Chapter 10: Comparing multiple independent populations

(Ott & Longnecker Sections: 14.2 and 14.5)

Duzhe Wang

Part 2

https://dzwang91.github.io/stat371/

WISCONSIN

UNIVERSITY OF WISCONSIN–MADISON

Assumptions of ANOVA

- The data are independent within and between treatments. (check from the story.)
- The variances are the same for all treatments
- Each treatment has a normal distribution. (check using QQ plot.)

Assumptions of ANOVA

- The data are independent within and between treatments. (check from the story.)
- The variances are the same for all treatments
- Each treatment has a normal distribution. (check using QQ plot.)

How do we check the equal variance?

Checking equal variance

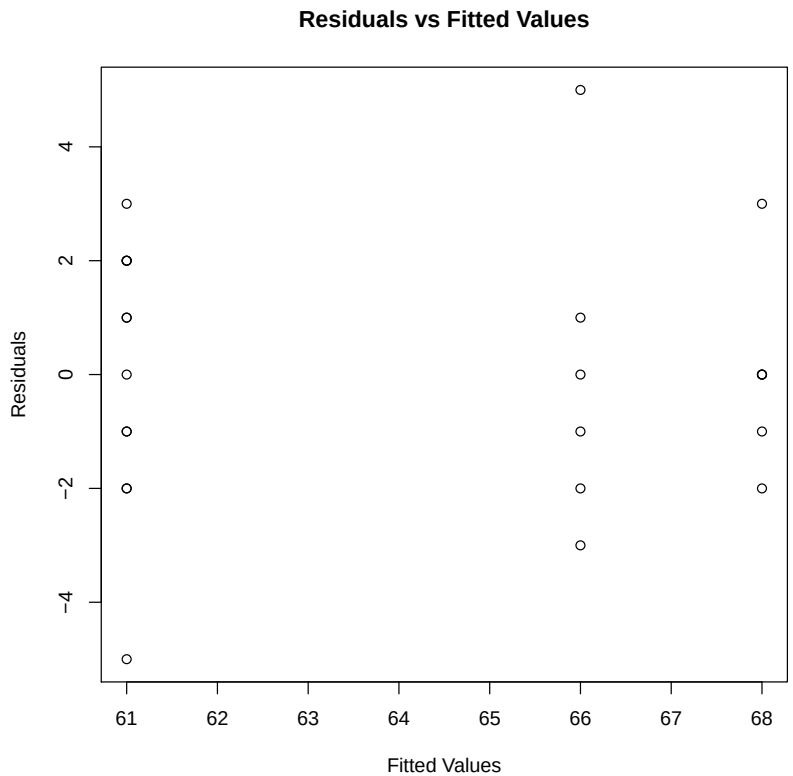
- The fitted values are the treatment means $\bar{y}_{i\cdot}$.
- The residuals are the differences between the observed data(y_{ij}) and the treatment means $y_{ij} - \bar{y}_{i\cdot}$.
- In fact, the sum of squares error (SSE) is the sum of the squares of the residuals.

Checking equal variance

- The fitted values are the treatment means $\bar{y}_{i\cdot}$.
- The residuals are the differences between the observed data(y_{ij}) and the treatment means $y_{ij} - \bar{y}_{i\cdot}$.
- In fact, the sum of squares error (SSE) is the sum of the squares of the residuals.
- **The approach to check equal variance:** If the spreads of residuals are about the same for each treatment, then we are safe to assume equal variance.
- So we need the Residuals vs Fitted values plot.

Residuals vs Fitted values plot

Assume there are 3 treatments. 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.

Residuals vs Fitted plot continued

Residuals vs Fitted plot continued

Residuals vs Fitted plot continued

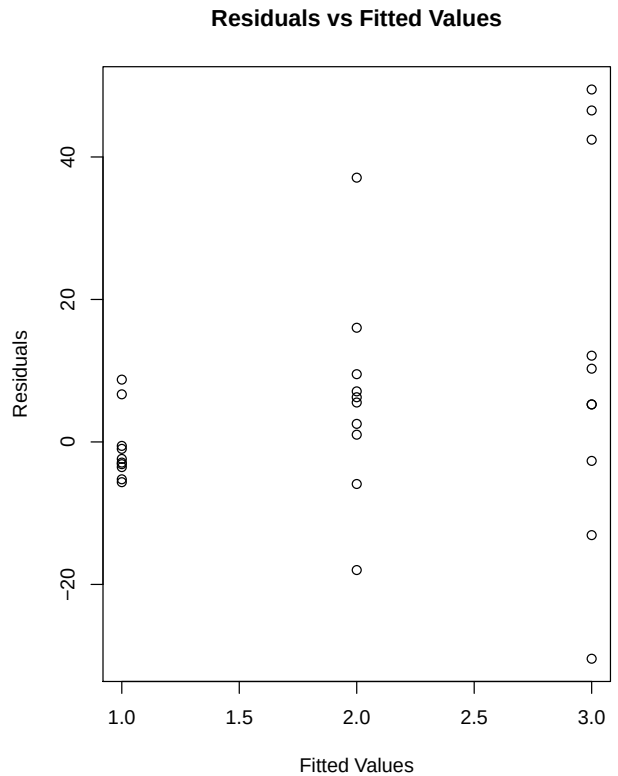
The SD of weights of 5 elephants will tend to be larger than the SD of weights of 5 hamsters.

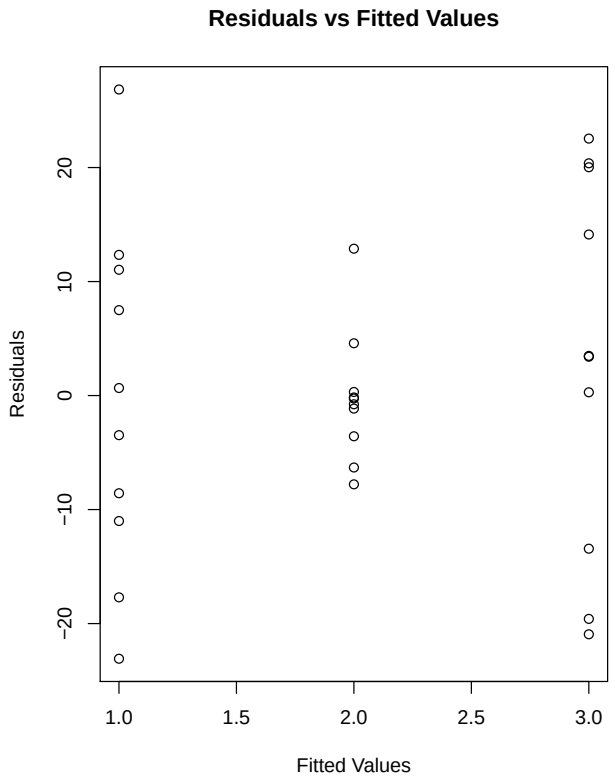
Residuals vs Fitted plot continued

- The reason we plot the residuals vs fitted values is that it is often the case that variability will increase with increasing fitted values.

Residuals vs Fitted plot continued

- The reason we plot the residuals vs fitted values is that it is often the case that variability will increase with increasing fitted values.
- Some bad examples: 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.





Checking equal variance

- Another option for checking equal variance 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.

Multiple comparisons following significant ANOVA

- 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.

Multiple comparisons following significant ANOVA

- 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.

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

which we would compare to a t_{df_E}.

Multiple comparisons following significant ANOVA

- 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.

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

  which we would compare to a t_{df_E}.
- Or 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.

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

CIs

- We will primarily work with CIs. Suppose we want 95% CIs, $t_{20,0.025} = 2.086$, and MSE $= 5.6$, so
  - 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.

Letter code

- The 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

Summary

If we end up doing a bunch of pairwise tests, why do we use ANOVA?

- 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.