$$= \frac{S_{xy}}{S_{xx}}$$

$$\beta = \frac{S_{xy}}{S_{xx}} \qquad S_{xy} = \Sigma xy - n\bar{x}\bar{y}$$

$$S_{xx} = \Sigma x^2 - n\bar{x}^2$$

$$\Sigma xy = 65.667$$

$$S_{xx} = \Sigma x^2 - n\bar{x}^2$$

$\div$ by $(n-1)$

$$\frac{S_{xx}}{n-1} = \frac{\Sigma x^2 - n\bar{x}^2}{n-1} \leftarrow \text{Sample variance}$$

$$\frac{S_{xx}}{n-1} = s_x^2 \qquad \therefore \underline{S_{xx} = s_x^2(n-1)}$$

Regression

1 (X, Y) are bivariate normal

$r \neq \rho \leftarrow$ Pop^n corr. coeff

$\uparrow$ sample corr. coeff

$H_0 : \rho = 0$

$H_1 : \rho < 0$

$\rho > 0$

$\rho \neq 0$

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

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

DC : Rej. H_0 if

$TS > t_{n-2,\alpha/2}$ ... 2-sided test

$TS > t_{n-2,\alpha}$ ... one-sided test

$H_0 : \rho = k$

$H_1 : \rho > k$

$\rho < k$

$\rho \neq k$

$$\left[\tanh^{-1}(r) - \tanh^{-1}(\rho)\right]\sqrt{n-3} \sim N(0,1)$$

$$TS = \left[\tanh^{-1}(r) - \tanh^{-1}(k)\right]\sqrt{n-3}$$

$Z_{\alpha/2}$

Z_{α}

TS $> t_{n-2,\alpha/2}$ ... 2-sided test $\qquad Z_{\alpha/2}$

TS $> t_{n-2,\alpha}$ ... one-sided test $\qquad Z_{\alpha}$

Eg :

x	2	4	8	12	5
y	3	1	6	2	0

$r = \underline{0.2117}$, $n = 5$, $\alpha = 5\%$

$H_0 : \rho = 0 \quad$ v/s $\quad H_1 : \rho > 0$

$H_0 : \rho = 0.4 \quad$ v/s $\quad H_1 : \rho \neq 0.4$

① $H_0 : \rho = 0$

TS : $\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.2117 \times \sqrt{3}}{\sqrt{1 - 0.2117^2}} = \underline{\underline{0.3752}}$

$t_{3,5\%} = \underline{\underline{2.353}}$

TS $\not> t_{n-2,\alpha}$

Insufficient evidence to reject $H_0 : \rho = 0$

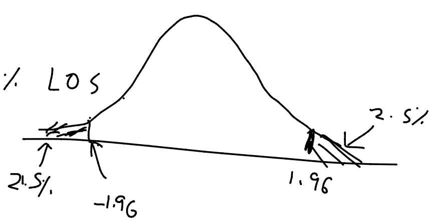
② $H_0 : \rho = 0.4$ (hyp fn)

TS $= [\tanh^{-1} r - \tanh^{-1}(0.4)]\sqrt{5-3} = -\underline{\underline{0.295}}$

$Z_{2.5\%} = 1.96$

$|TS| \not> Z_{2.5\%}$

Insuff evidence to rej H_0 @ 5% LOS



TS < -1.96

or

TS > 1.96

* Test for β

$H_0 : \beta = k \quad$ v/s $\quad H_1 : \beta > k, \ \beta < k, \ \beta \neq k$

TS : $\dfrac{\beta - k}{\sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2}$

DC : Same as for $\rho = 0$ (t distn)

$\hat{\sigma}^2 = \dfrac{1}{n-2}\left[S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}\right]$

95 % CI for β —

$$\hat{\beta} \pm t_{n-2,\alpha/2} \times \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

Eg : $H_0: \beta = 0.3$ v/s $H_1: \beta \neq 0.3$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left[S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right] = \frac{1}{6-2}\left[2.85484 - \frac{7.69189^2}{56.2249}\right]$$

$$\sum y^2 - n\bar{y}^2 = (n-1) S_y^2 = S_{yy}$$

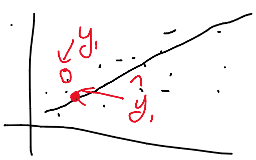
$$\hat{\sigma}^2 = 0.4506$$

$$TS = \frac{0.13681 - 0.3}{\sqrt{0.4506/56.2249}} = \underline{\underline{-1.82289}}$$

$$t_{4,2.5\%} = 2.776$$

$$|TS| \not> t_{4,2.5\%}$$

$\therefore$ Insuff. evidence to reject H_0.



$y \Rightarrow$ Actual / True values

$\hat{y} \rightarrow$ Fitted values

Error or residual $= y_1 - \hat{y}_1 = e_1$

Total Variation

Explained by regression

Unexplained by regression

Total variation : SS_{TOT}

$=$

Expl. by Reg $= SS_{REG}$

$+$

Not expl. by reg $= SS_{RES}$

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y}_i)^2$$

$$\sum (y - \bar{y})^2 = \sum y^2 - n\bar{y}^2 = \underline{\underline{S_{yy}}} \quad \text{——} \quad (1)$$

$$SS_{REG} = \frac{S_{xy}^2}{S_{xx}} \quad \text{——— (2)}$$

$$SS_{RES} = S_{yy} - \frac{S_{xy}^2}{S_{xx}} \quad \text{——— (3)}$$

$$\hat{\sigma}^2 = \frac{SS_{RES}}{n-2}$$

* . Error variance $\hat{\sigma}^2$

$H_0 : \sigma^2 = k$ v/s $H_1 : \sigma^2 > k$

$\sigma^2 < k$

$\sigma^2 \neq k$

$$TS = \frac{(n-2)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

$$\text{Comp. TS} = \frac{(n-2)\,\hat{\sigma}^2}{k}$$ compare with $\chi^2_{n-2,\, LOS}$

* How to test strength of regression ?

(1) Coefficient of determination (R^2)

Change in $X \xrightarrow{60\%}$ Change in Y

$$SS_{TOT} = SS_{REG} + SS_{RES}$$

$$R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{\frac{S_{xy}^2}{S_{xx}}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}}$$

$$\boxed{R^2 = r^2}$$

* . Prediction

$\hat{y} = 2.5 + 1.5x$

Predict y when $x = 4$

$\hat{y} = 2.5 + 1.5(4) = \underline{\underline{8.5}}$

$Y | X = x \sim N(\alpha + \beta x, \sigma^2)$

(1) Mean response (2) Individual response

$Y | X = x_i \sim N(\alpha + \beta x, \sigma^2)$

$Y_i = \alpha + \beta x_i + e_i$

Assump : $e_i \sim N(0, \sigma^2)$

$Y_i = \underline{\underline{\alpha + \beta x_i}} + N(0, \sigma^2) \longrightarrow$ Const + Random variation

$Y_i | X = x_i \sim \underline{N(\alpha + \beta x_i, \sigma^2)} \Leftarrow$ Ind. Response variable

$E(y | X = x) = E[\text{Normal dist}^n]$

$= \mu$

$= \alpha + \beta x_i \Leftarrow$ Mean response variable

Mean response	Ind. response
$\hat{\mu}_0 = \hat{\alpha} + \hat{\beta} x_0$	$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0$
$V(\hat{\mu}_0) = \left[\frac{1}{n} + \frac{(x_0 - \bar{x}^2)}{Sxx}\right]\sigma^2$	$V(\hat{y}_0) = \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx} + 1\right]\sigma^2$
	$= V(\hat{\mu}_0) + \sigma^2$
$\frac{\hat{\mu}_0 - \mu_0}{SE(\hat{\mu}_0)} \sim t_{n-2}$	$\hat{y}_0 - y_0$

$$\frac{\mu_0 - \mu_0}{SE(\hat{\mu}_0)} \sim t_{n-2}$$

$$\frac{\hat{y}_0 - y_0}{SE(\hat{y}_0)} \sim t_{n-2}$$