Numerical Summaries: For a r.s. $(y_1, \ldots, y_n)$ of size n

- Sample Mean: $\hat{\mu}_Y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
- Sample Standard Deviation: $s_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$
- Sample Variance: s_Y^2

Sampling Distribution:

- The mean and standard deviation of the sample mean, $\bar{Y}$, based on a random sample of size n, from a population with mean μ_Y and standard deviation σ_Y are
  1. mean($\bar{Y}$) $= \mu_{\bar{Y}} = \mu_Y$ (unbiased estimator)
  2. S.D.($\bar{Y}$) $= \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$
- If $Y \sim N(\mu_Y, \sigma_Y)$, then $\bar{Y} \sim N\left(\mu_{\bar{Y}} = \mu, \sigma_{\bar{Y}} = \sigma/\sqrt{n}\right)$
- Central Limit Theorem:
  If $Y \sim ?(\mu_Y, \sigma_Y)$, then for large n,
  $$\bar{Y} \mathrel{\dot\sim} N\left(\mu_{\bar{Y}} = \mu, \sigma_{\bar{Y}} = \sigma/\sqrt{n}\right)$$

General T-tools for θ:

- Estimate: $\hat{\theta}$
- Sampling Distribution: If is $\hat{\theta}$ normal and unbiased, then
  $$T = \frac{Estimate - Parameter}{S.E.(Estimate)} \sim t_{\text{\# of obs} - \text{\# unknown parameters}}$$
- For H_0: $\theta = \theta_0$:
  $$t_0 = \frac{\hat{\theta} - \theta_0}{S.E.(\hat{\theta})}$$
- A $(1-\alpha)100\%$ Confidence Interval (CI) for θ :
  Estimate ± (Critical value) × $S.E.(Estimate)$
  → $\hat{\theta} \pm [t_{df, \alpha/2}] \times [S.E.(\hat{\theta})]$

One Population Mean:

- For H_0: $\mu_Y = \mu_0$:
  $$t_0 = \frac{\bar{y} - \mu_0}{s_Y / \sqrt{n}} \sim t_{n-1}$$
- A $(1-\alpha)100\%$ CI for μ_Y :
  $$\bar{y} \pm [t_{n-1, \alpha/2}] \times [s_Y / \sqrt{n}]$$

Two Means – Paired Samples:

- For H_0: $\mu_d = D_0$:
  $$t_0 = \frac{\bar{d} - D_0}{s_d / \sqrt{n}} \sim t_{n-1}$$
- A $(1-\alpha)100\%$ CI for μ_Y :
  $$\bar{d} \pm [t_{n-1, \alpha/2}] \times [s_d / \sqrt{n}]$$

Two Means – Independent Samples:

Assuming $\sigma_2^1 = \sigma_2^2 = \sigma^2$:
$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

- For H_0: $\mu_1 - \mu_2 = 0$:
  $$t_0 = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$
- A $(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$:
  $$\bar{y}_1 - \bar{y}_2 \pm [t_{n_1+n_2-2, \alpha/2}] \times \left(s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Extra-Sum-of-Squares F-test:

H_0: reduced model
H_A: full model

- Test statistic:
  $$F_0 = \frac{(SSR(r) - SSR(f))/(df(r) - df(f))}{SSR(f)/df(f)}$$
- $F_0 \sim F_{df(f)}^{df(r)-df(f)}$

ANOVA for Several Means:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$ (One-Mean model)
H_A: $\mu_1, \mu_2, \ldots, \mu_I$ (I-Mean model)

- Test statistic:
  $$F_0 = \frac{(SSR(1) - SSR(I))/(df(1) - df(I))}{SSR(I)/df(I)} = \frac{SST/(I-1)}{SSE/(N-I)} = \frac{MST}{MSE}$$
- $F_0 \sim F_{N-I}^{I-1}$

Linear Combinations of Group Means:

- Parameter: $\gamma = C_1\mu_1 + C_2\mu_2 + ... + C_I\mu_I$
- Estimate: $g = C_1\overline{Y}_1. + C_2\overline{Y}_2. + ... + C_I\overline{Y}_I.$
- $S.E.(g) = \sqrt{MSE}\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + ... + \frac{C_I^2}{n_I}}$

  where $\sqrt{MSE} = \sqrt{\frac{(n_1-1)s_1^2 + ... + (n_I-1)s_I^2}{N-I}}$
- For H_0: $\gamma = 0$:

  $t_0 = \frac{g-0}{S.E.(g)} \sim t_{N-I}$
- A $(1-\alpha)100\%$ CI for γ:

  $g \pm t_{N-I,\,\alpha/2} \times S.E.(g)$

Simple Linear Regression:

- Y – response variable, X – explanatory variable
- Model: $\mu(Y\,|\,X) = \beta_0 + \beta_1 X$

  where β_0 – intercept, β_1 – slope
- Estimated Model: $\hat{y} = \hat{\mu}(Y\,|\,X = x) = \hat{\beta}_0 + \hat{\beta}_1 x$

  $\hat{\beta}_0, \hat{\beta}_1$ – least squares estimates
- σ^2 = the variance of Y at a fixed X

  $\hat{\sigma}^2 = \frac{SS(\text{Residuals})}{n-2} = MSE$

  The standard error of the model is $\hat{\sigma} = \sqrt{\frac{SS(\text{Residuals})}{n-2}}$

Inferential Tools for Regression Coefficients in SLR:

- For H_0: $\beta_i = 0$: $(i = 0, 1)$

  $t_0 = \frac{\hat{\beta}_i - 0}{S.E.(\hat{\beta}_i)} \sim t_{n-2}$
- A $(1-\alpha)100\%$ CI for β_i: $(i = 0, 1)$

  $\hat{\beta}_i \pm t_{\alpha/2,\,n-2} \times S.E.(\hat{\beta}_i)$

Coefficient of Determination:

$$R^2 = 1 - \frac{\text{SS(Residuals)}}{\text{SS(Total)}}$$

$$R^2_{Adjusted} = 1 - \frac{\text{MS(Residuals)}}{\text{MS(Total)}}$$

Interpretation of the following models:

"ln" = the natural logarithm

1. $\mu(\ln(Y)\,|\,X) = \beta_0 + \beta_1 X$

"An additive change in X by k units is associated with a multiplicative change of $e^{k\beta_1}$ in Median$(Y\,|\,X)$."

2. $\mu(Y\,|\,\ln(X)) = \beta_0 + \beta_1\ln(X)$

"A multiplicative change in X by a factor of k is associated with an additive change of $\beta_1 \ln(k)$ in $\mu(Y\,|\,X)$."

3. $\mu(\ln(Y)\,|\,\ln(X)) = \beta_0 + \beta_1\ln(X)$

"A multiplicative change in X by a factor of k is associated with a multiplicative change of k^{β_1} in Median$(Y\,|\,X)$."

CI's for the response for a given *X* in SLR:

- For the mean response, $\mu_{Y|x^*}$:

  $S.E.(\hat{\mu}_{Y|x^*}) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{(n-1)s_x^2}}$

  $\therefore$ A $(1-\alpha)100\%$ CI for $\mu_{Y|x^*}$ is:

  $\hat{\mu}_{Y|x^*} \pm t_{\alpha/2,\,n-2} \times S.E.(\hat{\mu}_{Y|x^*})$
- For an individual predicted response, Y:

  $S.E.(\hat{y}) = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{(n-1)s_x^2}}$

  $\therefore$ A $(1-\alpha)100\%$ PI for $\hat{y}$ is:

  $\hat{y} \pm t_{\alpha/2,\,n-2} \times S.E.(\hat{y})$

Two Proportions:

- For H_0: $\pi_1 - \pi_2 = 0$

  $$Z_0 = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}_c(1-\hat{\pi}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim Z$$

  $$\hat{\pi}_c = \frac{\text{"total\_succ esses"}}{n_1 + n_2} = \frac{n_1\hat{\pi}_1 + n_2\hat{\pi}_2}{n_1 + n_2}$$
- An *approximate* $(1-\alpha)100\%$ CI for $\pi_1 - \pi_2$:

  $$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \times \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

Multiple Linear Regression:

- Y – response variable, $X_1, \ldots, X_p$ – explanatory variables
- Model: $\mu(Y \mid X) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$
  where $\beta_0, \beta_1, \ldots, \beta_p \rightarrow$ Regression Coefficients
- Estimated Model:
  $\hat{\mu}(Y \mid X_1, \ldots, X_p) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_p X_p$
  $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p \rightarrow$ Estimated Regression Coefficients
- σ^2 = the variance of the model
  The standard error of the model is $\hat{\sigma} = \sqrt{\dfrac{SS(\text{Residuals})}{n-(p+1)}}$

Inference for Single Coefficients:

- For H_0: $\beta_i = 0$: $(i = 0, 1, \ldots, p)$
  $t_0 = \dfrac{\hat{\beta}_i - 0}{S.E.(\hat{\beta}_i)} \sim t_{n-(p+1)}$
- A $(1-\alpha)100\%$ CI for β_i: $(i = 0, 1, \ldots, p)$
  $\hat{\beta}_i \pm t_{\alpha/2,\, n-(p+1)} \times S.E.(\hat{\beta}_i)$

Inference for a Subset of Coefficients:

An Extra-Sum-of-Squares F-Test for k of the p+1 β's :
H_0: all k beta's are equal to zero (reduced model)
H_A: not all are equal to zero (full model)
F-stat = $F_0 \rightarrow$ see Extra-Sum-of-Squares F-Test
$F_0 \sim F^{k}_{n-(p+1)}$

PI for the response in MLR:

- For an individual predicted response, Y :
  $S.E.(\hat{y}) \approx \hat{\sigma}$
  $\therefore$ A $(1-\alpha)100\%$ PI for $\hat{y}$ is :
  $\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_p X_p\right) \pm t_{\alpha/2,\, n-(p+1)} \times \hat{\sigma}$

Logistic Regression:

Model: $\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k = \eta$

Logistic function: $\pi = \dfrac{e^{\eta}}{1+e^{\eta}}$

Odds: $\omega = e^{\eta}$

Odds at $X_1 = A$ relative to the odds at $X_1 = B$, for fixed values of the other X's,

$$\frac{\omega_A}{\omega_B} = e^{\beta_1(A-B)}$$

Two-Way ANOVA:

Factors: A with a – levels
B with b – levels
Data: n' observations for each ab factor combination
$\therefore$ total = $n = abn'$

A test for additivity :
H_0: *additive* model (reduced model)
$\mu(Y \mid A, B) = \beta_0 + A + B$
H_A: *non-additive* model (full model)
$\mu(Y \mid A, B) = \beta_0 + A + B + AB$
$F_0 = \dfrac{MSAB}{MSE} \sim F^{(a-1)(b-1)}_{n-ab}$
Overall test for any effects :
$F_0^* \sim F^{ab-1}_{n-ab}$

Additive Table for two-way ANOVA:

Overall test for any effects:
$F_0^* \sim F^{a+b-2}_{n-a-b+1}$

Odds and Odds Ratio:

- Odds: $\omega = \dfrac{\pi}{1-\pi}$
- Odds ratio: $\phi = \dfrac{\omega_1}{\omega_2} = \left(\dfrac{\pi_1}{1-\pi_1}\right) \Big/ \left(\dfrac{\pi_2}{1-\pi_2}\right)$
- For H_0: $\log(\phi) = 0$
  $Z_0 = \dfrac{\log(\hat{\phi}) - 0}{S.E.(\log(\hat{\phi}))} \sim Z$
  $\text{S.E.}(\log(\hat{\phi})) = \sqrt{\dfrac{1}{n_1 \hat{\pi}_c (1-\hat{\pi}_c)} + \dfrac{1}{n_2 \hat{\pi}_c (1-\hat{\pi}_c)}}$
- An *approx.* $(1-\alpha)100\%$ CI for $\log(\phi)$:
  $\log(\hat{\phi}) \pm z_{\alpha/2} \times \sqrt{\dfrac{1}{n_1 \hat{\pi}_1 (1-\hat{\pi}_1)} + \dfrac{1}{n_2 \hat{\pi}_2 (1-\hat{\pi}_2)}}$

Logistic Regression (continued):

For individual regression coefficients.: H_0: $\beta_j = 0$

$Z_0 = \dfrac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}$, use output

A CI for individual regression coefficient
$\hat{\beta}_j \pm z_{\alpha/2} \times SE(\hat{\beta}_j)$