Significance tests

Part 3: Analysis of variance (ANOVA)

By: Noureddin Sadawi, PhD

University of London

ANOVA

- ‘Suppose that, instead of an A/B test, we had a comparison of multiple groups, say A/B/C/D, each with numeric data.
- The statistical procedure that tests for a statistically significant difference among the groups is called analysis of variance, or ANOVA.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Key terms for ANOVA

- ‘**Pairwise comparison:** A hypothesis test (e.g., of means) between two groups among multiple groups.
- **Omnibus test:** A single hypothesis test of the overall variance among multiple group means.
- **Decomposition of variance:** Separation of components contributing to an individual value (e.g., from the overall average, from a treatment mean, and from a residual error).
- **F-statistic:** A standardized statistic that measures the extent to which differences among group means exceed what might be expected in a chance model.
- **SS:** “Sum of squares,” referring to deviations from some average value.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Example

- 'Instead of worrying about all the different comparisons between individual pages we could possibly make, we can do a single overall test that addresses the question, "Could all the pages have the same underlying stickiness, and the differences among them be due to the random way in which a common set of session times got allocated among the four pages?"
- The procedure used to test this is ANOVA.'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

	Page 1	Page 2	Page 3	Page 4
	164	178	175	155
	172	191	193	166
	177	182	171	164
	156	185	163	170
	195	177	176	168
Average	172	185	176	162
Grand average				173.75

The table shows the stickiness of four web pages, defined as the number of seconds a visitor spent on the page. The four pages are switched out so that each web visitor receives one at random.

Procedure

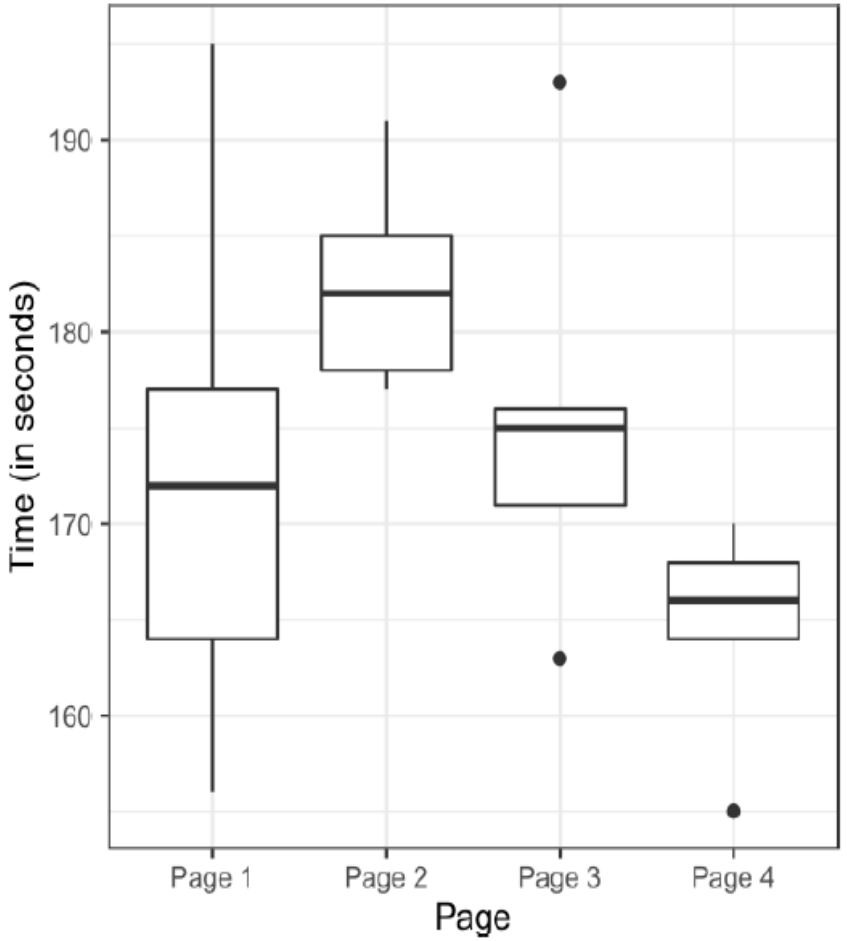
- ‘The basis for it can be seen in the following resampling procedure (specified here for the A/B/C/D test of web page stickiness):
    1. Combine all the data together in a single box.
    2. Shuffle and draw out four resamples of five values each.
    3. Record the mean of each of the four groups.
    4. Record the variance among the four group means.
    5. Repeat steps 2–4 many (say, 1,000) times.
- What proportion of the time did the resampled variance exceed the observed variance? This is the p-value.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Results

- ‘The p-value is 0.09278.
- In other words, given the same underlying stickiness, 9.3% of the time the response rate among four pages might differ as much as was actually observed, just by chance.
- This degree of improbability falls short of the traditional statistical threshold of 5%, so we conclude that the difference among the four pages could have arisen by chance.’

  (Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).



F-statistic

- ‘Just like the t-test can be used instead of a permutation test for comparing the mean of two groups, there is a statistical test for ANOVA based on the F-statistic.
- The F-statistic is based on the ratio of the variance across group means (i.e., the treatment effect) to the variance due to residual error.
- The higher this ratio, the more statistically significant the result.
- If the data follows a normal distribution, then statistical theory dictates that the statistic should have a certain distribution.
- Based on this, it is possible to compute a p-value.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Key ideas

- ‘ANOVA is a statistical procedure for analyzing the results of an experiment with multiple groups.
- It is the extension of similar procedures for the A/B test, used to assess whether the overall variation among groups is within the range of chance variation.
- A useful outcome of ANOVA is the identification of variance components associated with group treatments, interaction effects, and errors.’

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).