Lecture 16

Multiple Comparisons and Linear Contrasts

Readings: IntroStat Chapter 9; OpenIntro Chapter 7.5

STAT 8010 Statistical Methods I
June 7, 2023

Whitney Huang
Clemson University

Facebook Friends Example

CLEMSON UNIVERSITY

Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook

Stephanie Tom Tong
Brandon Van Der Heide
Lindsey Langwell

Department of Communication

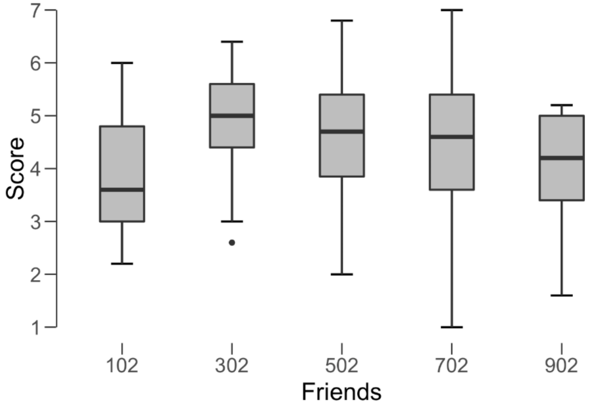
Joseph B. Walther

Departments of Communication and Telecommunication, Information Studies & Media
Michigan State University

A central feature of the online social networking system, Facebook, is the connection to and links among friends. The sum of the number of one's friends is a feature displayed on users' profiles as a vestige of the friend connections a user has accrued. In contrast to offline social networks, individuals in online network systems frequently accrue friends numbering several hundred. The uncertain meaning of friend status in these systems raises questions about whether and how sociometric popularity conveys attractiveness in non-traditional, non-linear ways. An experiment examined the relationship between the number of friends a Facebook profile featured and observers' ratings of attractiveness and extraversion. A curvilinear effect of sociometric popularity and social attractiveness emerged, as did a quartic relationship between friend count and perceived extraversion. These results suggest that an overabundance of friend connections raises doubts about Facebook users' popularity and desirability.

Facebook Friends Example Cont'd

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.



Facebook 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

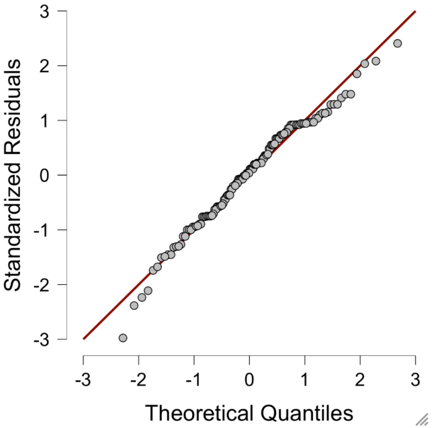
Example: Checking Model Assumptions

Assumption Checks ▼

Test for Equality of Variances (Levene's)

F	df1	df2	p
2.607	4.000	129.000	0.039

Q-Q Plot ▼



Facebook Friends: Overall F-Test

Question: Are Facebook attractiveness affected by # of friends?

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$
H_a: At least one group mean is different from others

CLEMSON UNIVERSITY

Facebook Friends: Overall F-Test

Question: Are Facebook attractiveness affected by # of friends?

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$
H_a: At least one group mean is different from others

```
Analysis of Variance Table

Response: Score
           Df Sum Sq Mean Sq F value
Friends     4  19.89  4.9726   4.142
Residuals 129 154.87  1.2005
           Pr(>F)
Friends   0.00344 **
Residuals
```

Facebook Friends: Overall F-Test

Question: Are Facebook attractiveness affected by # of friends?

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$
H_a: At least one group mean is different from others

```
Analysis of Variance Table

Response: Score
           Df Sum Sq Mean Sq F value
Friends     4  19.89  4.9726   4.142
Residuals 129 154.87  1.2005
           Pr(>F)
Friends   0.00344 **
Residuals
```

Next, we need to figure out where these differences occur

Facebook Example: Fisher's LSD

CLEMSON UNIVERSITY

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

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

Facebook Example: Fisher's LSD

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

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

```
> LSD_none$groups
       Score groups
302 4.878788      a
502 4.561538     ab
702 4.406667    abc
902 3.990476     bc
102 3.816667      c
```

CLEMSON UNIVERSITY

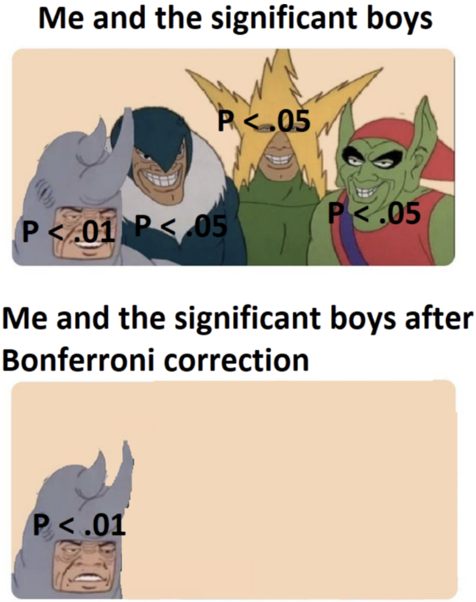
Facebook Example: Fisher's LSD

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

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

```
> LSD_none$groups
       Score groups
302 4.878788      a
502 4.561538     ab
702 4.406667    abc
902 3.990476     bc
102 3.816667      c
```

```
> LSD_bon$groups
       Score groups
302 4.878788      a
502 4.561538     ab
702 4.406667     ab
902 3.990476      b
102 3.816667      b
```


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

CLEMSON UNIVERSITY

Facebook Example: Tukey's HSD Test

Yet there is another method to deal with multiple testing: Tukey's Honest Significant Difference (HSD) test. We conclude that μ_i and μ_j differ at α familywise level if $|\bar{X}_i - \bar{X}_j| > \omega$, where

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

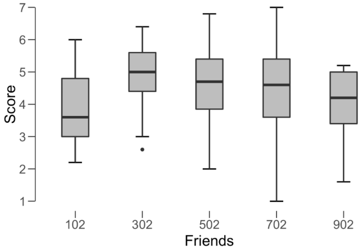
$q_\alpha(J, N-J)$ can be obtained from the studentized range table

Critical Values of Studentized Range Distribution(q) for Familywise ALPHA = .05.

Denominator	**Number of Groups (a.k.a. Treatments)**							
DF	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
51	3.414	3.756	3.999	4.187	4.340	4.469	4.580	4.677
52	3.412	3.753	3.996	4.184	4.337	4.465	4.576	4.673
53	3.410	3.751	3.994	4.181	4.334	4.462	4.572	4.669
54	3.408	3.749	3.991	4.178	4.331	4.459	4.569	4.666
55	3.406	3.747	3.989	4.176	4.328	4.455	4.566	4.662
56	3.405	3.745	3.986	4.173	4.325	4.452	4.562	4.659
57	3.403	3.743	3.984	4.170	4.322	4.449	4.559	4.656
58	3.402	3.741	3.982	4.168	4.319	4.447	4.556	4.652
59	3.400	3.739	3.979	4.165	4.317	4.444	4.553	4.649
60	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646

Facebook Example: Tukey's HSD Test

	diff	lwr	upr	p adj
302-102	1.0621212	0.2488644	1.87537798	0.003889635
502-102	0.7448718	-0.1132433	1.60298691	0.121456224
702-102	0.5900000	-0.2402014	1.42020143	0.288431585
902-102	0.1738095	-0.7320145	1.07963355	0.984016816
502-302	-0.3172494	-1.1121910	0.47769215	0.804080046
702-302	-0.4721212	-1.2368466	0.29260420	0.432633745
902-302	-0.8883117	-1.7345313	-0.04209203	0.034535577
702-502	-0.1548718	-0.9671402	0.65739661	0.984391504
902-502	-0.5710623	-1.4604793	0.31835479	0.391768065
902-702	-0.4161905	-1.2787075	0.44632652	0.669927748



Linear Contrasts

Suppose we have J populations (e.g. response for J different treatments) of interest. We have seen how to perform multiple comparisons. For example, the comparison between μ_1 and μ_2 can be conducted using the test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$. This comparison is actually a special case of linear contrasts

Linear Contrasts

Let $c_1, c_2, \cdots, c_J$ are constants where $\sum_{j=1}^{J} c_j = 0$, then $\sum_{j=1}^{J} c_j \mu_j$ is called a **linear contrast** of the population means.

Example: Suppose $J = 4$

1. $\mu_1 - \mu_3 : c_1 = 1, c_2 = 0, c_3 = -1, c_4 = 0$
2. $\mu_2 - \mu_4 : c_1 = 0, c_2 = 1, c_3 = 0, c_4 = -1$
3. $\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 : c_1 = 1, c_2 = c_3 = c_4 = -\frac{1}{3}$

Inferences for Linear Contrasts

If we want to make a inference about $L = \sum_{i=1}^{J} c_j \mu_j$. Then we use

$$\hat{L} = \sum_{j=1}^{J} c_j \bar{X}_j$$

as the point estimate. Furthermore, we can construct a $100(1-\alpha)$% CI for L:

$$(\hat{L} - t_{(\alpha/2, df=N-J)} \hat{se}_{\hat{L}}, \hat{L} + t_{(\alpha/2, df=N-J)} \hat{se}_{\hat{L}}),$$

where $\hat{se}_{\hat{L}} = \sqrt{\text{MSE}\left(\frac{c_1^2}{n_1} + \cdots + \frac{c_J^2}{n_J}\right)}$

To test whether L is significantly different from 0, we can conduct the following test:

$$H_0 : \sum_{j=1}^{J} c_j \mu_j = 0 \text{ vs. } H_a : \sum_{j=1}^{J} c_j \mu_j \neq 0$$

Hypothesis Testing for Linear Contrasts

1. Null and Alternative Hypotheses:

$$H_0 : \sum_{j=1}^{J} c_j\mu_j = 0 \text{ vs. } H_a : \sum_{j=1}^{J} c_j\mu_j \neq 0$$

2. Test Statistic:

$$t_{obs} = \frac{\hat{L} - 0}{\hat{se}_{\hat{L}}} = \frac{\sum_{j=1}^{J} c_j\bar{X}_j}{\sqrt{\text{MSE}\left(\frac{c_1^2}{n_1} + \cdots + \frac{c_J^2}{n_J}\right)}}$$

3. Decision:

Reject H_0 if $|t_{obs}| > t_{\alpha/2, df=N-J}$ (or p-value $< \alpha$)

CLEMSON UNIVERSITY

Facebook Example: Linear Contrast

Suppose we'd like to compare μ_1 vs. $\frac{\mu_3+\mu_4}{2}$. Let $L = 1\mu_1 - \frac{1}{2}\mu_3 - \frac{1}{2}\mu_4$. Then the above comparison is equivalent to test whether L is different from 0

1. $H_0 : L = 0$ vs. $H_a : L \neq 0$
2. $t_{obs} = \frac{\hat{L}}{\hat{se}_{\hat{L}}} = \frac{1\times3.817-0.5\times4.562-0.5\times4.407}{\sqrt{1.2005\times(\frac{1^2}{24}+\frac{0.5^2}{26}+\frac{0.5^2}{30})}} = \frac{-0.6674}{0.2675} = -2.495$
3. Since $|t_{obs}| = |-2.495| = 2.495 > t_{0.025,df=129} = 1.9785$. We reject H_0 at 0.05 level

> **Note**: If we are performing several tests for different linear contrasts simultaneously, we'll need to adjust α level accordingly to control the FWER

Summary

In this lecture, we learned

- Multiple Comparisons
- Fisher's LSD & Tukey's HSD
- Inference for Linear Contrasts