Statistics 2

Regression Modeling with Categorical Predictors: Multiple comparisons, contrasts

Casper Albers & Jorge Tendeiro

Lecture 9, 2019 – 2020

university of groningen

Overview

Literature for this lecture

Read:

Agresti, Section 12.2

Example – Preventing flashbacks (CONDITION)

Recall the James et al. (2015) data.

For any coding system, the same F test for the overall effect applies:

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	114.8	3	38.27	3.795	0.014
	Residual	685.8	68	10.09		
	Total	800.7	71			

$F(3, 68) = 3.795$, $p = .014$: We reject $\mathcal{H}_0$ at 5% significance level, where

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \iff \quad \mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0.$$

Rejecting $\mathcal{H}_0$ means that there is evidence that not all population group means are equal.

Q: But, which groups differ from which?

A: Further investigation is required:

- Visually — use plots.
- Perform statistical inference:
  - Planned comparisons: Contrasts.
  - Post hoc comparisons: Multiple comparisons.

Why not multiple t tests?

Why are contrasts and multiple comparisons needed?
For example, why not performing multiple t tests?

Example: 4 groups $\Rightarrow$ 6 t-tests:

- $\mathcal{H}_0 : \mu_1 = \mu_2$ vs $\mathcal{H}_a : \mu_1 \neq \mu_2$
- $\mathcal{H}_0 : \mu_1 = \mu_3$ vs $\mathcal{H}_a : \mu_1 \neq \mu_3$
- $\mathcal{H}_0 : \mu_1 = \mu_4$ vs $\mathcal{H}_a : \mu_1 \neq \mu_4$
- $\mathcal{H}_0 : \mu_2 = \mu_3$ vs $\mathcal{H}_a : \mu_2 \neq \mu_3$
- $\mathcal{H}_0 : \mu_2 = \mu_4$ vs $\mathcal{H}_a : \mu_2 \neq \mu_4$
- $\mathcal{H}_0 : \mu_3 = \mu_4$ vs $\mathcal{H}_a : \mu_3 \neq \mu_4$

Problem: Too large overall error rate, or experiment-wise error rate.

Probability of at least one Type I error in the set of tests.

Why not multiple t tests?

If $\alpha = 5\%$ for each test, then for 6 tests:

$$\begin{aligned} \text{overall error rate} &= \text{probability of at least one false rejection} \\ &= 1 - (\text{probability of no false rejection}) \\ &\approx 1 - (1 - .05)^6 \\ &= .265. \end{aligned}$$

Conclusion: There is a 26% probability of at least one false rejection $\longrightarrow$ overrejecting.

This problem is known by chance capitalization:

> *The probability of making a Type I error increases with the number of tests to be performed.*

Chance capitalization

# Groups	# Pairwise Comparisons	Overall α (in %)	
		$\alpha = 5\%$ per test	$\alpha = 1\%$ per test
2	1	5.0	1.0
3	3	14.3	3.0
4	6	26.5	5.9
5	10	40.1	9.6
8	28	76.2	24.5
10	45	90.1	36.4
100	4950	100.0	100.0

Avoid chance capitalization by using one of the following inference procedures:

- Planned comparisons: Contrasts.
- Post hoc comparisons: Multiple comparisons.

(Note: Contrasts are not covered in the book, but it is crucial to learn about them!).

Contrasts

Contrasts, or planned comparisons:

Hypotheses constructed prior to data collection.

Q: How do constrasts look like?

A: Contrasts are written as linear combinations of group means.

Example: Experiment with three groups.

- Group 1 $=$ Treatment 1.
- Group 2 $=$ Treatment 2.
- Group 3 $=$ Control group.

The researcher wants to know whether each treatment works.

Assume that larger scores $\Longrightarrow$ better treatment effect.

Contrasts

Interesting research hypotheses:

- **Contrast 1:** Is Treatment 1 more effective than Treatment 2?
  Test:

$$\mathcal{H}_{01} : \mu_1 = \mu_2 \quad \text{versus} \quad \mathcal{H}_{a1} : \mu_1 > \mu_2.$$

$$\boxed{\mathcal{H}_{01} : \mu_1 - \mu_2 = 0} \quad \text{versus} \quad \mathcal{H}_{a1} : \mu_1 - \mu_2 > 0.$$

$$\mathcal{H}_{01} : 1\mu_1 + (-1)\mu_2 + 0\mu_3 = 0.$$

$$\downarrow$$

$$\mathcal{H}_{01} : \psi_1 = 0, \text{ with } \psi_1 = 1\mu_1 + (-1)\mu_2 + 0\mu_3.$$

So,

- ✓ Contrast $1 = 1\mu_1 + (-1)\mu_2 + 0\mu_3$.
- ✓ Coefficients of contrast 1: 1, -1, 0.

Contrasts

Interesting research hypotheses:

- **Contrast 2:** Is the treatment effect (groups 1 and 2 combined) effective? Test:

$$\mathcal{H}_{02}: \frac{\mu_1+\mu_2}{2} = \mu_3 \quad \text{versus} \quad \mathcal{H}_{a2}: \frac{\mu_1+\mu_2}{2} > \mu_3.$$

$$\boxed{\mathcal{H}_{02}: \tfrac{\mu_1+\mu_2}{2} - \mu_3 = 0} \quad \text{versus} \quad \mathcal{H}_{a2}: \frac{\mu_1+\mu_2}{2} - \mu_3 > 0.$$

$$\mathcal{H}_{02}: .5\mu_1 + .5\mu_2 + (-1)\mu_3 = 0.$$

$$\boxed{\mathcal{H}_{02}: \psi_2 = 0, \text{ with } \psi_2 = .5\mu_1 + .5\mu_2 + (-1)\mu_3.}$$

✓ Contrast 2 $= .5\mu_1 + .5\mu_2 + (-1)\mu_3$.
✓ Coefficients of contrast 2: $.5$, $.5$, -1.

Contrasts

In general: A contrast is a linear combination of group means

$$\psi = \sum_i a_i \mu_i = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_g \mu_g .$$

Coefficients of the contrast

Contrasts cannot be computed:

- Known a_i's (defined by researcher), <u>but</u>
- Unknown μ's.

Solution: Estimate from sample (μ's $\longrightarrow$ $\overline{y}$'s).

$$\text{Sample contrast} = c = a_1 \overline{y}_1 + a_2 \overline{y}_2 + \cdots + a_g \overline{y}_g .$$

Contrasts

$$c = \sum_i a_1\overline{y}_1 + a_2\overline{y}_2 + \cdots + a_g\overline{y}_g$$

Inference using contrasts:

- Standard error of c: $SE_c = s_p\sqrt{\sum_i \frac{a_i^2}{n_i}}$.
- Statistical testing: $\mathcal{H}_0 : \psi = 0$ ($\mathcal{H}_a$ one- or two-sided).

$$t = \frac{c}{SE_c} \underset{\text{Under } \mathcal{H}_0}{\sim} t(N-g)$$

- Confidence interval:

$$c \pm t^* SE_c$$

$t^* =$ critical value from $t(N-g)$.

Contrasts

A note on the coefficients of contrasts.

$$\boxed{\psi = 1\mu_1 + (-1)\mu_2 + 0\mu_3 \longrightarrow \text{coefficients: } 1,\ -1,\ 0}$$

- A multiple of these coefficients could be used, e.g.
  - ✓ $(2, -2, 0)$, $(-.5, .5, 0)$, ...

  This does not affect the t test (t statistic & p-value the same).
- However, this does affect the computation of CIs!

Q: How to compute contrasts?

A: Contrasts are often tested in combination with ANOVA, as we will see.

However, contrasts can also be tested using regression with code variables (recall last lecture)!

Example – Preventing flashbacks (CONDITION)

CONDITION

1 = No-task control
2 = Reactivation + Tetris
3 = Tetris only
4 = Reactivation only

Group	z_1	z_2	z_3
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

$$\longleftrightarrow \begin{cases} \beta_1 = \mu_1 - \mu_4 \\ \beta_2 = \mu_2 - \mu_4 \\ \beta_3 = \mu_3 - \mu_4 \\ \alpha = \mu_4 \end{cases}$$

						95% CI	
	Unstd. Coef.	SE	Std. Coef	t	p	Lower	Upper
(Intercept)	4.833	0.749		6.457	$< .001$	3.340	6.327
z1	0.278	1.059	0.036	0.262	0.794	−1.835	2.390
z2	−2.944	1.059	−0.382	−2.781	0.007	−5.057	−0.832
z3	−0.944	1.059	−0.123	−0.892	0.375	−3.057	1.168

- $\psi_1 = \mu_1 - \mu_4$, $t(68) = 0.262$, $p = .794$, 95% CI $= (-1.835, 2.390)$.
- $\psi_2 = \mu_2 - \mu_4$, $t(68) = -2.781$, $p = .007$, 95% CI $= (-5.057, -0.832)$.
- $\psi_3 = \mu_3 - \mu_4$, $t(68) = -0.892$, $p = .375$, 95% CI $= (-3.057, 1.168)$.

Note: $df = N - g = 72 - 4 = 68$.

Example – Preventing flashbacks (CONDITION)

Let's look at one of the contrasts by hand: $\psi_2 = \mu_2 - \mu_4$.

	N_INTR			
	1	2	3	4
Valid	18	18	18	18
Mean	5.111	1.889	3.889	4.833
Std. Deviation	4.227	1.745	2.888	3.330

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	114.8	3	38.27	3.795	0.014
	Residual	685.8	68	10.09		
	Total	800.7	71			

- ψ_2 coefficients: $(a_1, a_2, a_3, a_4) = (0, 1, 0, -1)$.
- $c = \sum_i a_i \overline{y}_i = 1 \times 1.889 - 1 \times 4.833 = -2.944$ (note: This is b_2!!).
- $s_p = \sqrt{MSE} = \sqrt{10.09} = 3.1765$.
- $SE_c = s_p \sqrt{\sum_i \frac{a_i^2}{n_i}} = 3.1765 \sqrt{\frac{1}{18} + \frac{1}{18}} = 1.0588$.
- $t = \frac{c}{SE_c} = \frac{-2.944}{1.0588} = -2.781$.
- $95\%\text{CI} = c \pm t^*_{df=68} SE_c = c \pm 1.9955 \times 1.0598 = (-5.06, -0.83)$.

Contrasts

Q: What if I want to define my own contrasts, for example,

$$\mathcal{H}_0 : \mu_1 = \frac{\mu_2 + \mu_3 + \mu_4}{3}$$
$$\mathcal{H}_0 : \mu_2 = \frac{\mu_3 + \mu_4}{2}$$
$$\mathcal{H}_0 : \mu_3 = \mu_4?$$

A: You have two options:

- Use a predefined family of contrasts available in your software of choice. For example, the set of contrasts above is known as the Helmert type of contrast.
- Create your own set of code variables that reflects the exact contrasts you want to test. This uses the so-called contrast coding system – material of Stats III!

Multiple comparisons

Multiple comparisons: Tests (CIs) for differences between all pairs of group means.

- Only to be used if, prior to the analysis, no specific hypotheses can be (or have been) defined.

Post-hoc tests

- Use t tests (two-sided), but:
  - ✓ Caution: Chance capitalization.
- Tests (CIs) must therefore be adjusted.

CI for $(\mu_i - \mu_j)$:

Based on all groups!

$$CI_{ij} = (\overline{y}_i - \overline{y}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

There are many ways to find suitable critical values t^{**}. We mention two:

- Bonferroni method.
- Tukey method.

Multiple comparisons – Bonferroni

CI $(\mu_i - \mu_j)$:

$$CI_{ij} = (\overline{y}_i - \overline{y}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

- Adjusts Type I error level per test to α/k (k = total number of tests) so that the overall error rate $\leq \alpha$.
- Motivation: The Bonferroni inequality:

$$P(\text{at least one Type I error}) \leq$$

$$\leq \sum_{i=1}^{k} \underbrace{P(\text{Type I error in test } i)}_{\alpha/k} = k \times \frac{\alpha}{k} = \alpha.$$

- Conclusion: The overall error rate (i.e., the probability of at least one false rejection) is never larger than α (but, it can be quite smaller than α).

$$\boxed{t^{**} = t_{(\alpha/k)/2}, \text{ with } df = df_{\text{DFE}}}$$

Multiple comparisons – Tukey

CI $(\mu_i - \mu_j)$:

$$CI_{ij} = (\overline{y}_i - \overline{y}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

- Tukey's method tries to approximate the overall error rate, instead of only not being larger than it (like Bonferroni).
- Because of this, Tukey's CIs tend to be narrower than the Bonferroni CIs.
- The computation of Tukey's CI relies on the so-called 'studentized range' distribution, which we don't explicitly use in this course.
- Hence, rely on software only to derive Tukey's CIs.

Example – Preventing flashbacks (CONDITION)

Group i	Group j	Mean Difference	95% CI Lower	95% CI Upper	SE	t	p_{tukey}	p_{bonf}
1	2	3.222	0.434	6.010	1.059	3.044	0.017	0.020
	3	1.222	−1.566	4.010	1.059	1.155	0.657	1.000
	4	0.278	−2.510	3.066	1.059	0.262	0.994	1.000
2	3	−2.000	−4.788	0.788	1.059	−1.889	0.242	0.379
	4	−2.944	−5.733	−0.156	1.059	−2.781	0.034	0.042
3	4	−0.944	−3.733	1.844	1.059	−0.892	0.809	1.000

Note. Confidence intervals based on Tukey's HSD.

JASP (version 0.9.2) currently only offers Tukey's CIs.
It is easy to compute Bonferroni's CIs by hand (exam material!).
For example, Bonferroni's 95% CI for $(\mu_1 - \mu_2)$:

- $s_p = 3.1765$ (see previous example for contrasts)
- $n_1 = n_2 = 18$
- $\alpha = .05$, $k = 6$ (total number of paired comparisons), $df = 68$
- $t^{**} = t_{(.05/6)/2} = 2.7176$
- $95\%\text{CI} = 3.222 \pm t^{**} s_p \sqrt{\frac{1}{18} + \frac{1}{18}} = (0.345, 6.099)$.

Contrasts versus multiple comparisons

Advantages of contrasts over multiple comparisons:

- Fewer tests:
  - ✓ Smaller overall error rate.
  - ✓ Less chance capitalization.
  - ✓ Fewer Type I errors.
- More power:
  - ✓ Better probability to reject $\mathcal{H}_0$ when it is not true. (fewer Type II errors)

Disadvantage of contrasts:

- Researchers don't always know beforehand which comparisons should be made.

Next lecture – Regression Modeling with Categorical Predictors: ANOVA

Agresti, Section 12.3