# Two-sample inferences

## Confidence intervals for the two-sample problem

Bruce M. Boghosian

Department of Mathematics

Tufts University

# Outline

# Two-sample hypothesis testing with normal r.v.s



- Meaningful $H_0$ can always be defined for two-sample tests
- Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be independent random samples drawn from normal distributions with means $\mu_X$ and $\mu_Y$, respectively, and with the same standard deviation $\sigma$.
- Let $s_p$ denote the pooled standard deviation.
- A $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is given by

$$\left(\overline{x} - \overline{y} - t_{\alpha/2,n+m-2}s_p\sqrt{\frac{1}{n}+\frac{1}{m}}, \overline{x} - \overline{y} + t_{\alpha/2,n+m-2}s_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$$

# Two-sample hypothesis testing with normal r.v.s



- **Pf.:**
  - We know $\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$ is Student $T$ distributed with $n+m-2$ df, so

$$P\left(-t_{\alpha/2,n+m-2} \leq \frac{\overline{X}-\overline{Y}-(\mu_x-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \leq +t_{\alpha/2,n+m-2}\right) = 1-\alpha$$

  - Rearrange inequality to isolate $\mu_X - \mu_Y$ to obtain confidence interval.

# Confidence intervals for the variance ratio



- Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be independent random samples drawn from normal distributions with standard deviations $\sigma_X$ and $\sigma_Y$, respectively. A $100(1-\alpha)\%$ confidence interval for the variance ratio $\sigma_X^2/\sigma_Y^2$ is

$$\left(\frac{s_X^2}{s_Y^2} F_{\alpha/2,m-1,n-1}, \frac{s_X^2}{s_Y^2} F_{1-\alpha/2,m-1,n-1}\right)$$

- **Pf.:**
  - Note that $\frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2}$ is $F$ distributed with $m-1$ and $n-1$ df.
  - Same strategy: Write probability

$$P\left(f_{1-\alpha/2,m-1,n-1} \leq \frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} \leq f_{\alpha/2,m-1,n-1}\right) = 1-\alpha.$$

  - Isolate $\sigma_X^2/\sigma_Y^2$ in inequality.

# Confidence intervals for two-sample Bernoulli trials



- Let $x$ and $y$ denote the number of successes in two independent sets of $n$ and $m$ Bernoulli trials, respectively.
- If $p_X$ and $p_Y$ denote the true success probabilities, an approximate $100(1-\alpha)\%$ confidence interval for $p_X - p_Y$ is given by

$$\left(\frac{x}{n} - \frac{y}{m} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n} + \frac{\frac{y}{m}\left(1-\frac{y}{m}\right)}{m}},\right.$$

$$\left.\frac{x}{n} - \frac{y}{m} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n} + \frac{\frac{y}{m}\left(1-\frac{y}{m}\right)}{m}}\right)$$

# Summary



- We have studied two-sample confidence intervals for $\sigma_X^2/\sigma_Y^2$.
- We have studied them for both Bernoulli trials and normally distributed data.