(a1) (i) level - level. regression model.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

this Simple regression model follows all the Gauss markers Assumptions

$$\frac{d}{dx_i}(y_i = \beta_0 + \beta_1 x_i + u_i)$$

$$\frac{dy_i}{dx_i} = \beta_1$$

$$dy_i = \beta_1 dx_i \Rightarrow \frac{dy}{dx_i} = \beta_1$$

Thus it is clear that for one unit of change in x_i, there will be β_1 change in y_i,

(ii) log- level.

Simple regression model $\rightarrow \log(y_i) = \beta_0 + \beta_1 x_i + u_i$

This follows all the gauss markov assumptions

partial derivative w.r.t X_i

$$\frac{d}{dx_i}(\log(y_i)) = \frac{d(\beta_0 + \beta_1 x_i + u_i)}{dx_i}$$

$$\frac{1}{y}\frac{dy}{dx_i} = \beta_1$$

$$\frac{dy}{y} = \beta_1 dx_i$$

$$100 \frac{dy}{y} = (100)(\beta_i dx_i)$$

$$\Delta y\% = 100 \beta_i dx$$

$\therefore$ for one unit change in x_i there would be $100 \beta_i$ percentage change

(iii) level-log:

$$y = \beta_0 + \beta_1 \log(x_i) + u_i$$

This model follows all the Gauss Markov Assumptions.

$$\frac{d(y)}{dx_i} = \frac{(\beta_0 + \beta_1 \log(x_i) + u_i)}{dx_i}$$

$$\frac{dy}{dx_i} = \beta_1 \cdot \frac{1}{x_i}$$

$$dy = \beta_1 \times \frac{dx}{x} \Rightarrow dy = \frac{\beta_1}{100} \times \frac{dx}{x} \times 100$$

$$dy = \frac{\beta_1}{100} \times \Delta x\%$$

$\therefore$ for 1 percent change in x, y will change by $\frac{\beta_1}{100}$ units.

(iv) log - log.

$$\log(y_i) = \beta_0 + \beta_1 \log(x_i) + u_i$$

This model follows gaus markov Assumptions

$$\frac{d}{dx_i}\left(\log(y_i) = \beta_0 + \beta_1 \log(x_i) + u_i\right)$$

$$\frac{1}{y}\frac{dy}{dx_i} = \frac{\beta_1}{x}$$

$$\frac{dy}{y} = \beta_1 \times \frac{dx_i}{x}$$

$$100 \times \frac{dy}{y} = \beta_1 \times 100 \times \frac{dx}{x}$$

$$\Delta y\% = \beta_1 \; \Delta x\%$$

$\therefore$ Thus for 1 percent change in x, there will be β_1 percent of change

(Q5) Given that

(a) X is the duration of a phone call in minutes. And X is a random variable with probability density function $f(x) = c\,e^{-x/10}$, where C is constant & $x \geq 0$

Now, $f(x) = \begin{cases} c\,e^{-x/10} & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

$$\int_x f(x)\,dx = 1$$ [$\because$ Since total probability is always equal to 1]

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

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

$$c\left[\frac{e^{-x/10}}{-1/10}\right]_0^\infty = 1$$

$$-(10)c\left[e^{-\infty/10} - e^{0}\right] = 1$$

$$[0 - 1] = 1$$

$$10(c) = 1 \quad \boxed{c = 1/10}$$

(5b) $P(x = 7) = 0$

Since point on n. pdf curve has 0 probability at any point

from the definition of probability density function

ECO

(Q6) For every trial, the probability of number being greater than 4 (5, 6) $= \frac{2}{6} = 1/3$.

Thus, we repeat this experiment untill we get 5 or 6. probability of Sucess $= 1/3$.

Thus, this is a Bernoulli ~~random vari~~ principle in geometric random variable. with success probability $p = 1/3$.

$$P(N=k) = \begin{cases} \frac{1}{3}\left(\frac{2}{3}\right)^{k-1} & \text{for all } k \geq 1. \\ 0 & \text{otherwise.} \end{cases}$$

for $k=1$ $\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{0} = 1/3$

$k=2$ $\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{1} = \frac{2}{9}$

$k=3$ $\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{2} = \frac{1}{3} \times \frac{4}{9} = \frac{4}{27}$.

(Q2) $y = \beta_0 + \beta_1 x_i + u_i \quad i = 1, 2, \dots n$

To prove that OLS is given by

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Taking the estimations of u to be zero.

$$E(u) = 0$$

$$Cov(x, u) = E(xu) = 0$$

From equations,

$$y - \beta_0 - \beta_1 x = u$$

$$E(y - \beta_0 - \beta_1 x) = 0$$

These equations below

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \text{— (1)}$$

And $$\frac{1}{n} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \text{— (2)}$$

by solving (1) we get

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ — (3)

(3) — in (2)

$$\frac{1}{n} \sum_{i=1}^{n} x_i \left(y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} \left(x_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i (x_i - \bar{x} \right)$$

We knows that

$$\sum_{i=1}^{n} x_i (x_i - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\& \sum_{i=1}^{n} \left(x_i (y_i - \bar{y} \right) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

As, square number is always positive.

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

Thus,

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$y \rightarrow$ test scores for each student

$x \rightarrow$ no. of hours invested by the student

(a) minutes invested instead of hours.

$$\underset{(hrs)}{x_i} \longrightarrow \underset{(mins)}{60\, x_i}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (60 x_i - 60\bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{n} (60 x_i - 60\bar{x})^2}$$

$$\hat{\beta}_1 = \frac{60}{3600} \hat{\beta}_1$$

$$\hat{\beta}_1 = \frac{1}{60} \hat{\beta}_1$$

Thus $\hat{\beta}_1$ changes by a factor of $1/60$.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Now $\hat{\beta}_0$,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times 60\bar{x}$$

$$\hat{\beta}_0 = \bar{y} - \frac{1}{60} \hat{\beta}_1 \, 60\bar{x}$$

$$\hat{\beta}_0' = \bar{y} - \hat{\beta}_1 \bar{x} = \hat{\beta}_0$$

$\therefore$ there is no change in $\hat{\beta}_0$.

(b) seconds invested

$x_i \longrightarrow 3600\, x_i$ (seconds)

now

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{(3600\, x_i - 3600\bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (3600(x_i) - 3600\,\bar{x})^2}$$

$$= \frac{3600}{(3600)^2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = \frac{1}{3600} \hat{\beta}_1$$

Thus, β_1 changes by a factor of $\frac{1}{3600}$ "

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times 3600\,\bar{x}$$

$$= \bar{y} - \frac{1}{\cancel{3600}} \times \hat{\beta}_1 \times \cancel{3600}\,\bar{x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x =$$

Thus there is no changen in $\hat{\beta}_0$.

(2c) $\log(x_i)$ or $\log(x_i)$

$$\hat{\beta}_1 (\text{changed}) = \frac{\log(x_i) - \log(\bar{x}) \; (\bar{y} - y)}{\sum_{i=1}^{n} \log(x_i) - \log \bar{x}}$$

$$\hat{\beta}_0 (\text{changed}) = \bar{y} - \hat{\beta}(\text{changed}) \log x_i)$$

$$\hat{\beta}_0 (\text{changed}) = \bar{y} - \frac{\log(x_i - \log \bar{x}) \; (\bar{y} - y)}{\sum_{i=1}^{n} (\log x_i - \log \bar{x})^2} (\log x_i)$$

$\therefore \; \bar{y} - \log x(x_i)$

$$= \frac{\sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^{n} \log x_i}{n}}{\sum_{i=1}^{n} \log(x_i) - \sum_{i=1}^{n} \left(\log \frac{x_i}{n}\right)^2}$$

Q3

In a study of wage differences b/w naive & non native workers of similar age and similar training the following equations is estimated

$$W_i = \alpha + \beta D_i + u_i \quad \text{— (1)}$$

For non-native workers

$D = 1$

$w_i = \alpha + \beta + u_i$

For native workers

$D_i = 0$

Thus $= W_i = \alpha + u_i$

Average wage of native workers $= W_{nat}$.

" " non " " $= W_{non}$.

" " " " $= n_{nom} = n_1$

No. of native workers $= n_{nat} = n_0$

No. of

(i) Total wage of native workers $= \sum_{i=C_1} W_i$

$= \sum_{i=1..} (\alpha + u_i)$

Average wage of native workers $= \frac{1}{n_0} \times$ Total wage no. of workers

$= \frac{1}{n_0} \times \sum_{i=(1..n_0)} (\alpha + u_i)$

$\bar{W}_{nat} = \frac{1}{n_0} \times \sum_{i=(1..)} (\alpha + u_i)$

$\bar{w}_{nat} \cdot n_0 = \sum_{i=(1..n_0)} (\alpha + u_i)$.

total wage of non-native workers $= \sum_{i=(1..n)} w_i$

$= \sum_{i=(1..n)} (\alpha + \beta + u_i)$

Avg wage of non-native workers $= \frac{1}{n_1} \times$ Total wage of native workers

$\bar{w}_{non} = \frac{1}{n_1} \times \sum_{i=(1..n)} (\alpha + \beta + u_i)$.

$\bar{w}_{non} \cdot n_1 = \sum_{i=(1..n_1)} (\alpha + \beta + u_i)$.

Avg wage of all workers $= \frac{1}{n} \times$ total wage of all workers

$$\bar{w} = \frac{\bar{w}_{nat} \cdot n_{nat} + \bar{w}_{non} \cdot n_{non}}{n}$$

(ii) We know, for non-native workers

$D_i = 0$

and for non-native workers

total D for native worker $= \sum D_i = 0$

" " " non-native " $= \sum D_i = n_1$

total D for all worker $= 0 + n_1$

Average D $= \bar{D} = \frac{0 + n_1}{n_0 + n_1} = \frac{\cancel{n_{on}}\, n_{non}}{n_{0t} + n_{non}}$.

(c) To prove

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

We know, $\bar{D} = \frac{n_{on}}{n_{non} + n_{nat}} = \frac{n_1}{n_0 + n_1}$

Calculating the summation term for native and non-native separately.

$$\sum_{i=1}^{n} (D_i - \bar{D})^2 = \sum_{i=(1..n_0)} (D_i - \bar{D})^2 + \sum_{i=(1..n_1)} (D_i - \bar{D})^2$$

$$\sum_{i=1..n_0} (D_i - \bar{D})^2 \quad , \quad [\because D_i = 0) \text{ for native}$$

$$\Rightarrow \sum_{i=1...n_0} (0 - \bar{D})^2 = n_0 . (\bar{D})^2$$

$$= n_0 \cdot \left(\frac{n_1}{n_0 + n_1}\right)^2$$

$$\sum_{i=1..n_1} (D_i - \bar{D})^2 \quad D_i = 1$$

$$D_i - \bar{D} = 1 - \bar{D} = 1 - \frac{n_1}{n_0 + n_1} = \frac{n_0}{n_0 + n_1}$$

$$\sum_{i=(1..n_1)} (D_i - \bar{D})^2 = n_1 \times \left(\frac{n_0}{n_0 + n_1}\right)^2$$

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

$$= n_0\left(\frac{n_1}{n_0+n_1}\right)^2 + n_1\left(\frac{n_0}{n_1+n_0}\right)^2$$

$$= \frac{n_0 n_1}{(n_0+n_1)^2}(n_0+n_1)$$

$$\sum_{i=1}^{n}(D_i - D)^2 = \frac{n_0 n_1}{n_0+n_1} = \frac{n_{om}\, n_{nat}}{n_{on}+n_{nat}}$$

Hence proved.

3b) OLS of α & β of eq (1)

To prove that $\hat{\alpha} = \bar{W}_{nat}$ and.

using $\hat{\alpha} = \bar{W} - \hat{\beta}\bar{D}$

for $\hat{\beta}_1$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(D_i - \bar{D})(W_i - \bar{W})}{\sum_{i=1}^{n}(D_i - D)^2}$$

We already know that

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

Calculate the number

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

$$\frac{n_0}{n} \sum_{i=1}^{n} (D_i - \bar{D})(w_i - \bar{w})$$

$$\left(\sum \frac{D_i w_i}{n} - \bar{D}\bar{w} - \bar{D}\bar{w} + \bar{D}\bar{w} \right)$$

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

$$\Rightarrow \sum_{i=1}^{n} D_i w_i - n \cdot \bar{D}\bar{w}$$

Separating for native and non-native

$$\Rightarrow \sum_{i=0} 0 \cdot w_i + \sum_{i=\{1..n\}} 1 \cdot w_i - \bar{D}\bar{w}_n$$

$$n_1 \bar{N}_{non} - \bar{D}\bar{w}_n$$

$$\Rightarrow \frac{n_{non} \bar{w}_{non} - \bar{D}\bar{w}_n}{\frac{n_{non} \cdot n_{nat}}{n_{non} + n_{nat}}}$$

$$\frac{n_{non}\,\bar{w}_{non} - \left(\frac{n_{on}}{n}\right)\left(\frac{n_{non}\,\bar{w}_{non} + n_{nat}\,\bar{w}_{nat}}{\cancel{n}}\right) \times \cancel{n}}{\frac{n_{non} \cdot n_{nat}}{n}}$$

$$\Rightarrow \frac{(n)\left(n_{non} \cdot \bar{w}_{non} - \frac{n_{on}}{n}\right)\left(n_{non}\,\bar{w}_{non} + n_{nat}\,\bar{w}_{nat}\right)}{n_{on} \cdot n_{nat}}$$

$$\Rightarrow \frac{(n_{non} + n_{nat})\cancel{(n_{on})}\,\bar{w}_{non} - \cancel{(n_{on})}\left(n_{non}\,\bar{w}_{non} + n_{nat}\,\bar{w}_{nat}\right)}{\cancel{n_{on}} \cdot n_{nat}}$$

$$\Rightarrow \frac{\cancel{w_{non}\,n_{non}} + n_{nat}\,w_{non} - \cancel{n_{on}\,w_{non}} + n_{nat}\,w_{nat}}{n_{nat}}$$

$$= \frac{\cancel{(n_{nat})}\left(\bar{w}_{non} - w_{nat}\right)}{\cancel{n_{nat}}}$$

$$\bar{\beta} \Rightarrow \bar{w}_{non} - w_{nat}$$

$$\hat{\alpha} = \bar{w} - \hat{\beta}\bar{D}$$

$$\alpha = \frac{n_{non}\bar{w}_{non} + n_{nat}\bar{w}_{nat}}{n} - \frac{(\bar{w}_{non} - \bar{w}_{nat})\, n_{non}}{n}$$

$$\hat{\alpha} = \frac{1}{n}\left(\cancel{n_{non}\bar{w}_{non}} + n_{nat}\bar{w}_{nat} - n_{non}\bar{w}_{non} + n_{non}\bar{w}\right.$$

$$\hat{\alpha} = \frac{1}{\cancel{n}}\left(\cancel{(n_{nat} + n_{non})}\right) \cdot (\bar{w}_{nat})$$

$[\because n = n_{nat} + n_{non}$

$$\hat{\alpha} = \bar{w}_{nat}$$

$\therefore$ $\hat{\alpha}$ is equal to the average wage of native workers.

(4) ~~A0e0~~

Two R.V A & B . A is the response variable that is assumed to be related to the predictor B through a function f such that f(B) approximates A. In the regression form we specify this relationship

as $A = f(B) + u = y_1 B + u$.

$E U = 0$ & y_1 is the regression parameter.

(a) coefficient of correlation, defined as

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

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

$$\rho^2_{A,\hat{A}} = \left(\frac{\mathrm{Cov}(A,\hat{A})}{\frac{\sum (A_i - \bar{A})^2 \sum_{i=1} (\hat{A} - \bar{A})}{(N-1)}} \right)^2$$

$$= \frac{\mathrm{Cov}(A,\hat{A})^2}{\frac{\sum_{i=1}^{n} (A_i - \bar{A})^2 \left(\sum_{i=1}^{n} (\hat{A}_i - \bar{A})^2\right)}{(N-1)^2}}$$

$$\Rightarrow \frac{Cov(A, \hat{A})}{Var(A) \times Var(\hat{A})} \quad (1)$$

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

$$\frac{Cov(A, \hat{A}_i)\ Cov(A, \hat{A}_i)}{Var(A)\ Var(\hat{A})} \quad (2)$$

assuming (2) $\cancel{\quad}$

$Cov(\hat{A}, u) = 0$ or $Cov(u, \hat{A}) = 0$

$A = \hat{A} + u$

from (2) $$\frac{Cov(\hat{A} + u, \hat{A})\ Cov(\hat{A} + u, \hat{A})}{Var(A)\ Var(\hat{A})}$$

$$\frac{\left(Cov(\hat{A}, \hat{A}) + Cov(\hat{A}, u)\right)\left(Cov(\hat{A}, \hat{A}) + Cov(\hat{A}, u)\right)}{Var(A)\ Var(\hat{A})}$$

$Cov(x, y+z) = Cov(x, y) + Cov(y, z)$

$Cov(\hat{A}, u) = 0$ $\quad Cov(y, y) = var(y)$

$$\frac{var(\hat{A}) \cancel{\times var(\hat{A})}}{var(A) \cdot \cancel{var(\hat{A})}} \Rightarrow \frac{\dfrac{\sum_{i=1}^{n} (\hat{A}_i - \bar{A})^2}{N-1}}{\dfrac{\sum_{i=1}^{n} (A_i - \bar{A})^2}{N-1}}$$

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

Hence proved.