Chapter 10: Comparing Multiple Independent Populations
Part 2

1. Assumptions of ANOVA

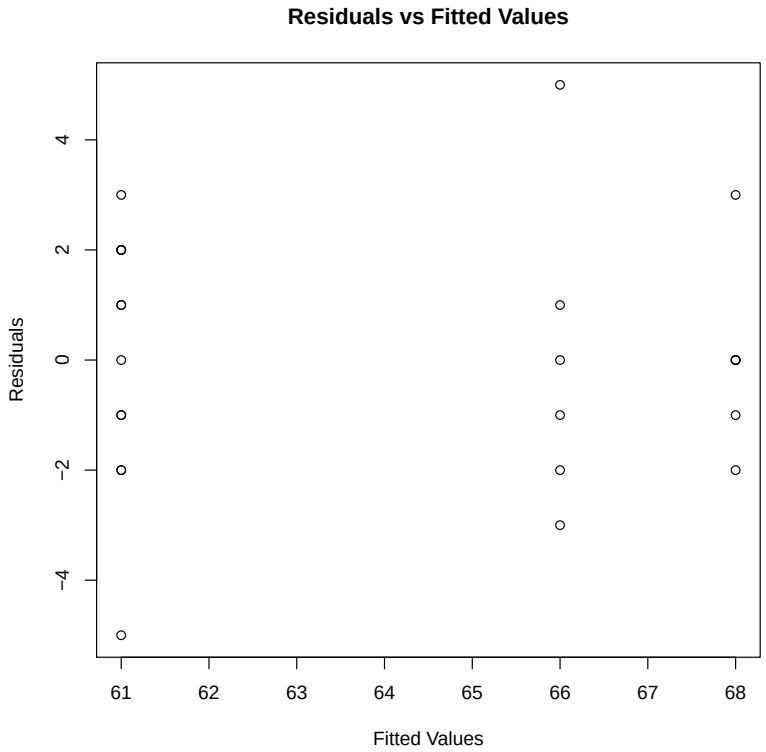
   ANOVA assumes:

   - The data are independent within and between treatments
   - The variances are the same for all treatments
   - Each treatment has a normal distribution,

   We focus on checking the second assumption(constant variance). For an ANOVA, the **fitted values** are the treatment means($\bar{y}_{i.}$), and the **residuals** are the differences between the observed data(y_{ij}) and the treatment means. Intuitively, if the residuals are usually small, then the model does a good job of predicting. In fact, the SSE is just the sum of the squares of the residuals.

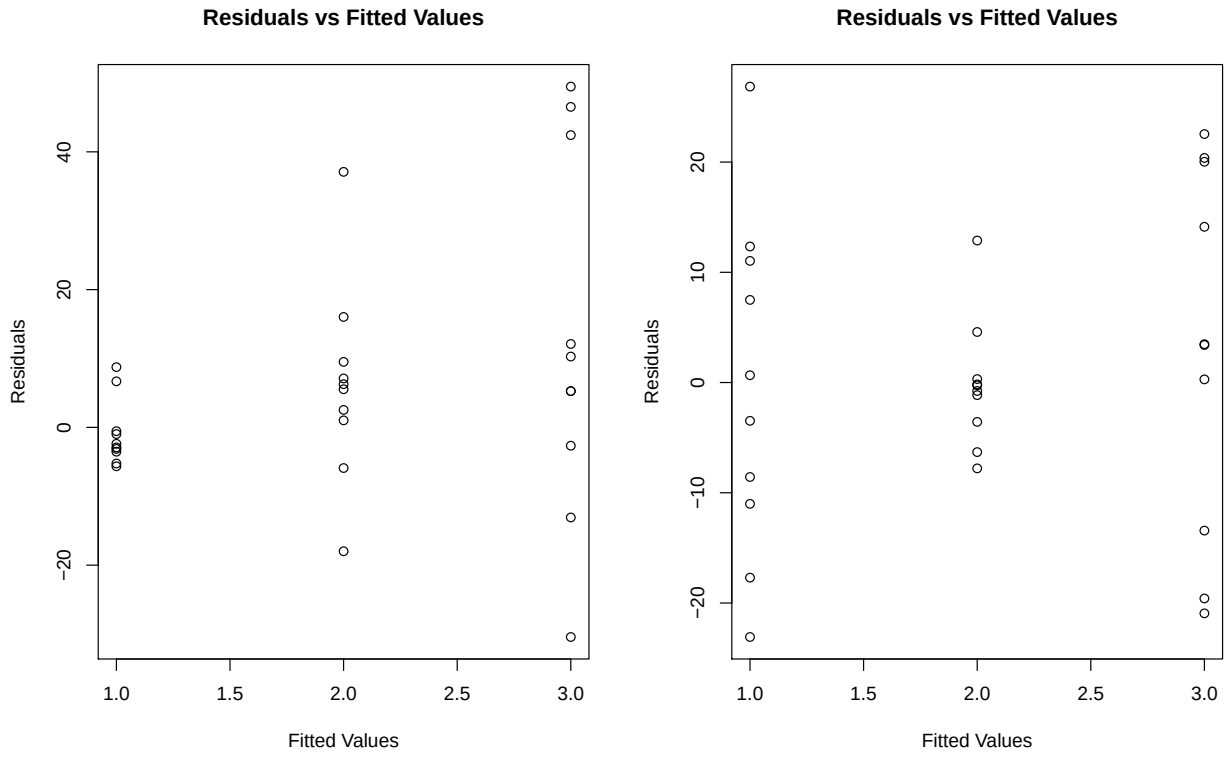
   Since the second assumption requires the variances are the same for all treatments. Hence, we are looking to verify that the spreads of residuals are about the same for each treatment.

   Assume there are 3 treatments. We plot the following graph. The y-axis is called 'Residuals' and the x-axis is called 'Fitted Values'. There might be slightly less spread in the last group, but they're close enough.



   The reason we plot the residuals vs the fitted values is that it is often the case that variability will increase with increasing fitted values. As an extreme example, the SD of weights of 10 elephants will tend to be larger than the SD of weights of 10 hampsters. So, we like to plot against the fitted values to make it easier to see changes in variability of this type. Here are some examples of residuals vs fitted plots that would indicate violation of constant variance. In the first, we see the classic funnel pattern where

the variability increases with the fitted value. In the second, we still see non-constant variance even though there is no pattern relative to the fitted value.



Another option for checking constant variances is to use the ratio of SDs guideline first mentioned when comparing two populations. Since there are now more than two groups, it is typical to take the ratio of the largest and smallest sample SDs - if this ratio passes the test, then every other pair will as well. For the example in Part 1 of Chapter 10, the sample SDs for the four groups are 1.83, 2.83, 1.67, and 2.62. The ratio of the smallest to the largest is $1.67/2.83 = 0.59$, which falls between 0.5 and 2.0, so assuming the variances equal should be safe.

2. Multiple Comparisons Following Significant ANOVA

The concepts on multiple comparisons in this section are covered in sections 9.3-9.5 of Ott and Longnecker.

When running an ANOVA, if we do not reject the null, we're done. But if we reject, we only know that at least one mean differs from at least one other mean, but not how many means differ, or which ones, or by how much. One way to figure these things out is with post-hoc pairwise tests. These pairwise tests are very similar to the two-sample t-tests that we considered using from the very beginning, with one very important difference.

The procedure is to compare treatments i and i' with either a special T-test:

$$t = \frac{\bar{y}_{i.} - \bar{y}_{i'.}}{\sqrt{MSE(1/n_i + 1/n_{i'})}}$$

which we would compare to a t_{df_E}. Or, to make a $100(1-\alpha)\%$ CI on the difference as follows. If the CI contains 0, then we don't reject the null hypothesis, which means the treatment has the same effect.

$$\bar{y}_{i\cdot} - \bar{y}_{i'\cdot} \pm t_{df_E,\alpha/2}\sqrt{MSE(1/n_i + 1/n_{i'})}$$

Either works fine and will give the same conclusions. In this problem, we will primarily work with CIs. We must then create these CIs for every pair of treatments. Supposing we want 95% CIs, $t_{20,0.025} = 2.086$, and MSE = 5.6, so they are as follows:

Trt 1 vs Trt 2: $61 - 66 \pm 2.086\sqrt{5.6(1/4 + 1/6)} = -5 \pm 3.19 = (-8.19, -1.81)$
Trt 1 vs Trt 3: $61 - 68 \pm 2.086\sqrt{5.6(1/4 + 1/6)} = -7 \pm 3.19 = (-10.19, -3.81)$
Trt 1 vs Trt 4: $61 - 61 \pm 2.086\sqrt{5.6(1/4 + 1/8)} = 0 \pm 3.02 = (-3.02, 3.02)$
Trt 2 vs Trt 3: $66 - 68 \pm 2.086\sqrt{5.6(1/6 + 1/6)} = -2 \pm 2.85 = (-4.85, 0.85)$
Trt 2 vs Trt 4: $66 - 61 \pm 2.086\sqrt{5.6(1/6 + 1/8)} = 5 \pm 2.67 = (2.33, 7.67)$
Trt 3 vs Trt 4: $68 - 61 \pm 2.086\sqrt{5.6(1/6 + 1/8)} = 7 \pm 2.67 = (4.33, 9.67)$

The conclusion is that treatments 2 and 3 are the same, and 1 and 4 are the same, but 2 and 3 differ from 1 and 4. Often this information is summarized by sorting the treatment means from largest to smallest, and then adding letter codes. **Two treatments share a letter if they do not differ significantly**:

Treatment	Sample Mean	Letter Code
3	68	A
2	66	A
1	61	B
4	61	B

But wait. If we end up doing a bunch of pairwise tests anyway, why bother with ANOVA at all? There is one theoretical reason, and one practical reason:

- **Theoretically, the F-test is the most powerful test for the hypotheses we specified, provided all of our assumptions are met.**

- **Practically, if there are a lot of treatments, if there are really no differences between the treatments, doing one ANOVA could save time over doing many pairwise tests.**