlation

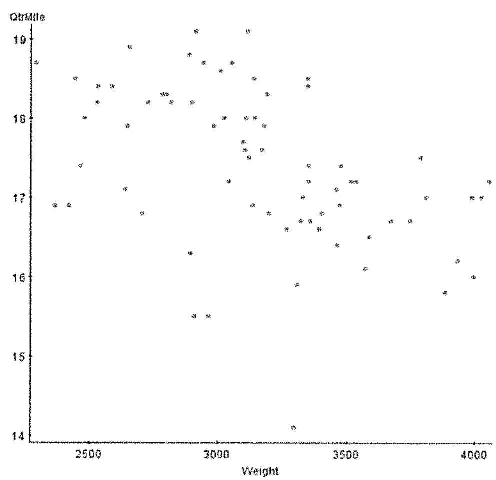
Best predicted values is mean fuel capacity around 16 gallons
R (correlation coefficient) = -0.0815 is close to Zero

Regression equations should not be used to make predictions.



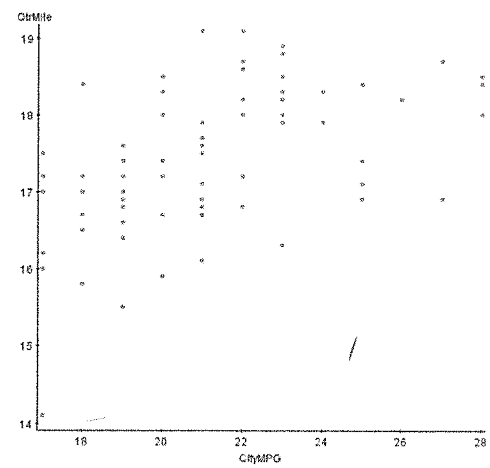
Weak Negative

Correlations is significant by regression equation will not give very reliable answers
R (correlation coefficient) = -0.4502



Weak Positive

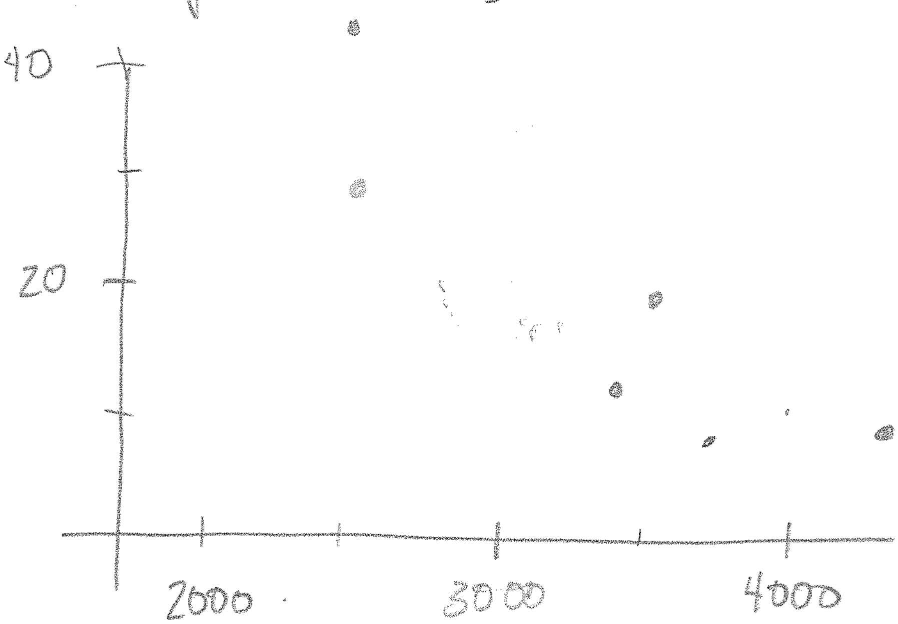
R (correlation coefficient) = 0.5101



Example Car Weights vs. MPG

$(x, y) = $ (weight, miles per Gallon)

$r \approx -.8$



Ex Eleven Randomly Selected Cars

a) 10.1 → What is the Correlation Coefficient.

	Weight	CMPG	
	3350	20	
	3325	19	
	2980	23	
	3345	18	
	3140	20	
Max	3990	17	Min
Min →	2365	25	Max
	3455	21	
	2790	23	
	2650	23	
	2645	24	

r = LinReg TTest

$r = -.9189$ $H_1: \rho \neq 0$

Support there is a linear Correlation

CV r^* = Table A-6

$r^* = .602 < |-.9189|$

We can use Regression Line

b) Find the Regression Line (10.2)

$\hat{y} = a + bx$

$\hat{y} = 37.1186 + -.00515x$

MPG = 37.1186 − .00515(weight)

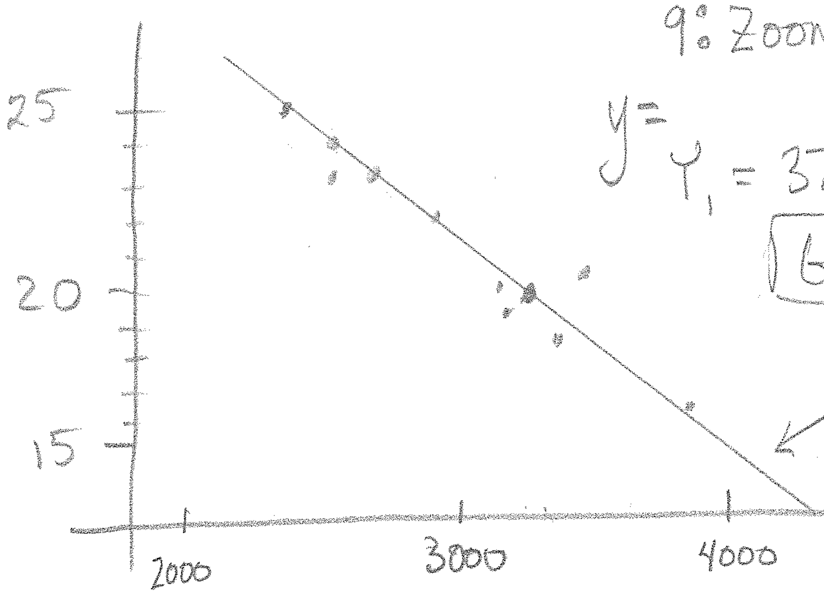
c) Predict the gas milage of a 3000 lb car

10.2 $y = 37.1186 - .00515(3000) = 21.7$ mpg

d) Scatter Plot (Plot the Points)

d) Scatter Plot $\boxed{\text{ZOOM}}$ Middle top Button

9: Zoom Stat



y= $Y_1 = 37.1186 - .00515X$

$\boxed{\text{Graph}}$ top Rt

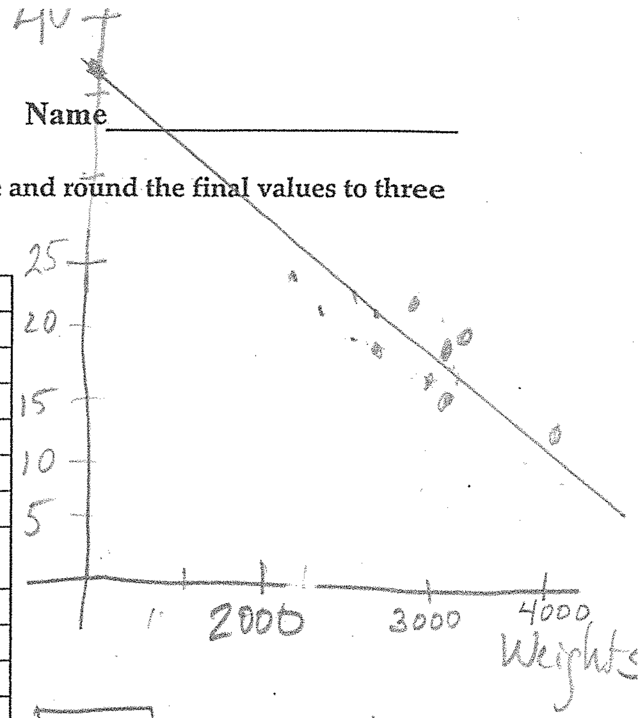
line that fits Data the Best

Lecture Chapter 10 **Math 15** **Jones** **Name**__________________

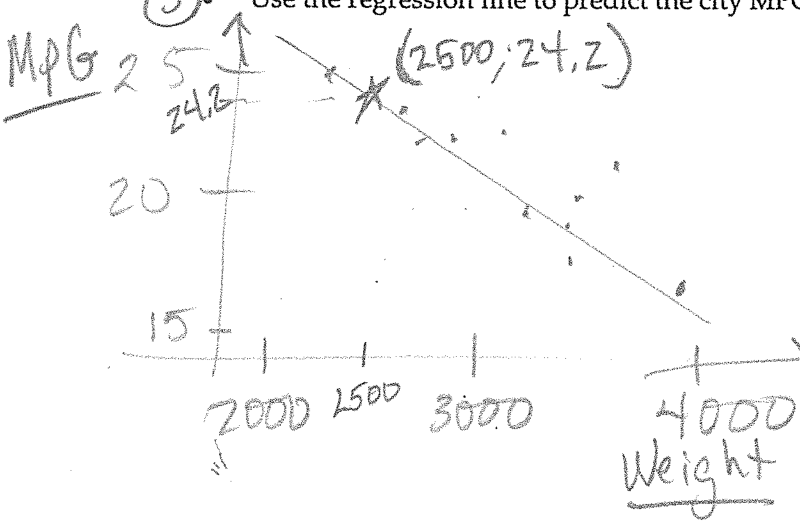
1 Use the given data to find the equation of the regression line and round the final values to three significant digits, if necessary.

Eleven randomly selected cars were chosen from the cars data.

Model		Weight	CityMPG	Type
Buick	Century	3350	20	family
Buick	Regal	3325	19	family
Subaru	Legacy	2980	23	family
Audi	A4	3345	18	upscale
Mercury	Cougar	3140	20	sports
Oldsmobile	Aurora	3990	17	large
Mazda MX-5	Miata	2365	25	sports
Dodge	Intrepid	3455	21	large
BMW	318Ti	2790	23	sports
Kia	Sephia	2650	23	small
Mazda	Protege	2645	24	small



1. Make a scatter plot of the data. STATPLOT → $\boxed{\text{ZOOM}}$ 9: ZoomStat
2. Graph the regression line one the same graph. LinReg TTest
3. Find the value of the linear correlation coefficient r.
4. Use Table A-6 to determine if there is enough evidence to support the claim that there is a linear correlation between the two variables. @ .05
5. Use the regression line to predict the city MPG of a car that weighs 2500 lb.



(2) LinReg TTest

$y = a + bx$

$a = 37.118$

$b = -.00515$

$r^2 = .8444$

(3) $\boxed{r = -.9189}$

(2) $\boxed{y = 37.118 - .00515x}$

(4) A-6 Get CV: $r^* = .602$

$n = 11$ $|r| = .9189$ Since $|r| > r^*$ there is a Significant Linear correlation.

(5) $y = 37.118 - .00515(2500) = 24.2$ mpg

So (2500, 24.2) is a point on the regression line

§10.2 Regression Equation

the line that fits the Data Values
Paired Data the Best

$\hat{y} = a + bx$

$r^2 = .844$

$a = 37.118$

$r = -.919$

$b = -.00515$

$y = mx + b$

$\hat{y} = 37.118 + (-.00515)x$

Use to make prediction when weight = 2500 lb = X

$\hat{y}(2500) = 37.118 - .00515(2500)$

$\hat{y}(2500) = 24.243$ mpg

We expect a 2500 lb car to get 24.2 mpg

What is the slope = $b = -.00515$ mpg/lb $= \frac{\text{rise}}{\text{run}} = \frac{\Delta y}{\Delta x}$

Unit for slope = $\frac{\text{units y-axis}}{\text{units x-axis}}$

* Interpret slope

For each pound of increase in weight of car the mpg decreases by .00515 mpg.

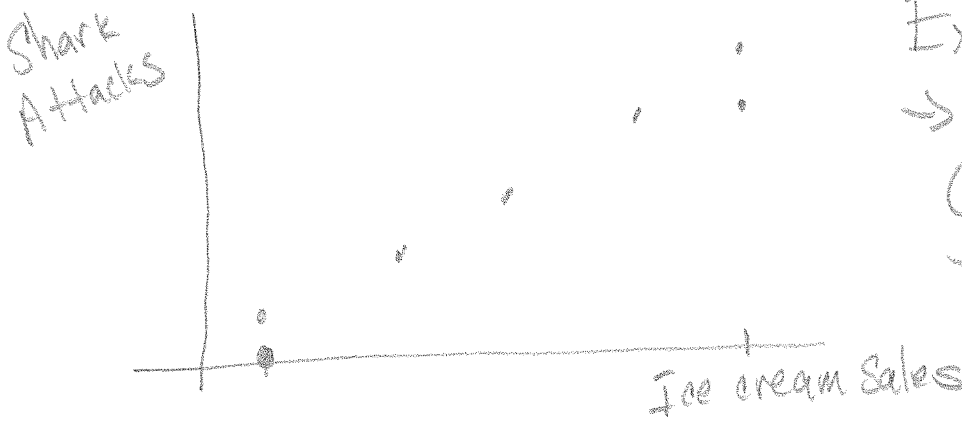
* Note card

for each x-units of increase in x variable the y-variable dec/inc ($b-$ / $b+$) by slope y-units.

Correlation Warnings

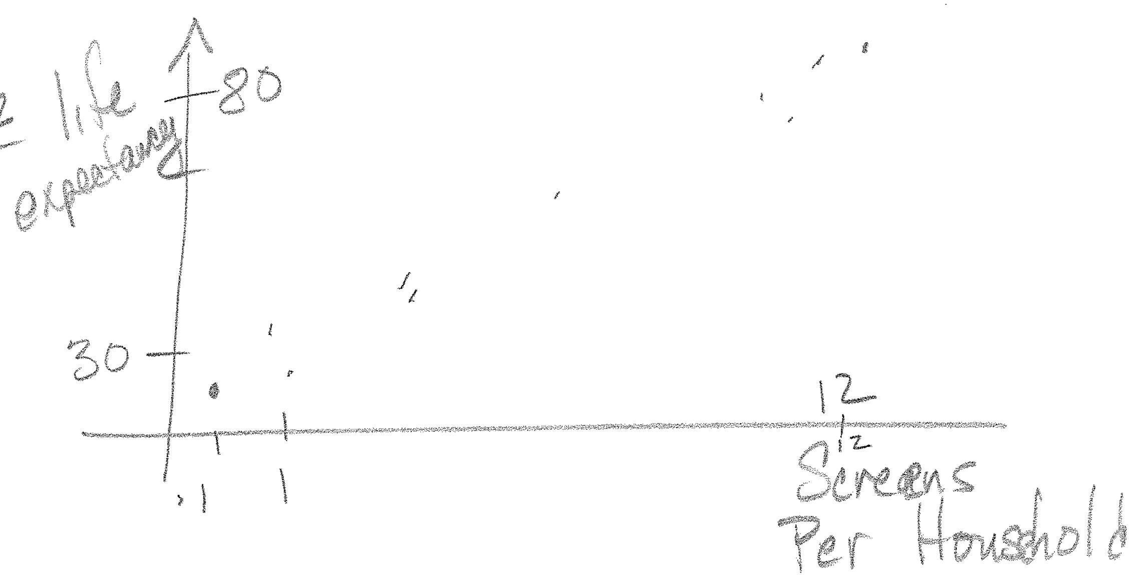
When **NOT** to use regression line

① Correlation Does Not imply Cause



Exreous Variable
→ Weather is Causeng more people to buy ice cream & go swim.

Ex2



Money → Causes screens & Healthcare to increase

3 Describe the error in the stated conclusion.

3) Given: There is a significant linear correlation between the number of homicides in a town and the number of movie theaters in a town.

Conclusion: Building more movie theaters will cause the homicide rate to rise.

No! Correlation does Not imply cause.
the extraneous variable of population
is causing both to increase.

4 Use the given data to find the equation of the regression line

Managers rate employees according to job performance and attitude. The results for several randomly selected employees are given below.

Performance	59	63	65	69	58	77	76	69	70	64
Attitude	72	67	78	82	75	87	92	83	87	78

- Make a scatter plot of the data.
- Graph the regression line one the same graph.
- Find the value of the linear correlation coefficient r.
- Use Table A-6 to determine if there is enough evidence to support the claim that there is a linear correlation between the two variables.
- Predict an employees attitude if their performance is 60.

Correlation Warning #2

Don't extrapolate → Go beyound given data

Predict Gas milage of 8000lb car

$y = 37.12 - .00515(5000)$

$y = 11.3$ OK

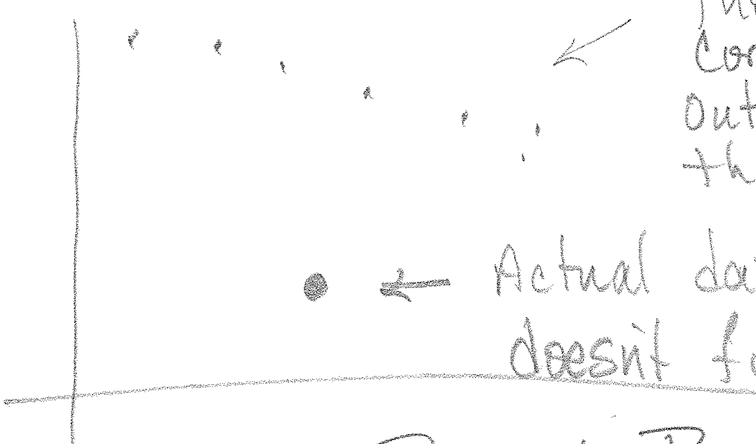
$y(8000) = -4.08$ mpg

Too far from given Data

Warning #3 catch

Outliers and Non linear Data with a Scatter plot.

Data is exponential
Should Not use linear Correlation



There is a linear Correlation but the Outlier makes it look like there isn't

Actual data Value that doesn't fit pattern

Best: Report Results with and without data Value

Warning #4

Outliers have a large effect on r

No linear Correlation but outlier makes it look like there is

Warning #3

Always Make Scatter Plot

r only Measures Linear Correlation

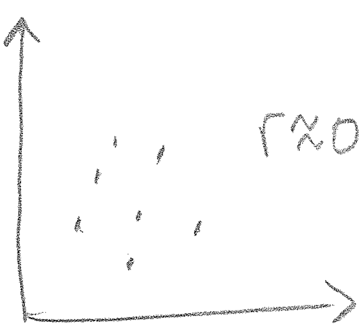
There may be a strong relationship between x and y that is Not linear.

$r \approx 0$
Not a linear correlation but there is a Quadratic Correlation

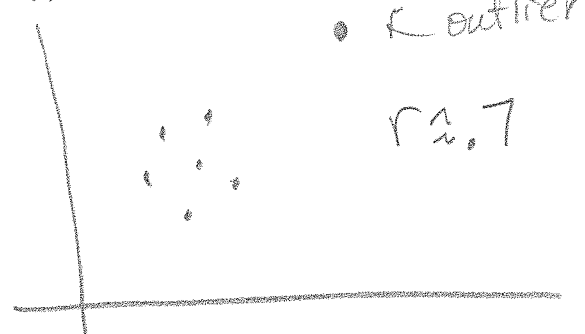
There is a Correlation between time and height it is just Not linear

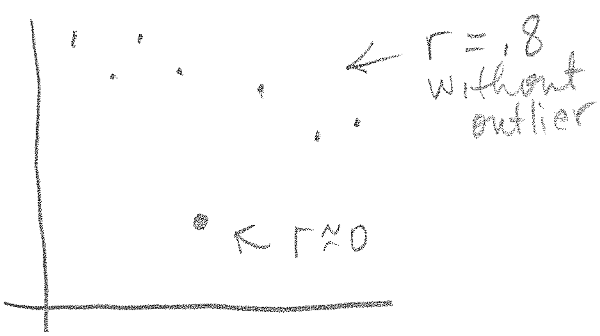
Warning #4

Outliers have a strong effect on Linear correlation



Add an outlier





Always Calculate With & Without Outlier and explain Outlier

Warning #5

Do Not Use Regression Line to Make Predictions if Correlation is Not Significant $|r| < r^*$

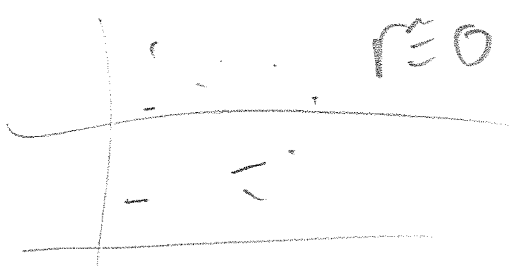
Calculator ↑ $|r|$ Table ↑ r^*

$|TS| < CV$

Use $\bar{y}$ or CI for $\bar{y}$ to make prediction.

Ex $\quad X$ = heights = 69 67 72 72 73 65

y = Test grades = 90 91 96 74 71 72



$\cancel{y = 79 + .03X}$ LinRegTTest

$r \approx 0$

use $\bar{y}$ = Mean of y-values

1-VarStat L_2 = $\bar{x}$

Correlation Warning #5

If $|r| < r^*$ do Not use $\hat{y} = mx + b$

use $\bar{y}$ = mean of y-values (Ranges)

Quiz 9 Math 15 Jones Name________________

1) (18 Points) Periodically during the last two and a half years my husband has gone out to do capacity runs to determine the range of his Electric Mustang. The paired data below consist of battery age in years and the range of my husband's EV in miles.

AGE (years)	0.2	0.3	0.8	1.0	1.2	1.5	1.7	1.8	2.0	2.5
Range (miles)	37	35	35	34	10	29	31	33	32	29

a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between the age of the batteries and the range of the car? r =________ r* =________

Is there a significant linear corrolation? Yes No

b) (2 Points) Make a scatter plot of this data.

Dont use

c) (1 Points) Find the equation for and graph the regression line.________________ $\bar{y}$ → 1Varstats L_2

d) (1 Points) Based on the above data what is the best predicted range for the car after 3 years. ________________

One of the batteries died and had to be exchanged. One of the capacity runs was done right before this happened. The data point that corresponds to this run is an outlier. Remove it and redo the test for correlation.

Years	0.2	0.3	0.8	1.0	1.5	1.7	1.8	2.0	2.5
Range	37	35	35	34	29	31	33	32	29

e) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between age and range.

r = ________ r* =________

Is there a significant linear corrolation? Yes No

f) (1 Points) Find the equation for the regression line.________________

g) (1 Points) Using the data without the outlier, find the best predict ed range for the car after 3 years. ________

h) (3 Points) Interpret the slope of this regression line.

i) (3points) Find r^2 and interpret what it means. r^2=________

x A

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	$r^* =$ .632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

NOTE: To test $H_0: \rho = 0$ against $H_1: \rho \neq 0$, reject H_0 if the absolute value of r is greater than the critical value in the table.

$H_0: \rho = 0$
$H_1: \rho \neq 0$

$\boxed{\text{Stat}}$ $\boxed{\text{Edit}}$

L_1 = Data x-values
L_2 = Data y-values

$\boxed{\text{STAT}}$ ▶▶ LinReg TTest

RegEQ: y_1 $\boxed{\text{VARS}}$ D $\boxed{\text{Y-VARS}}$ 1: Function

r^* = Value in Table

$y_1 = a + bx$

r = Correlation Coefficient

r^2 = prop. of error in y due to Linear correlation with x

$|r| > |r^*|$ Reject $\rho = 0$
there is a linear correlation

$|r| < r^*$ fail to Reject $\rho = 0$
there is Not a Significant Linear relationship
Do Not use $y = a + bx$ to Make predictions
1 - Varstats L_2 ⟹ use $\bar{y}$ as best predicted value

2 Correlation Caution #3

Do not use regression equation if the correlation is not significant. Use the mean of the y-variables instead.

Make a scatter plot of the fallowing data and use your calculator to find the value of the linear correlation coefficient r.

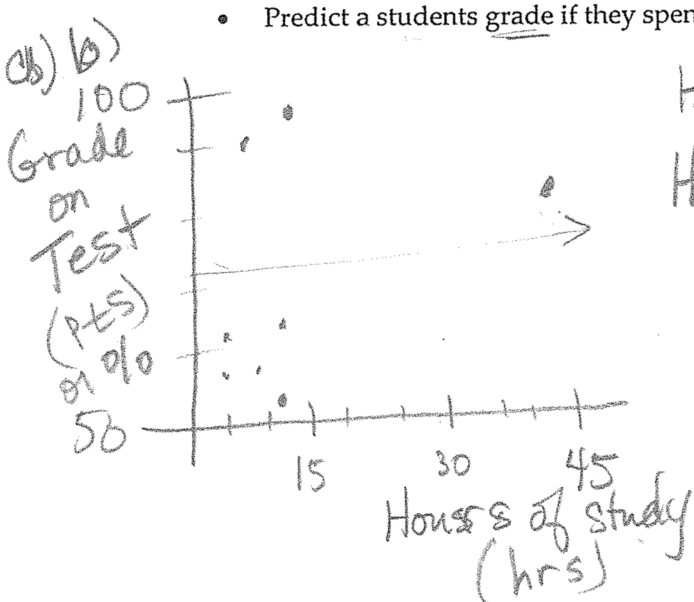
A study was conducted to compare the average time spent in the lab each week versus course grade for computer students. The results are recorded in the table below.

Number of hours spent in lab Grade (percent)

$x \to$ Hours Studying	10	11	16	9	7	45	16	10
$y \to$ Grade on Test	96	51	62	58	89	81	46	51

- Make a scatter plot of the data.
- Graph the regression line on the same graph.
- Find the value of the linear correlation coefficient r.
- Use Table A-6 to determine if there is enough evidence to support the claim that there is a linear correlation between the two variables.
- Predict a students grade if they spent 12 hours in the lab.

(a) b)



$H_0: \rho = 0$

$H_1: \rho \neq 0$

a) Graph
- scale label Axes
- Variable & Units

c) $r = .164$ $r^* = .707$

fail to reject H_0

$\Rightarrow$ Not a significant linear correlation

$\Rightarrow$ Do Not use Reg. line

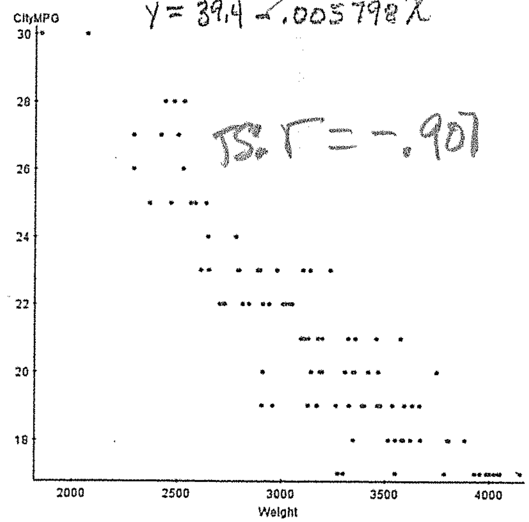
d) Make prediction

Do 1-VarStat on L_2 Says $\bar{x}$ but it is really $\bar{y} = 66.75$ = class average Test score.

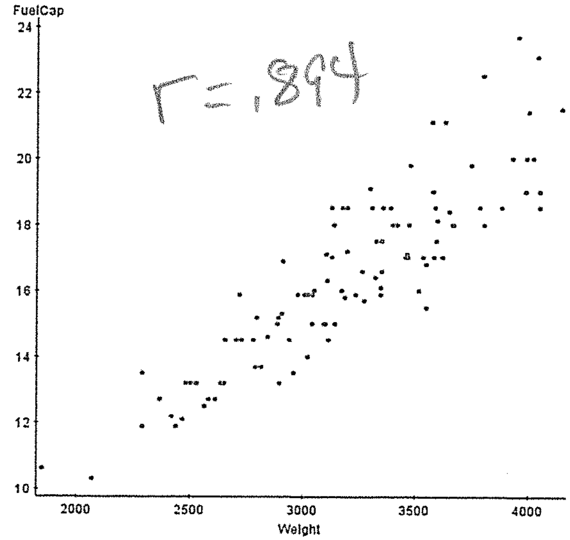
MPG VS. Weight
Strong Negative Correlation
CityMPG = 39.411663 - 0.00579757 Weight
R (correlation coefficient) = -0.9069

Good prediction

$y = 39.4 - .005798x$

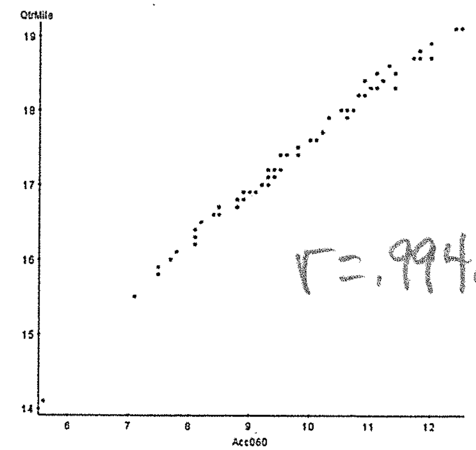


Weight Vs. Fuel Capacity
Strong positive correlation
R (correlation coefficient) = 0.894



Time to accellerate 0 to 60 and Quarter Mile Time

R (correlation coefficient) = 0.9942
Use regression equation
Very strong correlation regression equation will make reliable predictions

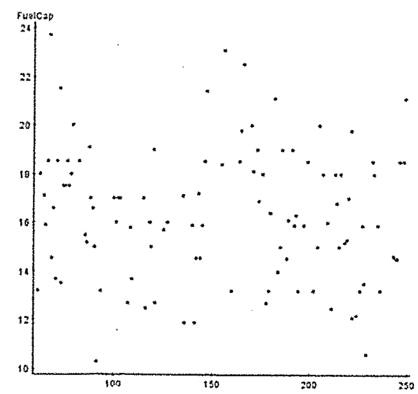


Use Regression Equation

No correlation

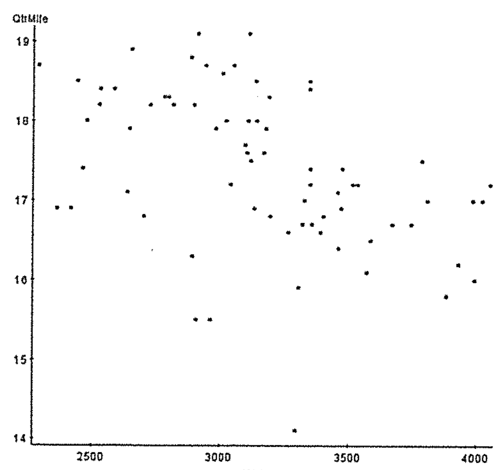
Best predicted values is mean fuel capacity around 16 gallons
R (correlation coefficient) = -0.0815 is close to Zero

Regression equations should not be used to make predictions.



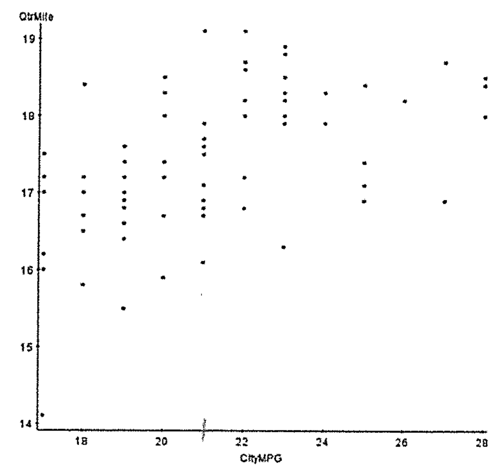
Weak Negative

Correlations is significant by regression equation will not give very reliable answers
R (correlation coefficient) = -0.4502



Weak Positive

R (correlation coefficient) = 0.5101



Do Not Use Regression Equation

x A

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	α = .05	α = .01
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

NOTE: To test H_0: $\rho = 0$ against H_1: $\rho \neq 0$, reject H_0 if the absolute value of r is greater than the critical value in the table.