Hypothesis Testing in Regression:

1- Overall Significance of Regression (ANOVA) Approach

1. H_0: Regression is insignificant ($\beta_1 = 0$)
   H_1: Regression is significant ($\beta_1 \neq 0$)

   Anova only allows two sided alternative to test against zero

2. $\alpha = 5\%$

3. Test-Statistics

$$F = \frac{MSR}{MSE} = \frac{\text{Mean Square Regression}}{\text{Mean Square error}}$$

4. Critical Region: (p-value based)

   if $p < \alpha$ Reject H_0
   if $p > \alpha$ Accept H_0

   Reject H_0 if
   $F > F_{\alpha}(v_1, v_2)$
   $F > F_{\alpha}(1, n-2)$

5. Conclusion:

ANOVA Table

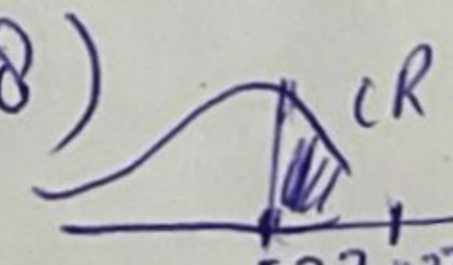
SoV	df	SS	MS	F-ratio	P-value
Regression	1	$SSR = \Sigma(\hat{Y} - \bar{Y})^2$	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$	
Error	$n-2$	$SSE = \Sigma(Y - \hat{Y})^2$	$MSE = \frac{SSE}{n-2}$		
Total	$n-1$	$SST = \Sigma(Y - \bar{Y})^2$			

Sol: (Question shipment time & Distance).

SoV	df	SS	MS	F-test
Reg	1	33.2815	33.2815	11.2731
error	8	23.6185	2.9523	
Total	9	56.9		

$\boxed{\text{Reject } H_0}$

CR: $F > F_{\alpha}(1,8) = 11.2731 > F_{0.05}(1,8)$
$= 11.2731 > 5.32$



Conclusion:

Reg is significant that is the significant amount of variability in response variable Y accounted for by the postulated model, the straight line function.

$\boxed{t^2 = F}$

Testing hypothesis about regression coefficient.

T-Test allows for testing of hypothesis one & two tail alternatives against a specific value and zero as well. This Procedure ($\beta_1 = 0, \beta_1 \neq 0$) is equivalent of testing for overall significance of Regression using ANOVA Approach (F-Test) in case of simple linear regression. Also note that F-test is restricted to testing against a two sided Alternatives. Moreover,

$t^2 = F \longrightarrow$ with $(1, n-2)$ df.

$\hookrightarrow$ with $n-2 = v$ df

(1) $H_0: \beta_1 = 0$ / or $\beta_1 = 0$: $\beta_1 \leq \beta_{10}$ / or $\beta_1 = 0$: $\beta_1 \geq \beta_{10}$ — β_{10}: specific value.

$H_1: \beta_1 \neq 0$ / $\beta_1 > \beta_{10}$ / $\beta_1 < \beta_{10}$

(2) $\alpha = 5\%$.

(3) test - statistics: $t = \frac{b_1 - \beta_1}{s/\sqrt{S_{xx}}}$

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \Rightarrow \frac{s}{\sqrt{S_{xx}}} = \frac{\sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-2}}}{\sqrt{\Sigma(X-\bar{X})^2}} = \frac{\sqrt{\Sigma(Y-\hat{Y})^2}}{\sqrt{(n-2)}\sqrt{(\Sigma(X-\bar{X})^2}}$$

with $n-2$ df.

(4) CR:

if $\beta_1 \neq 0$

$t < -t_{\alpha/2}(v)$ & $t > t_{\alpha/2}(v)$

if $\beta_1 > \beta_{10}$

$t > t_{\alpha}(v)$

if $\beta_1 < \beta_{10}$

$t < -t_{\alpha}(v)$

(5) conclusion:

when $\beta_1 = 0$, $\beta_1 \neq 0$.

then $t = \frac{b_1 - 0}{s/\sqrt{S_{xx}}}$

$= b_1 / s/\sqrt{S_{xx}}$.

Question (shipping time & distance). Using the estimated slope 0.0217, test whether $\beta_1 = 1.0$ against $\beta_1 < 1.0$.

(1) $H_0: \beta_1 = 1.0$

$H_1: \beta_1 < 1.0$

(2) $\alpha = 5\%$.

(3) Test - statistics:

$$t = \frac{b_1 - \beta_1}{s/\sqrt{S_{xx}}} \Rightarrow$$

$s/\sqrt{s_{xx}} \Rightarrow \sqrt{S_{xx}} = \sqrt{\Sigma(X-\bar{X})^2} = \sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$

$= \sqrt{5453565 - \frac{(7337)^2}{10}} = 265.3452$

$s = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-2}} = \sqrt{\frac{23.6185}{8}}$

$= 1.7182$

$s_b = s/\sqrt{s_{xx}} = \frac{1.7182}{265.3452} = \boxed{0.006475}$.

$t = \frac{0.0217 - 1.0}{0.006475} = -151.0888$

(4) CR: $t < -t_{\alpha(v)}$

$-151.0888 < -t_{0.05(8)}$

$-151.0888 < -1.860$

(5) Conclusion:

Reject Ho.

The Failure to Reject Ho: $\beta_1 = 0$ suggests that there is no linear relationship b/w x & Y; It may mean that changing x has little impact on changes in Y. However, it also indicates that true relationship is non linear. When $\beta_1 = 0$ is Rejected, it means that the linear term in x residing in the model explains a sig Portion of variability in Y.

Testing hypothesis about correlation coefficient

A test of special hypothesis $\rho = 0$ versus an appropriate alternative equivalent to testing $\beta_1 = 0$ for the simple linear regression model and therefore one can choose b/w t-test with $n-2$ df and F-test with 1 & $n-2$ df. if one wishes to avoid ANOVA Procedure and compute only sample correlation coefficient, It can be stated as

$$t = \frac{b_1}{s/\sqrt{S_{xx}}} \Rightarrow \text{can be written as}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim \nu = n-2 \text{ df}$$

① $H_0 : \rho = 0 \;/\; \rho = 0 \;/\; \rho = 0$
$H_1 : \rho \neq 0 \;/\; \rho > 0 \;/\; \rho < 0$

② $\alpha = 5\%$

③ test - statistics.

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

For any sample size this t-test can be applied.

CR: $t \leq -t_{\alpha/2(v)}$ & $t > t_{\alpha/2(v)}$ if $\rho \neq 0$

$v = n-2$. $t < -t_{\alpha(v)}$ if $\rho < 0$

$t > t_{\alpha(v)}$ if $\rho > 0$

Conclusion:

Question (shipping time & miles distance)

① H_0: There is no linear correlation b/w X & Y : i.e $\rho = 0$ or ($\beta = 0$)

H_1: There is linear correlation b/w X & Y i.e $\rho \neq 0$ or ($\beta \neq 0$).

② $\alpha = 5\%$

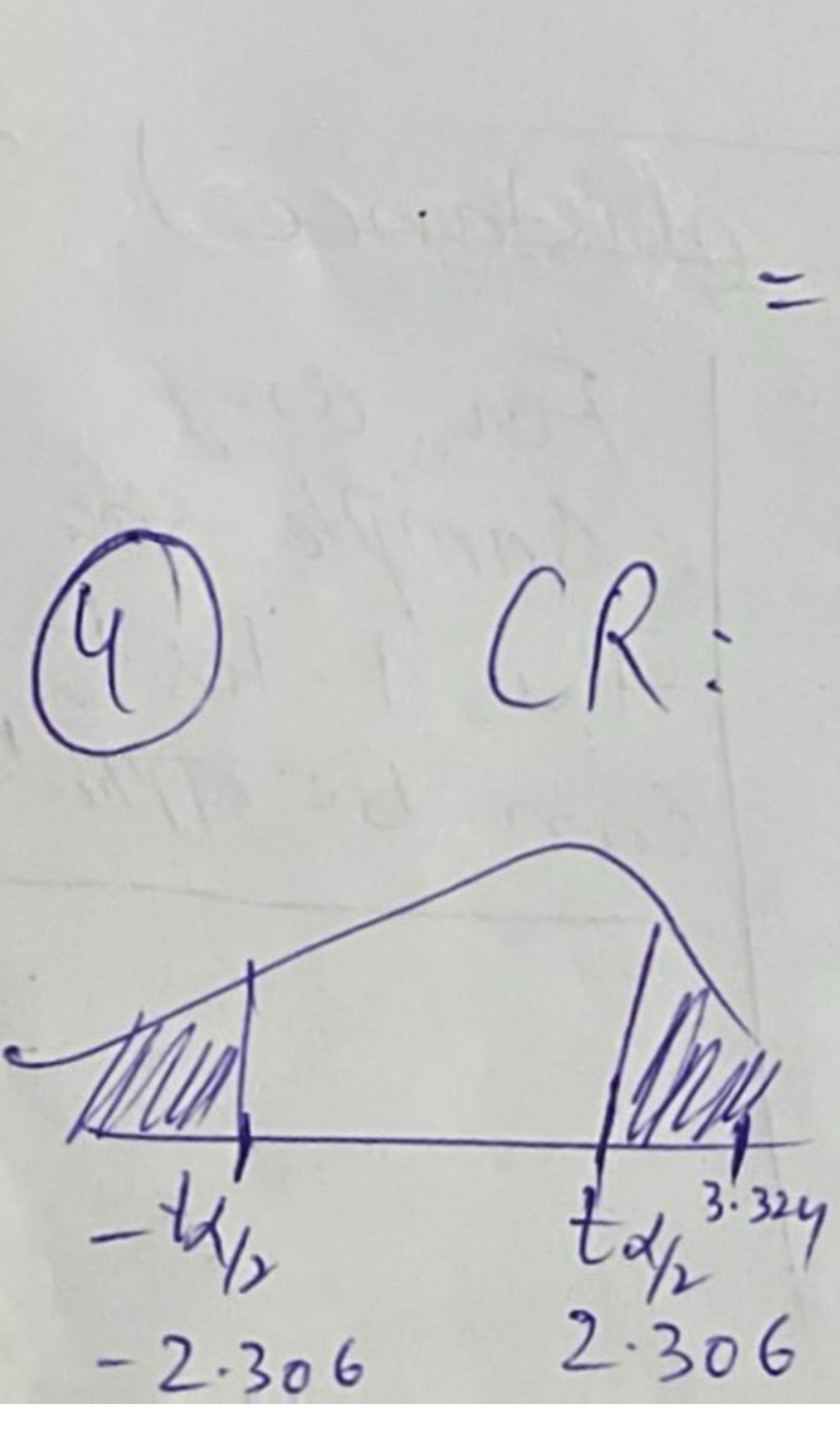
③ Test-statistics.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.7616\sqrt{10-2}}{\sqrt{1-(0.7616)^2}}$$

$$= \frac{2.1541}{0.6480} = 3.324$$

④ CR: $t < -t_{\alpha/2(v)}$ & $t > t_{\alpha/2(v)}$

$3.324 < -t_{0.025(8)}$ & $3.324 > t_{0.025(8)}$.



$3.324 < -2.306$. $3.324 > 2.306$

---

⑤ conclusion: Reject H_0. (Related linear X & Y) (