Social Robotics

Albert-Ludwigs-Universität Freiburg

Felix Lindner, Laura Wächter, Bernhard Nebel
SoSe 2019

2 / 22

One-Way ANOVA

Outline

- Alpha Adjustment Methods
- One-Way ANOVA
- Planned Contrasts

General Setting: Example
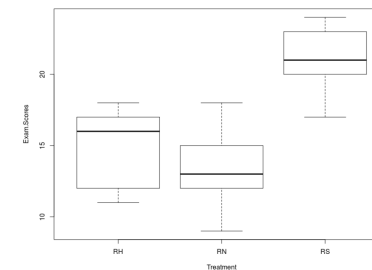
- Independent variable: Treatment, Levels: Robot.Humor (RH), Robot.Neutral (RN), Robot.Strict (RS)
- Dependent variable: Exam.Scores, Levels: $1, \ldots, 25$
- Research question: Is there an effect of emotionality of the robot on the exam scores?
- Quiz: How can we find this out using hypothesis tests?

Alpha Adjustment: The Problem

- Family-wise error rate: The probability of committing at least one Type-1 error in a family of hypothesis tests.
  - Consider a set of m true H_0 hypotheses. What is the probability of rejecting one (or more) of them (false positive)?
  - $\alpha_{total} = 1 - (1 - \alpha)^m$
  - Example: A family of 10 hypotheses, $\alpha = 0.05$. The probability of a false positive is $1 - (1 - 0.05)^{10} = 0.4$!
- Per-Family error rate: The expected number of Type-1 errors per family of hypothesis tests of size m.
  - $\sum_k^m k \binom{n}{k} (\alpha^k (1 - \alpha)^{m-k})$
  - Example: A family of 10 hypotheses, $\alpha = 0.05$. On average, one commits 0.5 Type-1 errors per family.
- Question: What constitutes a family?

Bonferroni Correction

- The Bonferroni correction adjusts the α by dividing it by the number of hypothesis tests to run, i.e., $\alpha' = \alpha/m$ with m being the number of hypothesis tests.
- Only reject those hypotheses, whose p-value is below or equal to α'.
- Bonferroni correction controls FWER and PFER[1]

[1] A. V. Frane. Are per-family type I error rates relevant in social and behavioral science? Journal of Modern Applied Statistical Methods, 14(1), pp. 12–23, 2015.

Holm-Bonferroni Method

- Let $H_1, \ldots, H_m$ be a family of hypotheses and $p_1, \ldots, p_m$ the corresponding p-values.
  1. Start by ordering the p-values (from lowest to highest) $p_{1'}, \ldots, p_{m'}$ and let the associated hypotheses be $H_{1'}, \ldots, H_{m'}$
  2. For a given significance α, let k be the minimal index such that $p_k > \frac{\alpha}{m-k+1}$
  3. Reject the null hypotheses $H_{1'}, \ldots, H_{k-1}$ and do not reject $H_k, \ldots, H_{m'}$
  4. If $k = 1$ then do not reject any of the null hypotheses and if no such k exist then reject all of the null hypotheses.
- Controls the $FWER \leq \alpha$, Proof: `https://en.wikipedia.org/wiki/Holm-Bonferroni_method`


Frane Method[2]

- Sometimes, a family of hypotheses consists of a primary hypothesis and several secondary hypotheses. Frane (2015) proposes to run a two-step procedure that allows to test the primary hypothesis at the unadjusted α:
  1. Step 1: Conduct the primary test at the unadjusted alpha level.
  2. Step 2: If the primary test is significant, then conduct the secondary tests using testwise alpha levels adjusted for the number of secondary tests (Bonferroni correction). But if the primary test is not significant, then forfeit the significance of the secondary tests.

[2] `http://jrp.icaap.org/index.php/jrp/article/view/514/417`

Simulations

See R-Skript `lecture10.Rmd` in git repository `slides`.

Analysis of Variance (ANOVA)

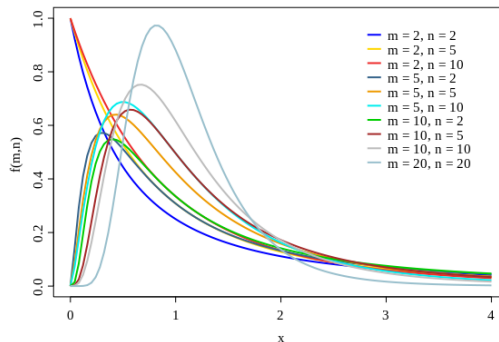- ANOVA makes it possible to test the hypothesis
  $H_0 : \mu_1 = \ldots = \mu_p$ in only one run.
- ANOVA therefore is also often called an omnibus test.

Reconsidering the Example
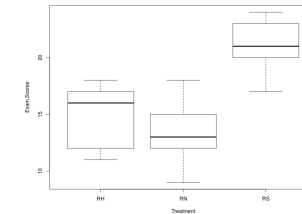
- Intuitively, does H_0 hold here or not? Why? Can we
  formulate the answer using the concept of variance?

F-Distribution

$$F = \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}} \sim F_{m;n}$$

Let's build an F statistics I

- To apply the F-statistics, we need two χ^2-distributed
  variables. To this end, we assume that all p samples are of
  size n and are drawn from a distribution with variance σ^2.
  Thus, all in all, we have drawn $N = np$ scores.
- We already know that, generally, variables of the form
  $\frac{(n-1)s^2}{\sigma^2}$ are χ_{n-1}^2-distributed. Therefore:
  - Let s_{Treat}^2 be the variance of the p group means, then
    $\frac{(p-1)s_{Treat}^2}{\frac{\sigma^2}{n}}$ is χ_{p-1}^2-distributed.
  - Let s_{Res}^2 be the pooled variance of the p samples, then
    $\frac{(N-p)s_{Res}^2}{\sigma^2}$ is χ_{N-p}^2-distributed.
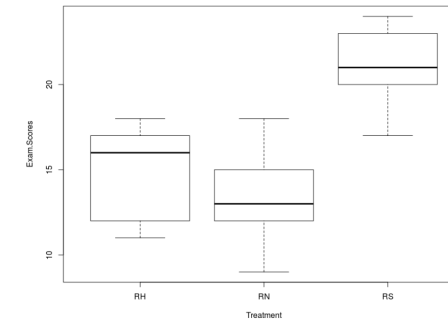
Let's build an F statistics II

- Goal: $F = \dfrac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}} \sim F_{m;n}$

- Just plug it in: $F = \dfrac{\frac{(p-1)s_{Treat}^2}{\sigma^2}}{\frac{(p-1)}{(N-p)s_{Res}^2}} \sim F_{p-1;N-p}$

- $(p-1)s_{Treat}^2$ is called the sum of squares of treatments
  ss_{Treat}, and $(N-p)s_{Res}^2$ is the sum of squares of residuals
  ss_{Res}. Hence, after doing some algebra, we can write:

$$F = n \times \frac{\frac{ss_{Treat}}{(p-1)}}{\frac{ss_{Res}}{(N-p)}} = n \times \frac{s_{Treat}^2}{s_{Res}^2}$$

ANOVA: Example I

- RN: (13, 12, 15, 18, 9)
- RH: (12, 16, 17, 18, 11)
- RS: (20, 21, 17, 23, 24)

ANOVA: Example II

- Scores and statistics:
  - RN: (13, 12, 15, 18, 9), $\overline{X}_{RN} = 13.4$, $s_{RN}^2 = 11.3$
  - RH: (12, 16, 17, 18, 11), $\overline{X}_{RH} = 14.8$, $s_{RH}^2 = 9.7$
  - RS: (20, 21, 17, 23, 24). $\overline{X}_{RS} = 21.0$, $s_{RS}^2 = 7.5$
- Computing the F-value
  - $n = 5$ (size of each sample)
  - $N = 5 \times 3$ (sum of sample sizes)
  - $\overline{X} = \frac{\overline{X}_{RN}+\overline{X}_{RN}+\overline{X}_{RN}}{3} = 16.4$
  - $s_{Treat}^2 = \frac{(\overline{X}_{RN}-\overline{X})^2+(\overline{X}_{RH}-\overline{X})^2+(\overline{X}_{RS}-\overline{X})^2)}{3-1} = 16.36$
  - $s_{Res}^2 = \frac{(5-1)s_{RN}^2+(5-1)s_{RH}^2+(5-1)s_{RS}^2}{(5-1)+(5-1)+(5-1)} = \frac{(5-1)(s_{RN}^2+s_{RH}^2+s_{RS}^2)}{15-3} = \frac{(s_{RN}^2+s_{RH}^2+s_{RS}^2)}{3} = 9.5$, (the s_i^2 are the unbiased variances)
  - $F = 5 \times \frac{16.36}{9.5} = 8.61$
  - $F_{2;12;0.95} = 3.88 \leq F = 8.61$
  - $p = 0.0048 \leq 0.05$
- $\Rightarrow$ Check `lecture10.Rmd` to see how this works in R.

Relation between ANOVA and t-Test

- Reconsider the t-statistics again: $t = \dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{s_1^2+s_2^2}{N}}}$

- It holds that $t^2 = n \times \dfrac{\frac{(\overline{X}_1-\overline{X})^2+(\overline{X}_2-\overline{X})^2}{2-1}}{\frac{ss_{Res}}{(N-2)}} \sim F_{df1=1,df2=N-2}$

- (I spare you the algebra here, but it is instructive to do it manually once.)

- Hence, for the case of two levels, running ANOVA and running t-Test is equivalent. Because ANOVA can handle more than two levels, ANOVA counts as a generalization of the t-Test.

- Assumptions for t-Tests thus also constitute assumptions for ANOVA.

Planned Contrasts

- An ANOVA model can be explored in more detail to test more specific hypotheses, so-called planned contrasts between two (blended) means.
- Using contrast coefficients c_p (one for each $\overline{X}_p$), one can specify which means one is interested in, and which means get blended.
- The F-value is then computed like this:

$$F = \frac{ss_{Con}}{s^2_{Res}} = \frac{n \times \frac{(\Sigma_p c_p \overline{X}_p)^2}{\Sigma_p c_p^2}}{s^2_{Res}}$$

- The degree of freedom of the nominator is always 1 (it involves two means), and the degree of freedom of the denominator is still $N - p$. Thus, the situation is similar to t-Tests.
- Actually, a lot more could be said about this, but we want to just use it to make more out of the ANOVA results.

Planned Contrasts: Example I
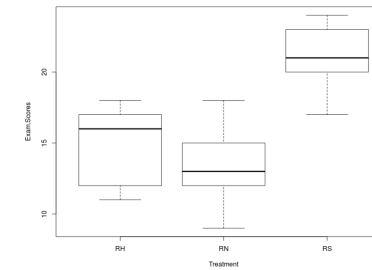
- Say we are interested in two null hypotheses:
  - $H0_1$: The treatment conditions (humor, strict) have no effect compared to the neutral condition, formally $H0_1 : \mu_{RN} = (\mu_{RH} + \mu_{RS})/2$.
  - $H0_2$: There is no difference between conditions humor and strict, formally $H0_2 : \mu_{RH} = \mu_{RS}$.

Planned Contrasts: Example II

- $H0_1$: The treatment conditions (humor, strict) have no effect compared to the neutral condition, formally $H0_1 : \mu_{RN} = (\mu_{RH} + \mu_{RS})/2$.
  - Coefficients: $c_{RN} = 1, c_{RH} = -1/2, c_{RS} = -1/2$
- $H0_2$: There is no difference between conditions humor and strict, formally $H0_2 : \mu_{RH} = \mu_{RS}$.
  - Coefficients: $c_{RN} = 0, c_{RH} = 1, c_{RS} = -1$
- Computing the F-values:
  - $s^2_{Res} = 9.5, n = 5, \overline{X}_{RN} = 13.4, \overline{X}_{RH} = 14.8, \overline{X}_{RS} = 21.0$
  - $F = \frac{ss_{Con}}{s^2_{Res}} = \frac{5 \times (13.4 - 7.4 - 10.5)^2 / 1.5}{9.5} = 7.105 \geq F_{1;12;0.95} = 4.75$
  - $F = \frac{ss_{Con}}{s^2_{Res}} = \frac{5 \times (14.8 - 21.0)^2 / 2}{9.5} = 10.116 \geq F_{1;12;0.95} = 4.75$
- $\Rightarrow$Check `lecture10.Rmd` to see how it works in R.

Summary

- α-Error cumulation calls for α adjustments
  - Bonferroni correction
  - Holm-Bonferroni method
  - Two-Step procedure (Frane, 2015)
- ANOVA as an omnibus test
- Using contrasts to test specific hypotheses

Notes

- Further ANOVA methods one can explore (but we won't in this lecture):
  - Two-Way ANOVA
  - Repeated-Measures ANOVA
  - MANOVA
- Next we will learn about analyzing data that involves ordinal variables using so-called non-parametric methods

Sketches
Intentionally left blank :-)