MTL390: Statistical Methods

Assignment #3
by

Rishit Sanmukhani
(2013CS10255)

Q-1 Note: My data is $X \sim U(5,14)$

Taking $\mu_0 = \bar{x}$

$$= \frac{5+14}{2} = 9.5$$

$$\therefore \boxed{\mu_0 = 9.5}$$

step-1

$H_0: \mu = 9.5$

$H_1: \mu \neq 9.5$

step-2

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

But as $n = 5000$ (large), we can approximate the 't' distribution with standard normal distribution 'z'.

$$\therefore \boxed{Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}} \quad (\because n = \text{large})$$

step-3

$$Z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Here, $\mu_0 = 9.5$, $\bar{x} = 9.5151$, $S = 2.5949$

$$Z_0 = \frac{9.5151 - 9.5}{\frac{2.5949}{\sqrt{5000}}}$$

$$\boxed{Z_0 = 0.4118}$$

Step-4 $\alpha = 0.05$

$Z_{\alpha/2} = Z_{0.025} = 1.96$ (i.e $P(Z \geq Z_{0.025}) = 0.025$)

$-Z_{\alpha/2} = -1.96$

Thus, critical region:

$Z_0 \leq -Z_{\alpha/2}$ OR $Z_0 \geq Z_{\alpha/2}$

$\Leftrightarrow$ $Z_0 \leq -1.96$ OR $Z_0 \geq 1.96$

Step-5 But, $-Z_{\alpha/2} < Z_0 < Z_{\alpha/2}$ ($\because Z_0 = 0.4118$)

$\therefore$ $\boxed{\text{Fail to reject } H_0}$

$\therefore$ Using the given sample, we could not reject hypothesis $H_0: \mu = 9.5$.

Q-2

Note: My data $X \sim U(5,14)$

Taking, $\sigma_0^2 = \frac{1}{12}(14-5)^2 = 6.75$

$$\boxed{\sigma_0^2 = 6.75}$$

step-1

$H_0: \sigma^2 = 6.75$

$H_1: \sigma^2 \neq 6.75$

step-2

$$D = \frac{(n-1)\, s^2}{\sigma^2} \sim \chi^2(n-1)$$

But since $n \to$ very large $(\because n = 5000)$

$$\chi^2(n-1) \sim N(n-1, 2(n-1))$$

Thus, $\boxed{Z = \frac{D-(n-1)}{\sqrt{2(n-1)}}}$ → standard normal distribution

step-3

$$Z_0 = \frac{D_0 - (n-1)}{\sqrt{2(n-1)}}$$

$$= \sqrt{\frac{n-1}{2}}\left(\frac{s^2}{\sigma_0^2} - 1\right)$$

Here, $n = 5000$

$s^2 = 6.7337$

$\sigma_0^2 = 6.75$

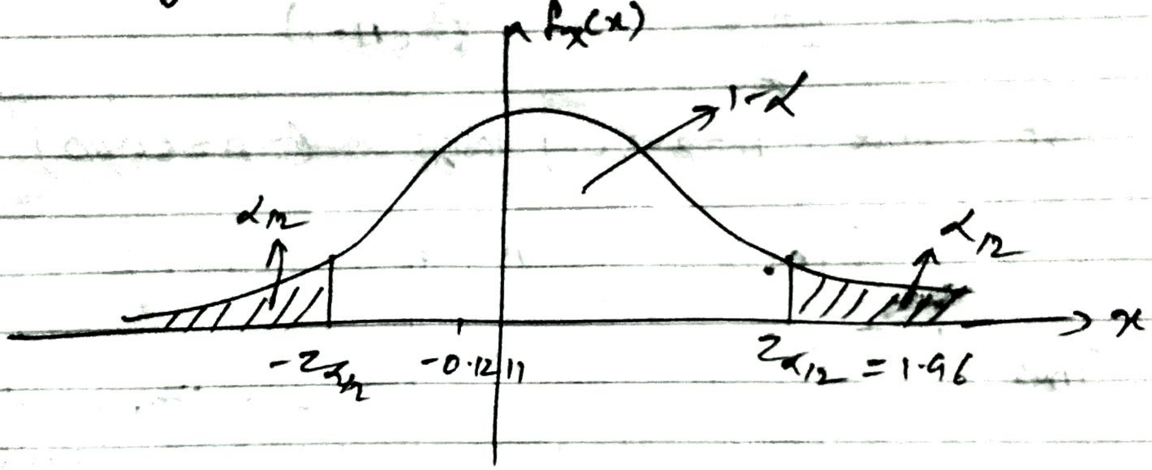
$$\therefore \boxed{Z_0 = -0.1211}$$

Step-4 $\alpha = 0.05$

$-z_{\alpha/2} = -1.96$

$z_{\alpha/2} = 1.96$

Critical region $\Rightarrow$ $Z_0 \le -1.96$ or $Z_0 \geq 1.96$



Step-5 Here, $Z_0 = -0.1211$

$\therefore$ $-z_{\alpha/2} < Z_0 < z_{\alpha/2}$

$\therefore$ Fail to reject H_0.

$\therefore$ Using the given sample, we fail to reject hypothesis $H_0 : \sigma^2 = 6.75$

Q-3 The distribution that best fits the data is Uniform distribution.

proof I used "allfitdist.m" matlab file for the same. The distribution was "generalized pareto".

Param: $K = -1.005$

$\text{sigma} = 9.041$

$\text{theta} = 5.003$

Placing in general pareto eq^n,

$$Y = f(x) = \frac{1}{9.0417}$$

$$\therefore \boxed{\text{Distribution is } U(5, 14)}$$

* Goodness of fit:

Data:

X_i	f_i
5 – 6	545
6 – 7	542
7 – 8	589
8 – 9	522
9 – 10	566
10 – 11	555
11 – 12	549
12 – 13	574
13 – 14	558
	$n = 5000$

groups = $K = 9$

$n = 5000$

Step-1 $H_0 : X \sim U(5,14)$
$H_1 : X \nsim U(5,14)$

Note: Here I am assuming both the parameters are already known to us.
- So I am not obtaining the parameters from the sample.

Step-2 $$D = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k-1)$$

O_i = observed freq
E_i = expected freq

Step-3 $P_i = P(X \in A_i)$ where A_i = ith group
Since uniform $U(5,14)$, $P_i = \frac{1}{9}$, $i = 1, 2, \ldots, 9$

$$\left(\because P_1 = \frac{6-5}{9},\ P_2 = \frac{7-6}{9}, \ldots, P_9 = \frac{14-13}{9}\right)$$

Now, we have,

$$E_i = nP_i, \quad n = 5000$$
$$= n \times \frac{1}{9}$$
$$\boxed{E_i = \frac{5000}{9}}, \quad i = 1, 2, \ldots, 9$$

$\therefore D = \dfrac{1}{\left(\frac{5000}{9}\right)} \left[0.2006 + 0.3308 + 2.0134 + 2.0268 + 0.1964 + 0.0006 + 0.0774 + 0.6124 + 0.0108 \right]$

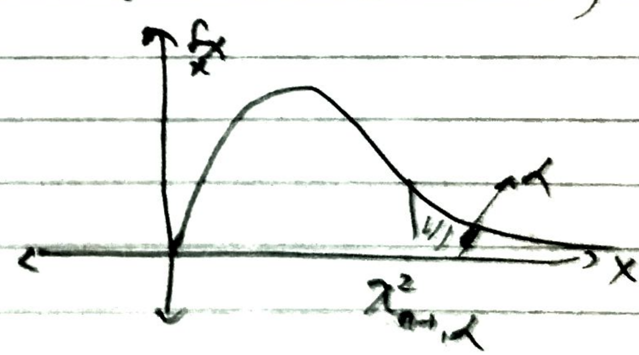
$$\boxed{D = 5.4688}$$

Step-4 Now, $D = \sum_{i=1}^{k} \dfrac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1}$

Here $k = 9$

$\therefore$ Critical Region:

$D_0 > \chi^2_{k-1, \alpha}$ (one-sided test)



$\therefore \chi^2_{8, 0.05} = 15.51$

Step-5 $\boxed{D_0 < \chi^2_{8, 0.05}}$

$\therefore$ Fail to reject H_0.

$\therefore$ Using given sample, we cannot reject null hypothesis $H_0: X \sim U(5, 14)$

parameters

Note: in above, if $U(5, 14)$, parameters were estimated, then,

$$\chi^2_{8-2, 0.05} = \chi^2_{6, 0.05} = 12.59$$

Still we fail to reject H_0

Q-4 Here, $m = \#$ samples

$\therefore$ $Y = [y_i]_{m\times 1}$

$X = [x_{(i)}]_{m\times(n+1)}$ where $x_{(i)} = [1 \; x_1 \; x_2 \cdots x_n]$

$$\therefore X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & x_{31} & & & \\ & \vdots & & & \\ 1 & x_{m1} & \cdots & & x_{mn} \end{bmatrix}_{m\times(n+1)}$$

* Analytical solution (least squares regression):

Let $\theta = (\theta_0 \; \theta_1 \cdots \theta_n)_{(n+1)\times 1}$ = set of parameters

$\therefore Y_p = \theta^T X$

$\therefore \min \|Y - Y_p\|^2 = \min \|Y - \theta^T X\|^2$

$\therefore$ Solving this we get, $\boxed{\theta = (X^T X)^{-1} X^T Y}$

Part (a) $\theta = [5 \quad 25.8567 \quad 49.2719]^T$

Intercept $= 5$

point estimates $= \beta_1 = 25.8567$

$\beta_2 = 49.2719$

Eqn: $\boxed{Y = 5 + 25.8567\, x_1 + 49.2719\, x_2}$

part-c Now, we have $y_p = \theta^T x$ — (1)

Residual sum of squares $RSS = \sum_{i=1}^{m} (y_i - y_{p_i})^2$

Now, $\frac{RSS}{\sigma^2} \sim \chi^2(m-n-1)$

$$\therefore \boxed{\sigma^2 = E\left[\frac{RSS}{m-n-1}\right]}$$

Here, $RSS = 9.0191 \times 10^{-19}$

$m = 100$

$n = 2$

$$\therefore \boxed{\sigma^2 = 9.298029 \times 10^{-21}}$$

part-d $R^2 = 1 - \frac{SSE}{SST}$

where SSE = residual sum of squares

$$= \sum_{i=1}^{m} (y_i - y_{p_i})^2$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Here, $SSE = 9.0191 \times 10^{-19}$

$SST = 2.4434 \times 10^{5}$

$\therefore \boxed{R^2 \approx 1}$ (very close to 1)

$$\text{Adjusted } R^2 = 1 - (1-R^2)\frac{m-1}{m-n-1}$$

But, $1-R^2 = \frac{SSE}{SST}$

$$\therefore \text{Adj-}R^2 = 1 - \frac{SSE}{SST}\,\frac{m-1}{m-n-1}$$

$$= 1 - \frac{\left(\frac{SSE}{m-n-1}\right)}{\left(\frac{SST}{m-1}\right)}$$

$$\boxed{\text{Adj-}R^2 = 1 - \frac{MSE}{MST}}$$

where MSE = mean sq error $= 9.2980 \times 10^{-21}$

MST = ~~mean~~ variance of y_i $= 2.4681 \times 10^{3}$

$$\therefore \text{Adj-}R^2 = 1 - \frac{9.2980\times 10^{-21}}{2.4681\times 10^{3}}$$

Again, $\boxed{\text{Adj-}R^2 \simeq 1}$

part - e $\quad$ t - test for significance of partial slope coeff:

step-1 $\quad H_0: \beta_1 = 0 \qquad\qquad H_0: \beta_2 = 0$

$\qquad\quad H_1: \beta_1 \neq 0 \qquad\qquad H_1: \beta_2 \neq 0$

step-2 $\quad t_1 = \dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{C_{22}}}$ $\quad$ where $\boxed{C = \hat{\sigma}^2 (X^T X)^{-1}}$

$\left(\text{Here } \hat{\sigma}^2 = MSE = \dfrac{SSE}{m-n-1} \right)$

$t_2 = \dfrac{\hat{\beta}_2 - \beta_2}{\sqrt{C_{33}}}$

step-3 In our case, $C = \begin{bmatrix} 0.0945 & 0.0022 & -0.0126 \\ 0.0022 & 0.1111 & 0.0102 \\ -0.1026 & 0.0102 & 0.1095 \end{bmatrix} \times 10^{-21}$

$\therefore \quad C_{22} = 0.1111 \times 10^{-21}$

$\qquad C_{33} = 0.1095 \times 10^{-21}$

$\therefore \quad t_1 = 2.4533 \times 10^{12} \quad \left(\because \beta_1 = 0, \ \hat{\beta}_1 = 25.8587 \right)$

$t_2 = 4.7094 \times 10^{12} \quad \left(\because \beta_2 = 0, \ \hat{\beta}_2 = 49.2719 \right)$

Step-4 Critical region:

$$t_{\alpha/2,\, m-n-1} = t_{0.025,\, 97} \approx 1.98$$

$$-t_{\alpha/2,\, m-n-1} = -t_{0.025,\, 97} = -1.98$$

Step-5 Clearly, $t_1 > t_{\alpha/2,\, m-n-1}$

$t_2 > t_{\alpha/2,\, m-n-1}$

$\therefore$ $\boxed{\text{Both the hypothesis are rejected.}}$

($H_0 : \beta_1 = 0 \longrightarrow$ Rejected

$H_0 : \beta_2 = 0 \longrightarrow$ Rejected) part-f F-test for overall significance of Regression

step-1
$H_0 : \beta_1 = 0, \beta_2 = 0$
$H_1 : \beta_i \neq 0$ for atleast one $i = 1, 2$

step-2

$$F = \frac{MST - MSE}{MSE} \sim F(n, m-n-1)$$

where, $$MST = \frac{\sum_{i=1}^{m} (y_i - \bar{y})^2}{m-1}$$

$$MSE = \frac{\sum_{i=1}^{m} (y_i - y_{p_i})^2}{m-n-1}$$

step-3

$$F_0 = \frac{2.4681 \times 10^3 - 9.298 \times 10^{-4}}{9.298 \times 10^{-4}}$$

$$\boxed{F_0 = 2.6544 \times 10^{23}}$$

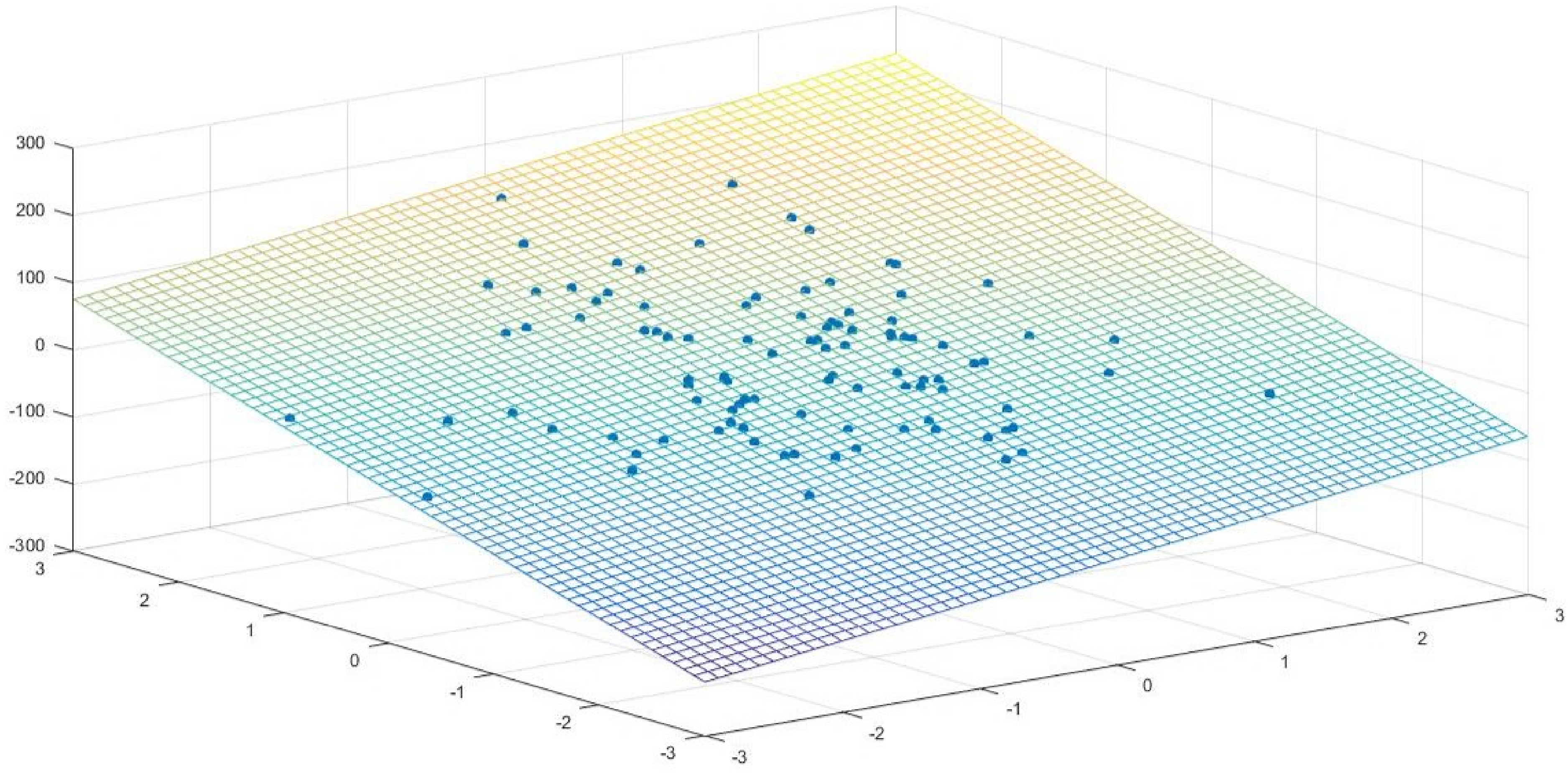
step-4 Critical Region :

$$f_{\alpha, n, m-n-1} = f_{0.025, 2, 97} = \approx 39.4$$

Step-5

$$\underbrace{F_0}_{2.6544\times10^{23}} > \underbrace{f_{\alpha, n, m-n-1}}_{\cong 39.4}$$

$\therefore$ $\boxed{\text{clearly } H_0 : \beta_1 = \beta_2 = 0 \text{ is rejected.}}$


300
200
100
0
-100
-200
-300
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3