- Define **deterministic** and **probabilistic** mathematical models. Give an example of each.

Determinlnistic models: models doesn't allow error in predicting Y as a function of x.

$Y = \beta_0 + \beta_1 x$

Ex: $F = ma$

Probabilistic models: model allow Error, ε is R.V with $E(\varepsilon)=0$

$Y = \beta_0 + \beta_1 x + \varepsilon$ Ex: Weather modeling

- Write the general equation for a **simple linear regression** model.

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

- Describe, in your own words, the overall concept of the **method of least squares**.

To obtain a predictive line that go through all given points with smallest Area formed by vertical distance between point and Line. which is to Minize the error.

- State the **least-squares estimators** for the simple linear regression model.

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

β_0 = A number, intersection

- State the means and variances of the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in simple linear regression.

$$E(\hat{\beta}_1) = \hat{\beta}_1 \quad , \quad V(\hat{\beta}_1) = \left[\frac{1}{S_{xx}}\right]^2 \Sigma (x_i - \bar{x})^2 V(Y_i)$$

where $V(Y_i) = \sigma^2$ for $i = 1, 2, 3, 4, 5 \ldots n$.

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

$$E(\hat{\beta}_0) = \beta_0 \quad , \quad V(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2 \left(\frac{\sigma^2}{S_{xx}}\right) - 0$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) = \frac{\sigma^2 \Sigma x_i^2}{n S_{xx}}$$

- State a pair of null and alternative hypotheses for making inferences about **single regression parameters** and **linear functions of the parameters**.

$\boxed{\text{Single Regression Parameter}}$

$H_0: \beta_i = \beta_{i0}$

Test-Statistic: $T = \dfrac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$

$H_a: \begin{cases} \beta_i > \beta_{i0} & \text{(upper tail Rejection Region)} \\ \beta_i < \beta_{i0} & \text{(lower tail Rejection Region)} \\ \beta_i \neq \beta_{i0} & \text{(two-tail Rejection Region)} \end{cases}$

Where $C_{00} = \dfrac{\Sigma x_i^2}{n S_{xx}}$ and $C_{11} = \dfrac{1}{S_{xx}}$ df: $n-2$

$\boxed{\text{Linear functions of the Parameter.}}$

Test for $\theta = a_0\beta_0 + a_1\beta_1$

$H_0: \theta = \theta_0$

$H_a: \begin{cases} \theta > \theta_0 \\ \theta < \theta_0 \\ \theta \neq \theta_0 \end{cases}$

Test-Statisstic: $T = \dfrac{\hat{\theta} - \theta_0}{S\sqrt{\dfrac{a_0^2 \frac{\Sigma x_i^2}{n} + a_1^2 - 2a_0 a_1 \bar{x}}{S_{xx}}}}$

Rejection Region: $\begin{cases} t > t_\alpha \\ t < -t_\alpha \\ |t| > t_{\alpha/2} \end{cases}$

Q1 a) Fit model : $Y = \beta_0 + \beta_1 x + \varepsilon$ to these data, using least square

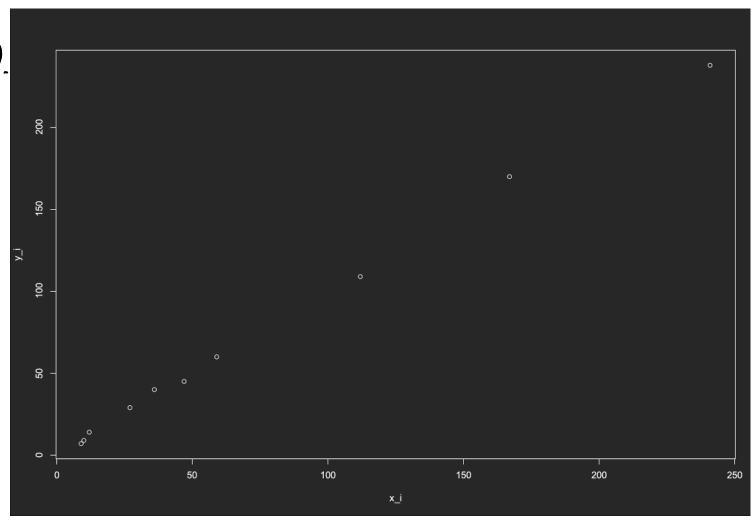
Since $E[\varepsilon] = 0 \Rightarrow \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{Cov(xy)}{Var(x)} = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = 0.9914$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.7198$$

$$Y = 0.9914\ x + 0.7198$$

b).



c). $SSE = \sum_{i=1}^{10} (y_i - \hat{y})^2 = 56.85$

$s^2 = \left(\frac{1}{10-2}\right) 56.85 = 7.10625$

d) $H_0: \mu = \beta_1 = 0$ $\quad H_A: \mu_1 = \beta_1 \neq 0$ $\quad$ 2-tail

$$T\text{-test} = \frac{\hat{\beta}_1 - \mu_0}{S\sqrt{C_{11}}} = \frac{0.9914}{2.665 \cdot \sqrt{\frac{1}{S_{xx}}}} = \frac{0.9914}{2.665 \cdot \sqrt{\frac{1}{54.74}}} = 87.016$$

$|t| > t_{0.05/2}, df = 8 \Rightarrow 2.306 < |t| = 87.016$

$\Rightarrow$ Reject $H_0 : \beta_1 = 0$, there is positive Relationship within the data

e) $E(Y) = \beta_1 x + \beta_0$

$= 0.9914 x + 0.7198$

Audit Value changes 1.7112 per 1 unit change in x.

f). $E(Y) = 0.9914 \times 100 + 0.7198$

$= 99.14 + 0.7198$

$= 99.8598$

2. Let β_0 and β_1 be the least-squares estimates for the intercept and slope in a simple linear regression model. Show that the least-squares equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ will always go through the point $(\bar{x}, \bar{y})$.

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

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} = \bar{y} - \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{(y_i - \bar{y})} \bar{x}$$

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

Now let $x_i = \bar{x} \Rightarrow$

$$\hat{y} = \bar{y} \underbrace{- \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{(y_i - \bar{y})} \bar{x} + \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{(y_i - \bar{y})} \bar{x}}_{\text{Cancel out}}$$

$\hat{y} = \bar{y} \Rightarrow \hat{y} = \beta_1 x + \beta_0$ go through $(\bar{x}, \bar{y})$

3. Suppose that the model $y = \beta_0 + \beta_1 x + \epsilon$ is fit to the n data points $(y_1, x_1), ..., (y_n, x_n)$. At what value of x will the length of the prediction interval for y be minimized?

$$\frac{d\,SSE}{d\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

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

$\Rightarrow$ when $x_i = \bar{x}$ the Prediction interval for y will be minized.