Describing Data

Mean: $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

Sample SD: $s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$

Probability

Factorials: $k! = \begin{cases} k \times (k-1) \times \cdots \times 1 & \text{if } k \geq 1 \\ 1 & \text{if } k = 0 \end{cases}$

Binomial

—coefficients: ${}_nC_j = \frac{n!}{j!(n-j)!}$

—probabilities: $\Pr\{Y = j\} = {}_nC_j p^j (1-p)^{n-j} \quad (0 \leq j \leq n)$

—mean: $\mu = np$

—SD: $\sigma = \sqrt{np(1-p)}$

Normal

—z-scores: $z = \frac{y - \mu}{\sigma}$

—quantiles: $y = \mu + z\sigma$

Sampling Distribution of $\bar{Y}$

Mean: $\mu_{\bar{Y}} = \mu$

Standard deviation: $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$

Confidence intervals for μ

SE: $\text{SE}_{\bar{y}} = \frac{s}{\sqrt{n}}$

Interval: $\bar{y} \pm t\frac{s}{\sqrt{n}}$

Degrees of freedom: $n - 1$

Sample size: $n \geq \left(\frac{\text{Guessed SD}}{\text{Desired SE}}\right)^2$

95% Confidence interval for p

Estimate: $\tilde{p} = \frac{y+2}{n+4}$

SE: $\text{SE}_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$

Interval: $\tilde{p} \pm 1.96 \times \text{SE}_{\tilde{p}}$

Sample size: $n \geq \frac{(\text{Guessed } p)(1 - (\text{Guessed } p))}{(\text{Desired SE})^2} - 4$

Confidence interval for $\mu_1 - \mu_2$

SE: $\text{SE}_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Interval: $(\bar{y}_1 - \bar{y}_2) \pm t \times \text{SE}_{(\bar{y}_1 - \bar{y}_2)}$

Degrees of freedom: $\text{df} = \frac{\left(\text{SE}_1^2 + \text{SE}_2^2\right)^2}{\frac{\text{SE}_1^4}{(n_1-1)} + \frac{\text{SE}_2^4}{(n_2-1)}}$

—where: $\text{SE}_1 = \frac{s_1}{\sqrt{n_1}}$ and $\text{SE}_2 = \frac{s_2}{\sqrt{n_2}}$

Hypothesis test for $\mu_1 - \mu_2$

Null hypothesis: $H_0\colon \mu_1 = \mu_2$

Alternative hypothesis: $H_A\colon \mu_1 \neq \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$

Test statistic: $t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\text{SE}_{(\bar{y}_1 - \bar{y}_2)}}$

Degrees of freedom: as above

Significance: $\alpha = \Pr\{\text{reject } H_0 | H_0 \text{ is true}\}$

Paired tests and confidence intervals

Standard error: $\text{SE}_{\bar{d}} = \frac{s_{\bar{d}}}{\sqrt{n_d}}$

Null hypothesis: $H_0\colon \mu_d = 0$

Test statistic: $t = \frac{\bar{d} - 0}{\text{SE}_{\bar{d}}}$

Degrees of freedom: $n_d - 1$

Confidence interval: $\bar{d} \pm t \times \text{SE}_{\bar{d}}$

Analysis of Categorical Data

The χ^2 test statistic is

$$\sum_{\text{categories}} \frac{(O-E)^2}{E} \quad .$$

χ^2 *goodness-of-fit test*

Expected counts: np_i

Degrees of freedom: (# categories - 1)

χ^2 *test of independence*

Expected counts: $\frac{(\text{Row total}) \times (\text{Column Total})}{(\text{Grand Total})}$

Degrees of freedom: (# Rows - 1) × (# Columns - 1).