Linear Regression Formula Sheet

---

1. Beta Definitions

- **β0**: The <u>y intercept</u> of the population regression line
- **β1**: The <u>slope</u> of the population regression line
  $\beta_1 = Cov(x,y) Var(x)$ = Slope of the true model

- **β^0 (βhat 0)**: The <u>estimated intercept</u> of the regression line based on the sample data.
  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- **β^1 (βhat 1)**: The <u>estimated slope</u> of the regression line based on the sample data. This is calculated by minimizing the sum of squared errors.
  $\hat{\beta}_1 = S_{xy}/S_{xx} = \sum(x_i - \bar{x})(y_i - \bar{y}) \,/\, \sum_{i=1}^{n}(x_i - \bar{x})^2$

- **$\bar{\beta}_0$ & $\bar{\beta}_1$**: The population averages of β0 & β1 across different samples.

---

2. Sigma Definitions

- **σ**: The standard deviation of the error term ϵ - this measures the spread of the errors around the regression line.
  $\sigma = \sqrt{Var(\epsilon)}$

- **σ^2 (σ squared)**: The variance of the error term, representing the average squared deviation from the true regression line.
  $\sigma^2 = \{ \sum ((y_i - \beta_0 - \beta_1 x_i)^2) \} / n$

- **σ^^2 (σ hat squared)**: The estimated variance of the error term, calculated using sample data.
  $\hat{\sigma}^2 = S_{yy} / (n-2) = (\sum (y_i - \hat{y}_i)^2) / (n-2)$

---

3. Epsilon Definitions

- **ϵ**: The error term in the population model, the difference between real & predicted
  $\epsilon i = yi - (\beta 0 + \beta 1xi)$
- **ϵ^ (ϵ hat)**: The residual in the sample model, representing the difference between the observed yi and the predicted y^i (y hat i).
  $\epsilon\hat{}i = yi - y\hat{}i = yi - (\beta\hat{}0 + \beta\hat{}1xi)$
- **$\bar{\epsilon}$ (ϵ bar)**: The mean of the error terms, which we assume to be zero
  $\bar{\epsilon} = E(\epsilon) = 0$

---

4. Key Regression Concepts

- **Linear Regression Model:** $yi = \beta 0 + \beta 1\ xi + \epsilon i$
- **Fitted (Predicted) Values y^i:** $y\hat{}i = \beta\hat{}0 + \beta\hat{}1xi$

SST:

- **Sum of Squares Total (SST):** $SST = \sum (yi - \bar{y})\hat{}2$
- **Total Sum of Squares (TSS):** $TSS = \sum (yi - \bar{y})\hat{}2$

SSE:

- **Sum of Squares Explained (SSE):** $SSE = \sum (yi - y\hat{})\hat{}2$
- **Explained Sum of Squares (ESS):** $ESS = \sum (yi - y\hat{})\hat{}2$

SSR:

- **Sum of Squared Residual (SSR):** $SSR = \sum (y\hat{}i - \bar{y}i)\hat{}2$
- **Residual Sum of Squares (RSS):** $RSS = \sum (y\hat{}i - \bar{y}i)\hat{}2$

---

5. Other things

- **Sum of Residuals:** $\sum \epsilon\hat{}i = 0$
  This ensures that the residuals sum to zero in a least-squares regression.
- **Sum of Residuals Weighted by X:** $\sum xi\ \epsilon\hat{}i = 0$
  This ensures that the weighted sum of the residuals, using xxx, is zero.
- **Sum of Fitted Values and Residuals:** $\sum y\hat{}i\ \epsilon\hat{}i = 0$
  This ensures that there is no correlation between the fitted values and the residuals.

How to Prove things

- **Estimating β^0 & β^1 (β 0&1 hat):**
  - Use the formula β^1 = Sxy/Sxx to find the slope
  - Use the formula $\beta\hat{}0 = \bar{y} - \beta\hat{}1\ \bar{x}$
- **Proof of Zero Residual Sum:**
  - Show that $\sum \epsilon\hat{}i = 0$ follows from the least-squares estimation process.
- **Variance of β^1:**
  - From the formula for β^1 (hat)
  - Var(β^1) = σ*σ/Sxx