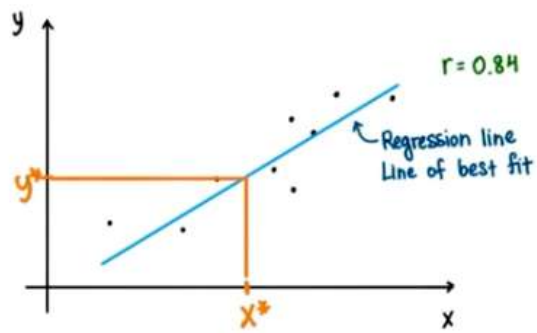
Dimensionless Technologies Private Limited
Visit us at: www.dimensionless.in
- info@dimensionless.in
- 9923170071, 8108094992

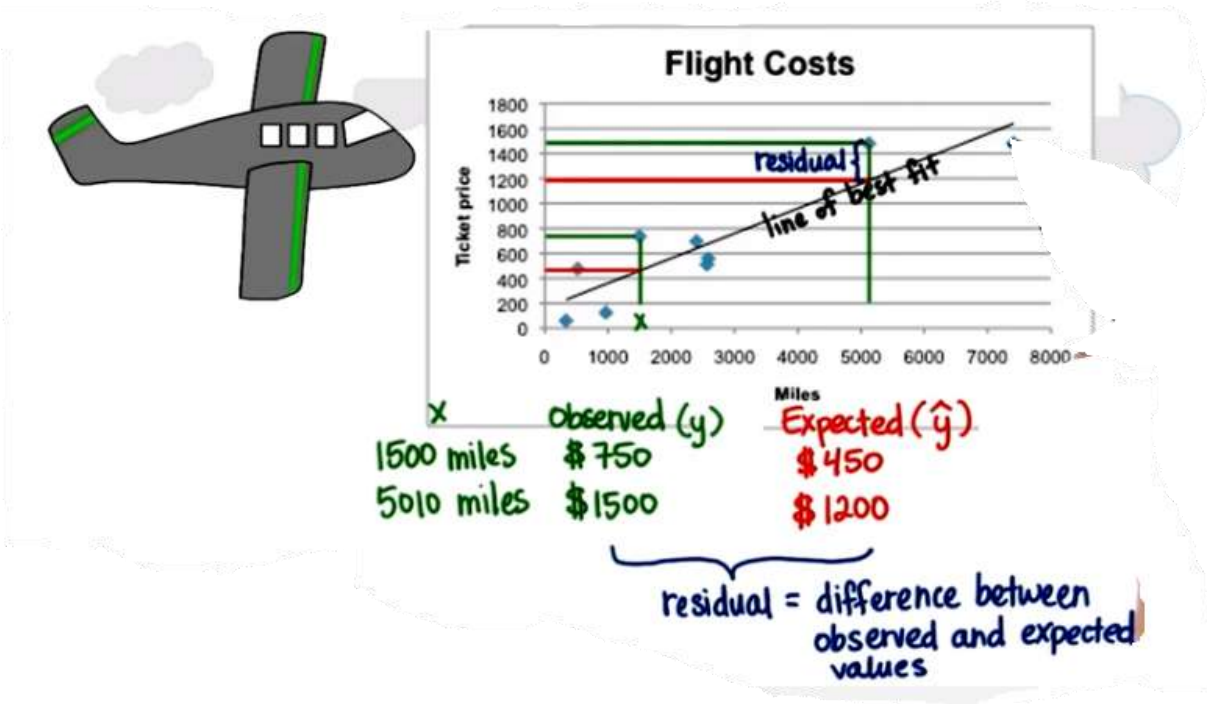
Notes for Students – Lesson 15
Regression



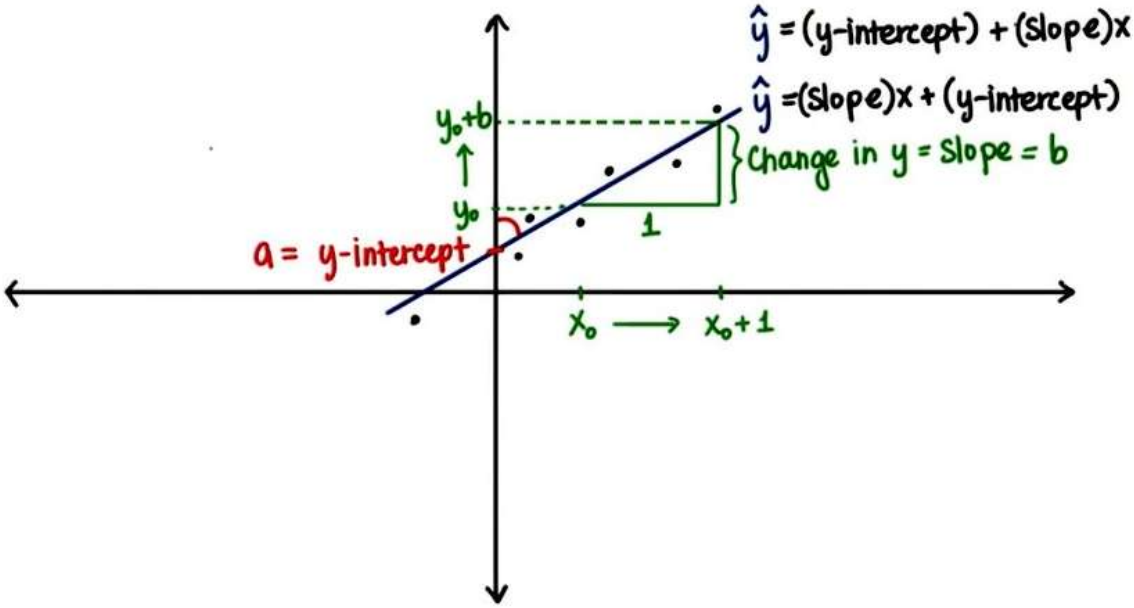
The line of best fit helps us:

- describe the data
- make predictions

Let's say you want to go on vacation, but you have a budget of $500.You decide to analyze some flights and plotted the number of miles they traveled and the costs of that plane ticket. Here's a scatter plot showing the nine flights you looked up. Here's the line of best fit.



Regression Line

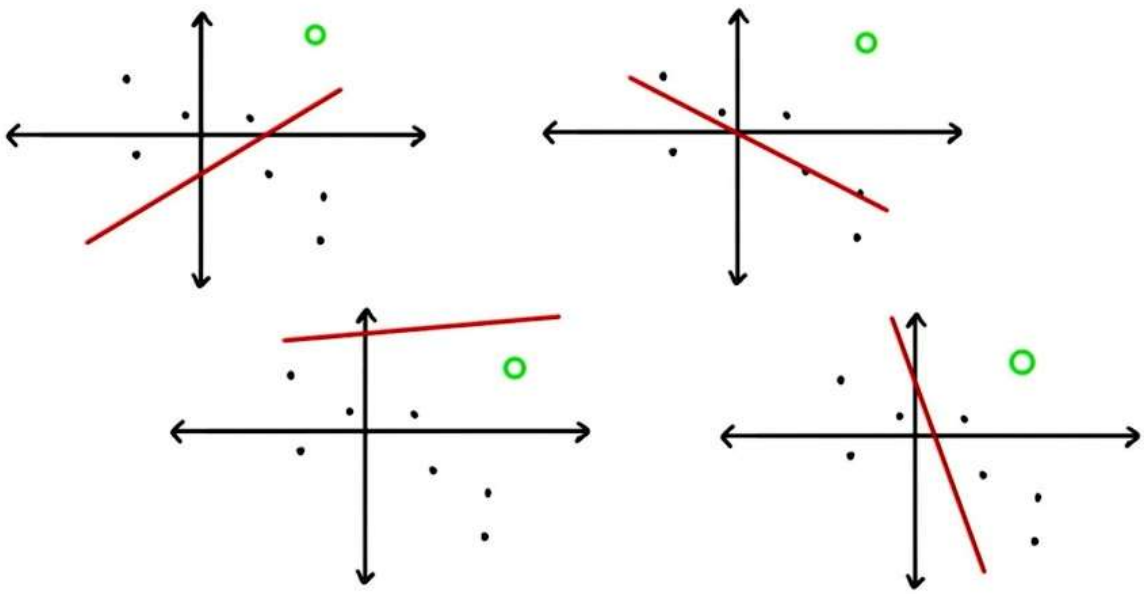


Q1.

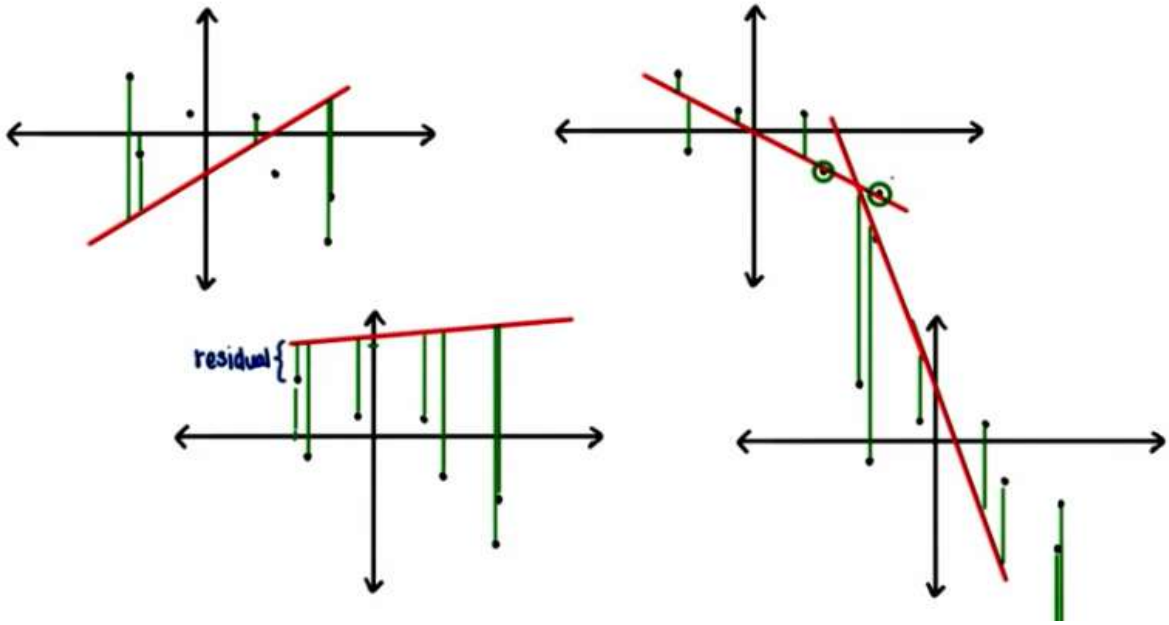
What will be our generic equation for the regression line?

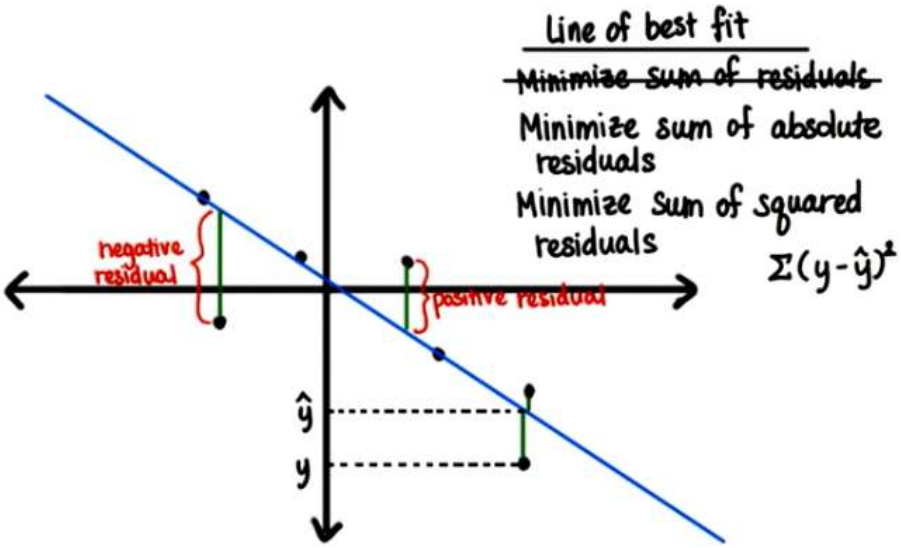
- ☐ $\hat{y} = ax + b$
- ☐ $\hat{y} = a + bx$
- ☐ $\hat{y} = bx + a$
- ☐ $\hat{y} = b + ax$

Q2. Guess the Best Fit Line



Calculating the residual



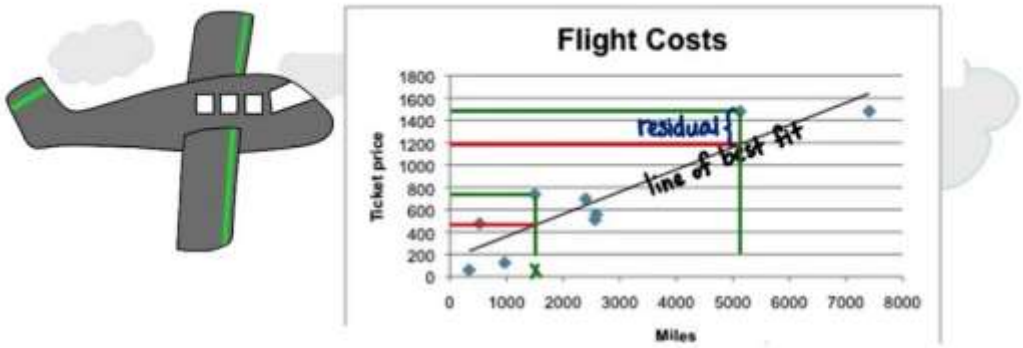


$$b = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$= r\left(\frac{S_y}{S_x}\right)$$

Q3. Use the formula to find b

Part 1



Open the data in the Google Doc. Calculate r.

= pearson (start $cell_x$: end $cell_x$, start $cell_y$: end $cell_y$)

r =

https://docs.google.com/spreadsheets/d/1NjZs8fKZy3a0pzCY17PqMZH1V3Run1kI_lQTqpKJqis/edit?usp=sharing

Part 2

$r = 0.91 \qquad r^2 = 0.83$

$y = bx + a$

$b = r\left(\frac{S_y}{S_x}\right)$

$=$

$S_y = 508.19$

$S_x = 2315.34$

Part 3

$r = 0.91 \quad r^2 = 0.83$

$y = 0.2x + a$?

Which of the following, if known, will enable us to find the y-intercept?

- A point on the regression line
- The standard deviation of x
- Any one of the data values

Q4.

$r = 0.91 \quad r^2 = 0.83$

$y = 0.2x + a$?

What point does the regression line go through?

Q5.

What point does the regression line go through?

Q6.

What is this point ?

$(\bar{x}, \bar{y})$ $\quad \bar{x} =$

$\bar{y} =$

Q7.

$r = 0.91$

$y = 0.2x + a$

$(2601.11, 680.35)$

$\Rightarrow \bar{y} - 0.2\bar{x} = a \Rightarrow a =$

Q8.

How much would you predict it costs to travel 4000 miles?

Q9.

What is the additional cost per mile?

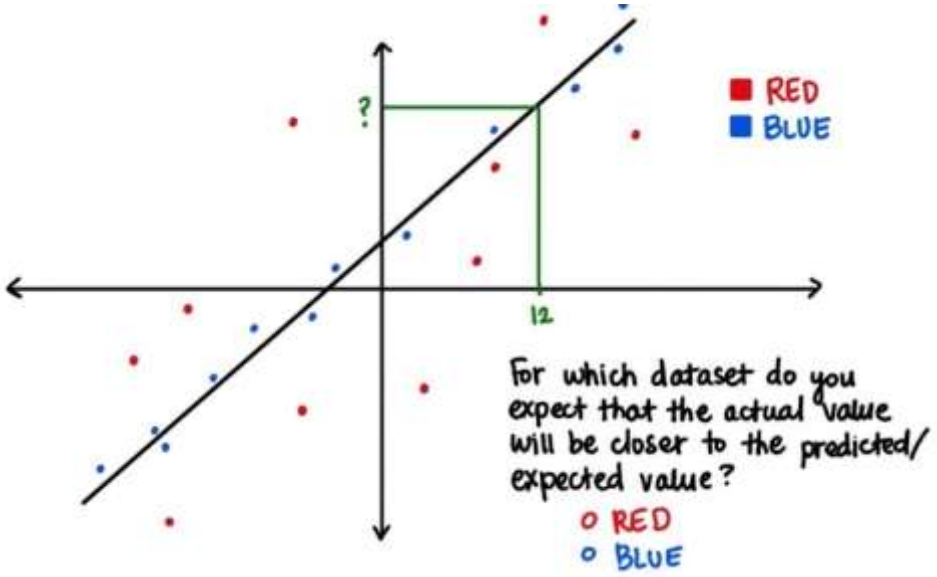
Q10.

What is the expected price for a flight that travels 0 miles?

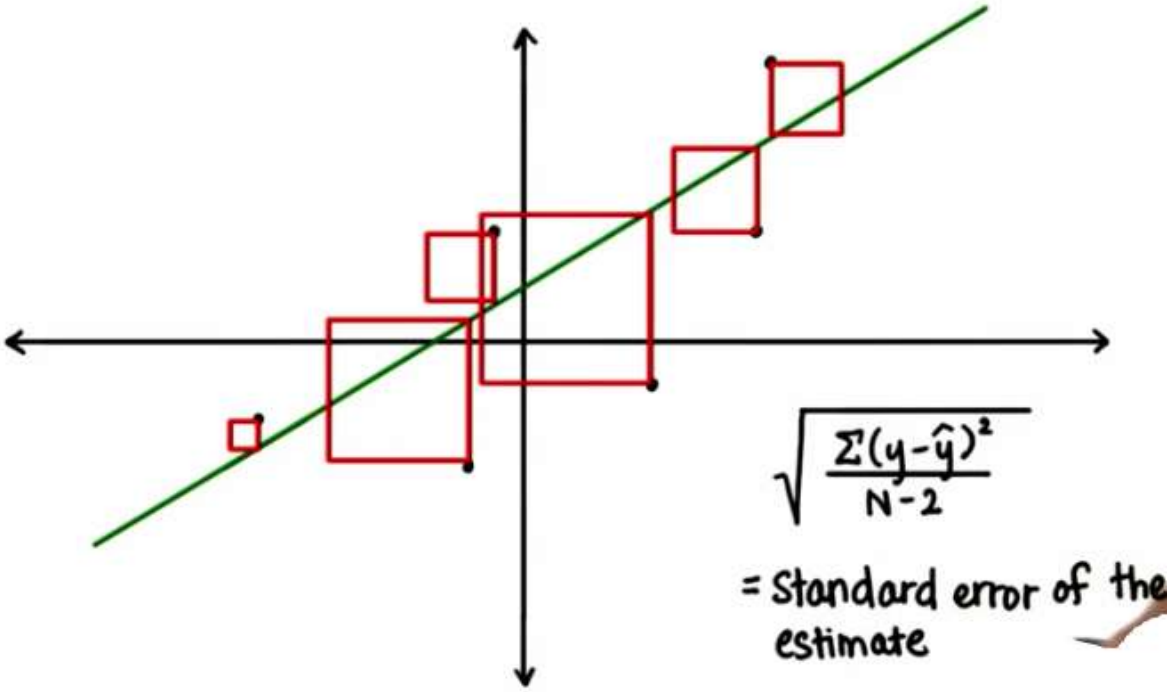
Q11.

On a budget of $500, what is the furthest distance you can travel?

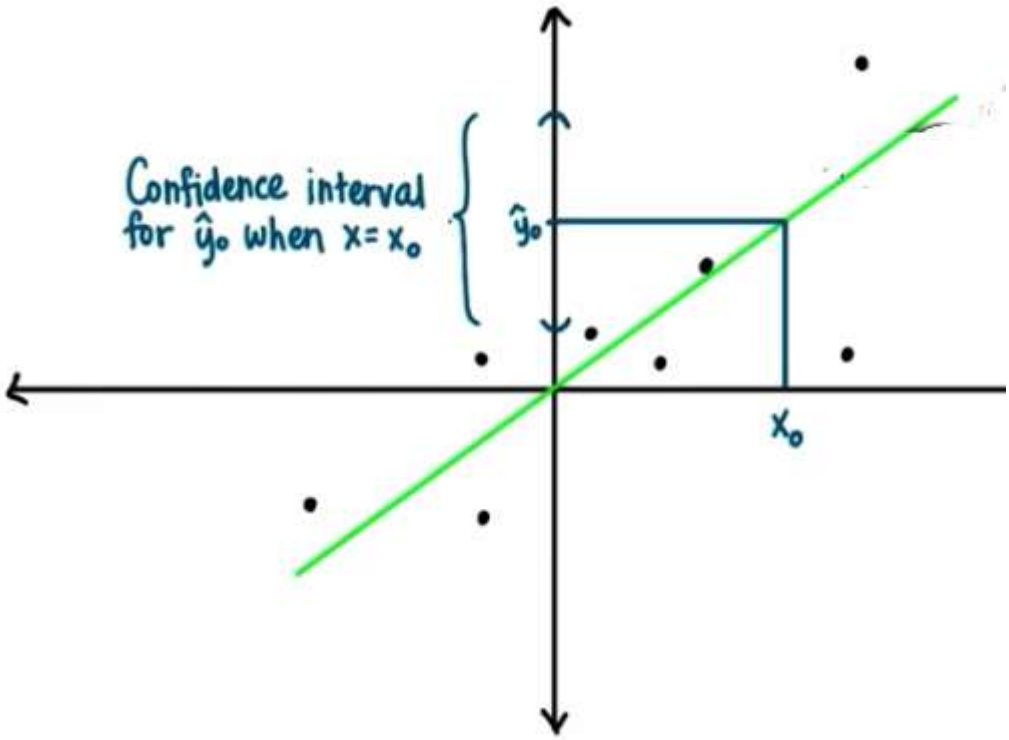
Q12.

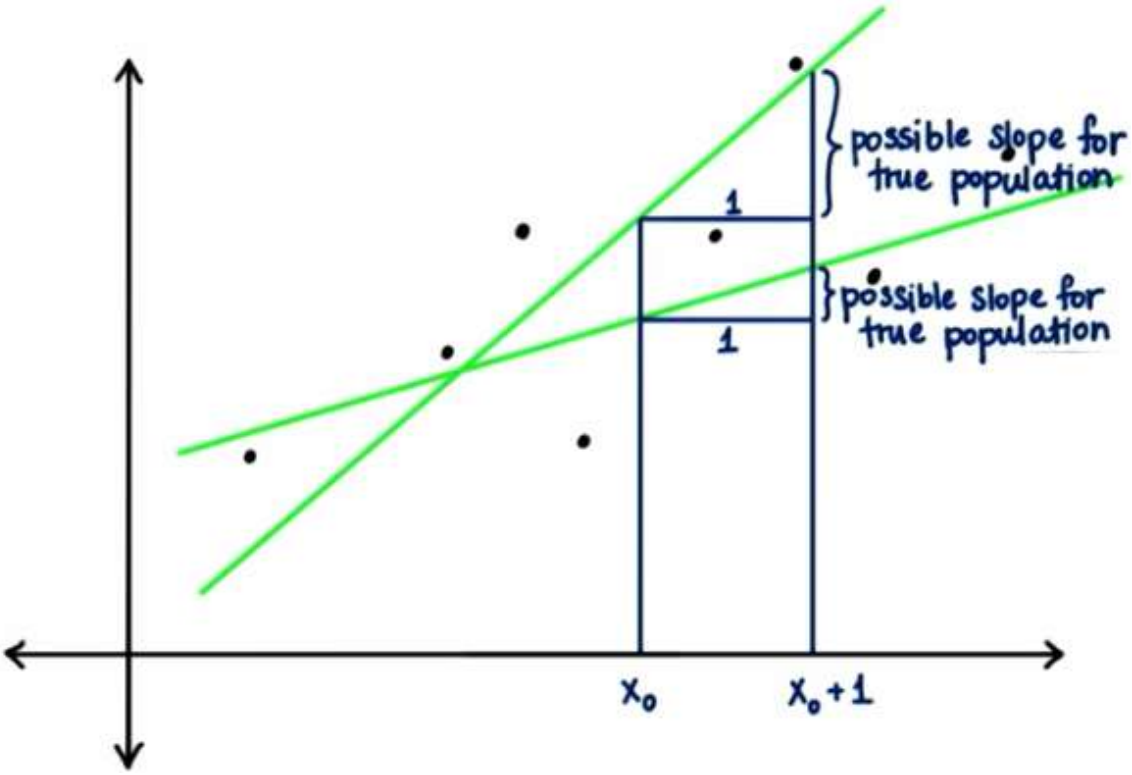


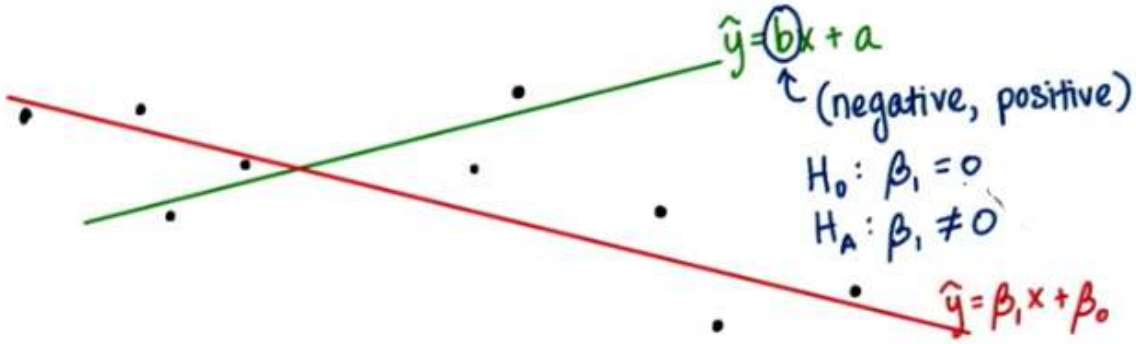
Standard Error of the Estimate



Confidence Interval







Hypothesis Testing for Slope

Q13.

Hypothesis Testing for Slope (Same outcome as hypothesis test for r)

Are x and y linearly related?

β_1 = population slope
β_0 = population y-int.
b = sample slope
a = sample y-int.

- $H_0: b = 0$; $H_A: b < 0$, $b > 0$, $b \neq 0$
- $H_0: \beta_1 = 0$; $H_A: \beta_1 < 0$, $\beta_1 > 0$, $\beta_1 \neq 0$
- $H_0: \beta_0 = 0$; $H_A: \beta_0 < 0$, $\beta_0 > 0$, $\beta_0 \neq 0$
- $H_0: \beta_1 < 0$, $\beta_1 > 0$, $\beta_1 \neq 0$; $H_A: \beta_1 = 0$

t-test for Slope

Q14.

Hypothesis Testing for Slope (Same outcome as hypothesis test for r)

Are x and y linearly related?

β_1 = population slope
β_0 = population y-int.
b = sample slope
a = sample y-int.

$H_0 : \beta_1 = 0$
$H_A : \beta_1 < 0$
$\beta_1 > 0$
$\beta_1 \neq 0$

df = N-2

N=9
α=0.05
t=5.77

There is/is not enough evidence to reject the null; there does/does not appear to be a significant relationship between x and y.

Linear Regression in R

```
> bestfit=lm(y~x,data=z)
> summary(bestfit)

Call:
lm(formula = y ~ x, data = z)

Residuals:
   Min     1Q Median     3Q    Max 
-229.8 -163.6 -117.8  209.0  296.7 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 161.38775  117.41071   1.375 0.211656    
x             0.19951    0.03458   5.770 0.000684 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 226.4 on 7 degrees of freedom
Multiple R-squared:  0.8263,	Adjusted R-squared:  0.8015 
F-statistic:  33.3 on 1 and 7 DF,  p-value: 0.000684

> |
```

Applying Regression model using EXCEL

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.909004
R Square	0.826288
Adjusted R Square	0.801472
Standard Error	226.4305
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1707137	1707137	33.2965	0.000684
Residual	7	358895.3	51270.76		
Total	8	2066032			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	161.3878	117.4107	1.374557	0.211656	-116.244	439.02	-116.244	439.02
X Variable 1	0.199515	0.034576	5.770312	0.000684	0.117755	0.281274	0.117755	0.281274

FACTORS THAT AFFECT SIMPLE LINEAR REGRESSION