ASSIGNMENT - 3

H) (a) 12 (4 + 4 + 4) + (b) 8 (4 + 4).

Q-1) In a study of wage differences b/w native & non-native workers of similar age & similar training the following equation is estimated:

$$\boxed{W_i = \alpha + \beta . D_i + u_i}$$

$W_i \rightarrow$ wage of worker i.

$D_i \rightarrow$ Dummy variables takes value 1 only if worker is non-native; OR zero otherwise.

$u_i \rightarrow$ stochastic error term.

$\overline{W}_{nat} \rightarrow$ Average wage of native workers.

$\overline{W}_{non} \rightarrow$ Average wage of non-native workers

$n_{nat} \rightarrow$ No. of native workers workers.

$n_{non} \rightarrow$ No. of non-native workers.

$\overline{W} \rightarrow$ average of W_i

$\overline{D} \rightarrow$ average of D_i.

(a) Show that the following relationships are

(1) $$\boxed{\bar{W} = \frac{n_{non}\bar{W}_{non} + n_{nat}\bar{W}_{nat}}{n}}$$

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\left[\sum_{i\in\{Nat\}} W_i + \sum_{i\in\{Non\}} W_i\right]$$

$$= \frac{1}{n}\left[n_{nat}\cdot\bar{W}_{nat} + n_{non}\bar{W}_{non}\right]$$

$\left(\because \sum_i x_i = n\bar{x}\right)$

$$\therefore \boxed{\bar{W} = \frac{n_{non}\bar{W}_{non} + n_{nat}\bar{W}_{nat}}{n}}$$

proved.

(2) $$\boxed{\bar{D} = \frac{n_{non}}{n_{non} + n_{nat}}}$$

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \frac{1}{n}\left[\sum_{i\in\{Nat\}} 0 + \sum_{i\in\{Non\}} 1\right]$$

$$= \frac{1}{n}\left[0 + n_{non}\cdot\bar{D}_{non}\right]$$

$$= \frac{1}{n} \left[n_{non} \cdot \left(\frac{\cancel{n_{non}}}{\cancel{n_{non}}} \right) \right]$$

$$= \frac{1}{n} (n_{non})$$

$$\boxed{\bar{D} = \frac{n_{non}}{n_{non} + n_{nat}}}$$

3

$$\boxed{\sum_{i=1}^{n} (D_i - \bar{D})^2 = \frac{n_{non} \cdot n_{nat}}{n_{non} + n_{nat}}}$$

$$\sum_{i=1}^{n} (D_i^2 + \bar{D}^2 - 2 D_i \bar{D}) = \sum_{i=1}^{n} (D_i - \bar{D})^2$$

$$= \sum_{i=1}^{n} D_i^2 + \sum_{i=1}^{n} \bar{D}^2 - 2\bar{D} \sum_{i=1}^{n} D_i$$

$$= n_{non} + n \left(\frac{n_{non}}{n} \right)^2 - 2 \left(\frac{n_{non}}{n} \right) (n_{non})$$

$$= n_{non} + \cancel{n} \frac{n_{non}^2}{n^{\cancel{2}}} - 2 \frac{n_{non}^2}{n}$$

$$= n_{non} - \frac{n_{non}^2}{n}$$

$$= \frac{n \cdot n_{non} - n_{non}^2}{n}$$

$$= \frac{n_{non}(n - n_{non})}{n}$$

$$= \frac{n_{non}(n_{nat} + \cancel{n_{non}} - \cancel{n_{non}})}{n}$$

$$\boxed{\sum_{i=1}^{n}(D_i - \bar{D})^2 = \frac{n_{non} \cdot n_{nat}}{n_{non} + n_{nat}}}$$

(b) OLS estimates of α & β.

(i) $$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n} (w_i - \bar{w})(D_i - \bar{D})}{\sum_{i=1}^{n} (D_i - \bar{D})^2}.$$

NR :

$$\hat{\beta}_{OLS}' = \sum_{i=1}^{n} (w_i - \bar{w}) D_i - \bar{D} \underbrace{\sum_{i=1}^{n} (w_i - \bar{w})}_{0}.$$

($\because$ sum of deviations from mean)

$$= \sum_{i=1}^{n} (w_i - \bar{w}) D_i$$

$\because D_i = 1$ for non- natives :

$$= \sum_{i \in \{non\}}^{n} (w_i - \bar{w}) . \quad = \sum_{i \in \{non\}} W_i - \sum_{i \in \{non\}} \bar{W}.$$

$$= (n_{non} \cdot \bar{W}_{non}) - n_{non} \cdot \bar{W}$$

$$= n_{non} (\bar{W}_{non} - \bar{W})$$

$$= n_{non} \left(\bar{W}_{non} - \left(\frac{n_{non} \bar{W}_{non} + n_{nat} \bar{W}_{nat}}{n} \right) \right)$$

$$= n_{non}\left(\frac{n\bar{W}_{non} - n_{non}\bar{W}_{non} - n_{nat}\bar{W}_{nat}}{n}\right)$$

$$\cancel{= n_{non}\left(\bar{W}_{non}(\quad n\right.}$$

$$= n_{non}\frac{\left(\bar{W}_{non}(n_{non}+n_{nat}) - n_{non}\bar{W}_{non} - n_{nat}\bar{W}_{nat}\right)}{n}$$

$$= n_{non}\cdot\frac{\left(\cancel{n_{non}\bar{W}_{non}} - \cancel{n_{non}\bar{W}_{non}} + n_{nat}\bar{W}_{non} - n_{nat}\bar{W}_{nat}\right)}{n}$$

$$= \frac{n_{non}\cdot n_{nat}(\bar{W}_{non} - \bar{W}_{nat})}{n}$$

$$= \left(\frac{n_{non}\, n_{nat}}{n}\right)(\bar{W}_{non} - \bar{W}_{nat})$$

$$= \sum_{i=1}^{n}(D_i - \bar{D})^2 \cdot (\bar{W}_{non} - \bar{W}_{nat}) \quad (a)\,③\,(Q-1)$$

D_R : $\Sigma(D_i - \bar{D})^2$.

$$\Rightarrow \boxed{\frac{N_R}{D_R} = \hat{\beta}_{OLS} = \bar{W}_{non} - \bar{W}_{nat}}$$

(iii) $\hat{\alpha}_{OLS} = \bar{W} - \hat{\beta}_{OLS}\bar{D}$

$$\hat{\alpha}_{OLS} = \bar{W} - (\bar{W}_{non} - \bar{W}_{nat}) \cdot \frac{n_{non}}{n}.$$

$$= \frac{n\bar{W} - n_{non}\bar{W}_{non} + n_{non}\bar{W}_{nat}}{n}.$$

$$= \cancel{\frac{1}{n}\left[(n_{non} + n_{nat})(\right.}$$

NR.:

$$= \cancel{n}\left(\frac{\cancel{n_{non}\bar{W}}_{non} + n_{nat}\bar{W}_{nat}}{\cancel{n}}\right) -$$

$$\cancel{n_{non}\bar{W}}_{non} + n_{non}\bar{W}_{nat}.$$

$$= \bar{W}_{nat}(n_{nat} + n_{non})$$

$$= \bar{W}_{nat}(n)$$

DR : n.

$$\Rightarrow \frac{NR}{DR} = \frac{\bar{W}_{nat}\,(\cancel{n})}{\cancel{n}} = \bar{W}_{nat}.$$

$$\boxed{\hat{\alpha} = \bar{W}_{nat}}$$

Q-2) Consider 2 r.v(s) A & B.

A → response variable.

B → predictor variable.

f → function.

$$A = f(B) + u. \rightarrow (1).$$

$$A = \gamma_1 B + u.$$

$$E(u) = 0$$

γ_1 → regn parameter.

2 performance metrics:

(a) coefficients of correlation:

$$\rho_{A,\hat{A}} = \sum_{i=1}^{n} (A_i - \bar{A})(\hat{A}_i - \bar{A}) \Big/ \sqrt{\sum_{i=1}^{n}(A_i-\bar{A})^2 \sum_{i=1}^{n}(\hat{A}_i-\bar{A})^2}$$

b) Goodness - of - fit parameter : ⑤

$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n} (\hat{A}_i - \bar{A})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2}.$$

To show that : $R^2 = (\rho_{A,\hat{A}})^2$.

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{A}_i - \bar{A})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2}.$$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{\gamma}_1 B_i - \bar{A})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2}.$$

$$= \frac{\sum_{i=1}^{n} (\hat{\gamma}_1 B_i - \hat{\gamma}_1 \bar{B})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2}.$$

$$= \hat{\gamma}_1^2 \cdot \frac{\sum_{i=1}^{n} (B_i - \bar{B})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2}$$

10

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} (A_i - \bar{A})(B_i - \bar{B})}{\sum_{i=1}^{n} (B_i - \bar{B})^2}.$$

10

$$\therefore R^2 = \frac{\left[\sum_{i=1}^{n} (A_i - \bar{A})(B_i - \bar{B})\right]^2}{\left[\sum_{i=1}^{n} (B_i - \bar{B})^2\right]^{\cancel{2}}} \cdot \frac{\cancel{\sum_{i=1}^{n} (B_i - \bar{B})^2}}{\sum_{i=1}^{n} (A_i - \bar{A})^2}.$$

$$= \frac{\left[\sum_{i=1}^{n} (A_i - \bar{A})(B_i - \bar{B})\right]^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2 \sum_{i=1}^{n} (B_i - \bar{B})^2}.$$

$$= \frac{\sum_{i=1}^{n} \left[(A_i - \bar{A}) \left(\frac{\hat{A}_i}{\hat{\gamma}_1} - \frac{\bar{A}}{\hat{\gamma}_1}\right)\right]^2}{\sum_{i=1}^{n} \left[\left(\frac{\hat{A}_i}{\hat{\gamma}_1} - \frac{\bar{A}}{\hat{\gamma}_1}\right)^2\right] \left[\sum_{i=1}^{n} (A_i - \bar{A})^2\right]}$$

$$\left[\begin{aligned} \bar{\hat{A}} &= \hat{\gamma}_1 \bar{B} \\ &= \bar{A} \quad (\because E(u) = 0) \end{aligned}\right]$$

$$= \frac{\cancel{\frac{1}{\hat{\gamma}_1^2}} \cdot \sum_{i=1}^{n} \left[(A_i - \bar{A})(\hat{A}_i - \bar{A})\right]}{\cancel{\frac{1}{\hat{\gamma}_1^2}} \left[\sum_{i=1}^{n} [\hat{A}_i - \bar{A}]^2\right] \left[\sum_{i=1}^{n} (A_i - \bar{A})^2\right]}$$

$$= \left(\rho_{A, \hat{A}}\right)^2.$$

Proved.

ASSIGNMENT - 3.
SOLUTIONS.

Q-1) & Q-2) → ~~[illegible]~~ separately done.

Q-3) The integral of a probability density
(a) function over its domain is 1.

(7.5)

$$\therefore \int_0^\infty f(x)\, dx = 1.$$

$$\int_0^\infty c \cdot e^{-x/10}\, dx = 1.$$

Putting $-x/10 = u$.

$$-\frac{dx}{10} = du.$$

$$-10c \int_0^{-\infty} e^{u}\, du = 1.$$

$$-10c \cdot \left[e^{u}\right]_0^{-\infty} = 1.$$

$$10c = 1.$$

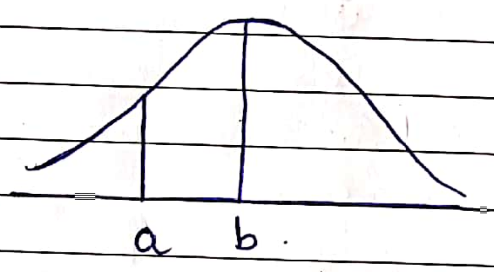
$$\boxed{c = 1/10}$$

(7.5)

(b) Probability that a call lasts exactly 7 minutes :

$f(x=7) = Pr(x=7)$ → Probability at a point for a continuous distribution.

In case of continuous distributions; probability = area under the curve.



$$Pr(a \le x \le b) = \int_a^b f(x)\,dx.$$

At a pt., area under the curve is zero.

$\therefore$ $\boxed{Pr(x=7) = 0}$

→ (10 for either solⁿ)

(2).

Q-4) In each trial, I may observe a no. larger than 4 with the probability $= \frac{2}{6} = \frac{1}{3}$.

$\therefore$ this is like a Bernoulli experiment w/ success probability, $p = \frac{1}{3}$ until you observe the first success.

$$P_N(k) = \begin{cases} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{k-1} & \text{for } k = 1,2,3 \ldots \\ 0 & \text{o/w.} \end{cases}$$

$\therefore$ N is a geometric random variable of probability of success $= \frac{1}{3}$.

$$N \sim \text{geometric}\left(\tfrac{1}{3}\right)$$

Geometric Distn = No. of trials before the first success.

Q-4). For $k=1$: Total no. of times a dice is rolled until number > 4 appears, is 1.

$S_1 = \{5, 6\}$.

$Pr(X=1) = Pr(S_1) = 2/6 = 1/3$.

For $k=2$: Total no. of times a dice is rolled until no. >4 appears, its 2.

1st roll results in $\{1, 2, 3, 4\} = S_1$
2nd " " " $\{5, 6\} = S_2$

$Pr(X=2) = P(S_1) \cdot P(S_2) = \frac{4}{6} \cdot 2/6 = 2/9$.

etc.

$\boxed{\text{Alternately.}}$

Q-1).

(a) → 1 (d) → 2 (g) → 1

(b) → 2 (e) → 2 (h) → 2

(c) → 2 (f) → 3

③

Marg. Distn of X

X \ Y	0	1	2	3	TOTAL
20	0.25	0.04	0.01	0.00	0.30
40	0.15	0.12	0.08	0.05	0.40
60	0.25	0.04	0.01	0.00	0.30
TOTAL	0.65	0.20	0.10	0.05	1.00

Marg. distn of Y.

(a) Marginal Distribution of X = $Pr(X=x)$

$$Pr(X=x) = \sum_{y \in Y} Pr(X=x, y)$$

$Pr(X=20) = 0.30$

$Pr(X=40) = 0.40$

$Pr(X=60) = 0.30$

Marginal Distribution of Y = $Pr(Y=y)$

$$Pr(Y=y) = \sum_{x \in X} Pr(x, Y=y)$$

$Pr(Y=0) = 0.65$

$Pr(Y=1) = 0.20$

$Pr(Y=2) = 0.10$

$Pr(Y=3) = 0.05$

(b) $E(X) = \sum_{x_i \in X} x_i \cdot Pr(x_i = X)$

$= 20 \cdot (0.30) + 40 \cdot (0.40) + 60 \cdot (0.30)$

$= 6 + 16 + 18$

$\boxed{E(X) = 40}$

Scanned by CamScanner

$$E(Y) = \sum_{y_i \in Y} y_i \Pr(Y = y_i)$$

$$= 0.(0.65) + 1.(0.20) + 2.(0.10) + 3.(0.05)$$

$$= 0.20 + 0.20 + 0.15$$

$$\boxed{E(Y) = 0.55}$$

(c) $\sigma_X^2 = E(X^2) - (E(X))^2$

$$E(X^2) = \sum_{x_i \in X} x_i^2 . \Pr(X = x_i)$$

$$= 400 . (0.30) + 1600 . (0.40) + 3600 . (0.30)$$

$$= 120 + 640 + 1080$$

$$= 1840$$

$$\sigma_X^2 = 1840 - (40)^2$$

$$= 1840 - 1600 = 240$$

$$\boxed{\sigma_X^2 = 240}$$

$$\sigma_Y^2 = E(Y^2) - (E(Y))^2$$

$$E(Y^2) = 0.(0.65) + 1.(0.20) + 4.(0.10) + 9.(0.05)$$

$$= 0.20 + 0.40 + 0.45$$

$$= 1.05$$

$$\sigma_Y^2 = 1.05 - (0.55)^2 = 1.05 - 0.3025 = 0.7475$$

$$\boxed{\sigma_Y^2 = 0.7475}$$

d) $\sigma_{xy} = E(XY) - E(X) \cdot E(Y)$.

$E(XY) = (0)(20)(0.25) +$
$(1)(20)(0.04) +$
$(2)(20)(0.01) +$
$(3)(20)(0) +$
$(0)(40)(0.15) +$
$(1)(40)(0.12) +$
$(2)(40)(0.08) +$
$(3)(40)(0.05) +$
$(1)(60)(0.04) +$
$(0)(60)(0.25) +$
$(2)(60)(0.01) +$
$(3)(60)(0)$.

$= 0.8 + 0.4 + 4.8 + 6.4 + 6 + 2.4$
$+ 1.2 - (40)(0.55)$

$= 22 - 22 = 0. \Rightarrow \boxed{\sigma_{xy} = 0}$

$\boxed{corr(X,Y) = \sigma_{xy} / \sigma_x \sigma_y = 0}$

(e) If X & Y are independent:

$$Pr(X = x, Y = y) = P(X = x) \cdot P(Y = y) \ \forall x, y.$$

at $X = 20$, $Y = 0$.

$Pr(X = 20, Y = 0) = 0.25$.

$Pr(X = 20) = 0.30$.

$Pr(Y = 0) = 0.65$.

$0.65 \times 0.30 \neq 0.25$.

$\therefore$ X & Y are <u>not</u> independent.

(f) To get to conditional means, we first need the conditional probability distributions.

$$f(y \mid X = x) = \frac{\text{Joint pmf } (X = x, y)}{(\text{Marginal pmf of } X = x)}$$

$$f(y \mid X = 20) = \begin{cases} \frac{0.25}{0.30} = 0.83 & \text{if } y = 0 \\ \frac{0.04}{0.30} = 0.13 & \text{if } y = 1. \\ \frac{0.01}{0.30} = 0.03. & \text{if } y = 2. \\ \frac{0.00}{0.30} = 0 & \text{if } y = 3. \end{cases}$$

⑥

Ily.

$$f(y \mid x=40) = \begin{cases} 0.375 & \text{if } y=0 \\ 0.3 & \text{if } y=1 \\ 0.2 & \text{if } y=2 \\ 0.125 & \text{if } y=3 \end{cases}$$

$$f(y \mid x=60) = \begin{cases} 0.83 & \text{if } y=0 \\ 0.13 & \text{if } y=1 \\ 0.03 & \text{if } y=2 \\ 0 & \text{if } y=3 \end{cases}$$

$$E(Y|X=20) = \sum_{y_i \in Y} y_i \cdot f(y|X=20) = 0(.83) + 1(0.13) + 2(0.03) + 3.(0).$$

$$= 0.13 + 0.06$$

$$\boxed{E(Y|X=20) = 0.19}$$

$$E(Y|X=40) = \sum_{y_i \in Y} y_i \cdot f(y|X=40) = 0(0.375) + 1(0.3) + 2(0.2) + 3(0.125).$$

$$= 0.3 + 0.4 + 0.375$$

$$\boxed{E(Y|X=40) = 1.075}$$

(g). $Pr(X=40 \mid Y=2) = ? = \dfrac{Pr(X=40, Y=2)}{Pr(Y=2)}$

$= \dfrac{0.08}{0.10} = 8/10 = 0.8.$

(h). $Z = 100 + 25Y.$

$E(Z) = E(100 + 25Y)$

$= 100 + E(25Y) = 100 + 25(E(Y))$

$= 100 + 25(0.55).$

$= 100 + 13.75.$

$= 113.75.$

$\boxed{\text{Avg. exp.} = \$113.75.}$

(i).

$V(Z) = V(100 + 25(Y))$

$= \cancel{V(100)}^{0.} + 625\,[V(Y)] + 2\,\cancel{cov.(100, 25Y)}^{0.}$

$= 625 \cdot V(Y).$

$\sigma_Z = \sqrt{625 \cdot V(Y)} = 25\sqrt{V(Y)} = 25\sqrt{0.7475}.$

$= 21.61.$

$\boxed{S.D_Z = \$21.61.}$

Scanned by CamScanner

(a), (b), (c), (d) → 2.5 each.

Q-2) (a) If $Y \sim N(2, 25)$. then, what is $Pr(Y>4)$

$\mu_Y = 2$.

$\sigma_Y = 5$.

$$Pr(Y>4) = 1 - Pr(Y \le 4)$$

$$= 1 - Pr\left(\frac{Y - \mu_Y}{\sigma_Y} \le \frac{4-2}{5}\right)$$

$$= 1 - Pr\left(Z \le 2/5\right)$$

$(Z \sim N(0,1))$

$$= 1 - Pr(Z \le 0.4)$$

$$= 1 - 0.6554$$

$$= 0.3446.$$

(b). $Y \sim N(7, 49)$, $Pr(Y<0) = ?$

$\mu_Y = 7$

$\sigma_Y = 7$.

$$Pr(Y<0) = Pr\left(\frac{Y - \mu_Y}{\sigma_Y} < \frac{0-7}{7}\right)$$

$$= Pr(Z < -1) \quad (Z \sim N(0,1))$$

$$= 0.1587.$$

(c) $Y \sim N(5,4)$; $Pr(3 \le Y \le 7) = ?$

$\mu_Y = 5$

$\sigma_Y = 2$

$Pr(3 \le Y \le 7)$

$$= Pr\left(\frac{3-5}{2} \le \frac{Y-\mu_Y}{\sigma_Y} \le \frac{7-5}{2}\right)$$

$= Pr(-1 \le Z \le 1)$ $(Z \sim N(0,1))$

$= Pr(Z \le 1) - Pr(Z \le -1)$

$= 0.8413 - 0.1587$

$= 0.6826$

(d) $Y \sim N(5,16)$, $Pr(3 \le Y \le 11) = ?$

$\mu_Y = 5$

$\sigma_Y = 4$

$$Pr(3 \le Y \le 11) = Pr\left(\frac{3-5}{4} \le \frac{Y-\mu_Y}{\sigma_Y} \le \frac{11-5}{4}\right)$$

$= Pr(-0.5 \le Z \le 1.5)$ $(Z \sim N(0,1))$

$= Pr(Z \le 1.5) - Pr(Z \le -0.5)$

$= 0.9332 - 0.3085 = 0.6247$

(a), (b), (c) (d) → 2.5 each.

Q-3) (a) $Y \sim \chi^2_{11}$; $Pr(Y > 19.68) = ?$

Note: $f(x) = \begin{cases} Y_0 \cdot x^{\nu/2 - 1} e^{-x/2} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$

Y_0 → constant. (dependent on degrees of freedom).
ν → degrees of freedom.

$Y \sim \chi^2_{11}$

$\nu = 11$

$Pr(Y > 19.68) = 1 - Pr(Y \leq 19.68)$
$= 1 - 0.95$
$= 0.05$.

(b). $Y \sim \chi^2_3$.
$Pr(Y > 11.34) = 1 - Pr(Y \leq 11.34)$.
$= 1 - 0.99$
$= 0.01$.

Scanned by CamScanner

(c) $Y \sim F_{4,20}$; $Pr(Y > 2.25) = ?$ (9).

$Pr(Y > 2.25) = 0.10$.

(d) $Y \sim F_{3,7}$.

$Pr.(Y > 8.45) = 0.01$.