Final term Preparation
STA404

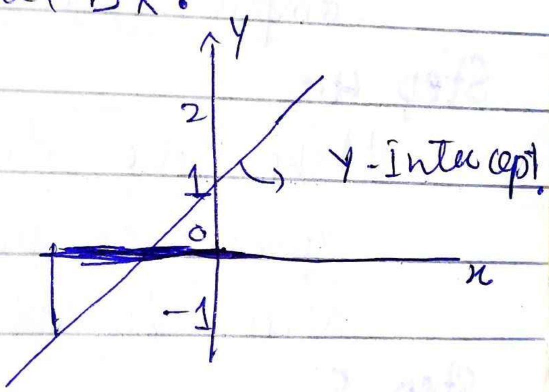
Lecture # 09.
Statistical Inference about Regression.

=> using the t-test for hypothesis testing of α - the Intercept of regression line.

(1) Intercept of regression line?

The Y - Intercept of a stright line is simply where the line crosses the Y- axis.

$$Y = a + bX.$$



Population Regression	Sample Regression.
$Y = \alpha + \beta X + \varepsilon$	$Y = a + b\cancel{X} + e$

Hypothesis Testing
General procedure.

(In this course we'll Study only ~~to~~ traditional Method)

Step 1:-

State the hypotheses and Identify the Claim.

Step 2:-

Find the Critical values from the appropriat table.

Step 3:-

Compute the test value.

Step 4:-

Make the decision to reject or not reject the Null hypothesis.

Step 5:-

Summarize the result.

Testing Hypothesis

Example:-

Step 1:-

State the Hypothesis.

$H_0 : \alpha = 32$

$\boxed{H_1 : \alpha \neq 32} \Rightarrow$ claim.

Step 2:-

Level of Significance $(\alpha) = 1\% = 0.01$

Step 3:-

Calculations:-

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

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

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

$$S.E(a) = S_a = S_{yx}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sqrt{\sum(X-\bar{X})^2}}}$$

$$S_{yx} = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n-2}}$$

$$(Y-\bar{Y})^2 = \sum Y^2 - a\sum Y - b\sum XY$$

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

Sta404;

Lecture(#10).

Hypothesis Testing about Simple Correlation

Notations:

* Sample Correlation coefficient	* Population Correlation coefficient
* English Latter "r"	* Greek Small Letter (rho) "ρ"

Two Cases

Testing population Correlation coefficient ρ

- $\rho = \rho_0$ = Some Specific value
  z-test
- $\rho = 0$
  T-Test

Some Important points;

* Sampling distribution of "r" is;
  - Neither normal
  - Nor become approximately normal for large Sample Size.

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

Do Transformation;

"r" to "Z_f"

formula for finding Z_f.

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

Testing formula,

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

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

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

Case - II;

when ($\rho = 0$)

(Main points):

* Sampling distribution is Not Symmetry But It is Interesting to Note that When $\rho = 0$, then its Symmetrical.
* Above make it possible to use the the ~~t~~ test.
* T- test can be applied for any Sample size.

* $H_0 : \rho = 0$ Means that there is No linear Correlation between two variables.
* The two variable are $\boxed{\text{Independent.}}$

Formula for t-test; when $\rho = 0$.

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

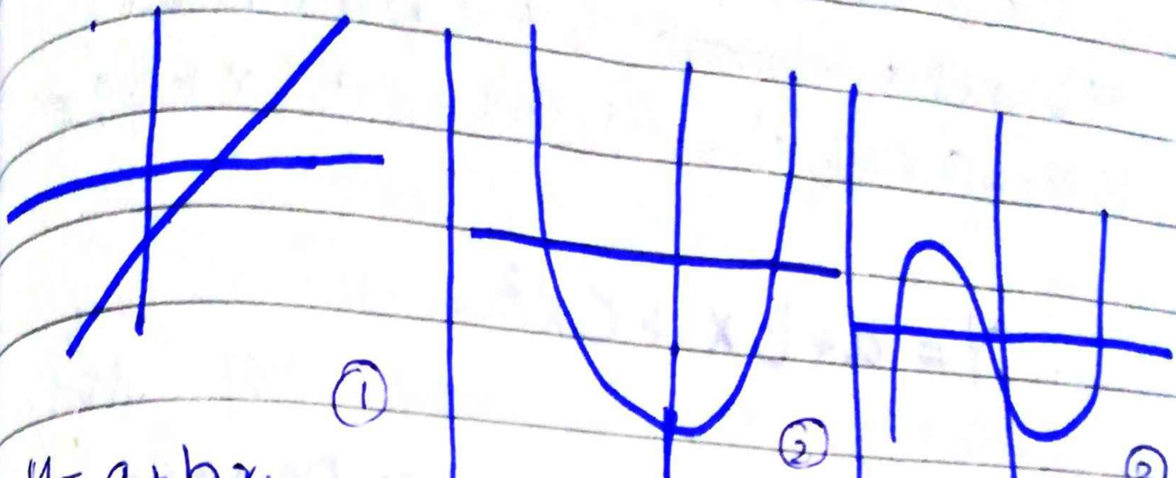
It follows t-distribution with dgree of freedom $\boxed{df = n-2}$

* Testing $H_0 : \rho = 0$ is equivalent to test $H_0 : \beta = 0$.

Sta404.

Lecture # 11.

Fitting a Second Degree Parabola



1. $y = a + bx$ Straight line
2. $y = a + bx + cx^2$ Second Degree or Parabola equation
3. $y = a + bx + cx^2 + dx^3$ 3rd degree equation

Normal equations;

* (Normal equations) are equations obtained by setting equal to zero the partial deriatives of the Sum of Squared errors. (Least Sq)

* It finds the regressions coefficients **analytically**.

How to Make Normal Equations;

$Y = a + bx$, $Y = a + bx + Cx^2$

The normal equations are ——

$Y = a + bx \longrightarrow \Sigma Y = na + b\Sigma X$

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

$$Y = a + bx + Cx^2$$

The Normal equations of a parabola are —.

$1(Y = a + bx + Cx^2) \quad \Sigma Y = na + b\Sigma x + C\Sigma X^2$

$x(Y = a + bx + Cx^2) \quad \Sigma XY = a\Sigma X + b\Sigma X^2 + C\Sigma X^3$

$X^2(Y = a + bx + Cx^2) \quad \Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + C\Sigma X^4$

Sta404 Lec #12

"How to Find the Standard Deviation of Regression"

"Standard Deviation of Regression is also called Standard Error of Estimate"

Formulas

For Population Data

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

For Sample Data.

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

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

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

$\boxed{e = Y - \hat{Y}}$ Residual Error.

$\sum(Y-\hat{Y})^2$ is called Sum of Square of Residuals.

Solving Examples of Standard error by using another formula;

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

Another form of this is :-

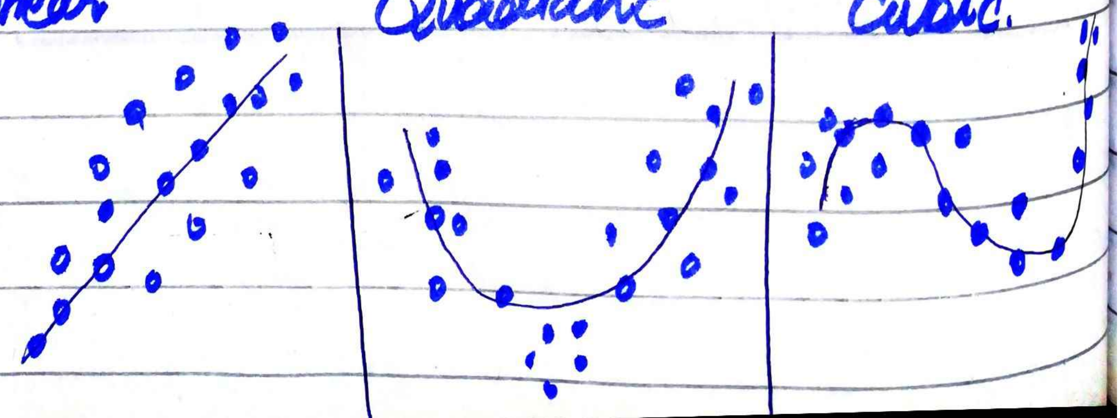
$$S_{y,x} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n-2}}$$

Lecture #13

Deciding the Types of Curve to Fit.

(which curve is suitable in which situtiow)

Linear Quadratic Cubic.



Two Check the Data Type

Two Methods usually used

(1) Check by (Difference) (Delta)
(2) Check by Graph.

Criteria for Suitable Curve.

* If the First difference of Y are approximately Constant, use a **Staright line.**

* if the 2nd difference of Y are approximately Constant, use a Second degree order equation (parabola).

* If the Third Difference of Y are approximately Constant, use a third degree order equation.

* If the 1st difference of logy are approximately Constant use a exponential curve.

* If the First Difference of logy & logx are approximately Constant, use a Geometric curve.

* If the First Difference of reciprocals of Y are approximately Constant, use a third degree of parabola.

Remember;

* **Straight line;**

$$Y = a + bx$$

* **Second Degree Line (Parabola)**

$$Y = a + bx + Cx^2$$

* **Third degree Line;**

$$Y = a + bx + Cx^2 + dx^3$$

Use graph to get the Idea of Suitable equation.

Points to Remember;

* If the graph of the value of X & Y gives a Straight line, use a Straight line.

* If the graph of the values of X & Y gives a Curve with only one bend, use a 2nd degree parabola.

* if the graph of the values of X & Y gives a Curve of the Shape of reverse "S", use 3rd degree parabola.

* If the graph of the values of X & Log Y gives a straight line, use an exponential Curve.

* If the graph of the valle of Log X & Log Y gives a Straight line use a geometric or Logarithmic Curve.

* if the graph of the values of x & $1/y$ gives a Straight line, use a hyperbola curve.

$Y = ab^x$ >>> $\log Y = \log a + X \log b$

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

$\frac{1}{Y} = a + bx$ ~~Log~~ $Y' = a + bx$

Transformation from Non-Linear to Linear Equations.

Non-Linear form.

$$Y = ax^b$$

Transformation:

Taking log on both sides;

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

$$\log Y = \log a + \log x^b$$

$$= \log a + b\log x$$

Let;

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

$$\boxed{Y' = A + bX'}$$

linear form.

Lecture # 14

Fitting of Exponential Curve.

$$Y = ab^x$$

The Curve to be fitted is

$$y = ab^x$$

Taking $\log_{10}$ on ;

$$\underbrace{\log_{10} Y}_{Y} = \underbrace{\log_{10} a}_{A} + x\underbrace{\log_{10} b}_{B}$$

$$Y = A + Bx$$

Normal equation ;

$$\boxed{\begin{aligned} \Sigma Y &= nA + B\Sigma x \\ \Sigma XY &= A\Sigma X + B\Sigma x \end{aligned}}$$

Example;

Find a Curve $y = ab^x$ to the following data using method of least Squares.

x	2	3	4	5	6
y	8.3	15.4	33.1	65.2	127.4

Ans:

x	y	$Y = \log_{10} y$	x^2	xY
2	8.3	0.9191	4	1.8382
3	15.4	1.1875	9	3.5625
4	33.1	1.5198	16	6.0792
5	65.2	1.8142	25	9.0710
6	127.2	2.1052	36	12.6312
$\Sigma x = 20$		$\Sigma Y = 7.5458$	$\Sigma x^2 = 90$	$\Sigma xY = 33.182$

Putting all values in equations

$$\left.\begin{aligned} 7.5458 &= 5A + 20B \\ 33.1821 &= 20A + 90B \end{aligned}\right\} \text{solve equations}$$

$A = 0.8096 = \log_{10} a$

$B = 0.2999 = \log_{10} b$

$a = \text{Antilog}(0.3096)$

$b = \text{Antilog}(0.2999)$

$a = 2.04$

$b = 1.995$

$$\boxed{\begin{aligned} y &= ab^x \\ y &= (2.04)(1.995)^x \end{aligned}}$$

Lecture # 15

Fitting Reciprocal Curve.

use reciprocal term when the effect of an Independent variable decreases as its values Increases. The value of this term decreases as the Independent variable (x) Increases It is the denominator.

Like; $y = \frac{A}{x}$, $y = \frac{1}{x}$

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

if we take some values for x;

x	-2	0	2
y	$-1/2$	∞	$1/2$

Put $x = -2$, $x = 0$, and $x = 2$ in $y = \frac{1}{x}$, we get;