Lesson 25. One-Way ANOVA – Confidence Intervals and Effect Sizes – Part 1

Note. Part 2 of this lesson shows how to do the calculations here in Part 1 using R.

1 Previously...

- One-way ANOVA model:

$$\begin{aligned} Y &= \mu + \alpha_k + \varepsilon \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2) \\ \mu_k &= \mu + \alpha_k \end{aligned}$$

- Parameter estimates:

$$\hat{\mu} = \bar{y} \qquad\qquad \hat{\alpha}_k = \bar{y}_k - \bar{y}$$

- One-way ANOVA F-test for K groups tests the following hypotheses:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K \qquad \text{vs.} \qquad H_A : \text{at least one of the } \mu_k \text{ is different}$$

  - This helps us answer:

    How strong is the evidence that the treatment has an effect on the response?

  - But it does not help us answer:

    If there is a difference due to treatment, how big is the effect?

2 Confidence intervals

- The $100(1-\alpha)\%$ **confidence interval for a group mean** μ_k is

$$\bar{y}_k \pm t_{\alpha/2, n-K} \cdot SD\sqrt{\frac{1}{n_k}}$$

- To compute the confidence interval for the difference of two group means, we need to factor in the sample sizes of both groups

- The $100(1-\alpha)\%$ **confidence interval for the difference of two group means** $\mu_1 - \mu_2$ is

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, n-K} \cdot SD\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

  - Note that two group means are significantly different at the $100(\alpha)\%$ level if and only if a $100(1-\alpha)\%$ confidence interval for their difference does not include 0

Example 1. Continuing with the `FatRats` example from Lessons 23 and 24...

Recall the setting: Thirty baby rats were fed high-protein diets with different sources of protein: beef, cereal, or pork. The rats were randomly assigned to one of these three diets; 10 rats got their protein from beef, 10 from cereal, and 10 from pork. Their weight gains were recorded.

The response variable is the weight gain in grams (*Gain*), and the explanatory variable is the type of protein (*Source*).

We found the following parameter estimates:

$$\hat{\mu} = \bar{y} = 95.1333 \qquad \hat{\alpha}_{\text{Beef}} = 4.8667 \quad \hat{\alpha}_{\text{Cereal}} = -9.2333 \quad \hat{\alpha}_{\text{Pork}} = 4.3667$$

The corresponding ANOVA table is below:

```
                  Df Sum Sq Mean Sq F value Pr(>F)
as.factor(Source)  2   1280   640.0   3.346 0.0503 .
Residuals         27   5165   191.3
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

a. Based on the parameter estimates above, what are the values of $\bar{y}_{\text{Beef}}$, $\bar{y}_{\text{Cereal}}$, and $\bar{y}_{\text{Pork}}$?

b. Find the 95% confidence interval for the mean weight gain for the Beef group.

c. Find the 95% confidence interval for the difference between the mean weight gains for the Beef and Cereal groups.

d. Are the mean weight gains for the Beef and Cereal groups significantly different at the 0.05 level?

3 Effect sizes

- Remember: statistical significance ≠ practical importance
- The **effect size** is one commonly used way to measure how much practical importance a numerical idfference might make in real life
- For one-way ANOVA: the effect size is a ratio of a difference to the SD within all groups
- For a single group:

$$D_k = \frac{\hat{\alpha}_k}{SD} = \frac{\bar{y}_k - \bar{y}}{SD}$$

- For any pair of groups:

$$D_{jk} = \frac{\bar{y}_j - \bar{y}_k}{SD}$$

Example 2. Continuing with the `FatRats` example from Example 1...

a. Estimate the effect of a beef-based protein diet for mean weight gain in rats.

b. Compare the effect of a beef-based protein diet versus a cereal-based protein diet for mean weight gain in rats.