**EX 12.2.1:** The systolic blood pressures of twelve adults were measured as a single group without intervention†:

| SYSTOLIC BLOOD PRESSURE (in mmHg) versus Age (in years) | | | | | | | | | | | | | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index ($i$) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | $n = 12$ |
| Age ($x_i$) | 40 | 41 | 43 | 44 | 44 | 46 | 48 | 50 | 52 | 54 | 54 | 55 | $\sum_i x_i = 571$ |
| BP ($y_i$) | 158.8 | 158.5 | 163.1 | 165.1 | 166.3 | 165.7 | 172.1 | 174.5 | 173.8 | 176.2 | 171.2 | 177.2 | $\sum_i y_i = 2022.5$ |

†(Simplification of data) Y.H. Chan, "Biostatistics 201: Linear Regression Analysis", *Singapore Med. J.*, **45** (2004), 55-61.

(a) Identify the response and the sole regressor in this observational study.

Response := Systolic Blood Pressure (in mmHg)

Regressor := Age (in years)

(b) Formulate this observational study as a linear model.

$Y_i = \beta_0 + \beta_1 x_i + E_i$ where $E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$

(c) Compute the response sum and regressor sum: $S_y$, $S_x$.

$S_y := \sum_i y_i \overset{TABLE}{=} 2022.5 \implies \boxed{S_y = 2022.5}$ $\quad S_x := \sum_i x_i \overset{TABLE}{=} 571 \implies \boxed{S_x = 571}$

(d) Compute the squared regressor sum, $S_{xx}$.

$S_{xx} := \sum_i x_i x_i = x_1 x_1 + \cdots + x_n x_n = (40)(40) + \cdots + (55)(55) = 27483 \implies \boxed{S_{xx} = 27483}$

(e) Compute the cross-termed regressor-response sum, $S_{xy}$.

$S_{xy} := \sum_i x_i y_i = x_1 y_1 + \cdots + x_n y_n = (40)(158.8) + \cdots + (55)(177.2) = 96596.6 \implies \boxed{S_{xy} = 96596.6}$

(f) Compute the cross-termed regressor-response centered sum, $SC_{xy}$.

$SC_{xy} := \sum_i (x_i - \overline{x})(y_i - \overline{y}) = S_{xy} - \frac{1}{n} S_x S_y = 96596.6 - \left(\frac{1}{12}\right)(571)(2022.5) \approx 359.308333333 \implies \boxed{SC_{xy} \approx 359.308}$

(g) Compute the squared regressor centered sum, $SC_{xx}$.

$SC_{xx} := \sum_i (x_i - \overline{x})(x_i - \overline{x}) = S_{xx} - \frac{1}{n} S_x S_x = 27483 - \left(\frac{1}{12}\right)(571)(571) \approx 312.916666667 \implies \boxed{SC_{xx} \approx 312.917}$

(h) Compute the ordinary least-squares (OLS) estimates of the regression parameters $\hat{\beta}_0, \hat{\beta}_1$.

Use the two centered sums to compute slope estimate: $\hat{\beta}_1 \overset{OLS}{=} \frac{SC_{xy}}{SC_{xx}} = \frac{359.308333333}{312.916666667} \approx 1.148255659$

Use two sums & $\hat{\beta}_1$ to compute intercept estimate: $\hat{\beta}_0 \overset{OLS}{=} \frac{S_y - \hat{\beta}_1 S_x}{n} = \frac{2022.5 - (1.148255659)(571))}{12} \approx 113.903834893$

Finally, round the estimates: $\boxed{\hat{\beta}_0 \approx 113.904}$ $\boxed{\hat{\beta}_1 \approx 1.148}$

(i) Determine the OLS best-fit line.

OLS Best-fit line: $y = \hat{\beta}_0 + \hat{\beta}_1 x \implies$ OLS Best-fit line: $y = 113.904 + 1.148x$

(j) Compute the squared response sum, $S_{yy}$.

$S_{yy} := \sum_i y_i y_i = y_1 y_1 + \cdots + y_n y_n = (158.8)(158.8) + \cdots + (177.2)(177.2) = 341342.31 \implies \boxed{S_{yy} = 341342.31}$

(k) Compute the squared response centered sum, $SC_{yy}$.

$SC_{yy} := \sum_i (y_i - \overline{y})(y_i - \overline{y}) = S_{yy} - \frac{1}{n} S_y S_y = 341342.31 - \left(\frac{1}{12}\right)(2022.5)(2022.5) \approx 466.789166667 \implies \boxed{SC_{yy} \approx 466.789}$

($\ell$) Compute the residual sum of squares, $\text{SS}_{res}$.

$\text{SS}_{res} = SC_{yy} - \hat{\beta}_1^2 SC_{xx} = 466.789166667 - (1.148255659)^2(312.916666667) \approx 54.211339634 \implies \boxed{\text{SS}_{res} \approx 54.211339634}$

(m) Compute the linear model's estimated variance, $\hat{\sigma}^2$.

$\hat{\sigma}^2 = \frac{\text{SS}_{res}}{n-2} = \frac{54.211339634}{12-2} \approx 5.421133963 \implies \boxed{\hat{\sigma}^2 \approx 5.421}$

(n) Compute $R^2$.

$R^2 = \frac{SC_{xy} \cdot SC_{xy}}{SC_{xx} \cdot SC_{yy}} = \frac{359.308333333 \cdot 359.308333333}{312.916666667 \cdot 466.789166667} = \frac{129102.478402538}{146066.110069704} \approx 0.88386333 \implies \boxed{R^2 \approx 0.884}$