Lecture 25

Multiple Testing & Family-Wise Error Rate

STAT 8010 Statistical Methods I
October 21, 2019

Whitney Huang
Clemson University

One-Way ANOVA & Overall F-Test

- We use one-way ANOVA to compare means of **J ($\geq$ 3) groups/conditions**

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$$
$$H_a : \text{at least a pair } \mu\text{'s differ}$$

- If H_0 is rejected, ANOVA just states that there is a significant difference between the groups **but not where those differences occur**

- We need to perform additional post hoc tests, multiple comparisons, to determine where the group differences are

Frame Title

- Suppose we have 4 groups, i.e. $J = 4$, then we need to perform $\binom{4}{2} = 6$ two-sample tests to locate where the group differences are

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_3 \text{ vs. } H_a : \mu_1 \neq \mu_3$$
$$H_0 : \mu_1 = \mu_4 \text{ vs. } H_a : \mu_1 \neq \mu_4$$
$$H_0 : \mu_2 = \mu_3 \text{ vs. } H_a : \mu_2 \neq \mu_3$$
$$H_0 : \mu_2 = \mu_4 \text{ vs. } H_a : \mu_2 \neq \mu_4$$
$$H_0 : \mu_3 = \mu_4 \text{ vs. } H_a : \mu_3 \neq \mu_4$$

- What if we simply perform these tests using, say, $\alpha = 0.05$ for each test?

  $\mathbb{P}(\text{making a least one type I error}) = 1 - (1 - 0.05)^6 = 0.265$

  if each test was independent

Family-Wise Error Rate (FWER)

> Family-Wise Error Rate (FWER) $\bar{\alpha}$: the probability of making 1 or more type I errors in a set of hypothesis tests

For m independent tests, each with individual type I error rate α, then we have

$$\bar{\alpha} = 1 - (1 - \alpha)^m$$

	α		
m	0.1	0.05	0.01
1	0.100	0.050	0.010
3	0.271	0.143	0.030
6	0.469	0.265	0.059
10	0.651	0.401	0.096
15	0.794	0.537	0.140
21	0.891	0.659	0.190

The Bonferroni Correction

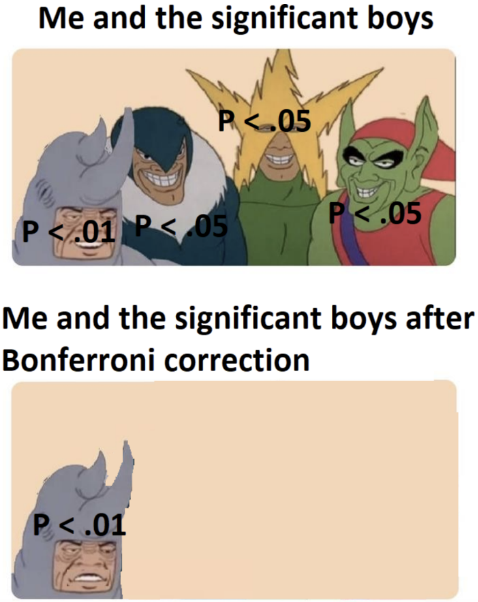
If we would like to control the FWER to be α, then we adjust the significant level for each of the m tests to be $\frac{\alpha}{m}$

$$FWER = \mathbb{P}(\cup_{i=1}^{m} p_i \leq \frac{\alpha}{m}) \leq \sum_{i=1}^{m} \mathbb{P}(p_i \leq \frac{\alpha}{m}) = m\frac{\alpha}{m} = \alpha$$

where p_i is the p-value for the i_{th} test

If we have 4 treatment groups, then we need to perform 6 tests $(m = 6) \Rightarrow$ will need to set the significant level for each individual pairwise t-test to be $0.05/6 = 0.0083$ to ensure that FWER is less than 0.05

Remark: Bonferroni procedure can be very conservative but gives guaranteed control over FWER at the risk of reducing statistical power. Does not assume independence of the comparisons.


Me and the significant boys
P < .01
P < .05
P < .05
P < .05
Me and the significant boys after Bonferroni correction
P < .01

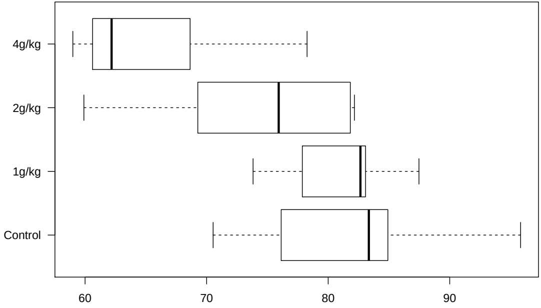
Example

A researcher who studies sleep is interested in the effects of ethanol on sleep time. She gets a sample of 20 rats and gives each an injection having a particular concentration of ethanol per body weight. There are 4 treatment groups, with 5 rats per treatment. She records Rapid eye movement (REM) sleep time for each rat over a 24-period.

Treatment	Control	1g/kg	2g/kg	4g/kg
Mean	82.2	81.0	73.8	65.7
Std	9.6	5.3	9.4	7.9

Recall in last lecture we reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ at 0.05 level. But where these differences are?

Example: Multiple Testing with Bonferroni Correction



P-value

Test	μ_1, μ_2	μ_1, μ_3	μ_1, μ_4	μ_2, μ_3	μ_2, μ_4	μ_3, μ_4
Pooled	0.816	0.202	0.018	0.175	0.007	0.179
Non-pooled	0.818	0.202	0.019	0.185	0.009	0.180

Fisher's Protected Least Significant Difference (LSD) Procedure

- We conclude that μ_i and μ_j differ at α significance level if $|\bar{X}_i - \bar{X}_j| > LSD$, where

$$LSD = t_{\alpha/2, df=N-J} \sqrt{\text{MSE} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

- This procedure builds on the equal variances t-test of the difference between two means

- The test statistic is improved by using MSE rather than s_p^2

Tukey's Honest Significance Difference (HSD) Test

- The test procedure:
  - Requires equal sample size n per populations
  - Find a critical value ω as follows:

$$\omega = q_\alpha(J, N-J)\sqrt{\frac{\text{MSE}}{n}}$$

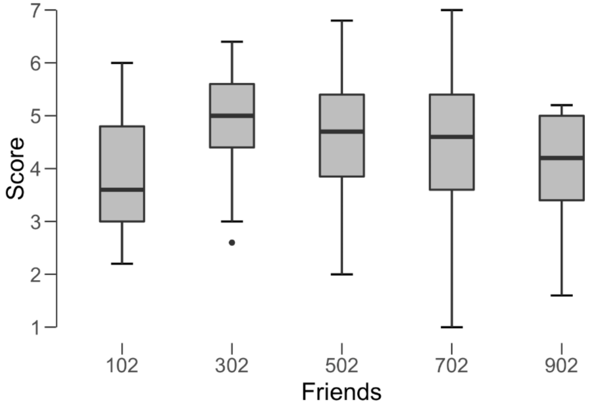
  where $q_\alpha(J, N-J)$ can be obtained from the studentized range table

- If $\bar{X}_{max} - \bar{X}_{min} > \omega \Rightarrow$ there is sufficient evidence to conclude that $\mu_{max} > \mu_{min}$

- Repeat this procedure for each pair of samples. Rank the means if possible

Facebook Friends Example

A researcher would like to investigate the relationship between Facebook social attractiveness and the number of Facebook friends. An experiment was conducted where five groups of participant judge the same Facebook profiles, except for the one aspect that was manipulated: the number of friends for that profile.



Example: Descriptive Statistics

	Score				
	102	302	502	702	902
Valid	24	33	26	30	21
Missing	0	0	0	0	0
Mean	3.817	4.879	4.562	4.407	3.990
Std. Deviation	0.999	0.851	1.070	1.428	1.023
Minimum	2.200	2.600	2.000	1.000	1.600
Maximum	6.000	6.400	6.800	7.000	5.200

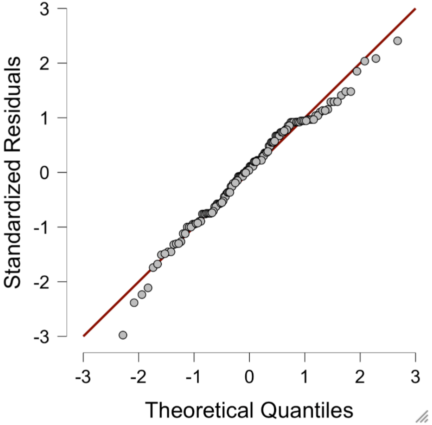
Example: Checking Model Assumptions

Assumption Checks ▼

Test for Equality of Variances (Levene's)

F	df1	df2	p
2.607	4.000	129.000	0.039

Q-Q Plot ▼



CLEMSON UNIVERSITY

Example: ANOVA Table

ANOVA - Score

Cases	Homogeneity Correction	Sum of Squares	df	Mean Square	F	p
Friends	None	19.890	4.000	4.973	4.142	0.003
Friends	Brown-Forsythe	19.890	4.000	4.973	4.184	0.003
Friends	Welch	19.890	4.000	4.973	5.445	< .001
Residual	None	154.867	129.000	1.201		
Residual	Brown-Forsythe	154.867	114.185	1.356		
Residual	Welch	154.867	61.144	2.533		

Note. Type III Sum of Squares

Example: Multiple Testing

CLEMSON UNIVERSITY

Post Hoc Tests

Post Hoc Comparisons - Friends

		Mean Difference	95% CI for Mean Difference Lower	Upper	SE	t	Cohen's d	p_{tukey}	p_{bonf}
102	302	−1.062	−1.875	−0.249	0.294	−3.613	−1.160	0.004	0.004
	502	−0.745	−1.603	0.113	0.310	−2.402	−0.718	0.121	0.177
	702	−0.590	−1.420	0.240	0.300	−1.966	−0.470	0.288	0.514
	902	−0.174	−1.080	0.732	0.327	−0.531	−0.172	0.984	1.000
302	502	0.317	−0.478	1.112	0.287	1.104	0.333	0.804	1.000
	702	0.472	−0.293	1.237	0.276	1.708	0.406	0.433	0.900
	902	0.888	0.042	1.735	0.306	2.904	0.964	0.035	0.043
502	702	0.155	−0.657	0.967	0.294	0.528	0.121	0.984	1.000
	902	0.571	−0.318	1.460	0.321	1.776	0.544	0.392	0.780
702	902	0.416	−0.446	1.279	0.312	1.335	0.326	0.670	1.000

Note. Cohen's d does not correct for multiple comparisons.

Note. Confidence interval adjustment: tukey method for comparing a family of 5 estimates

Summary

In this lecture, we learned

- Multiple Testing:
  - Family-Wise Error Rate
  - Bonferroni Correction
  - Fisher's LSD and Tukey's HSD

In next lecture we will learn

- Linear Contrasts
  Example: $H_0 : \mu_1 = \frac{\mu_2+\mu_3}{2}$ vs. $H_a : \mu_1 \neq \frac{\mu_2+\mu_3}{2}$