## Side-by-Side: Sample Mean v. OLS/SLR

| **1. Estimation** | **Sample Mean** | **OLS** |
|---|---|---|
| DGM: Data Generation Mechanism | $\{Y_i\}$ iid<br>$E(Y_i) = \mu$, $Var(Y_i) = \sigma^2$ | $\{Y_i\} = \{\beta_0 + \beta_1 x_i + U_i\}$ indept.<br>$E(Y_i \mid x_i) = \beta_0 + \beta_1 x_i$,<br>$Var(Y_i \mid x_i) = \sigma^2$ |
| Unknown parameter (to be estimated) | $\mu$ | $\beta_1$ |
| Linear Estimator | $b_0 + \sum b_i Y_i$ | $b_0 + \sum b_i Y_i$ |
| Linear Unbiased Estimator | $\sum b_i Y_i \mid \sum b_i = 1$ | $\sum b_i Y_i \mid \sum b_i = 0$ & $\sum b_i x_i = 1$ |
| Variance to be minimized | $\sigma^2 \sum b_i^2$ | $\sigma^2 \sum b_i^2$ |
| Constraints | $\sum b_i = 1$ | $\sum b_i = 0$ & $\sum b_i x_i = 1$ |
| ***BLUE*** | $b_i^* = \frac{1}{n}$ | $b_j^* = \frac{(x_j - \overline{x})}{\sum (x_i - \overline{x})^2}$ |
| Estimator | $M = \frac{1}{n} \sum Y_i$<br>(unweighted avg. of values) | $B_1 \mid x's = \sum \left\{ \frac{(x_i - \overline{x})^2}{(n-1)S_{xx}} \right\} \frac{(Y_i - \overline{Y})}{(x_i - \overline{x})}$<br>(weighted avg. of slopes) |
| Estimate | $\hat{\mu} = \frac{1}{n} \sum y_i$ | $\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_j - \overline{x})^2}$ |
| Estimator Variance | $\frac{\sigma^2}{n}$ | $\frac{\sigma^2}{\sum (x_j - \overline{x})^2}$ |
| … Standard Deviation | $sd(M) = \frac{\sigma}{\sqrt{n}}$ | $sd(B_1) = \frac{\sigma}{\sqrt{\sum (x_j - \overline{x})^2}}$ |

***Side by Side: Sample Mean v. OLS/SLR***

| | | |
|---|---|---|
| … Standard Error | $se(M) = \frac{S_Y}{\sqrt{n}}$ | $se(B_1) = \frac{RMSE}{\sqrt{\sum (x_j - \overline{x})^2}}$ |

| **2. Inference** | **Sample Mean** | **OLS** |
|---|---|---|
| Dist. of $Y_i$ | $Y_i \sim N(\mu, \sigma^2)$ | $Y_i \mid x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ |
| Degrees of Freedom | n-1 | n-2 |
| Distribution of Estimator | $\frac{M-\mu}{sd(M)} \sim N(0,1)$ where $sd(M) = \sigma / \sqrt{n}$ | $\frac{B_1 - \beta_1}{sd(B_1)} \sim N(0,1)$ where $sd(B_1) = \sigma / \sqrt{\sum (x_i - \overline{x})^2}$ |
| t statistic | $\frac{M-\mu}{se(M)} \sim t_{n-1}$ where $se(M) = \frac{S_Y}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n-1}}$ | $\frac{B_1 - \beta_1}{se(B_1)} \sim t_{n-2}$ where $se(B_1) = \frac{RMSE}{\sqrt{\sum (x_i - \overline{x})^2}}$ |
| Confidence Intervals | $\left[M \pm c\, se(M)\right]$, $P[\mid t_{n-1} \mid < c) = p$ | $\left[B_1 \pm c\, se(B_1)\right]$, $P[\mid t_{n-2} \mid < c) = p$ |
| Null Hypothesis: | $H_0 : \mu = 0$ | $H_0 : \beta_1 = 0$ |
| t statistic (under Ho) | $\frac{M}{se(M)}$ | $\frac{B_1}{se(B_1)}$ |
| t statistic (for the given sample) | $t_{\hat{\mu}} = \frac{\hat{\mu}}{se(\hat{\mu})}$ | $t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ |
| p value | $P[\mid t_{n-1} \mid > t_{\hat{\mu}})$ | $P[\mid t_{n-2} \mid > t_{\hat{\beta}_1})$ |
| Hypothesis Test: | Reject if $\left\lvert \frac{\hat{\mu} - 0}{se(\hat{\mu})} \right\rvert > c$ or $\left\lvert t_{\hat{\mu}} \right\rvert > c \ldots$ or if $p < \alpha$ | Reject if $\left\lvert \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)} \right\rvert > c$ or $\left\lvert t_{\hat{\beta}_1} \right\rvert > c$ or if $p < \alpha$ |