

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

STA404

Regression and Correlation Analysis

STA404 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

STA404

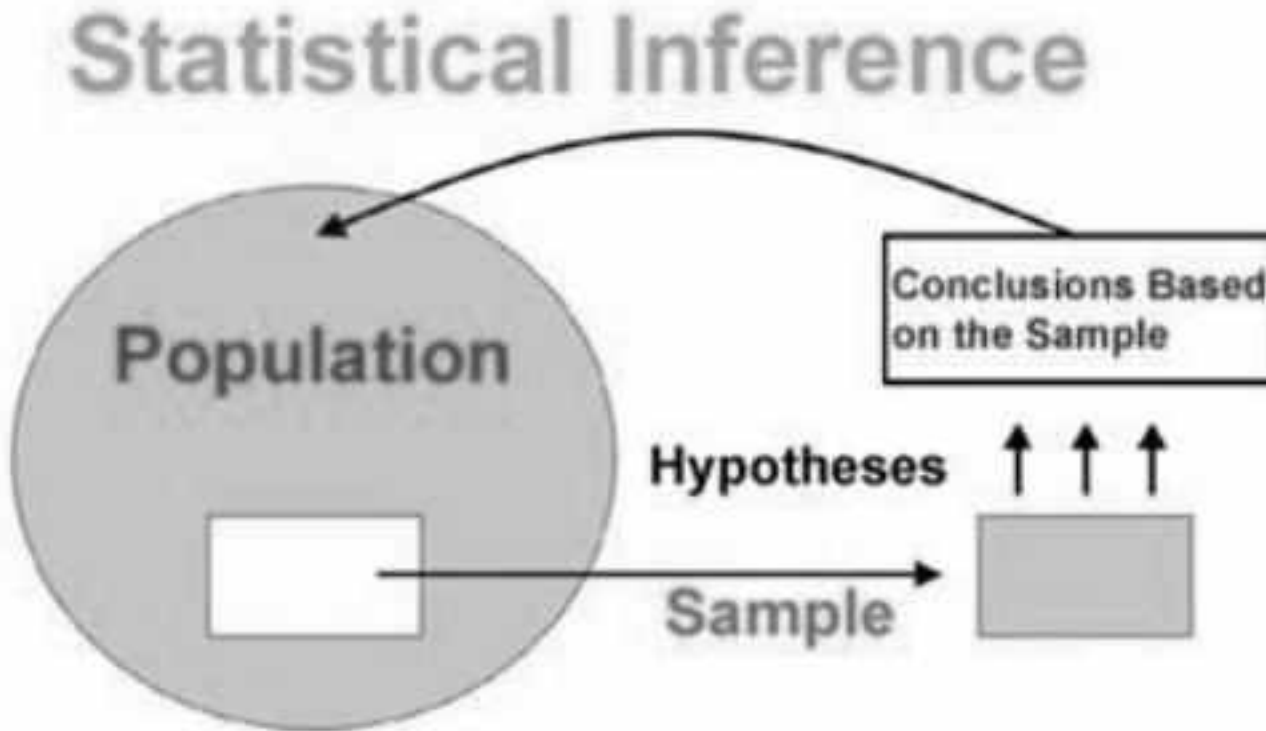
Regression & Correlation Analysis

Lesson 8-1

Statistical Inference about Regression

Explanation of the hypothesis testing procedure
for simple linear regression coefficient β

What is Statistical inference?



mcqs

Difference between Population & Sample Regression

- Population Regression

$$y = \alpha + \beta x + \varepsilon$$

Intercept

Slope

Error



- Sample Regression

$$\hat{y} = a + bx + e$$

a

b

e

show

Hypothesis Testing – General Procedure

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

- Step 1** State the hypotheses and identify the claim.
- Step 2** Find the critical value(s) from the appropriate table
- Step 3** Compute the test value.
- Step 4** Make the decision to reject or not reject the null hypothesis.
- Step 5** Summarize the results.

Testing Hypothesis – Procedure about Regression Coefficient β

Step 1

Stat the hypothesis.

$$H_0 : \beta = 0 \quad \text{and} \quad H_1 : \beta \neq 0$$

$$H_0 : \beta \leq 0 \quad \text{and} \quad H_1 : \beta > 0$$

$$H_0 : \beta \geq 0 \quad \text{and} \quad H_1 : \beta < 0$$

Step 2

Decide the level of significance.

Here, we set the value of $\alpha = 5\% = 0.05$

$$y = a + b_1x$$

M C Q

Su

Sh 6 v

General formula format

Many hypotheses are tested using a statistical test based on the following general formula:

$$\text{Test Value} = \frac{(\text{Observed value}) - (\text{Expected value})}{\text{Standard error}}$$

The observed value is the “statistic” (such as the mean) that is computed from the sample data.

The expected value is the “parameter” (such as the mean) that you would expect to obtain if the null hypothesis were true. In other words, the hypothesized value.

The denominator is the standard error of the statistic being tested (in this case, the standard error of the mean).

$$t = \frac{b - \beta_0}{S.E(b)}$$

It is t-test. It depends on *Degree of Freedom* = $df = n - 2$

Required calculations

Step 4

Know and do the required calculations and find the test value.

$$t = \frac{b - \beta_0}{S.E(b)} \quad \beta_0 \Rightarrow \text{Hypothesis / given}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

$$a = \bar{Y} - b\bar{X}$$

$$s_{YX} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n - 2}}$$

$$S.E(b) = s_b = \frac{s_{YX}}{\sqrt{\sum (X - \bar{X})^2}}$$

Short

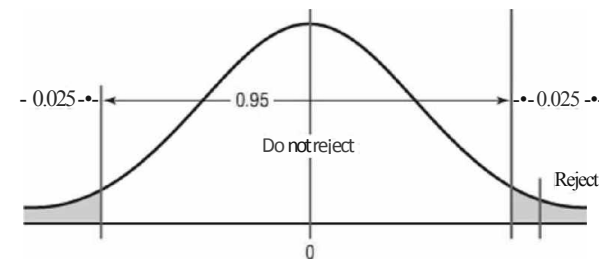
Σ m c D

Critical Value, Comparison and Conclusion

Step 5

- Use t-table to find the critical values.
- Plot these values on bell shaped curve.
- Compare calculated value with table value
- Conclude and summarize the result

	intervals	80%	90%	95%	98%	99%
	One tail, α	0.10	0.05	0.025	0.01	0.005
	Two tails, α	0.20	0.10	0.05	0.02	0.01
d.f.						
1		30.08	6.314	12.706	31.821	63.657
2		18.86	2.920	4.303	6.965	9.923
3		16.01	2.353	3.182	4.541	5.841
4		15.33	2.132	2.776	3.747	4.779
5		14.76	2.015	2.571	3.365	4.032
6		14.40	1.943	2.447	3.143	3.707
7		14.09	1.895	2.365	2.998	3.499
8		13.82	1.860	2.306	2.896	3.355
10		13.28	1.812	2.228	2.764	3.169
11		13.09	1.796	2.201	2.718	3.106
12		12.96	1.782	2.179	2.681	3.055
13		12.85	1.771	2.160	2.650	3.012
14		12.77	1.761	2.145	2.624	2.977
15		12.71	1.753	2.131	2.602	2.947
16		12.66	1.746	2.120	2.583	2.921
17		12.62	1.740	2.110	2.567	2.898
18		12.59	1.734	2.101	2.552	2.876
19		12.58	1.729	2.093	2.539	2.856
20		12.57	1.725	2.086	2.528	2.845
22		12.56	1.721	2.080	2.518	2.831
23		12.55	1.717	2.074	2.508	2.819
24		12.55	1.714	2.059	2.500	2.807



Testing Hypothesis – Example

- For the Given data:
- Estimate the regression line from the following data of height (X) and Weight (Y)
- Test the hypothesis that the height and weight are independent.

Height (X)	Weight(Y)
60	110
60	135
60	120
62	120
62	140
62	130
62	135
64	150
64	145
70	170
70	185
70	160

Testing Hypothesis – Example

Step 1

Stat the hypothesis.

$H_0 : \beta = 0$ (There is NO slop. There is NO linear relationship.)

$H_1 : \beta \neq 0$ (There exist linear relationship)

Step 2

We set the level of significance (alpha) $\alpha = 0.05$

Step 3

The test-statistic (formula) to be used is:

$$t = \frac{b - \beta_0}{S.E(b)}$$

It follows t-distribution with degree of freedom $df = n-2$

MCQ

Computation

X	Y	XX	YY	XY
60	110	3600	12100	6600
60	135	3600	18225	8100
60	120	3600	14400	7200
62	120	3844	14400	7440
62	140	3844	19600	8680
62	130	3844	16900	8060
62	135	3844	18225	8370
64	150	4096	22500	9600
64	145	4096	21025	9280
70	170	4900	28900	11900
70	185	4900	34225	12950
70	160	4900	25600	11200
766	1700	49068	246100	109380

What we need to calculate?

Step 4

Do necessary computations from data and solve the formula using these values.

• $t = \frac{b - \beta_0}{s_b}$

$s_b = \frac{s_{YX}}{\sqrt{\sum(X - \bar{X})^2}}$

$s_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n-2}}$

$a = \bar{Y} - b\bar{X}$

$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$

$\sum(X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$

Let's do calculations

Step 4

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} = \frac{(12)(109380) - (766)(1700)}{(12)(49068) - (766)^2} = \frac{10360}{2060} = 5.03$$

$$a = \bar{Y} - b\bar{X} = \left(\frac{1700}{12}\right) - (5.03)\left(\frac{766}{12}\right) = -179.41$$

$$\begin{aligned}\sum(Y - \hat{Y})^2 &= \sum Y^2 - a\sum Y - b\sum XY = 246100 - (-179.41)(1700) - (5.03)(109380) \\ &= 915.60\end{aligned}$$

$$\sum(X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n} = 49068 - \frac{(766)^2}{12} = 171.67$$

Let's do calculations

Step 4

$$s_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{915.60}{12 - 2}} = 9.5687$$

- $$s_b = \frac{s_{YX}}{\sqrt{\sum(X - \bar{X})^2}} = \frac{9.5687}{\sqrt{171.47}} = \frac{9.5687}{13.10} = 0.73$$

- $$t = \frac{b - \beta_0}{s_b} = \frac{5.03 - 0}{0.73} = 6.89$$

Calculated value = 6.89

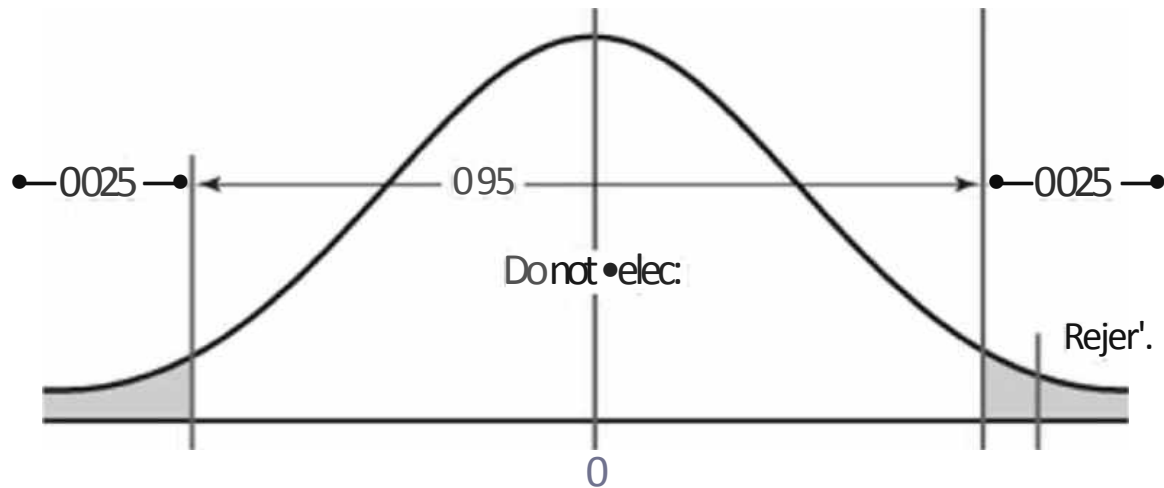
Critical Value from t-table

Step 5 : Find the 0.05 column in the top row and 16 in the left-hand column. Where the row and column meet, the appropriate critical value is found.

Table F The t Distribution						
	Confidence intervals	80%	90%	95%	98%	99%
	One tail, α	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.20	0.10	0.05	0.02	0.01
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807

Decision Rule

Step 5



Conclusion and Summarizing:

Since the computed value of $t=6.89$ falls in the critical region, so we reject the null hypothesis and may conclude that there is sufficient reason to say with 95% confidence that height and weight are related.

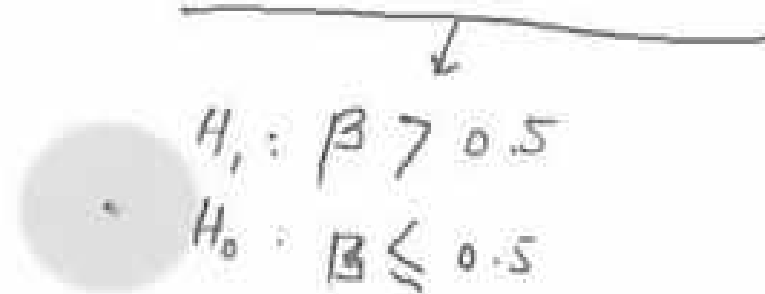
An other example – Do yourself

In linear regression problem, the following sums were computed from a random sample of size 10.

$$\Sigma X = 320, \Sigma Y = 250, \Sigma X^2 = 12400, \Sigma XY = 9415 \text{ and } \Sigma Y^2 = 7230$$

Using 5% level of significance, test the hypothesis that the population regression coefficient, β is greater than 0.5.

Hint:



A diagram showing a horizontal line with a downward-pointing arrow from the word 'greater' in the text above. Below the line, the hypotheses are written: $H_1 : \beta > 0.5$ and $H_0 : \beta \leq 0.5$.

$$H_1 : \beta > 0.5$$
$$H_0 : \beta \leq 0.5$$

Hint

$$H_0: \beta = 0$$

$$H_1: \beta > 0 \text{ (Given claim)}$$

T

$$\text{test formula } = t = \frac{b - \beta_0}{s_b}$$

- $$s_b = \frac{s_{YX}}{\sqrt{\sum (X - \bar{X})^2}}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

$$s_{YX} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n - 2}}$$

$$a = \bar{Y} - b\bar{X}$$

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

Summary

- Remember the all hypothesis-testing situations using the traditional method should include the

$$\text{Test Value} = \frac{(\text{Observed roluue}) - (\text{Expected value})}{\text{Standard error}}$$

1. State the null and alternative hypotheses and identify the claim.
2. State an alpha level and find the critical value(s).
3. Compute the test value.
4. Make the decision to reject or not reject the null hypothesis.
5. Summarize the results.

The END

VMLQ

Get Ready ...

- Check mic position
- Check resolution size
- Check pen
- Say something....

STA404

Regression & Correlation Analysis Lesson 9-1

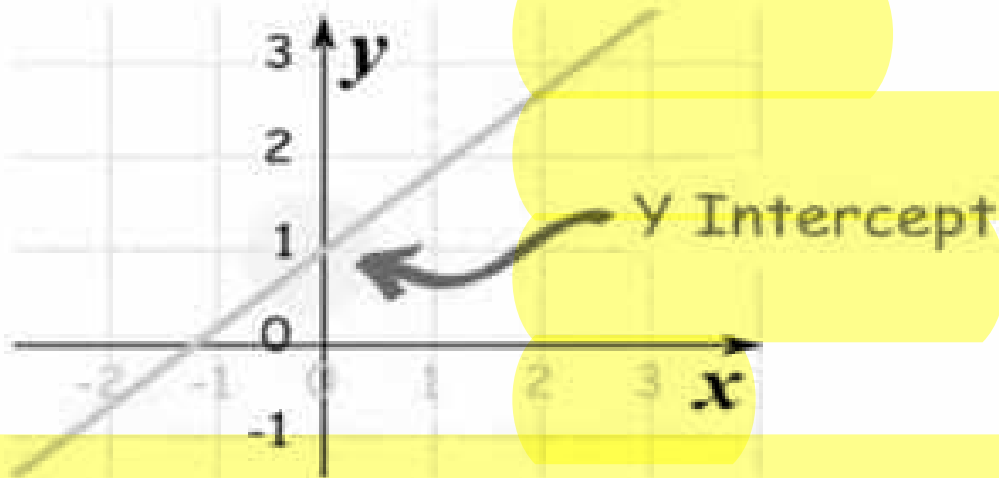
Department of
Statistics Virtual
University of Pakistan

Statistical Inference about Regression

Using the t-test for hypothesis testing of
– the intercept of regression line

What is intercept of regression line?

- The Y intercept of a straight line is simply where the line crosses (cut) the Y axis.

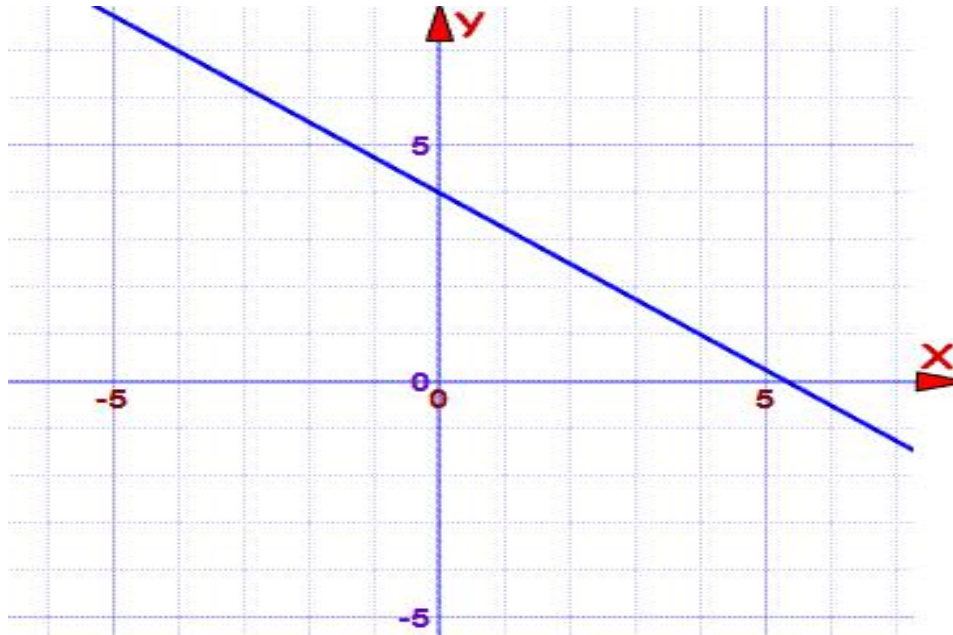


$$Y = a + bX$$

- In the above diagram the line crosses the Y axis at 1.
- So the Y intercept is equal to 1.

Can you tell?

- What is intercept of the given line?



↓
m d

Difference between Population & Sample Regression

- Population Regression

$$\underline{Y = \alpha + \beta X + \varepsilon}$$

French

- Sample Regression

$$\underline{Y = a + bX + e}$$

English

MLQ

Hypothesis Testing – General Procedure

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

- Step 1** State the hypotheses and identify the claim. ✓
- Step 2** Find the critical value(s) from the appropriate table. ✓
- Step 3** Compute the test value. ✓
- Step 4** Make the decision to reject or not reject the null hypothesis.
- Step 5** Summarize the results.

MO

Testing Hypothesis – Example

- For the Given data:
- Test the hypothesis that $H_0 : \alpha = 32$ against $H_1 : \alpha \neq 32$

X	Y
65	85
50	74
55	76
65	90
55	85
70	87
65	94
70	98
55	81
70	91
50	76
55	74

Testing Hypothesis – Example

Step 1

Stat the hypothesis.

$$H_0 : \alpha = 32$$

$$H_1 : \alpha \neq 32$$

Step 2

We set the **level of significance (alpha) $\alpha = 0.01$ (1%)**

Step 3

The test-statistic (formula) to be used is: $\text{Test Value} = \frac{(\text{Observed value}) - (\text{Expected value})}{\text{Standard error}}$

$$t = \frac{\alpha - \alpha_0}{S.E(\alpha)}$$

It follows t-distribution with degree of freedom $df = n-2$

What we need to calculate?

Do necessary computations from data and solve the formula using these values.

Handwritten notes and arrows indicate the relationships between the formulas:

- Top right: $x, y \Rightarrow y = a + bx$
- Top center: $t = \frac{a - a_0}{S.E(a)}$ (with a handwritten note "Given" and an arrow pointing to a)
- Bottom right: $b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$
- Bottom center: $a = \bar{Y} - b\bar{X}$
- Bottom left: $\Sigma(X - \bar{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$
- Center: $S_{YX} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\Sigma Y^2 - a \Sigma Y - b \Sigma XY}{n-2}}$
- Top left: $S.E(a) = s_a = S_{YX} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X - \bar{X})^2}}$

Handwritten notes include "Given data" and "Given" with arrows pointing to the formulas.

Computation

X	Y	XX	YY	XY
65	85	4225	7225	5525
50	74	2500	5476	3700
55	76	3025	5776	4180
65	90	4225	8100	5850
55	85	3025	7225	4675
70	87	4900	7569	6090
65	94	4225	8836	6110
70	98	4900	9604	6860
55	81	3025	6561	4455
70	91	4900	8281	6370
50	76	2500	5776	3800
55	74	3025	5476	4070
725	1011	44475	85905	61685

Let's do calculations

Step 4

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} = \frac{12(61685) - (725)(1011)}{12(44475) - (725)^2} = \frac{7245}{8075} = 0.897$$

$$a = \bar{Y} - b\bar{X} = \left(\frac{1011}{12}\right) - (0.897)\left(\frac{725}{12}\right) = 30.056$$

$$\begin{aligned}\Sigma(Y - \hat{Y})^2 &= \Sigma Y^2 - a\Sigma Y - b\Sigma XY = 85905 - (30.056)(1011) - (0.897)(61685) \\ &= 186939\end{aligned}$$

$$\Sigma(X - \bar{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n} = 44475 - \frac{(725)^2}{12} = 672.917$$

Let's do calculations

Step 4

$$s_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{186.939}{12 - 2}} = 4.3236$$

$$S.E(a) = s_a = s_{YX} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum(X - \bar{X})^2}} = (4.3236) \sqrt{\frac{1}{12} + \frac{(68.12)^2}{672.917}} = \dots = 11.47$$

$$t = \frac{a - \alpha_0}{S.E(a)} = \frac{30.056 - 32}{11.47} = -0.169$$

Critical Value from t-table

Step 5 : Find the 0.01 column in the top row and 10 in the left-hand column. Where the row and column meet, the appropriate critical value is found.

Table F The t Distribution						
	Confidence intervals	80%	90%	95%	98%	99%
	One tail, α	0.10	0.05	0.025	0.01	0.005
d.f. $\rightarrow$	Two tails, α	0.20	0.10	0.05	0.02	0.01
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.804
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.105
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807

Decision Rule

Step 5



Conclusion and Summarizing:

Since the computed value of $t = -0.169$ does NOT fall in the critical region, so we DO NOT reject the null hypothesis.

An other example – Do yourself

In linear regression problem, the following sums were computed from a random sample of size 10.

$$\Sigma X = 320, \Sigma Y = 250, \Sigma X^2 = 12400, \Sigma XY = 9415 \text{ and } \Sigma Y^2 = 7230$$

$$\hat{Y} = 4.04 + 0.655X$$

$$\Sigma(Y - \hat{Y})^2 = 53.175$$

$$\Sigma(X - \bar{X})^2 = 2160$$

Using 5% level of significance, test the hypothesis that the intercept of regression line, α is zero.

Hint:

Hint

$$H_0 : \alpha = 0 \text{ (Given claim)}$$

$$H_1 : \alpha \neq 0$$

T

$$\text{Test formula} = t = \frac{a - \alpha_0}{S.E.(a)}$$

$$S.E.(a) = s_a = s_{YX} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X - \bar{X})^2}}$$

$$s_{YX} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n - 2}}$$

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$\hat{Y} = 4.04 + 0.655X$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

$$a = \bar{Y} - b\bar{X} = 4.04$$

Summary

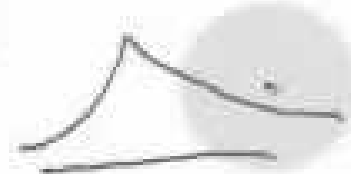
- Remember the all hypothesis-testing situations using the traditional method should include the

$$\text{Test Value} = \frac{(\text{Observed value}) - (\text{Expected value})}{\text{Standard error}}$$

$$\alpha = 0, \alpha > 0, \alpha < 0$$

- State the null and alternative hypotheses and identify the claim.
- State an alpha level and find the critical value(s).
- Compute the test value. ✓
- Make the decision to reject or not reject the null hypothesis.
- Summarize the results.

cricket



The END



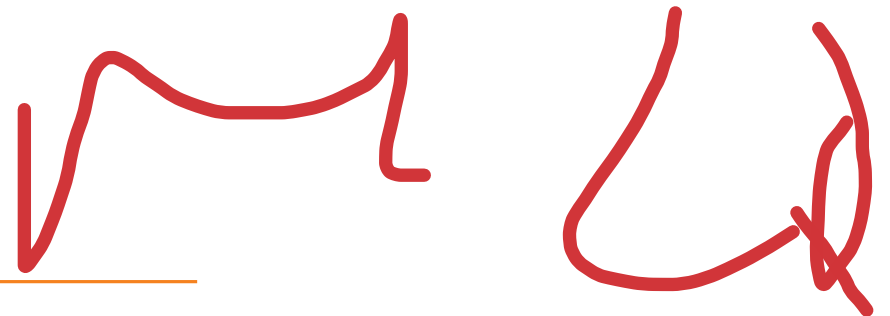
STA404

Regression & Correlation Analysis Lesson 10-1

Department of
Statistics Virtual
University of Pakistan

Statistical Inference about Simple Correlation

Testing hypothesis that the population correlation coefficient ρ is equal to some specified value other than zero



Hypothesis Testing – General Procedure

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

- | | |
|---------------|--|
| Step 1 | State the hypotheses and identify the claim. |
| Step 2 | Find the critical value(s) from the appropriate table |
| Step 3 | Compute the test value. |
| Step 4 | Make the decision to reject or not reject the null hypothesis. |
| Step 5 | Summarize the results. |

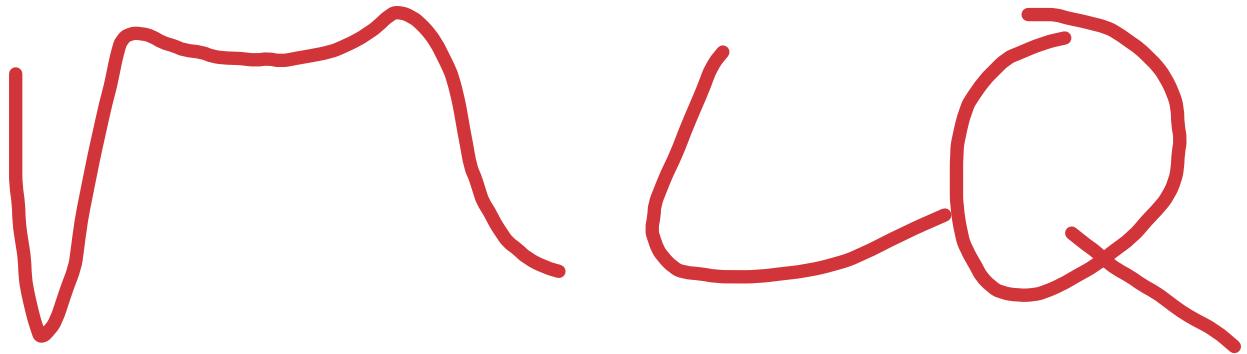
Notations:

- **Sample** Correlation Coefficient

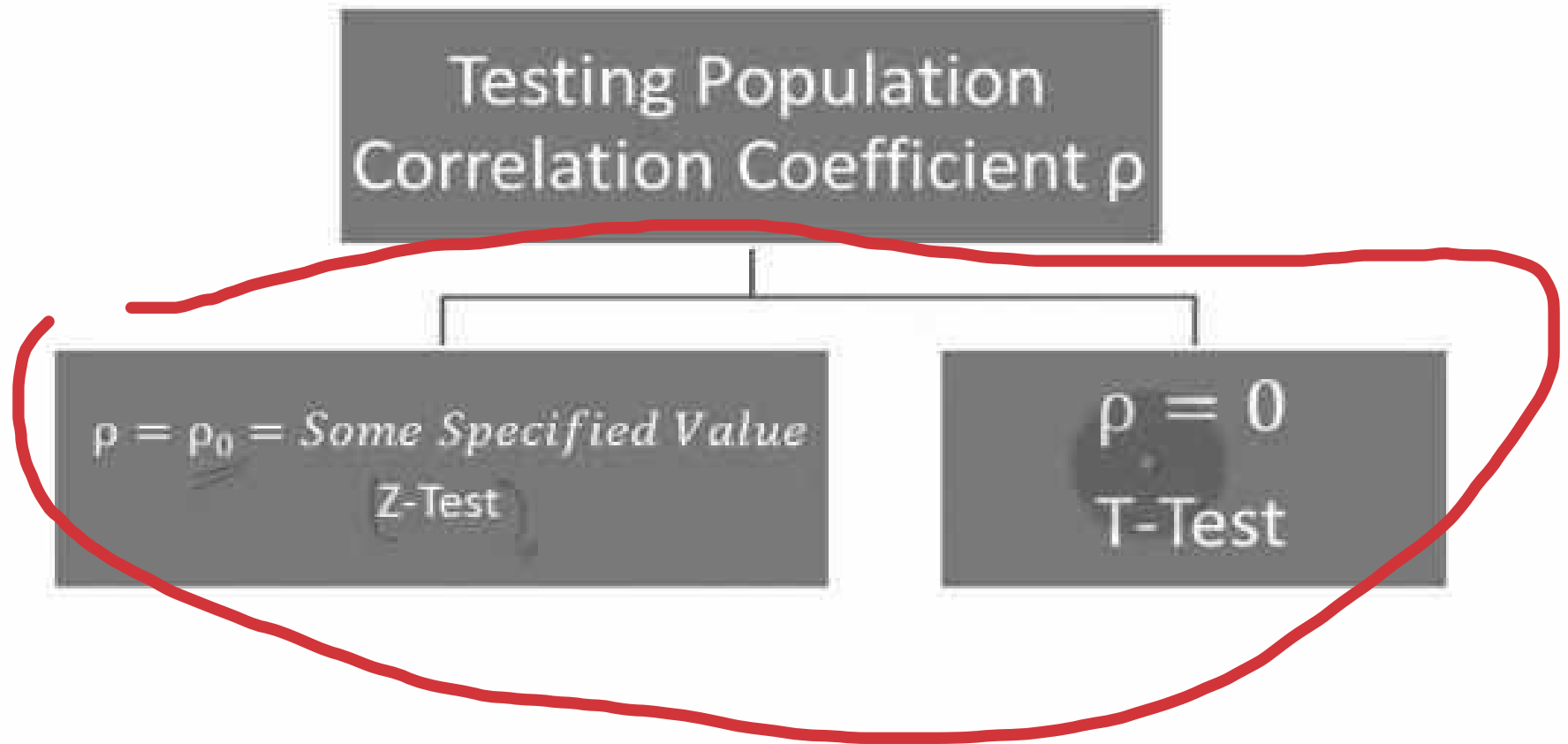
- **Population** Correlation Coefficient

- English letter “r”

- Greek small letter “ ρ ” (rho)



Two Cases



MLQ

Some Main Points

- **Problem:** Sampling distribution of "r" is
 - Neither normal
 - Nor becomes approximately normal for large sample size
- **Solution:** Do Transformation:
Old variable → new variable

"r" ----- "Z_f"

$$\text{"r"} \dots\dots\dots Z_f = 1.1513 \log \frac{1+r}{1-r}$$

ml

Testing formula

$$\text{Test Value} = \frac{\overset{\text{sample}}{\text{(Observed value)}} - \overset{\text{pop.}}{\text{(Expected value)}}}{\text{Standard error}}$$

$$Z = \frac{Z_f - \mu_z}{1 / \sqrt{n - 3}}$$

Where,

$$Z_f = 1.1513 \log \frac{1+r}{1-r}$$

$$\mu_z = 1.1513 \log \frac{1+\rho}{1-\rho}$$

MLQ

Testing Hypothesis – Example

- A random sample of 28 pairs from a bivariate normal population showed a correlation coefficient of 0.7.
- Is this value consistent with the assumption that the correlation coefficient in the population is 0.5?

Testing Hypothesis – Example

Step 1

Stat the hypothesis.

$$H_0 : p = 0.5 \quad (\text{Given claim})$$

$$H_1 : p \neq 0.5$$

Step 2

We set the level of significance (alpha) $\alpha = 0.05$ (5%)

Step 3

The test-statistic (formula) to be used is: $\text{Test Value} = \frac{(\text{Observed value}) - (\text{Expected value})}{\text{Standard error}}$

$$Z = \frac{Z_f - \mu_z}{\frac{1}{\sqrt{n-3}}}$$

– It is z-test.

Calculations

$$Z = \frac{Z_f - \mu_z}{\frac{1}{\sqrt{n-3}}}$$

Given : $r = 0.7$, $\rho = 0.5$ and $n = 28$

$$Z_f = 1.1513 \log \frac{1+r}{1-r} = 1.1513 \log \frac{1+0.7}{1-0.7} = 1.1513 \log \frac{1.7}{0.3}$$

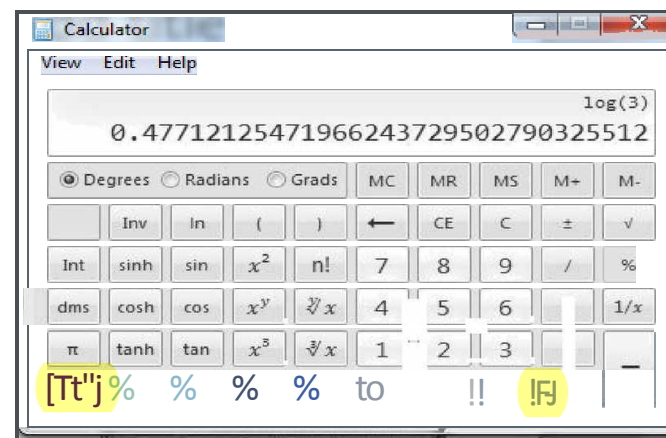
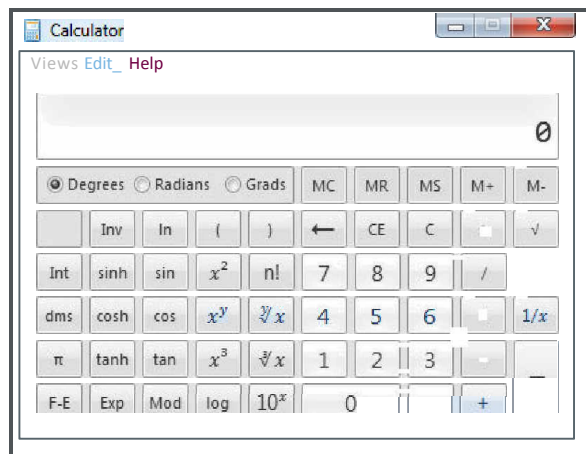
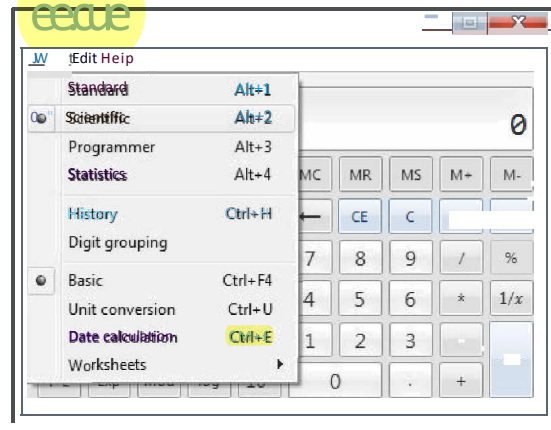
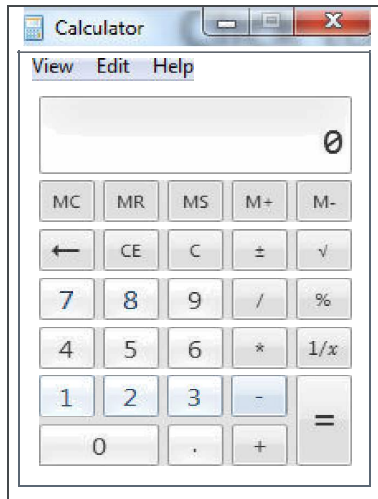
$$Z_f = 1.1513 \log(5.66) = 1.1513 (0.7528) = 0.87$$

$$\mu_z = 1.1513 \log \frac{1+\rho}{1-\rho} = 1.1513 \log \frac{1+0.5}{1-0.5} = 1.1513 \log \frac{1.5}{0.5} = 1.1513 \log(3)$$

$$= 1.1513 (0.4771) = 0.55$$

Calculating log Value

Windows calculator



Calculations

Now we have: $r = 0.7$, $n = 28$, $Z_f = 0.87$, $\mu_z = 0.55$

$$Z = \frac{Z_f - \mu_z}{\frac{1}{\sqrt{n-3}}}$$

$$Z = \frac{0.87 - 0.55}{1/\sqrt{28-3}} = \frac{(0.32)}{1/\sqrt{25}} = (0.32) \cdot (5) = 1.60$$

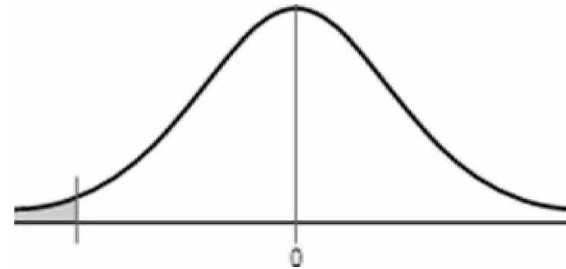
$$\Rightarrow Z_{\text{calculated}} = 1.60$$

Frequently used Z Critical Values

1°lypotksais Testing
wd Cnt'ical Values

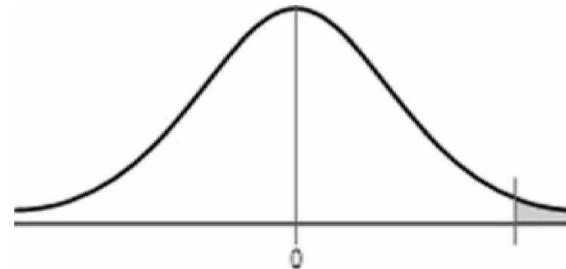
$$H_0: \mu = k \quad \begin{cases} \alpha = 0.10, \text{C.V.} = -1.28 \\ \alpha = 0.05, \text{C.V.} = -1.65 \\ \alpha = 0.01, \text{C.V.} = -2.33 \end{cases}$$

$$H_1: \mu < k$$



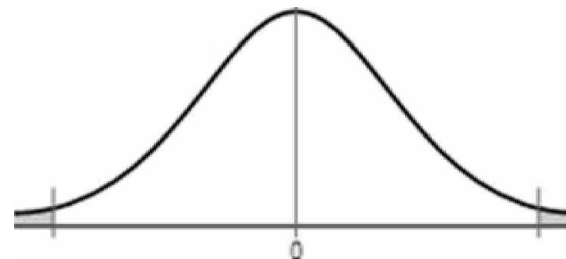
$$H_0: \mu = k \quad \begin{cases} \alpha = 0.10, \text{C.V.} = +1.28 \\ \alpha = 0.05, \text{C.V.} = +1.65 \\ \alpha = 0.01, \text{C.V.} = +2.33 \end{cases}$$

$$H_1: \mu > k$$



(6)Rly6t\sls6

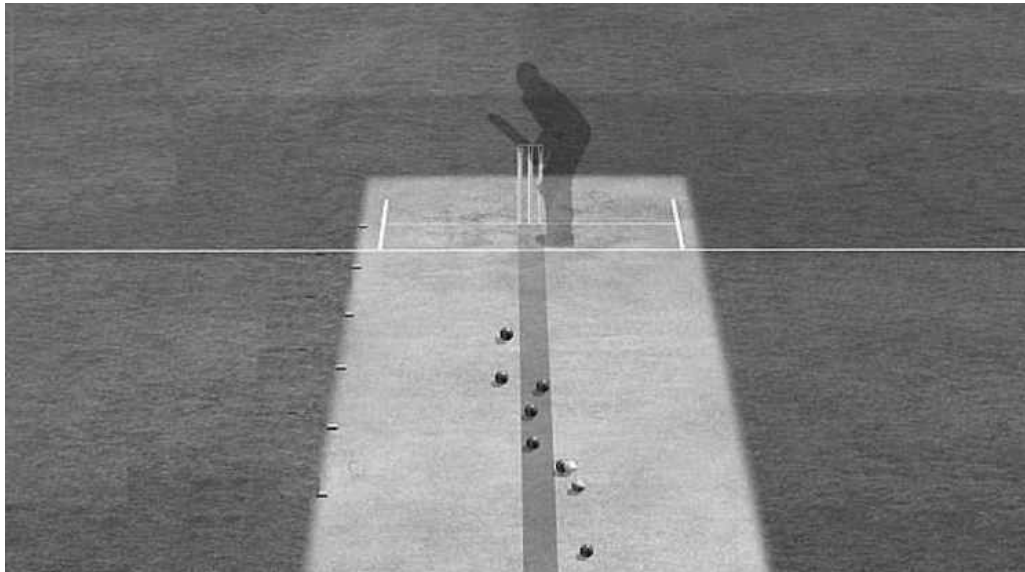
$$\alpha = 0.10, \text{I.V.} = \pm 1.65$$



Decision Rule

Compare Calculated value with Table Value

Step 5



Conclusion and Summarizing:

Since the computed value of $z = 1.60$ **does NOT fall** in the critical region, so we do not reject our null hypothesis and may conclude that the given claim is correct.

Summary

- When $\rho \neq 0$, then distribution of "r" is
 - Neither normal ✓
 - Nor becomes normal on large sample ✓

- We transform "r" into a new variable which is distributed normally. Hence we can use to for testing.

$$Z = \frac{Z_r - \mu_z}{\frac{1}{\sqrt{n-3}}}$$

- Its z-test.
- Rest of the testing procedure is same.

MLR

The END

Get Ready ...

- Check mic position
- Check resolution size
- Check pen
- Say something....

STA404

Regression & Correlation Analysis Lesson 10-2

Department of
Statistics Virtual
University of Pakistan

Statistical Inference about Simple Correlation

Testing hypothesis that the population correlation coefficient ρ is equal to zero

Hypothesis Testing – General Procedure

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

- | | |
|---------------|--|
| Step 1 | State the hypotheses and identify the claim. |
| Step 2 | Find the critical value(s) from the appropriate table |
| Step 3 | Compute the test value. |
| Step 4 | Make the decision to reject or not reject the null hypothesis. |
| Step 5 | Summarize the results. |

Notations:

- Correlation Coefficient
 - from SAMPLE data

Correlation Coefficient
from POPULATION data

- English letter “r”

- Greek small letter “ ρ ”
(rho)

ML

Two Cases

Testing Population
Correlation Coefficient ρ

$\rho = \rho_0 = \text{Some Specified Value}$
Z-Test

$\rho = 0$
T Test

Some Main Points

- Sampling distribution is NOT symmetry BUT it is interesting to note that when $p = 0$, then its symmetrical. 
- Above make it possible to use the t-test.
- T-test can be applied for any sample size
- What does it means No: $p = 0$
 - There is NO linear correlation between two variables

MLQ

Testing Hypothesis – Example

- A random sample of 20 pairs of observations gives a coefficient of 0.45.
- Test the hypothesis at the 0.05 level of significance that the correlation coefficient in the population is zero.

Testing Hypothesis – Example

Step 1

Stat the hypothesis.

$H_0 : \rho = 0$ (There is NO linear correlation between two variables)

$H_1 : \rho \neq 0$

Step 2

We set the level of significance (alpha) $\alpha = 0.05$ (5%)

Step 3

The test-statistic (formula) to be used is:

$$\text{Test Value} = \frac{(\text{Observed value}) - (\text{Expected value})}{\text{Standard error}}$$

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

It follows *t*-distribution with degree of freedom $df = n-2$

Calculations

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

$$t = \frac{0.45\sqrt{20-2}}{\sqrt{1-(0.45)^2}}$$

$$t = \frac{1.91}{0.89} = 2.14$$

$t = t\text{-Calculated Value} = 2.14$

Critical Value from t-table

Step 5 : Find the 0.05 column in the top row and 18 in the left-hand column.

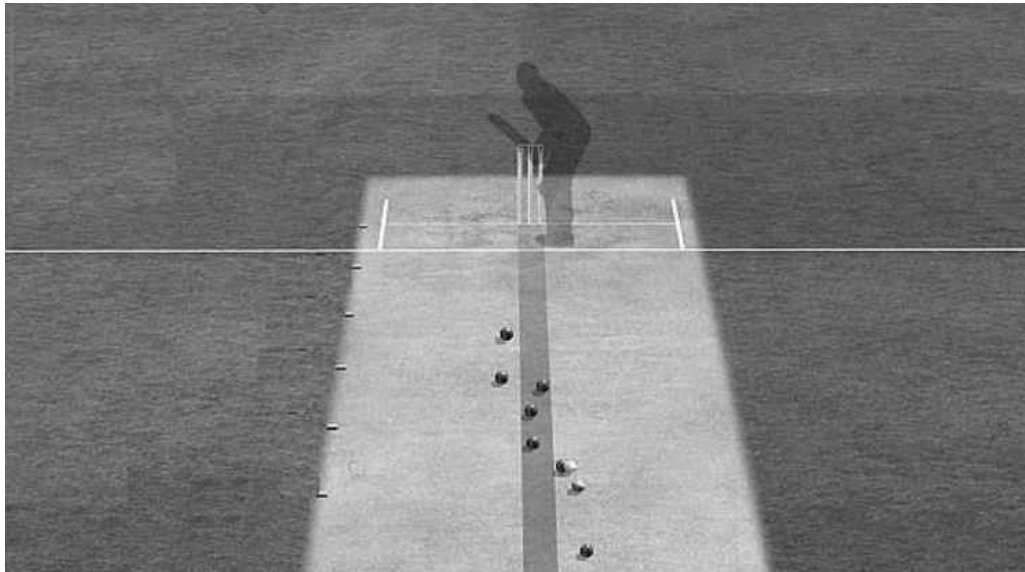
Where the row and column meet, the appropriate critical value is found.

Table F The t Distribution						
	Confidence intervals	80%	90%	95%	98%	99%
	One tail, α	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.20	0.10	0.05	0.02	0.01
1		30.8	6.314	12.006	31.821	63.657
2		1.638	2.920	4.032	6.965	9.925
3		1.533	2.353	3.182	5.841	8.451
4		1.476	2.132	2.776	5.408	7.171
5		1.440	2.015	2.571	5.208	6.908
6		1.415	1.943	2.447	5.143	6.708
7		1.397	1.895	2.365	5.099	6.581
8		1.383	1.861	2.306	5.041	6.495
9		1.372	1.833	2.262	5.000	6.426
10		1.363	1.812	2.228	4.963	6.377
11		1.356	1.796	2.201	4.937	6.335
12		1.350	1.782	2.179	4.917	6.299
13		1.345	1.771	2.160	4.901	6.270
14		1.341	1.761	2.145	4.888	6.250
15		1.338	1.753	2.131	4.877	6.239
16		1.335	1.746	2.120	4.868	6.232
17		1.333	1.740	2.110	4.860	6.226
18		1.331	1.735	2.101	4.854	6.222
19		1.330	1.731	2.093	4.850	6.219
20		1.329	1.728	2.088	4.847	6.217
21		1.328	1.726	2.084	4.845	6.216
22		1.327	1.724	2.081	4.844	6.215
23		1.326	1.723	2.079	4.843	6.214

Decision Rule

Compare Calculated value with Table Value

Step 5



Conclusion and Summarizing:

Since the computed value of $I = 2.14$ falls in the critical region, so we reject our null hypothesis and may conclude that the correlation coefficient in the population differ from zero.

It is similar to test the

- $H_0: \rho = 0$

- There is NO linear correlation between two variables.
- The two variables are INDEPENDENT

MLD

- $H_0: \beta = 0$

- There is NO linear correlation between two variables.
- The two variables are INDEPENDENT

Summary

- When $H_0: \rho = 0$, then distribution of " r " is symmetry. Therefore we can use t-test

- The formula is:

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \quad \checkmark$$

- Testing $H_0: \rho = 0$ is equivalent to test $H_0: \beta = 0$
- T-test can be applied to any sample size.

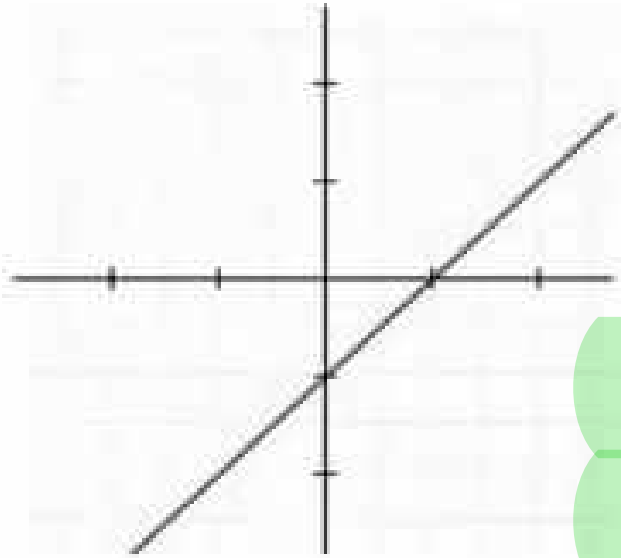
The END

Regression & Correlation Analysis

How to fit a Second Degree Parabola line?

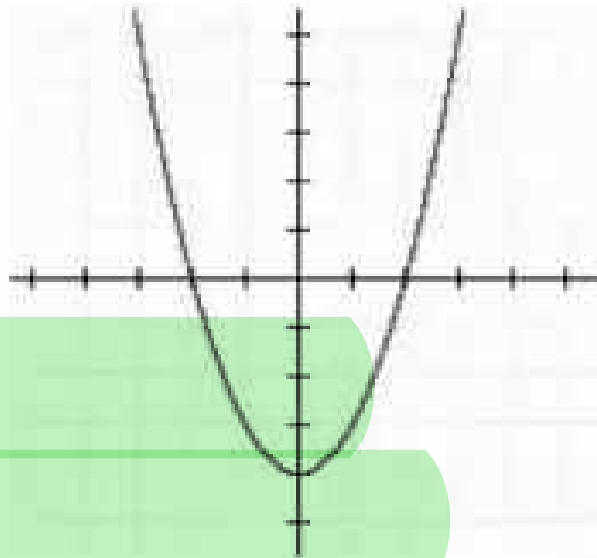
Setting up Normal Equations

Shape of equations/data ...



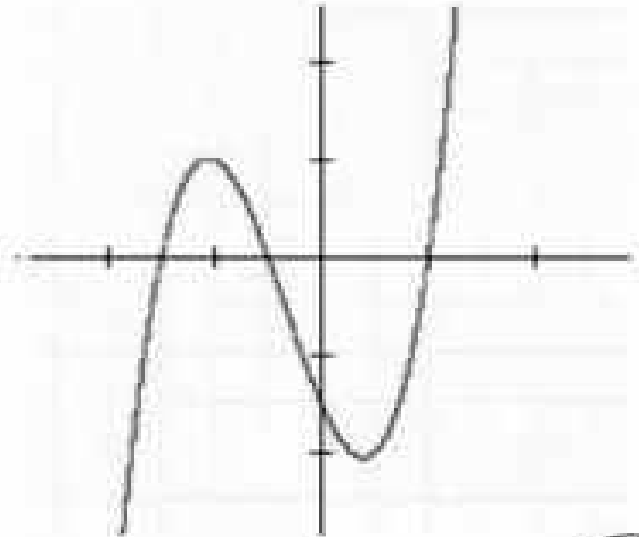
$$y = a + bx \text{ ①}$$

Straight line



$$y = a + bx + cx^2 \text{ ②}$$

Second degree Eq.
Parabola



$$y = a + bx + cx^2 + dx^3 \text{ ③}$$

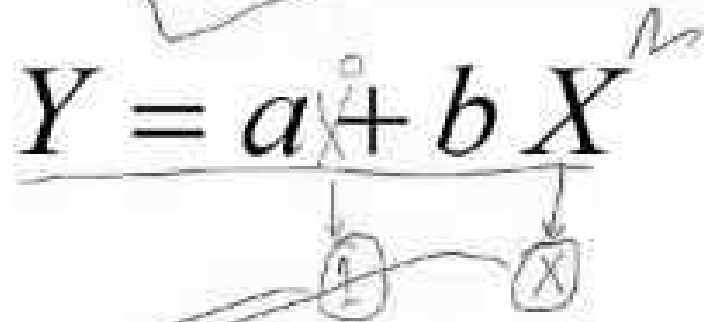
Third degree Eq.

Normal Equations?

- Normal equations are equations obtained by setting equal to zero the partial derivatives of the sum of squared errors (least squares)
- This is a technique for computing coefficients for equations
- It finds the regression coefficients *analytically*

How to make Normal Equations ...

$\boxed{n=1}$

$$Y = a + bX$$


$$Y = a + bX + cX^2$$

The normal Equations are

$$\sum (Y = a + bX) \longrightarrow \Sigma Y = \underline{na} + b\Sigma X$$

$$XY = aX + bX^2 \longrightarrow \Sigma XY = a\Sigma X + b\Sigma X^2$$

How to make Normal Equations ...

$$Y = a + bX + cX^2$$

The normal Equations are

$$1(Y = a + bX + cX^2)$$

$$\bullet \Sigma Y = na + b\Sigma x + c\Sigma x^2$$

$$X(Y = a + bX + cX^2)$$

$$\bullet \Sigma XY = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$X^2(Y = a + bX + cX^2)$$

$$\bullet \Sigma X^2Y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Test your Knowledge ...

Can you find normal equation for the following equation ?

$$Y = a + bX + cX^2 + dX^3$$

① ✓ ② ✓ ③ ✓ ④ ✓

$$\bullet \Sigma Y = na + b \Sigma X + c \Sigma X^2 + d \Sigma X^3$$

$$\bullet \Sigma XY = a \Sigma X + b \Sigma X^2 + c \Sigma X^3 + d \Sigma X^4$$

$$\bullet \Sigma X^2 Y = a \Sigma X^2 + b \Sigma X^3 + c \Sigma X^4 + d \Sigma X^5$$

$$\bullet \Sigma X^3 Y = a \Sigma X^3 + b \Sigma X^4 + c \Sigma X^5 + d \Sigma X^6$$

- The
END -

Regression & Correlation Analysis

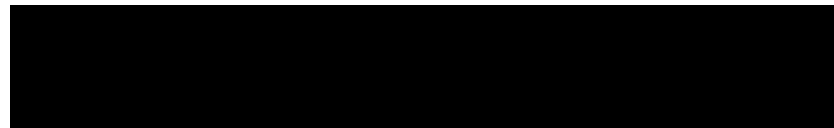
How to fit a Second Degree Parabola line?

Procedure and step by step
calculation with example

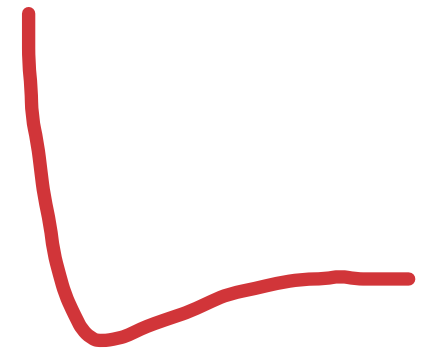
How to fit a Second Degree Parabola?

Example:

Fit a second degree parabola to the following data, taking X as independent variable.



0	1
1	1.8
2	1.3
3	2.5
4	6.3



Equation of Second Degree ...

$$Y = a + bX + cX^2 \checkmark$$

The normal Equations are

$$\bullet \Sigma Y = na + b\Sigma X + c\Sigma X^2 \checkmark$$

$$\bullet \Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3 \checkmark$$

$$\bullet \Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 \checkmark$$

What we need...?

$$\Sigma Y \quad \Sigma X \quad \Sigma X^2$$

$$\Sigma XY \quad \Sigma X^3$$

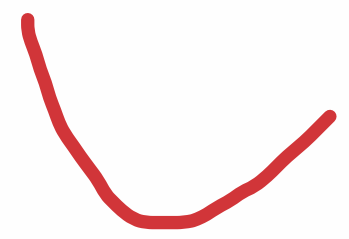
$$\Sigma X^2 Y \quad \Sigma X^4$$

n

Required Calculation

X	Y	XY	XX	XXY	XXX	XXXX
0	1.0	0	0	0	0	0
1	1.8	1.8	1	1.8	1	1
2	1.3	2.6	4	5.2	8	16
3	2.5	7.5	9	22.5	27	81
4	6.3	25.2	16	100.8	64	256
Total $\sum x = 10$	12.9	37.1	30	130 130.3	100	354

Substituting the values in the normal Equations.....



$$\Sigma Y = na + b\Sigma X + c\Sigma X^2$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3$$

$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4$$

$$129 = 5a + 10b + 30c \text{ — (1)}$$

$$371 = 10a + 30b + 100c \text{ — (2)}$$

$$130.3 = 30a + 100b + 354c \text{ — (3)}$$

Solve these equations to eliminate "a"

$$12.9 = 5a + 10b + 30c \quad \text{--- (1)}$$

$$37.1 = 10a + 30b + 100c \quad \text{--- (2)}$$

$$130.3 = 30a + 100b + 354c \quad \text{--- (3)}$$

② Multiply eq (1) by 2, and subtract it from eq (2)

$$37.1 = 10a + 30b + 100c$$

$$- 25.8 = -10a + 20b + 60c$$

$$11.3 = 10b + 40c \quad \text{--- (4)}$$

Multiply eq (2) by 3 and subtract it from eq (3)

$$130.3 = 30a + 100b + 354c$$

$$- 111.3 = -30a + 90b + 300c$$

$$19 = 0 + 10b + 54c$$

Now, Solve these equation to eliminate "b"

$$11.3 = 10b + 40c \quad \text{--- (4)}$$

$$19 = 10b + 54c \quad \text{--- (5)}$$

$$-7.7 = -14c$$

$$\Rightarrow c = \frac{-7.7}{-14} = 0.55 \quad \boxed{c = 0.55} \quad !$$

Now to put the value of c in eq (4)

$$11.3 = 10b + 40(0.55)$$

$$11.3 = 10b + 22$$

$$-10.7 = 10b \quad \Rightarrow b = -1.07$$

Now, put value of "b" and "c" in Eq...1

$$12.9 = 5a + 10b + 30c \quad \text{--- (1)}$$

put the value of b & c in eq

$$12.9 = 5a + 10(-1.07) + 30(0.55)$$

$$12.9 = 5a - 10.7 + 16.5$$

$$7.1 = 5a$$

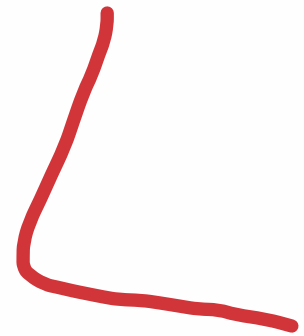
$$\Rightarrow \boxed{a = 1.42}$$

So by solving three normal equations simultaneously, we are able to find the values

$$a = 1.42$$

$$b = -1.07$$

$$c = 0.55$$

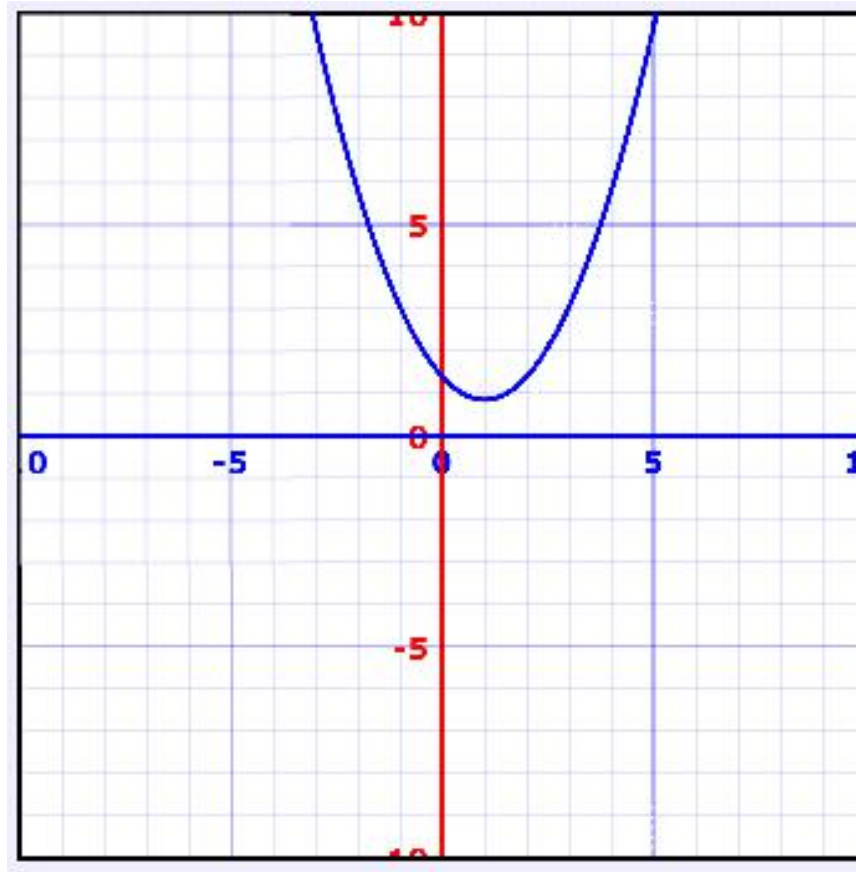


$$Y = a + bX + cX^2$$

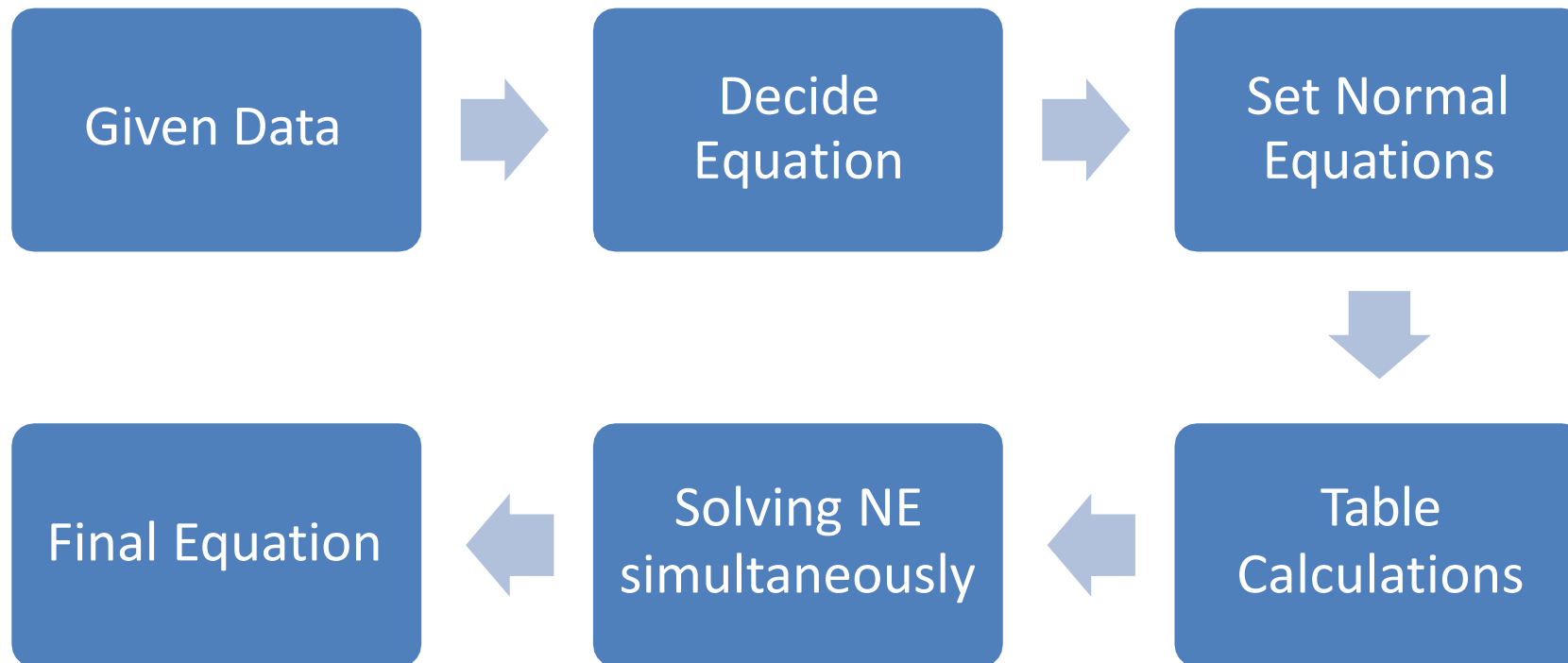
$$\hat{y} = 1.42 - 1.07X + 0.55X^2$$

Find fitted y 0.12

Just for fun ... Graph of this equation



Recap



- The
END -

Test 1...2...3

- Mic/Audio test
- Screen Area test
- Pen test

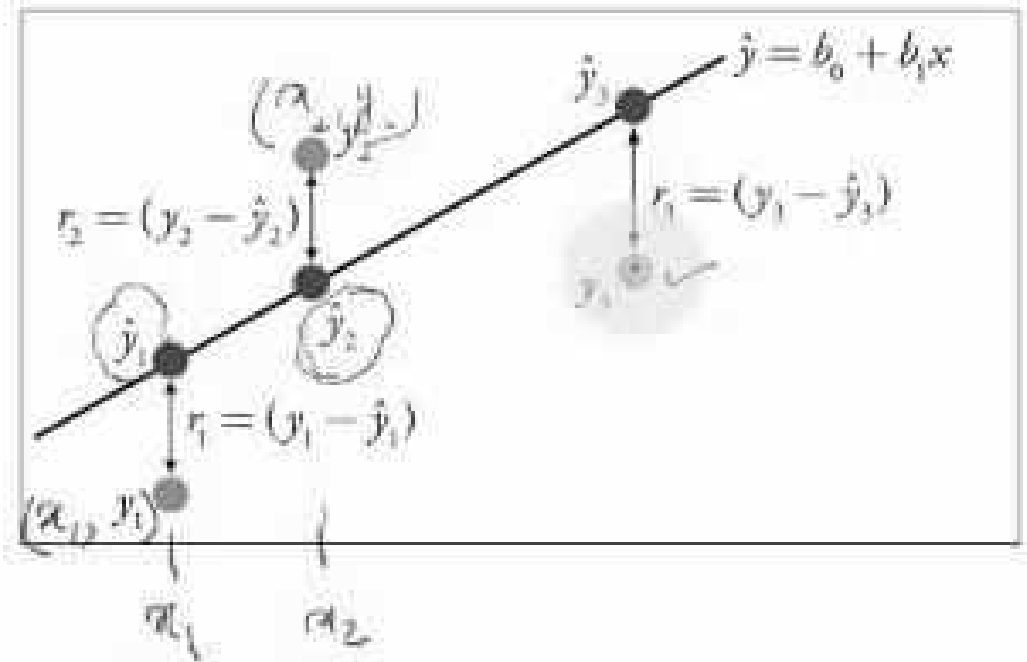
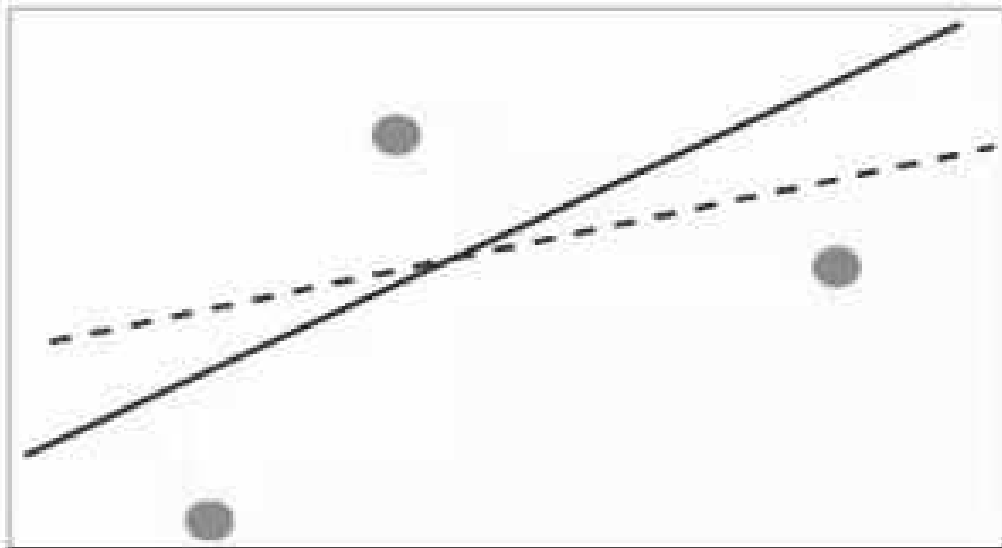
How to Find the Standard Deviation of Regression lecture 12

It is also called Standard Error of
Estimate

Basic Idea

- The observed values of (X, Y) do not fall on the regression Line but scatter away from it.
- The degree of scatter (dispersion) of the observed values about the regression line is measured by a method/formula.
- It is called standard deviation of regression
- It is also known as “Standard Error of Estimate”

Visual Seen ...



Formula ...

- For population data

$$\sigma_{Y,X} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{N}}$$

- For sample data

$$s_{Y,X} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n-2}}$$

MLQ

Example

Example:

- Find the values of $\hat{Y}$.
- Find the residual values "e".
- Show that $\sum(Y - \hat{Y}) = 0$.
- Compute the standard error of estimate s_{est} .

X	Y
5	16
6	19
8	23
10	28
12	36
13	41
15	44
16	45
17	50

Fitting regression line

You can do it by yourself ...

$$Y = a + bX$$

$b = \checkmark$

→ 2.831 ✓

Slop

$a = \checkmark$

→ a = 1.4

D.Y.S

Finding $\hat{y}$

X	Y	$\hat{y} = 1.47 + 2.831X$
5	16	$1.47 + 2.83(5) = 15.63$
6	19	$1.47 + 2.83(6) = 18.46$
8	23	24.12
10	28	✓
12	36	✓
13	41	✓
15	44	✓
16	45	✓
17	50	✓

Residual values and showing $\Sigma(Y - \hat{Y}) = 0$

X	Y	$\hat{Y}$	$e = Y - \hat{Y}$	
5	16	15.63	$16 - 15.63 = 0.38$	e_1
6	19	18.46	$19 - 18.46 = 0.54$	e_2
8	23	24.12	$23 - 24.12 = -1.12$	e_3
10	28	29.78	$28 - 29.78 = -1.78$	
12	36	35.44	0.56	
13	41	38.27	2.73	
15	44	43.94	0.07	
16	45	46.77	-1.77	
17	50	49.60	0.40	

$\Sigma(Y_i - \hat{Y}_i) = 0$
 $\Sigma e = \Sigma \hat{e}$

Showing Standard error

X	Y	$\hat{Y}$	$e = Y - \hat{Y}$	$(Y - \hat{Y})^2$
5	16	15.63	0.38	0.14
6	19	18.46	0.54	0.30
8	23	24.12	-1.12	1.25
10	28	29.78	-1.78	3.17
12	36	35.44	0.56	0.31
13	41	38.27	2.73	7.44
15	44	43.94	0.07	0.00
16	45	46.77	-1.77	3.12
17	50	49.60	0.40	0.16

SUM =

15.89

Formula ...

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - 2}}$$
$$= \sqrt{\frac{15.89}{9 - 2}}$$

$$s_{y \cdot x} = 1.51$$

Standard Error!

Recap

Given Data

Find a and
 b

Find $\hat{Y}$

Final Value

Square and
simplify

$(Y - \hat{Y})$

- The
END -

Test 1...2...3

- Mic/Audio test
- Screen Area test
- Pen test

Regression & Correlation Analysis

How to Find the Standard Deviation of Regression lecture12

Example with formula

Example #2

Example:

- Compute the standard error of estimate $\underline{s_{y.x}}$

X	Y
20	280
24	360
36	450
32	420
28	400
38	500
32	475
26	320

Formula for Standard Error...

Recall

$$s_{y,x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

So

$$s_{y,x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}}$$

already ✓

$$\hat{Y} = a + bx$$

or

$$s_{y,x} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$\sum(Y - \hat{Y})^2 = \sum(Y - a - bx)^2$$

First we fit the Regression line

$$\hat{y} = a + bx$$

Formula for a :

$$a = \bar{y} - b(\bar{x})$$

$$\bar{x} = ? \checkmark$$

$$\bar{y} = ? \checkmark$$

Formula for b :

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$\bar{y} = ?$$

$$\sum xy = ?$$

$$\sum x = ?$$

$$\sum y = ?$$

$$\sum x^2 = ?$$

Required Calculation

Day	X	Y	XY	X ²	Y ²
1	20	280	5600	400	78400
2	24	360	8640	576	129600
3	36	450	16200	1296	202500
4	32	420	13440	1024	176400
5	28	400	11200	784	160000
6	38	500	19000	1444	250000
7	34	475	16150	1156	225625
8	26	320	8320	676	102400
Total	238	3205	98550	7356	1324925

– From previous table we have:

$$n = 8$$

$$\Sigma x = 238$$

$$\Sigma y = 3205$$

$$\Sigma x^2 = 7356$$

$$\Sigma xy = 98550$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{3205}{8} = 400.62$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{238}{8} = 29.75$$

– Now putting these values in “b” and “a”, and by solving, we get:

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{8(98550) - (238)(3205)}{8(7356) - (238)^2} = \frac{25610}{2204} = \boxed{11.62}$$

$$a = \bar{y} - b(\bar{x}) = 400.62 - 11.62(29.75) = \boxed{54.92}$$

- Putting the values of “a” and “b” in the equation, we get:

$$\hat{y} = 54.92 + 11.62(x)$$

Using Formula ...

$$s_{y.x} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

Handwritten calculation:

$$s_{y.x} = \sqrt{\frac{1324925 - (54.92)(3205) - (11.62)(98550)}{8-2}}$$

Intermediate result:

$$= \sqrt{\frac{3755.4}{6}}$$

Final result:

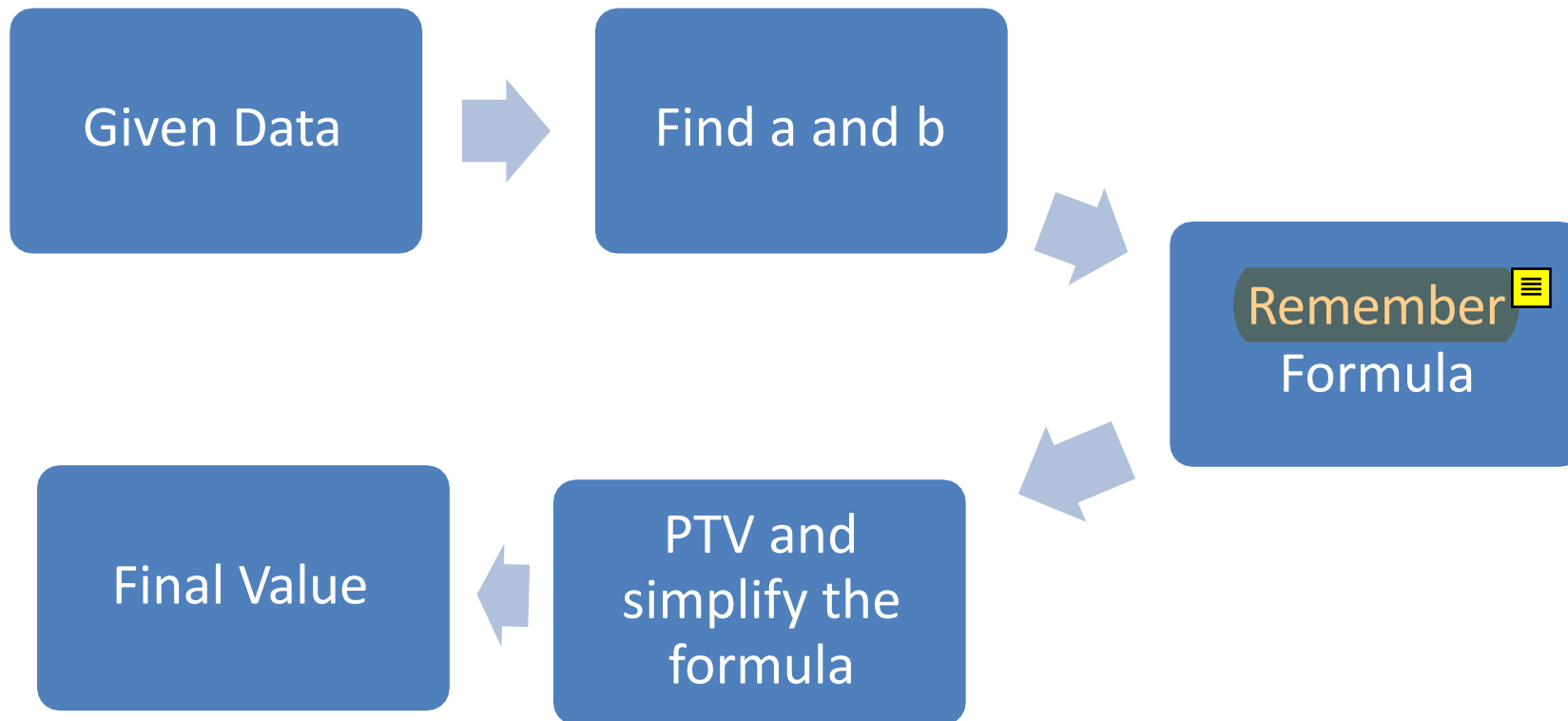
$$= 25.02$$

Summary of values:

- $n = 8$
- $a = 54.92$
- $b = 11.62$
- $\sum y = 3205$
- $\sum y^2 = 1324925$
- $\sum xy = 98550$

S. Eashankar

Recap



- The
END -

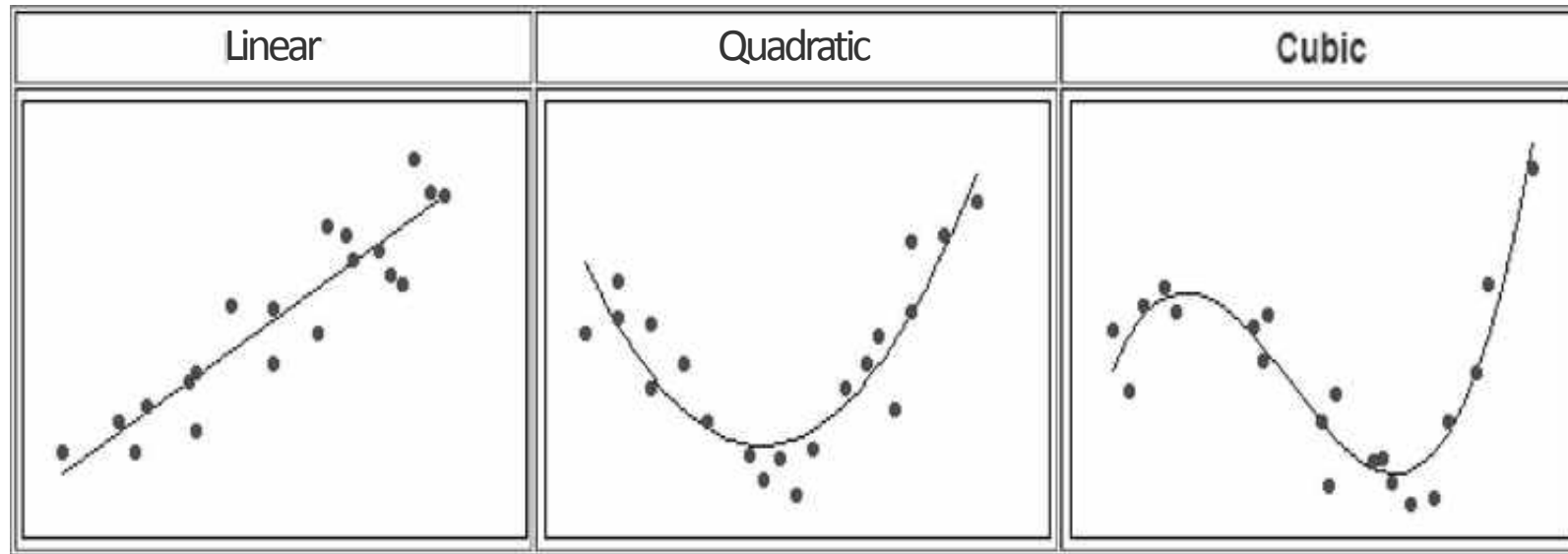
Test 1...2...3

- Mic/Audio test
- Screen Area test
- Pen test

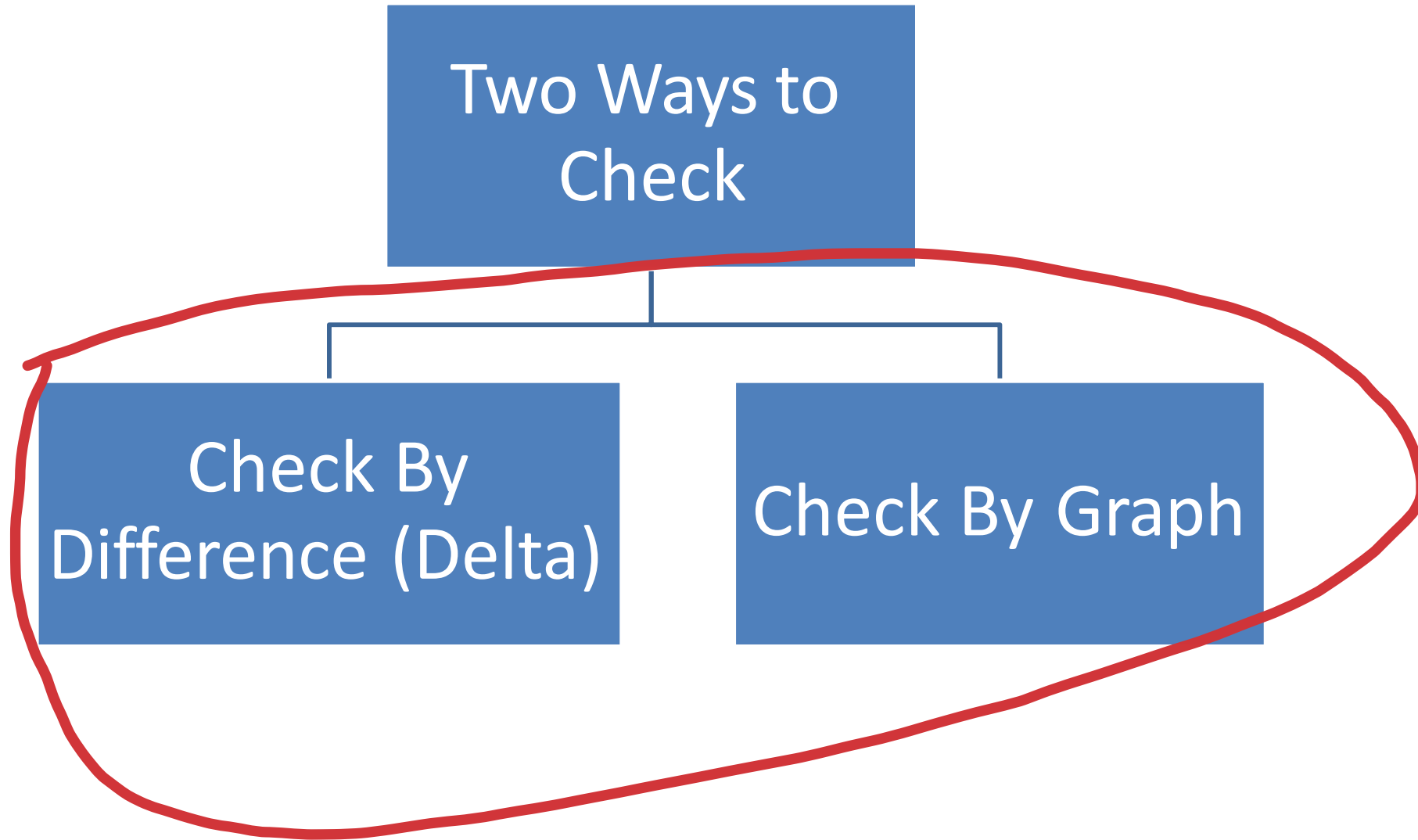
Deciding the Type of Curve to fit

Which curve is suitable in which situation

Lines/Curves come in different shapes ..



Choosing the right line...



Example #1

Question:

Decide the suitable curve for the following data.



1	12
2	15
3	18
4	21

Lets do it

Solution:

Let, first we take its first difference as below and see whether it is constant or not?

X	Y	First Difference
1	12	
2	15	$15 - 12 = 3$
3	18	$18 - 15 = 3$
4	21	$21 - 18 = 3$

Delta Y
 ΔY

ΔX

Criteria for suitable curve

Some more rules:

- If the FIRST DIFFERENCES of Y are *approximately* constant, use a straight line
- If the SECOND DIFFERENCES of Y are *approximately* constant, use a second degree order equation (parabola)
- If the THIRD DIFFERENCES of Y are *approximately* constant use a third degree order equation

Criteria for suitable curve

Here are few rules:

- If the FIRST DIFFERENCES of $\log Y$ are *approximately* constant, use a exponential curve
- If the FIRST DIFFERENCES of $\log Y$ and $\log X$ are *approximately* constant, use a Geometric curve
- If the FIRST DIFFERENCES of reciprocals of Y are *approximately* constant, use a third degree parabola

Example #2



Question:

Decide the suitable curve for the following data.



1	10
2	14
3	21
4	31
5	44

Lets do it

SOLUTION:

- we take its first difference
- and see whether it is constant or not?
- If not then go for second difference

X	Y	First Difference	Second Difference
1	10	$\rightarrow$ 4	=
2	14	$14 - 10 = 4$	-
3	21	$21 - 14 = 7$	$7 - 4 = 3$
4	31	$31 - 21 = 10$	$10 - 7 = 3$
5	44	$44 - 31 = 13$	$13 - 10 = 3$

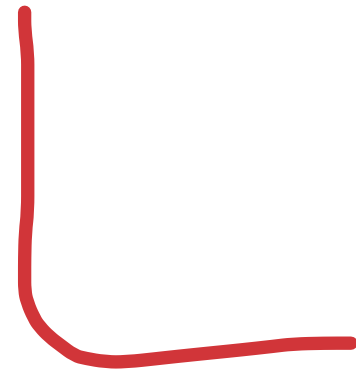


Example #3

Question:

Suggest a suitable curve for the given data?

0	10
1	17
2	28
3	43
4	62



Remember

- Straight Line:

$$Y = a + bX$$

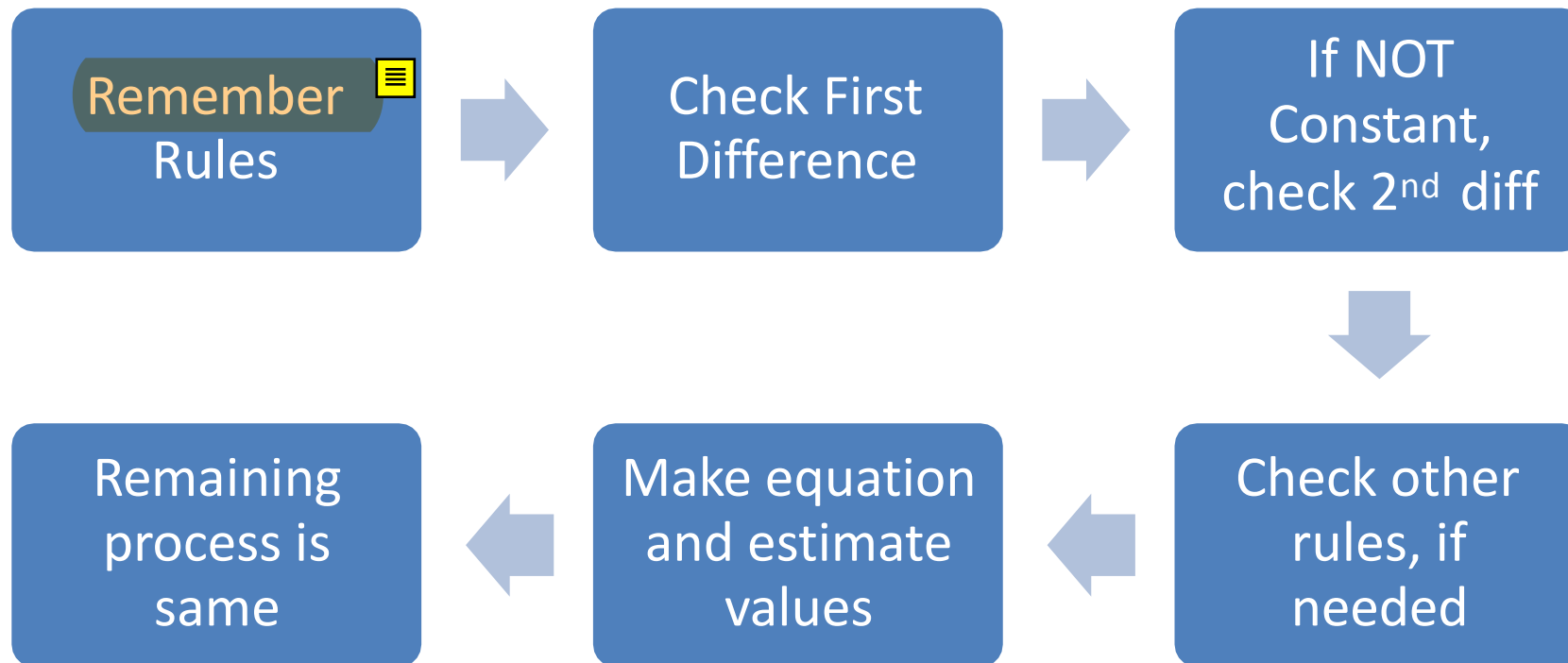
- Second Degree line (Also called Parabola)

$$Y = a + bX + cX^2$$

- Third Degree Line

$$Y = a + bX + cX^2 + dX^3$$

Recap



- The
END -

Test 1...2...3

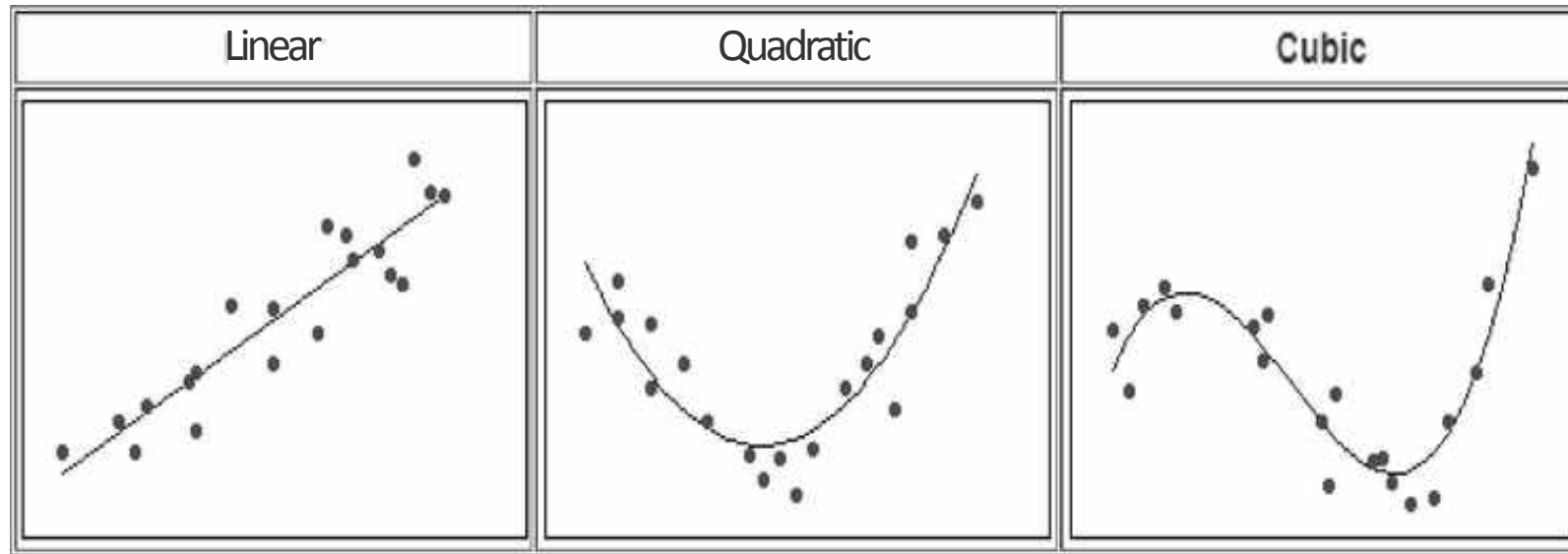
- Mic/Audio test
- Screen Area test
- Pen test

Regression & Correlation Analysis

Deciding the Type of Curve to fit

Use graph to get the idea of suitable equation

Lines/Curves come in different shapes ..



Criteria for suitable curve

Here are few guidelines:

- If the graph of the values of X and Y gives a straight line, use a straight line.
- If the graph of the values of X and Y gives a curve with only one bend, use a second degree parabola.
- If the graph of the values of X and Y gives a curve of the shape of reverse "S", use a third degree parabola.

Criteria for suitable curve

Few more guidelines:

- If the graph of the values of X and $\log Y$ gives a straight line, use an exponential curve .
- If the graph of the values of $\log X$ and $\log Y$ gives a straight line, use a geometric or logarithmic curve .
- If the graph of the values of X and $1/Y$ gives a straight line, use a hyperbola curve.

Lets explore few ... Example #1

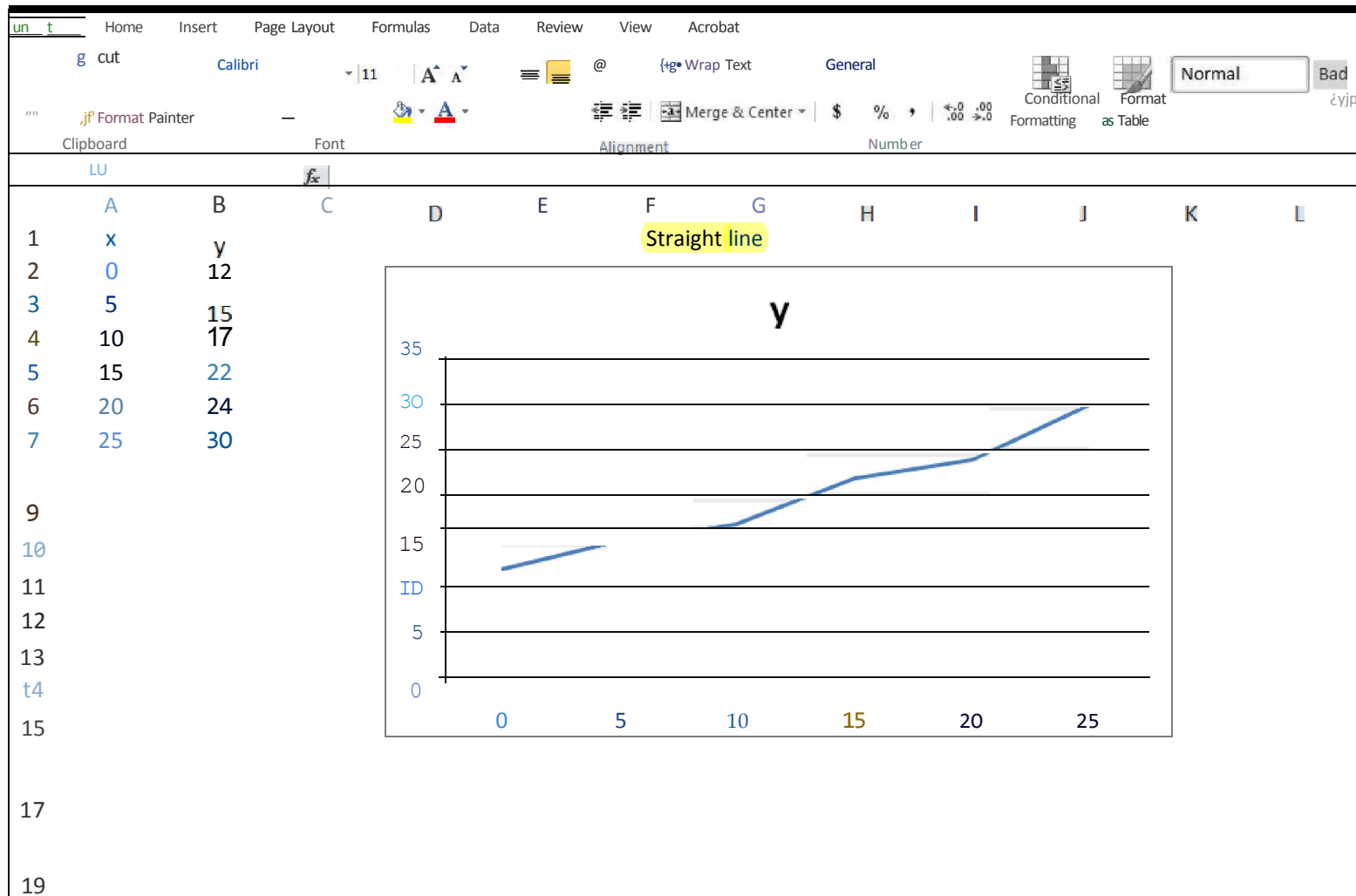
Question:

By plotting, which type of line/curve is appropriate for the data?

x	y
0	12
5	15
10	17
15	22
20	24
25	30



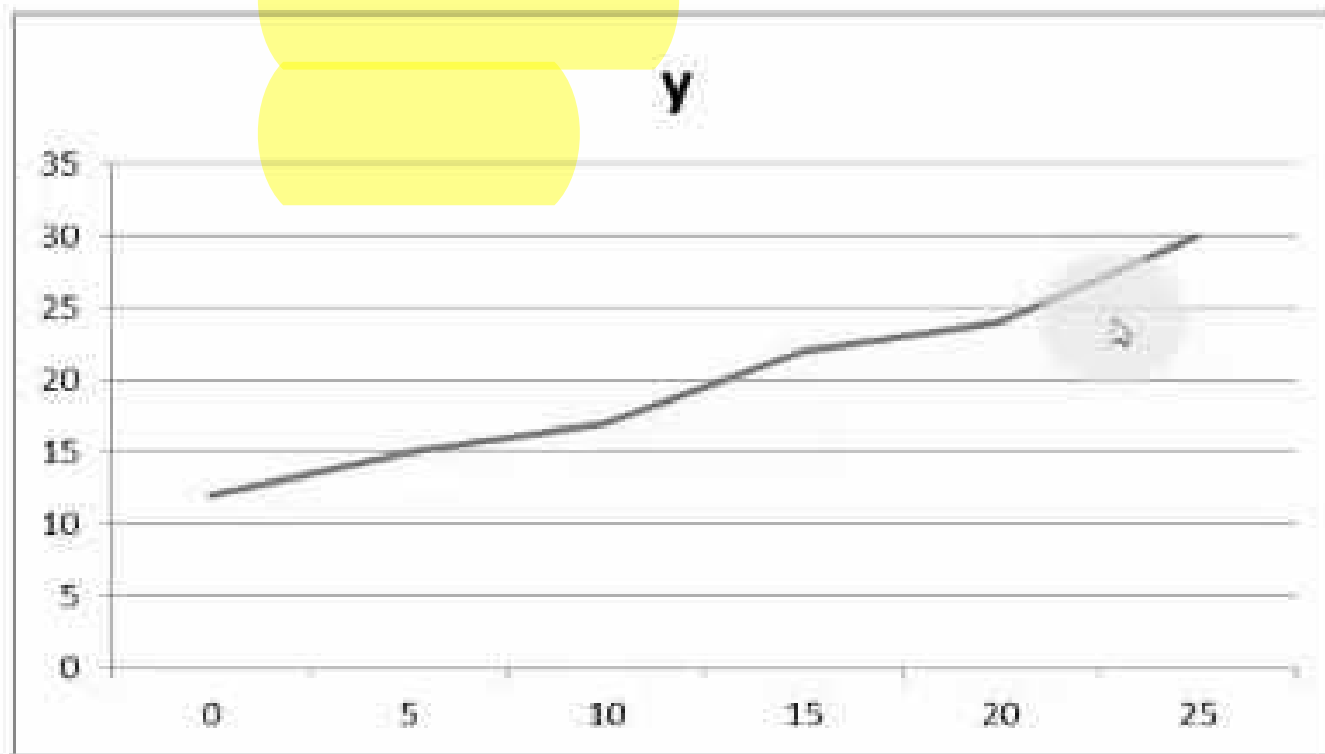
Lets Go To Excel



Straight line ...

Solution:

- If the graph of the values of X and Y gives a straight line, use a straight line .

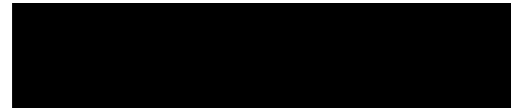


Example #2



Question:

Decide the suitable curve
for the following data.

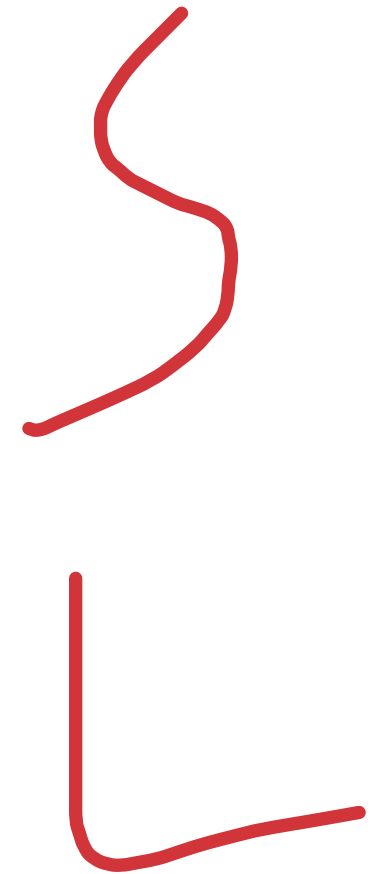
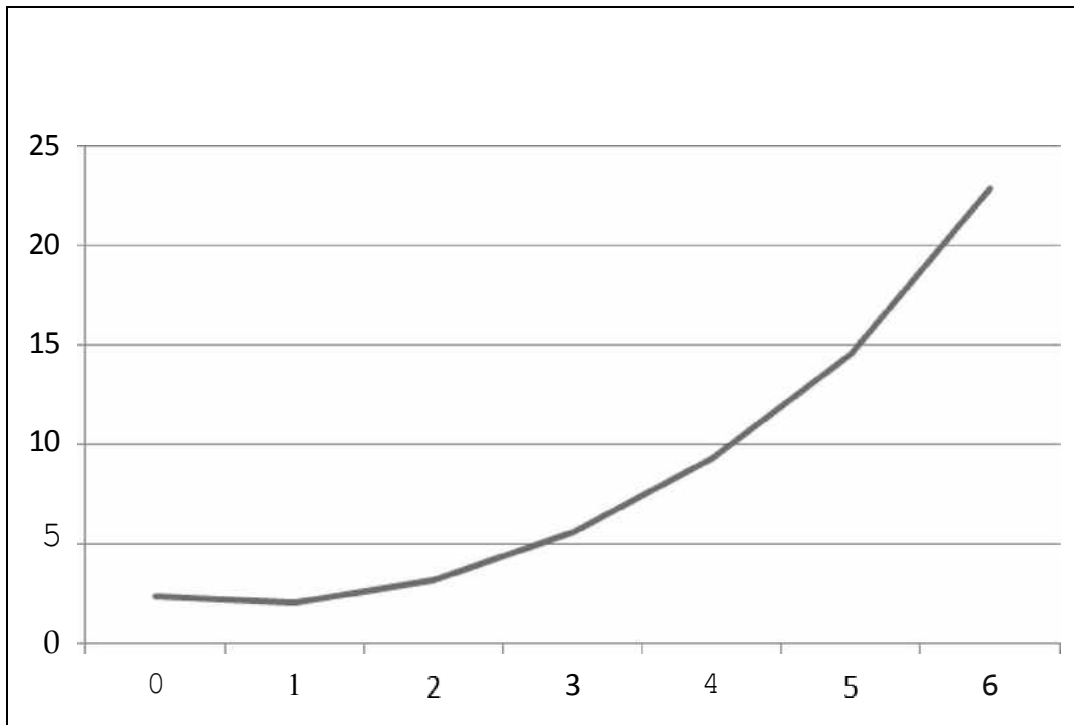


0	2.4
1	2.1
2	3.2
3	5.6
4	9.3
5	14.6
6	22.9

Second Degree Parabola ...

SOLUTION:

- If the graph of the values of _____ gives a curve with only one bend, use a second degree parabola.



Example #3

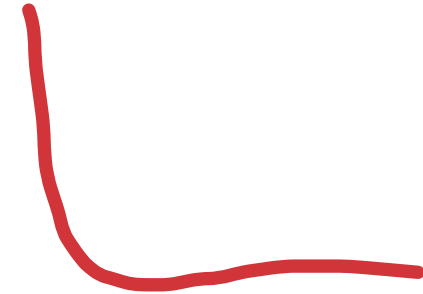


Question:

Decide the suitable curve
for the following data.



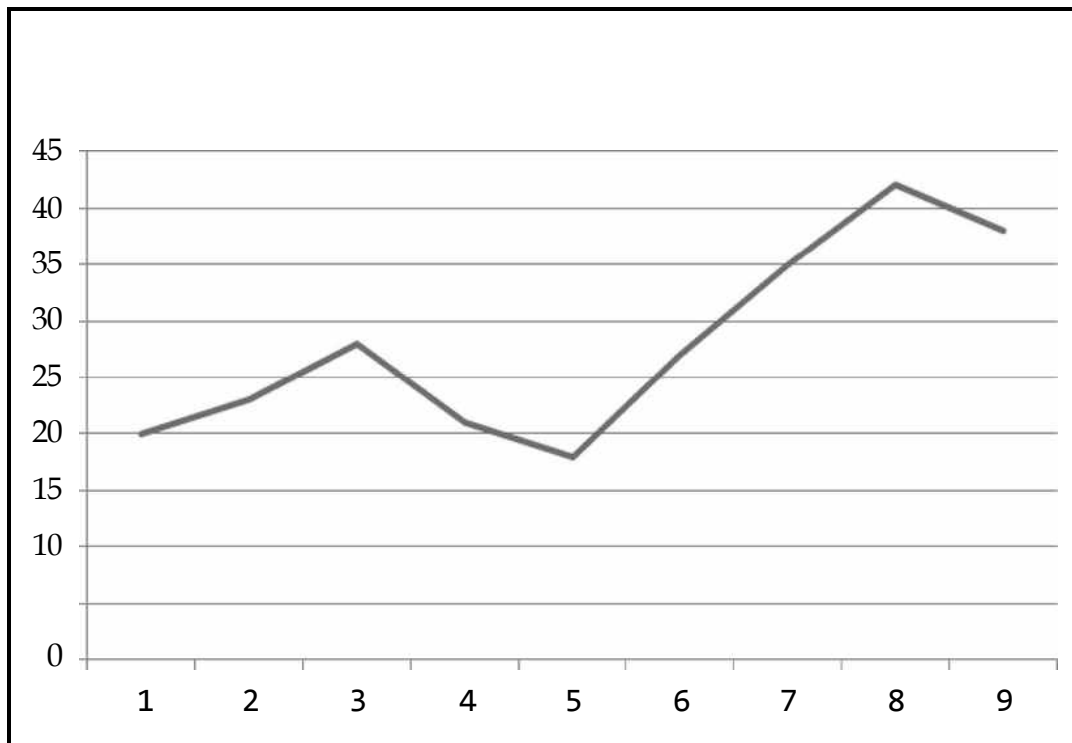
1	20
2	23
3	28
4	21
5	18
6	27
7	35
8	42
9	38



Third Degree Parabola...

SOLUTION:

- If the graph of the values of _____ gives a curve of the shape of reverse “S”, use a third degree parabola.

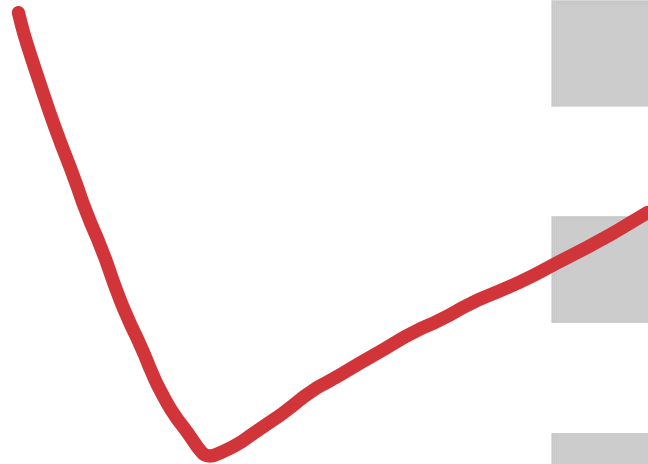


Example #4....



Question:

Suggest a suitable curve
for the given data?

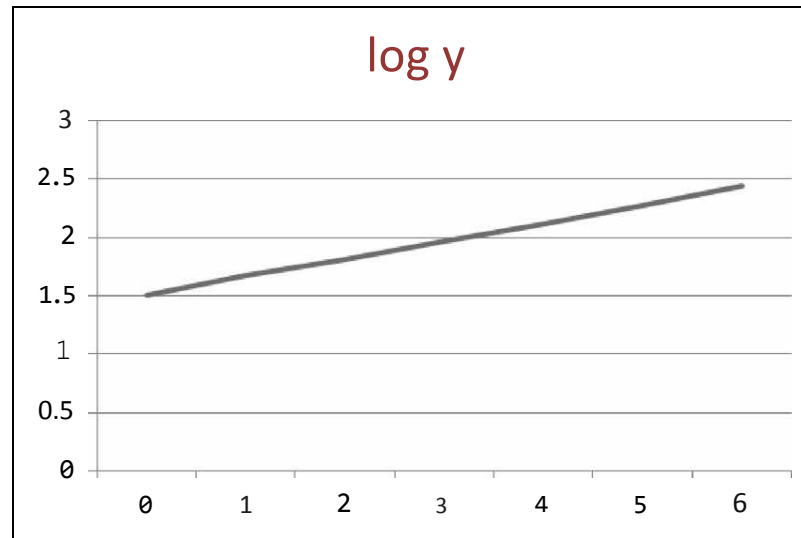
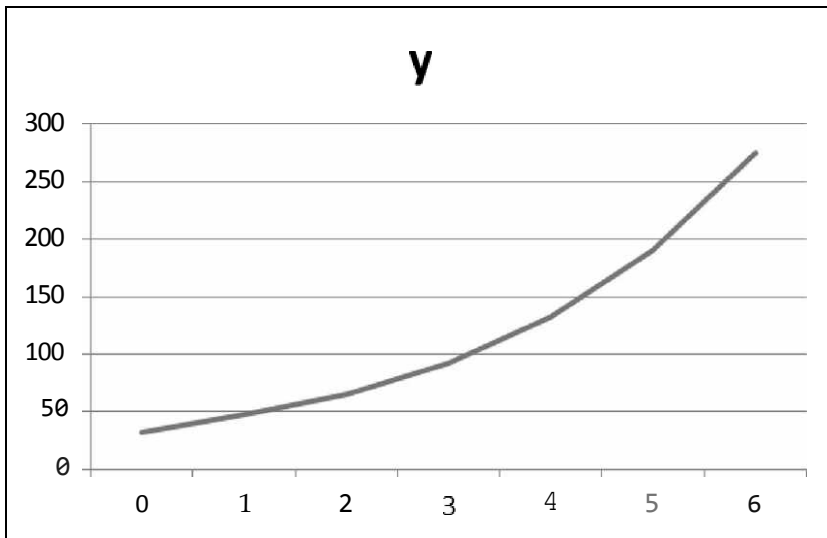


0	32
1	47
2	65
3	92
4	132
5	190
6	275

Exponential Curve ...

SOLUTION:

- If the graph of the values of X and $\log Y$ gives a straight line, use an exponential curve .



Remember

- Straight Line:

$$Y = a + bX$$

- Second Degree line (Also called Parabola)

$$Y = a + bX + cX^2$$

- Third Degree Line

$$Y = a + bX + cX^2 + dX^3$$

Remember

$$Y = ab^X \ggg \log Y = \log a + X \log b$$

$$Y = A + BX$$

$$Y = ae^{bX} \ggg \log Y = \log a + (b \log e)X$$

$$Y' = A + BX$$

$$\frac{1}{Y} = a + bX \ggg \underline{Y' = a + bX}$$

Recap

Remember Rules

Draw plot of data

Observe the plot

Also apply verify
from mathematical
tools

Graph only gives
you rough ideas

If not clear picture,
draw some other
values

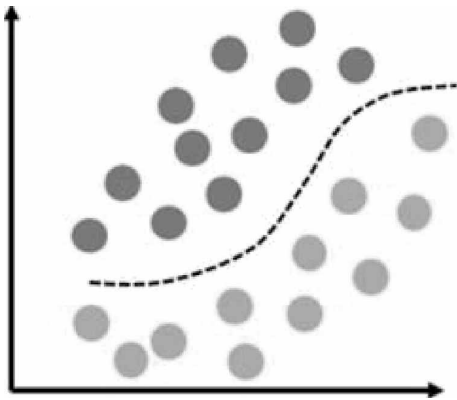
- The
END -

Converting non-linear into linear form

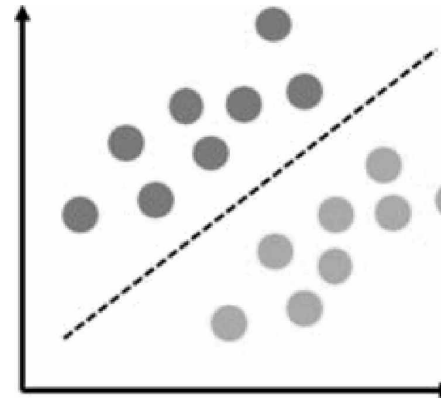
Some useful transformations to convert non-linear equations into linear form

Non-linear vs Linear

Nonlinear



Linear



#1

Non-linear
form

$$Y = aX^b$$

$$y = a + bx$$

Transformation

Taking log on both sides,

$$\log Y = \log(ax^b)$$

$$\log Y = \log a + \log(X^b)$$

$$\log Y = \log a + b \log X$$

Let $Y' = \log Y$, Let $A = \log a$, $\log X = X'$

Linear form

$$Y' = A + bX'$$

#2

Non-linear
form

$$Y = ab^X$$

Transformation

Taking log on both sides,

$$\begin{aligned}\log Y &= \log(ab^X) \\ &= \log a + \log b^X \\ &= \log a + X \log b \\ \log Y &= \log a + (\log b)X\end{aligned}$$

Linear form

$$Y' = A + BX$$

#3

Non-linear form	$\frac{1}{Y} = a + bX$
Transformation	Tranform $\frac{1}{Y}$ only...
Linear form	$Y' = a + bX$

#4

Non-linear
form

$$Y = \frac{1}{a + bX}$$

Transformation

Take Reciprocal / Transform Y only...

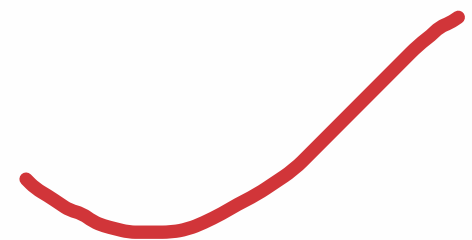
$$\left(\frac{1}{Y}\right)' = a + bX$$

$$Y' = a + bX$$

Linear form

$$Y' = a + bX$$

#5



Non-linear form	$\frac{1}{Y} = a + \frac{b}{1+X}$
Transformation	Change $\frac{1}{Y}$ and $\frac{1}{1+X}$ Let $\frac{1}{Y} = Y'$ Let $\frac{1}{1+X} = X'$ $Y' = a + bX'$
Linear form	$Y' = a + bX'$



#6

Non-linear form	$Y = a + b\sqrt{X}$
Transformation	Let $\sqrt{X} = X'$ So $Y = a + bX'$
Linear form	$Y = a + bX'$

#7

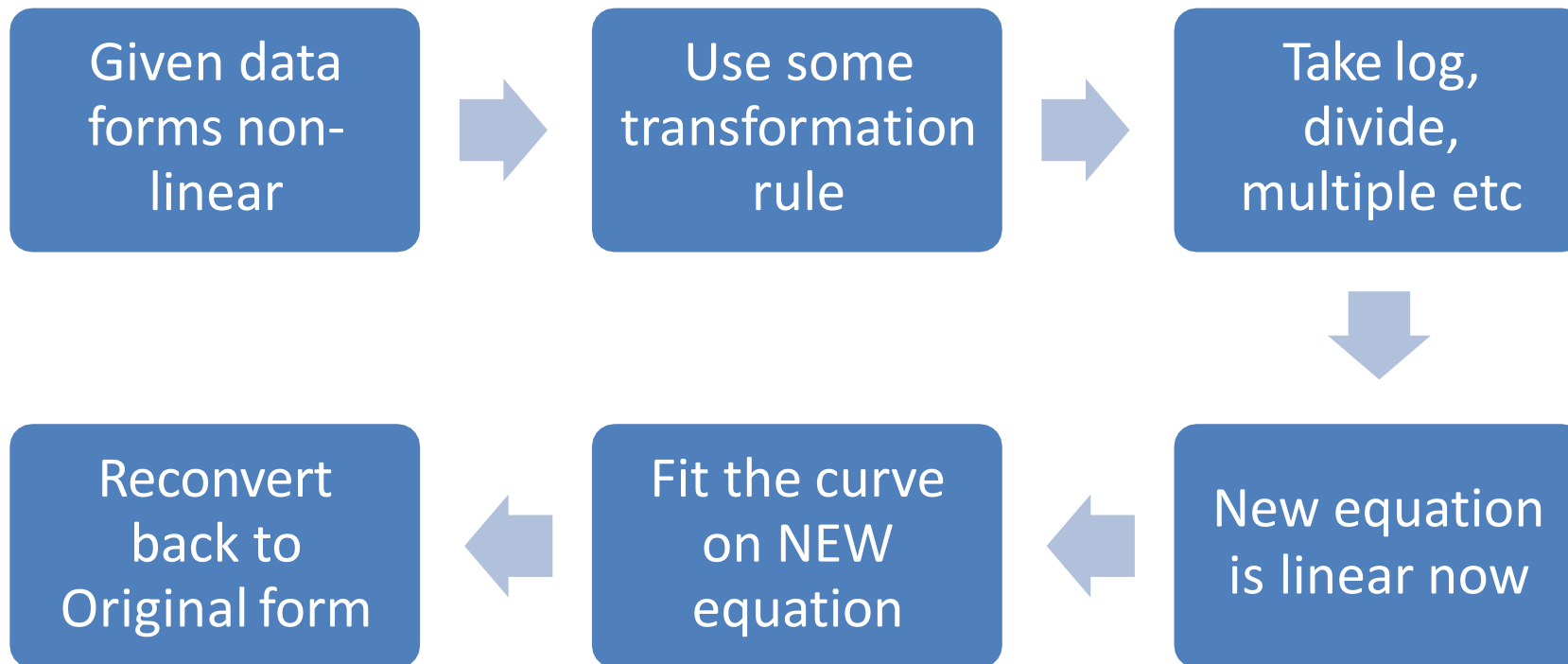


Non-linear form	$Y = aX^2 + bX$
Transformation	Divide all by X $\frac{Y}{X} = \frac{aX^2}{X} + \frac{bX}{X}$ Let $\left(\frac{Y}{X}\right) = Y'$ $Y' = aX + b$ Y-intercept $b = Y - b \cdot \text{slope} = \checkmark$ $A = \text{slope} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$
Linear form	slope = a $Y' = aX + b$

Summary

Non-linear form	Let/Suppose/Transformation	Linear form
$Y = aX^b$	$Y' = \log Y, A = \log a, X' = \log X$ ✓	$Y' = A + bX'$ ✓
$Y = ab^X$	$Y' = \log Y, A = \log a, B = \log b$ ✓	$Y' = A + BX$
$Y = \frac{1}{a + bX}$	$Y' = \frac{1}{Y}$ ✓	$Y' = a + bX'$
$\frac{1}{Y} = a + bX$	$Y' = \frac{1}{Y}$ ✓	$Y' = a + bX'$
$\frac{1}{Y} = a + \frac{b}{1+X}$	$Y' = \frac{1}{Y}, X' = \frac{1}{1+X}$ ✓	$Y' = a + bX'$
$Y = a + b\sqrt{X}$	$X' = \sqrt{X}$ ✓	$Y' = a + bX'$
$Y = aX^2 + bX$	$Y' = \frac{Y}{X}$ ✓	$Y' = aX + b$

Recap



- The
END -

How to fit an Exponential Curve?

Lecture 14

Procedure and step by step
calculation with example

How to fit an Exponential Curve

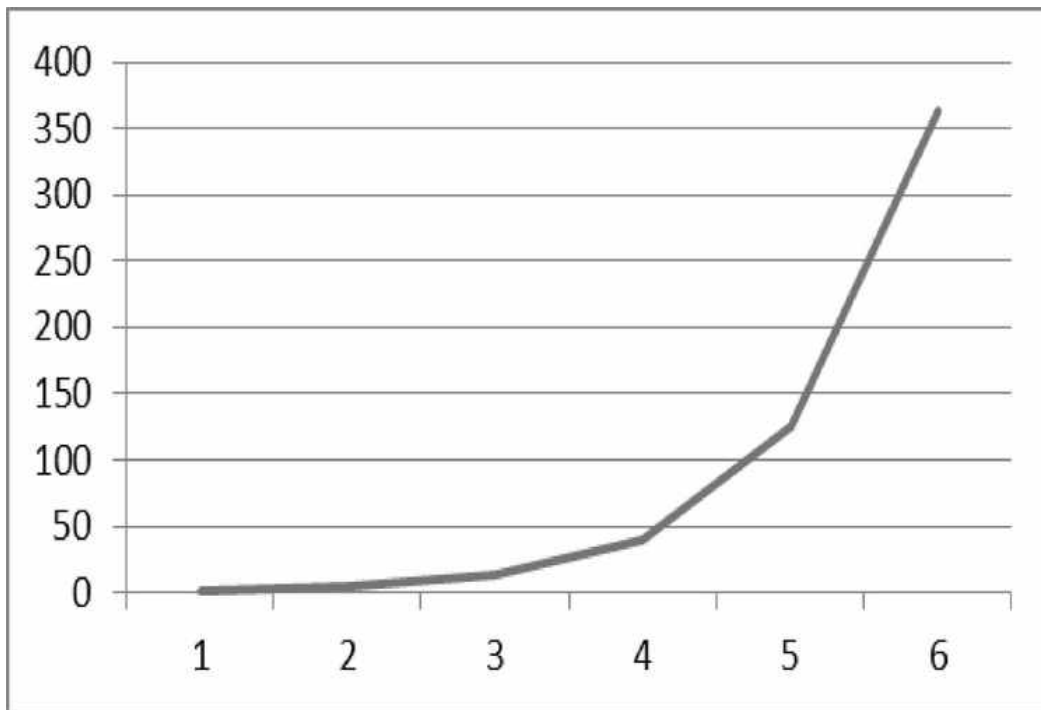
Example:

Fit an exponential curve $Y = ae^{bX}$ to data given below:

X	Y
1	1.6
2	4.5
3	13.8
4	40.2
5	125.0
6	363.0

Before we start ...

- What information you get from this graph?



Transformation of given equation

$$Y = ae^{bX}$$

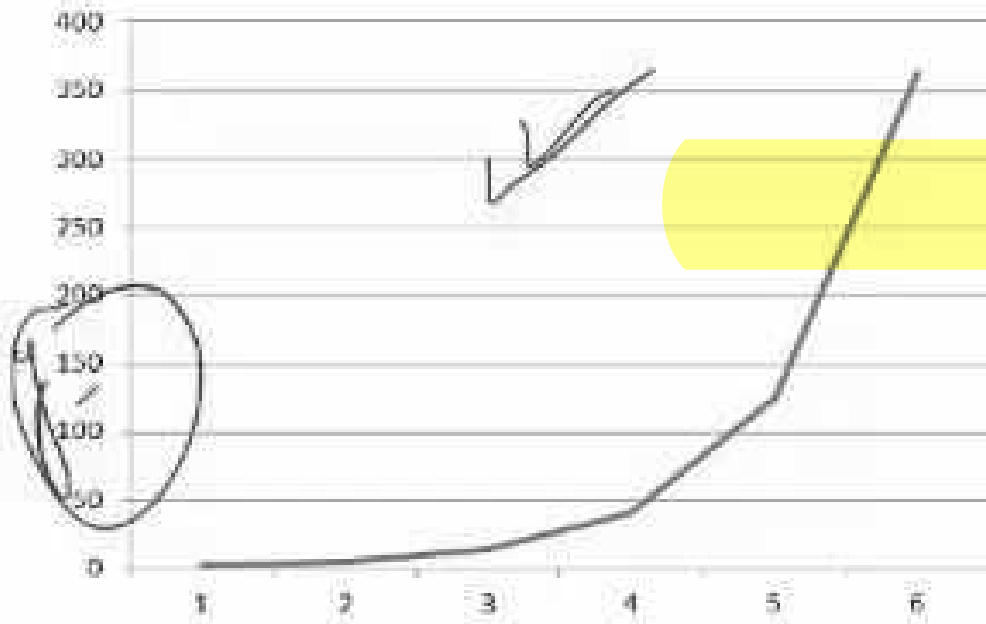
Taking log on both sides...

$$\begin{aligned}\log Y &= \log(ae^{bx}) \\ &= \log a + \log e^{bx} \\ &= \log a + bx \log e \\ \log Y &= \log a + (b \log e)X\end{aligned}$$

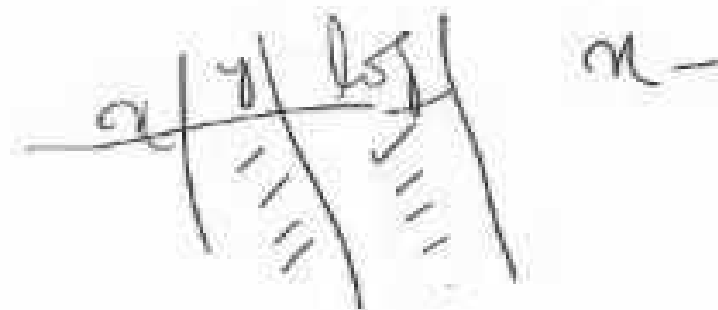
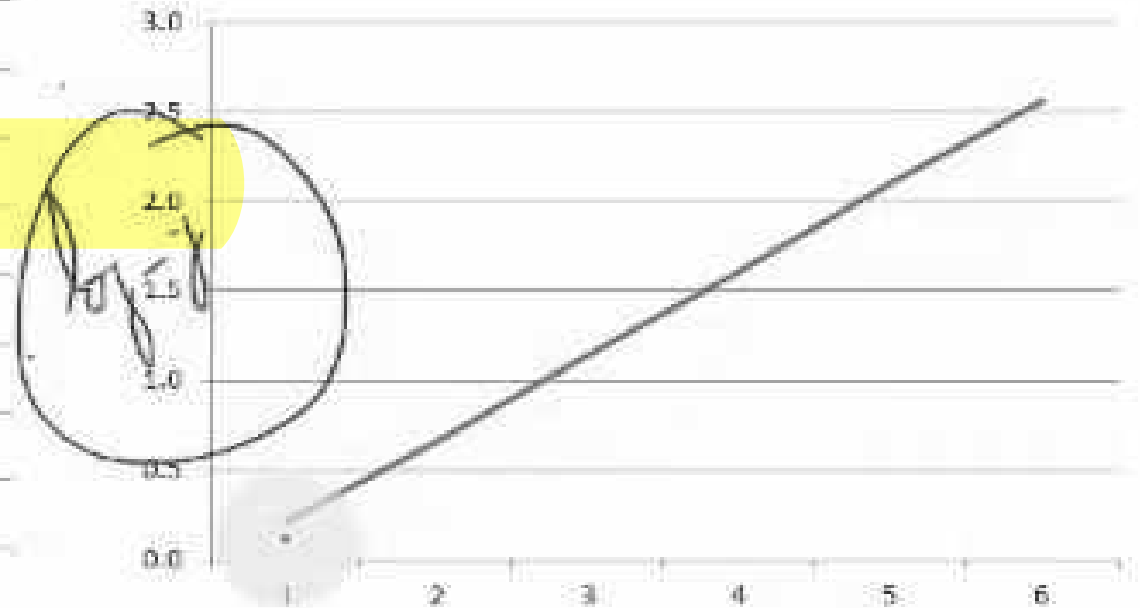
$$\underline{Y = A + BX}$$

Just for fun ... Graph of this equation

Original data

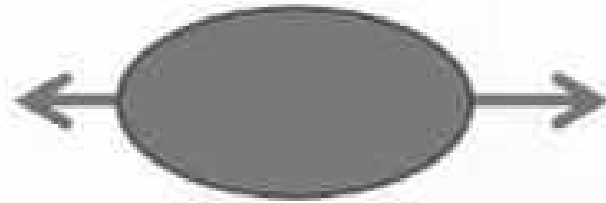


After Transformation



Transformation of given equation

Non-linear form



Linear form

$$Y = ae^{bX}$$

$$y' = A + BX$$

where

$$y' = \log Y$$
$$A = \log a$$
$$B = b \log e$$

Now it is linear form ...

$\frac{a}{\log y'}$

$$Y = a + bX$$

$$Y' = A + BX$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$B = \frac{n\sum XY' - (\sum X)(\sum Y')}{n\sum X^2 - (\sum X)^2}$$

$$a = \bar{Y} - b\bar{X}$$

$$A = \bar{Y}' - B\bar{X}$$

Required Calculation

X	Y	Log Y	(Log Y) ²	(X)(Log Y)	(X)(X)
1	1.6	0.2041	0.0417	0.2041	1
2	4.5	0.6532	0.4267	1.3064	4
3	13.8	1.1399	1.2993	3.4196	9
4	40.2	1.6042	2.5735	6.4169	16
5	125.0	2.0969	4.3970	10.4846	25
6	363.0	2.5599	6.5531	15.3594	36
21	-	8.2582	15.2914	37.1911	91

$n=6$, $\sum X=21$, $\sum Y^{\uparrow}$ $\sum Y'^2$ $\sum XY'$ $\sum X^2$

B formula

$$B = \frac{n(\sum XY') - (\sum X)(\sum Y')}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{6(37.1911) - (21)(8.2582)}{6(91) - (21)^2}$$

$$= \frac{49.72}{105}$$

$$B = 0.47$$

A formula

$$A = \bar{Y}' - B\bar{X}$$

$$= 1.3764 - (0.4735)(3.5)$$

$$A = -0.2811$$

$$Y' = A + BX$$

$$Y' = -0.2811 + 0.4735X$$

$$\bar{Y}' = \frac{\sum Y'}{n}$$
$$\bar{X} = \frac{\sum X}{n}$$

Converting back ...

Let

$$A = \log a$$

$$\Rightarrow a = \text{anti} - \log(A)$$

$$= \text{anti} - \log(-0.2811)$$

$$a = 0.5242$$

$$B = b(\log e)$$

$$0.4735 = b \log(2.7183)$$

$$0.4735 = b(0.4343)$$

$$\underline{0.4735} = b$$

$$0.4743$$

$$b = 109$$

$$e = 2.7183$$

How to Use Windows Calculator

- For Anti – log

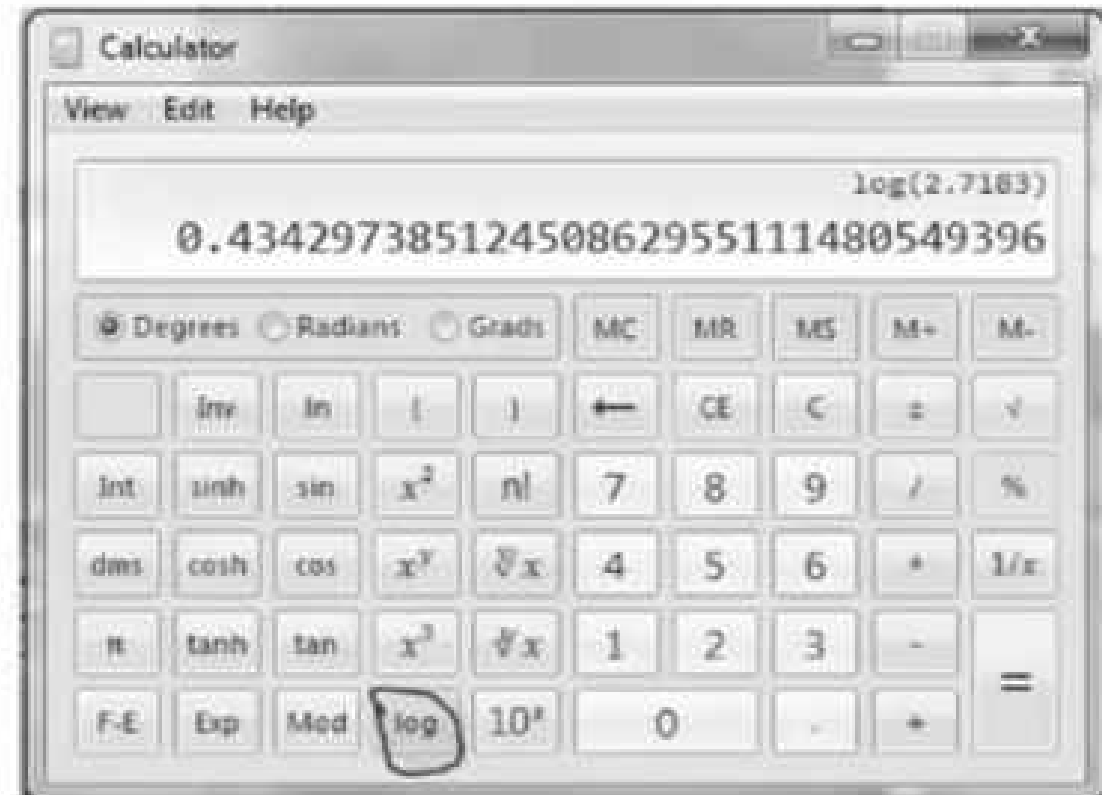
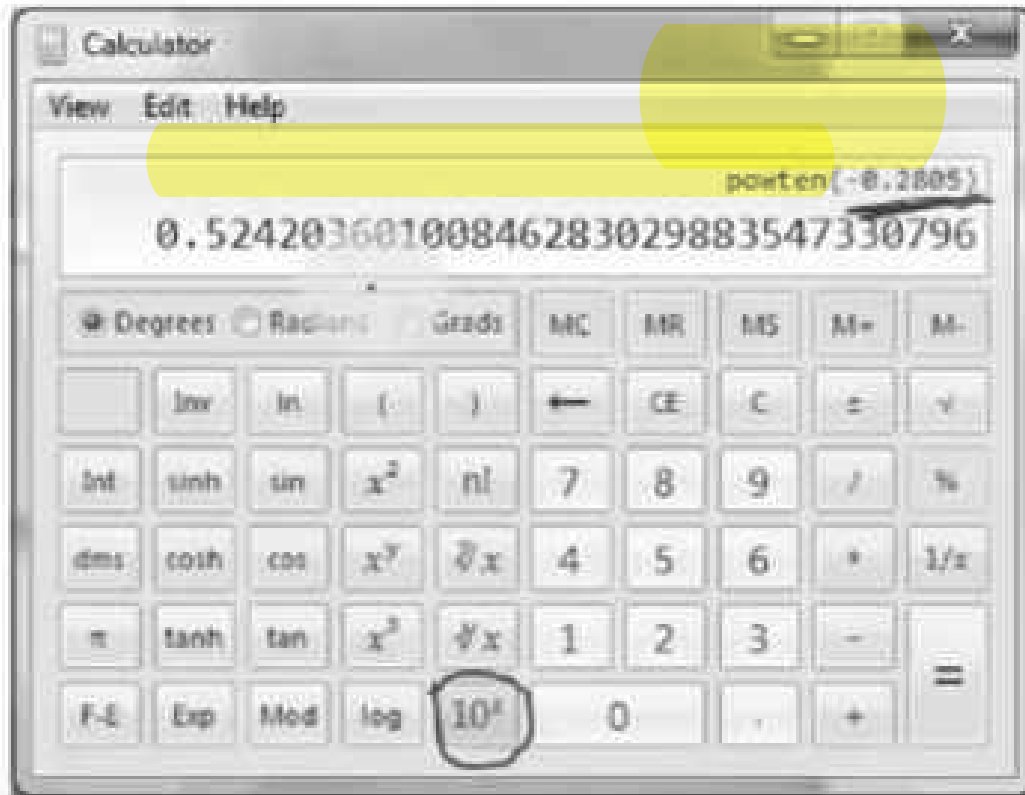
- Write value

- Then Press 

- For log

- Write value

- Then Press 



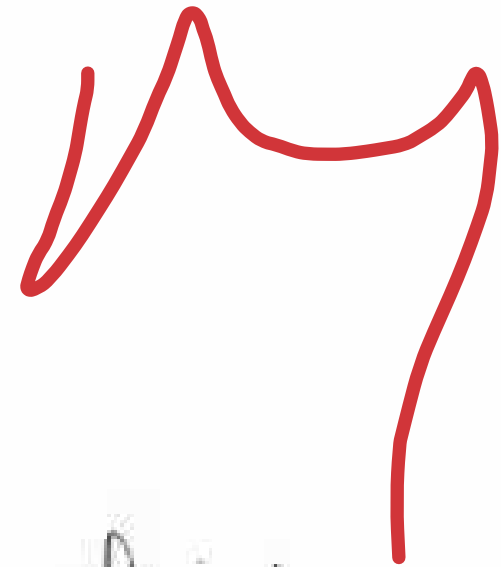
Hence the original equation ...

$$Y = a e^{bx}$$

Putting values

$$Y = 0.5242 e^{(1.09)X}$$

the required equation
0/0 ✓



Recap

Given Data

Make suitable
Transformation

Get new
equation

Final Equation

Converted back

Do Table
Calculations

- The
END -

Regression & Correlation Analysis

How to fit an Exponential Curve?

Procedure and step by step
calculation with example II

How to fit an Exponential Curve

Example:

Fit an exponential curve $Y = aX^b$ to data given below:

X	Y
1	2.98
2	4.26
3	5.21
4	6.10
5	6.80
6	7.50

Transformation of given equation

Non-linear form



Linear form

Y

How ...

Taking

$$Y = aX^b$$

How ?

Taking log on both sides...

$$\begin{aligned}\log Y &= \log (aX^b) \\ &= \log a + \log X^b \\ \log Y &= \log a + b \log X \\ Y' &= A + bX'\end{aligned}$$

$$\begin{aligned}\text{Let} \\ \log Y &= Y' \\ \log a &= A \\ \log X &= X'\end{aligned}$$

N

Now it is linear form ...

1st normal eq
(Apply Z on both sides)

$$Y' = A + bX'$$

A, b ✓
Normal Equation

1st normal equation:
(Apply Σ on both sides)

$$\underline{\underline{\Sigma Y' = nA + b \Sigma X'}}$$

2nd normal equation:
(Multiply with X' and then

Apply Σ on both sides)

$$\begin{aligned} X'Y' &= AX' + bX'^2 \\ \underline{\underline{\Sigma X'Y' = A \Sigma X' + b \Sigma X'^2}} \end{aligned}$$

Required Calculation

X

1

2

3

4

5

6

Total

X	Y	Log Y	Log X	(Log X) ²	(Log X)*(Log Y)
1	2.98	2.98	0.0000	0.0000	0.0000
2	4.26	4.26	0.3010	0.0906	0.1895
3	5.21	5.21	0.4771	0.2276	0.3420
4	6.10	6.10	0.6021	0.3625	0.4728
5	6.80	6.80	0.6990	0.4886	0.5819
6	7.50	7.50	0.7782	0.6055	0.6809
Total	-	4.3133	2.8574	1.7749	2.2672

n=6

$\Sigma Y'$

$\Sigma X'$

$\Sigma X'^2$

$\Sigma X'Y'$

Solving Normal Eqs...find b

$$\frac{6}{28574} = \frac{2.099}{8}$$

$$\Sigma Y' = nA + b\Sigma X' \text{ ----- (1)}$$

$$\Sigma X'Y' = A\Sigma X' + b\Sigma (X')^2 \text{ ----- (2)}$$

$$4.3133 = 6A + b(28574) \text{ ----- (1)}$$

$$2.2672 = (28574)A + (1.7749)b \text{ ----- (2)}$$

Now multiply Eq (2) with 2.0998,

$$4.7607 = 6A + 3.7270b \text{ ----- (3)}$$

Subtract Eq (1) from Eq (3)

$$0.4474 = 0 + 0.8696b \Rightarrow b = 0.5145$$



Now we put the

Putting "b" ... get value of "A"

$$\Sigma Y' =$$

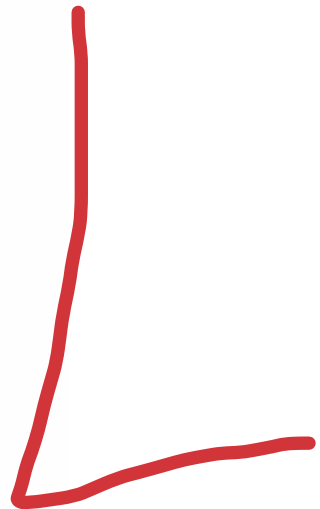
Now we put the value of $b = 0.515$ in Eq1 and solve as

$$\Sigma Y' = nA + b\Sigma X' \quad \text{--- Eq 1}$$

$$4.3133 = 6A + (0.5145)(2.8574)$$

$$2.8432 = 6A$$

$$\Rightarrow \boxed{A = 0.4739}$$



Converting back ...

$$A = \log a$$

$$\Rightarrow a = \text{anti} - \log(A)$$

$$= \text{anti} - \log(0.4739)$$

$$a = 1.61$$

How to Use Windows Calculator

- For Anti – log

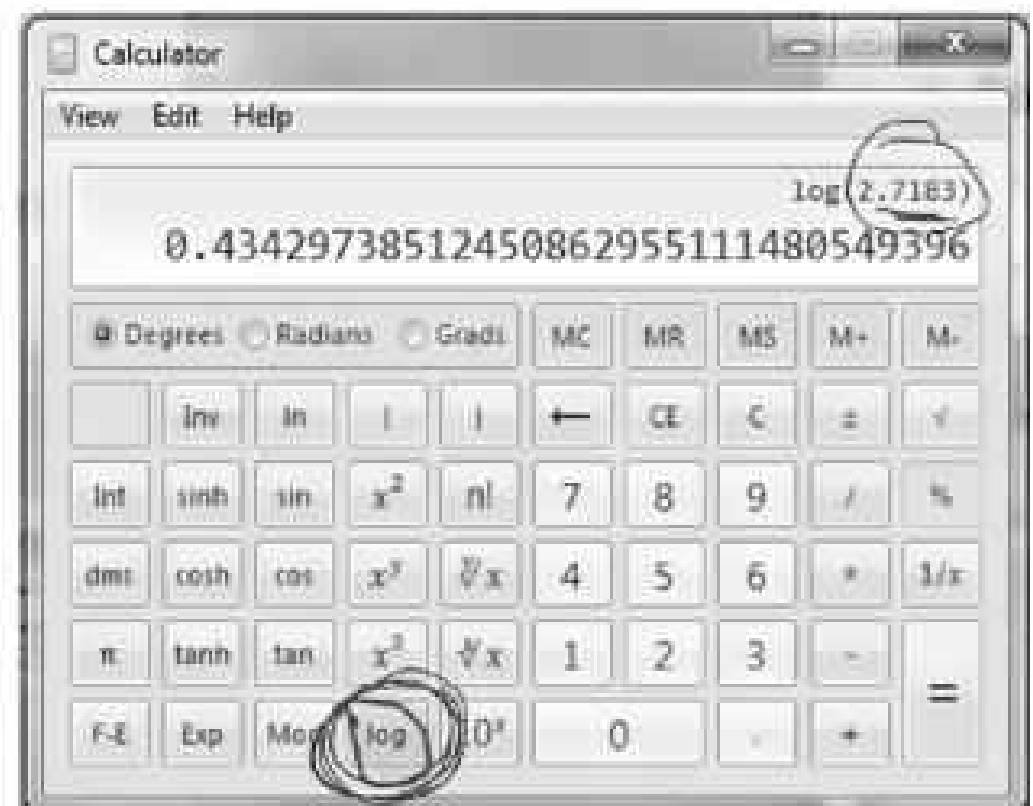
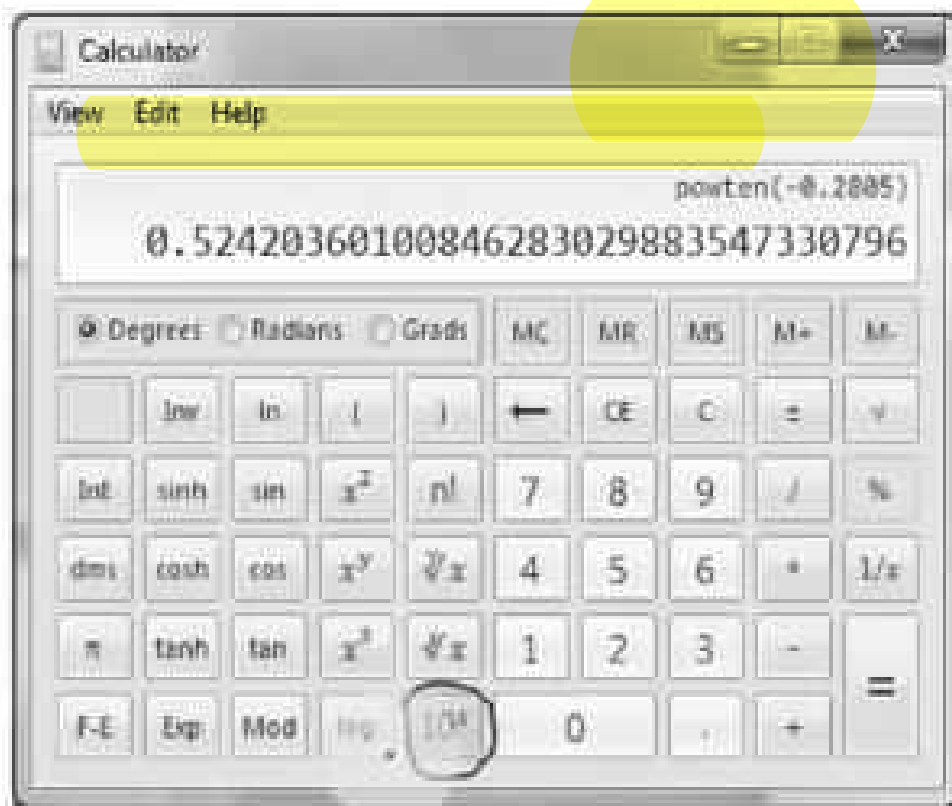
- Write value

- Then Press 10^x

- For log

- Write value

- Then Press log



Hence the fitted equation becomes...

$$Y = a X^b$$

$$Y = 161 X^{0.5145}$$

This is the required equation.



Recap

Given Data

Make suitable
Transformation

Get new
equation

Final Equation

Converted back

Solve Normal Eqs
and Do Table
Calculations

- The
END -

How to fit a Reciprocal Curve?

Lecture 15

Procedure and step by step
calculation with example

How to fit a Reciprocal Curve

Explanation by Example.

Example #1:

Fit a reciprocal curve $\frac{1}{y} = a + bX$ to data given below:

X	Y
0	10
1	8
4	5
6	4
12	2.5
16	2

Given equation ...

$$\frac{1}{Y} = \check{a} + \check{b}X$$

non-linear
equation
linear?

Let $y' = \frac{1}{Y}$ and make a new equation

$$y' = a + bx$$



Now it is linear form ...

$$Y = a + bX$$

$$Y' = a + bX$$

As we already know

New form

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$b = \frac{n \sum xy' - \sum x \sum y'}{n \sum x^2 - (\sum x)^2}$$

$$a = \bar{Y} - b\bar{X}$$

$$a = \bar{y}' - b\bar{x}$$

Required Calculation

x	y	$Y' = 1/y$	Xy'	xx
0	10	0.10	0	0
1	8	0.13	0.125	1
4	5	0.20	0.8	16
6	4	0.25	1.5	36
12	2.5	0.40	4.8	144
16	2	0.50	8	256
39		1.58	15.225	453

$n=6, \Sigma y$
 ΣY
 $\Sigma xy'$
 Σx^2

b formula

$$b = \frac{n(\sum XY') - (\sum X)(\sum Y')}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{6(15.225) - (39)(1.58)}{6(453) - (39)^2}$$
$$= \frac{29.73}{1197}$$

$$b = 0.0248$$



a formula

$$a = \bar{Y}' - b\bar{X}$$

$$= \left(\frac{1.58}{6} \right) - (0.0248) \left(\frac{39}{6} \right)$$

$$a = 0.1019$$

$$\bar{Y}' = \frac{\sum Y'}{n}$$

$$\bar{X} = \frac{\sum X}{n}$$

Hence the original equation ...

$$\frac{1}{y} = a + bX$$

Putting value of "a" and "b"

$$\frac{1}{y} = 0.1019 + 0.0248X$$

This is my final
answer to it

Recap

Given Data

Make suitable
Transformation

Get new
equation

Final Equation

Convert back, if
needed

Do Table
Calculations

- The
END -

Regression & Correlation Analysis

How to fit a Reciprocal Curve?

Lecture 15

Procedure and step by step
calculation with example

How to fit a Reciprocal Curve

Explanation by example:

Example #2:

Fit a line of the form $Y = \frac{1}{a + bX}$ to data given below:

X	Y
0	10
1	8
4	5
6	4
12	2.5
16	2

Given equation ...

$$Y = \frac{1}{a+bX}$$

→ Non-linear

Take the reciprocal on both sides

$$\frac{1}{Y} = \frac{a+bX}{1}$$

Let $\frac{1}{Y} = Y'$

$$= a + bX$$

$$Y' = a + bX$$

Transform

Linear form

Now it is linear form ...

$$Y = a + bX$$

As we already know

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$a = \bar{Y} - b\bar{X}$$



$$Y' = a + bX$$

New form

$$b = \frac{n\sum XY' - \sum X \sum Y'}{n\sum X^2 - (\sum X)^2}$$

$$a = \bar{Y}' - b\bar{X}$$

$$a, b \rightarrow$$

$$x, y'$$

Required Calculation

x	y	$y' = 1/y$	xy'	xx
0	10	0.10 ✓	0	0
1	8	0.13 ✓	0.125	1
4	5	0.20 ✓	0.8	16
6	4	0.25 ✓	1.5	36
12	2.5	0.40 ✓	4.8	144
16	2	0.50 ✓	8	256
39	=	1.58	15.225	453

$n=6$ $\sum x$

$\sum y =$

$\sum xy$

$\sum x^2$

b formula

$$b = \frac{n(\sum XY') - (\sum X)(\sum Y')}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{6(15.225) - (39)(1.58)}{6(453) - (39)^2}$$

$$= \frac{29.73}{1197} = 0.0248$$

a formula

$$a = \bar{Y}' - b\bar{X}$$

$$= \frac{1.58}{6} - (0.0248) \left(\frac{39}{6} \right)$$

$$a = 0.1019$$

$$\bar{Y}' = \frac{\sum Y}{n} = \frac{1.58}{6}$$

$$\bar{X} = \frac{\sum X}{n} = \frac{39}{6}$$

Hence the original equation ...

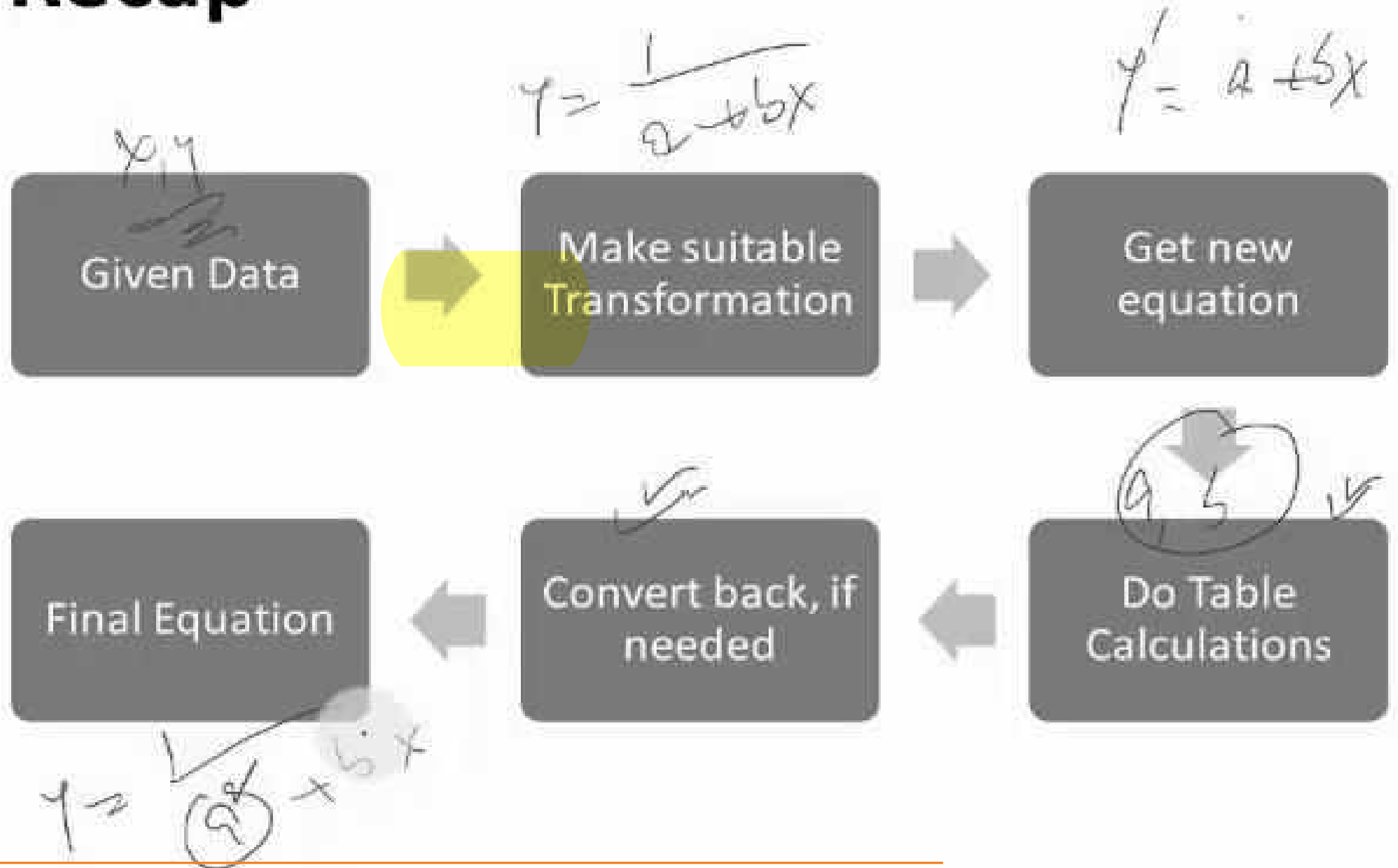
$$\frac{1}{y} = a + bx$$

Putting value of "a" and "b"

$$\frac{1}{y} =$$

$$y = \frac{1}{0.1019 + 0.0248x}$$

Recap



- The
END -