SIMPLE LINEAR REGRESSION:
MODEL INFERENCE PREREQUISITES [DEVORE 12.3]

SUMS (DEFINITION): Let vectors $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$. Then:

$$S_x := \sum_i x_i \qquad S_y := \sum_i y_i \qquad S_{xx} := \sum_i x_i x_i \qquad S_{yy} := \sum_i y_i y_i \qquad S_{xy} := \sum_i x_i y_i$$

CENTERED SUMS (DEFINITION): Let vectors $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$. Then:

$$SC_{xx} := \sum_i (x_i - \overline{x})^2 \qquad SC_{yy} := \sum_i (y_i - \overline{y})^2 \qquad SC_{xy} := \sum_i (x_i - \overline{x})(y_i - \overline{y})$$

CENTERED SUMS LEMMA:

$$(a)\ SC_{xx} = S_{xx} - \frac{1}{n} S_x S_x \qquad (b)\ SC_{yy} = S_{yy} - \frac{1}{n} S_y S_y \qquad (c)\ SC_{xy} = S_{xy} - \frac{1}{n} S_x S_y$$

OLS ESTIMATORS:

$$\begin{cases} \hat{\beta}_1 = \dfrac{SC_{xy}}{SC_{xx}} = \dfrac{S_{xy} - \frac{1}{n} S_x S_y}{S_{xx} - \frac{1}{n} S_x S_x} \\ \\ \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \dfrac{S_y - \hat{\beta}_1 S_x}{n} \end{cases}$$

SUMS OF SQUARES (DEFINITION):

$$SS_{total} := \sum_i (y_i - \overline{y})^2 \qquad SS_{res} := \sum_i (y_i - \hat{y}_i)^2 \qquad SS_{reg} := \sum_i (\hat{y}_i - \overline{y})^2$$

VARIATION PARTITIONING:

$$\underbrace{\sum_i (y_i - \overline{y})^2}_{SS_{total}} = \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{SS_{res}} + \underbrace{\sum_i (\hat{y}_i - \overline{y})^2}_{SS_{reg}}$$

EXPECTATION & VARIANCE OF $\hat{\beta}_1$:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 \qquad \mathbb{V}[\hat{\beta}_1] = \frac{\sigma^2}{SC_{xx}}$$

MEAN SQUARED RESIDUAL:

$$\mathrm{MS}_{res} := \frac{\mathrm{SS}_{res}}{n-2}$$

POINT ESTIMATOR OF σ^2:

$$\mathbb{E}\left[\mathrm{MS}_{res}\right] = \sigma^2 \implies \hat{\sigma}^2 = \mathrm{MS}_{res}$$

ESTIMATED VARIANCE OF $\hat{\beta}_1$:

$$\hat{\mathbb{V}}[\hat{\beta}_1] = \frac{\hat{\sigma}^2}{SC_{xx}} = \frac{MS_{res}}{SC_{xx}}$$

SIMPLE LINEAR REGRESSION: MODEL INFERENCE [DEVORE 12.3]

ESTIMATED STANDARD DEVIATION OF $\hat{\beta}_1$: $\hat{\mathbb{D}}[\hat{\beta}_1] = \sqrt{\hat{\mathbb{V}}[\hat{\beta}_1]} = \sqrt{\frac{\hat{\sigma}^2}{SC_{xx}}} = \sqrt{\frac{MS_{res}}{SC_{xx}}}$

t-CI FOR β_1: The $100(1-\alpha)\%$ **independent t-CI for** β_1 is: $\hat{\beta}_1 \pm t^*_{\nu_{res};\alpha/2} \cdot \hat{\mathbb{D}}[\hat{\beta}_1]$

MODEL UTILITY t-TEST

Linear Model:	**1NR(1NR) | RaP BaFL | I.N.EV** $Y_i = \beta_0 + \beta_1 x_i + E_i$ $E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$
Realized Linear Model:	$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, \cdots, n$
Model Parameter Estimates:	$\hat{\beta}_1 = \frac{SC_{xy}}{SC_{xx}}, \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$
Predicted Values & Residuals:	$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad y_i^{res} := y_i - \hat{y}_i$
Test Statistic Value:	$t = \frac{\hat{\beta}_1}{\hat{\mathbb{D}}[\hat{\beta}_1]}$

HYPOTHESIS TEST:	**REJECTION REGION AT LVL α :**
$H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$	$t \leq -t^*_{\nu_{res};\alpha/2}$ or $t \geq t^*_{\nu_{res};\alpha/2}$

HYPOTHESIS TEST:	**P-VALUE DETERMINATION:**
$H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$	P-value $= 2 \cdot [1 - \Phi_t(\lvert t \rvert; \nu_{res})]$

MODEL UTILITY F-TEST

1. Determine degrees of freedom: $\nu_{total} = n - 1, \quad \nu_{res} = n - 2, \quad \nu_{reg} = 1$
2. Compute Sums: $S_x := \sum_i x_i, \quad S_y = \sum_i y_i, \quad S_{xx} = \sum_i x_i x_i, \quad S_{yy} = \sum_i y_i y_i, \quad S_{xy} = \sum_i x_i y_i$
3. Compute Centered Sums: $SC_{xx} = S_{xx} - \frac{1}{n} S_x S_x, \qquad SC_{yy} = S_{yy} - \frac{1}{n} S_y S_y, \qquad SC_{xy} = S_{xy} - \frac{1}{n} S_x S_y$
4. Compute Parameter Estimates: $\hat{\beta}_1 = \frac{SC_{xy}}{SC_{xx}}, \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$
5. Compute Sums of Squares: $\text{SS}_{total} = SC_{yy}, \quad \text{SS}_{res} = SC_{yy} - \hat{\beta}_1^2 \cdot SC_{xx}, \quad \text{SS}_{reg} = \hat{\beta}_1^2 \cdot SC_{xx}$
6. Compute Mean Squares: $\text{MS}_{res} := \text{SS}_{res}/\nu_{res}, \qquad \text{MS}_{reg} := \text{SS}_{reg}/\nu_{reg}$
7. Compute Test Statistic Value: $f_{reg} = \text{MS}_{reg}/\text{MS}_{res}$
8. Compute F-cutoff/P-value: By hand, lookup $f^*_{\nu_{reg},\nu_{res};\alpha}$
By SW, compute $p_{reg} = 1 - \Phi_F(f_{reg}; \nu_{reg}, \nu_{res})$
9. Render Decision: If $f \geq f^*_{\nu_{reg},\nu_{res};\alpha}$, then reject H_0; else accept H_0.
If $p_{reg} \leq \alpha$, then reject H_0; else accept H_0.

Model Utility F-Test (Significance Level α)

Variation Source	**df**	**Sum of Squares**	**Mean Square**	**F Stat Value**	**P-value**	**Decision**
Regression	ν_{reg}	SS_{reg}	MS_{reg}	f_{reg}	p_{reg}	Acc/Rej H_0
Unknown	ν_{res}	SS_{res}	MS_{res}			
Total	ν_{total}	SS_{total}				

EX 12.3.1: 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) Compute all sums: S_x, S_y, S_{xx}, S_{yy}, S_{xy}.

(b) Compute all centered sums: SC_{xx}, SC_{yy}, SC_{xy}.

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

(d) Compute the residual sum of squares, SS_{res}.

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

(f) Compute the estimated variance of estimated OLS slope parameter, $\hat{\mathbb{V}}[\hat{\beta}_1]$.

(g) Compute the estimated standard deviation of estimated OLS slope parameter, $\hat{\mathbb{D}}[\hat{\beta}_1]$.

(h) Construct a 95% t-CI for slope parameter β_1.

(i) Perform a model utility F-test of slope parameter β_1 with $\alpha = 0.05$ significance level. Also, build its ANOVA table.