Distributions

- Sampling dist. for mean of sample of size n from population with mean μ, standard deviation σ

$$\overline{X} \sim \mathsf{Norm}(\mu, \sigma/\sqrt{n}).$$

- Sampling dist. for sample proportion has

$$\mu_{\hat{p}} = p, \qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

When 10% rule in play, $np \geq 10$, and $n(1-p) \geq 10$, then $\hat{p}$ has approx. dist. $\mathsf{Norm}(p, \sqrt{p(1-p)/n})$.

Inference Procedures

- Level C Confidence Intervals (general):

(estimate) ± (critical value)(approx. std. error)

- 1-sample proportion:
  - CIs for p

$$\text{SE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  - z-test (when $\hat{p}$ approx. normal)
    test stat. ($\mathbf{H}_0 : p = p_0$): $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

- 2-sample proportion:
  - Confidence intervals for $p_1 - p_2$ use

$$\text{SE} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

  - Hyp. tests: test statistic when $\mathbf{H}_0 : p_1 = p_2$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}}, \qquad \text{SE} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

where $\hat{p}$ is the "pooled sample proportion" obtained by considering the two samples to be one big sample

- 1-sample t: test statistic when $\mathbf{H}_0 : \mu = \mu_0$

$$t = \frac{\overline{x} - \mu_0}{\text{SE}}, \quad \text{SE} = \frac{s}{\sqrt{n}}, \quad df = n - 1$$

- 2-sample t: test statistic when $\mathbf{H}_0 : \mu_1 - \mu_2 = 0$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\text{SE}}, \quad \text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

As conservative estimate, take

$$df = \min(n_1 - 1, n_2 - 1)$$

- Chi-square test statistic:
$$\chi^2 = \sum \frac{[(\text{observed count}) - (\text{expected count})]^2}{\text{expected count}}$$
contingency table: df = (#rows - 1)(#columns - 1)
goodness-of-fit: df = (#groups) − 1 − (# est. params)

- Model utility test:

$$t = \frac{b_1}{\text{SE}_{b_1}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad \text{with} \quad df = n - 2$$

- F-test in ANOVA:

$$F = \frac{\text{MSG}}{\text{MSE}},$$

where

$$\begin{aligned} df_{\text{numer}} &= \text{(\# of groups) - 1,} \quad \text{and} \\ df_{\text{denom}} &= \text{(sample size) - (\# of groups)} \end{aligned}$$

Miscellaneous

- Determine sample size, 1-proportion settings:
To have margin of error no larger than a desired size ME, take,

$$n \geq \left(\frac{z^*}{ME}\right)^2 \tilde{p}(1-\tilde{p}),$$

where $\tilde{p}$ is an estimate of p (take $\tilde{p} = 0.5$ if no estimate is available)