

Lecture 32 ANOVA & *Post hoc* Multiple Comparisons

BIO210 Biostatistics

Xi Chen

Fall, 2023

School of Life Sciences
Southern University of Science and Technology

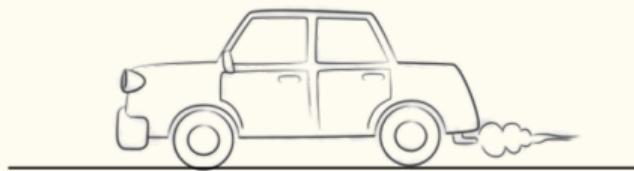


南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Stopping Distance of A Car - Data

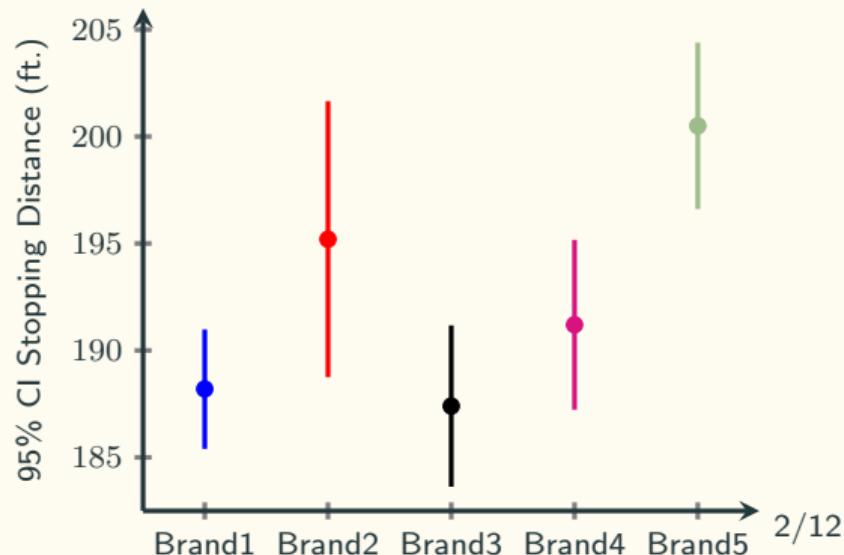
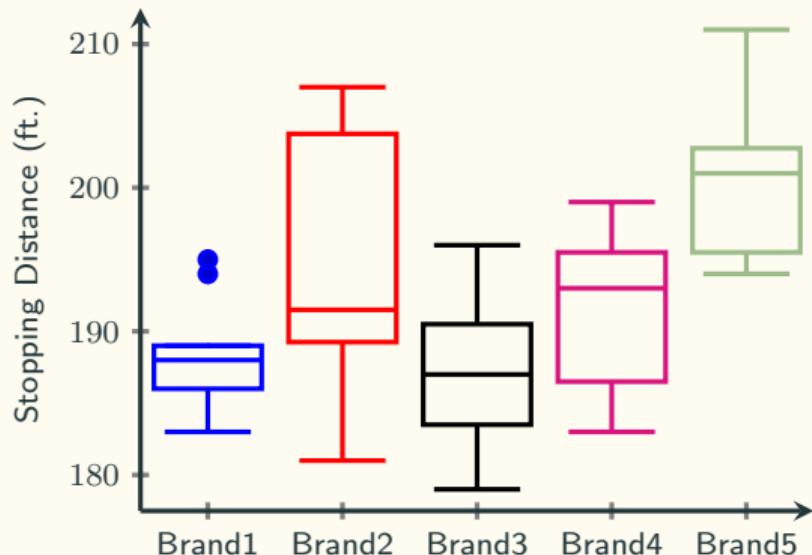
A researcher for an automobile safety institute was interested in determining whether or not the distance that it takes to stop a car going 60 miles per hour depends on the brand of the tire. The researcher measured the stopping distance (in feet) of ten randomly selected cars for each of five different brands. The researcher arbitrarily labeled the brands of the tires as Brand1, Brand2, Brand3, Brand4, and Brand5, so that he and his assistants would remain blinded. Here are the data resulting from his experiment:

Brand1	Brand2	Brand3	Brand4	Brand5
194	189	185	183	195
184	204	183	193	197
189	190	186	184	194
189	190	183	186	202
188	189	179	194	200
186	207	191	199	211
195	203	188	196	203
186	193	196	188	206
183	181	189	193	202
188	206	194	196	195



Stopping Distance of A Car - Descriptive Stats

	Brand1	Brand2	Brand3	Brand4	Brand5
n	10	10	10	10	10
Mean	188.2	195.2	187.4	191.2	200.5
Var	15.06	81.29	27.82	30.84	29.61



Stopping Distance of A Car - The ANOVA Table

	Brand1	Brand2	Brand3	Brand4	Brand5
n	10	10	10	10	10
Mean	188.2	195.2	187.4	191.2	200.5
Var	15.06	81.29	27.82	30.84	29.61
Source of Variation	SS	df	MS	F	p-value
Between	1174.8	4	293.7		
Within	1161.7	45	36.9	7.95	6.17×10^{-5}
Total	2836.5	49			

Assumptions When Using ANOVA

- Randomness, Independence
- Population normally distributed $\left(F = \frac{\text{MSB}}{\text{MSW}} \right)$
- Different groups have equal variance (classical ANOVA)

$$\text{MSW} = \frac{\text{SSW}}{n - k} = \frac{df_1 \cdot s_1^2 + df_2 \cdot s_2^2 + \cdots + df_k \cdot s_k^2}{n - k} = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2 + \cdots + (n_k - 1) \cdot s_k^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)}$$

- Unequal variance: Welch's ANOVA

The Relation Between F -test and t -test

- **Think:** What if the ANOVA method, i.e. using SSB, SSW and the F statistic, is used to compare means from two groups? Valid, or not ?
- t -test statistic with equal variance:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \nu = n_1 + n_2 - 2, s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- The ANOVA Table When $k = 2$

Source of Variation	SS	df	MS	F
Between	$n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2$	1	SSB	$\frac{SS_B(n_1 + n_2 - 2)}{SS_W}$
Within	$(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2$	$\frac{n_1 - 1}{n_2 - 1}$	$\frac{SSW}{n_1 + n_2 - 2}$	
Total	SSB + SSW	$n - 1$		

F-test vs *t*-test When There Are Two Groups

- Example: Brand 3 ($\bar{x}_1 = 187.4$, $s_1^2 = 27.82$) vs. Brand 4 ($\bar{x}_2 = 191.2$, $s_2^2 = 30.84$)
- $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = -1.57$, $p = \mathbb{P}(|t| \geq 1.57) = 2 \times \mathbb{P}(t \leq -1.57) = 0.134$
- $F_{1,18} = \frac{\text{MSB}}{\text{MSW}} = 2.46$, $p = \mathbb{P}(F \geq 2.46) = 0.134$

Post hoc Tests

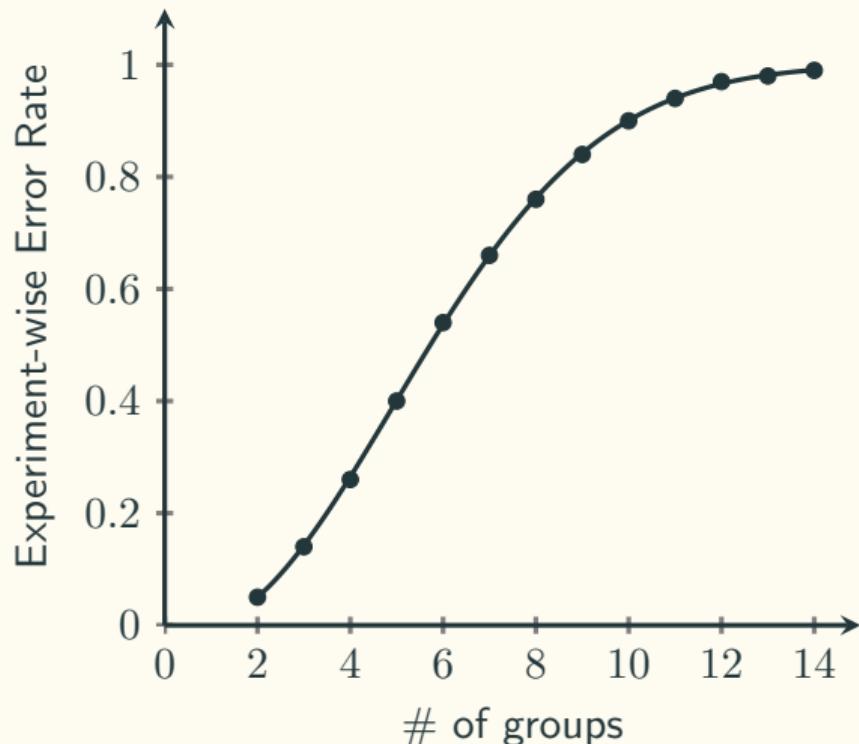
- ANOVA test tells me to reject $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, so what ?
- *Post hoc* tests - multiple pairwise comparisons. The following commonly-used tests have different ways of controlling type I error rate:
 - Bonferroni Procedure
 - Duncan's new multiple range test (MRT)
 - Dunn's Multiple Comparison Test
 - Fisher's Least Significant Difference (LSD)
 - Holm-Bonferroni Procedure
 - Newman-Keuls
 - Rodger's Method
 - Scheffé's Method
 - Tukey's Test (often used in classical ANOVA in stats software)
 - Dunnett's correction
 - Benjamini-Hochberg (BH) procedure

Post hoc Tests

Pairwise comparison $\alpha = 0.05$

# of groups	# of comparisons	Probability of making at least one type I error
2	1	0.05
3	3	0.14
4	6	0.26
5	10	0.4
6	15	0.54
7	21	0.66
8	28	0.76
9	36	0.84
10	45	0.9
11	55	0.94
12	66	0.97
13	78	0.98
14	91	0.99

$$1 - (1 - \alpha)^c$$



The Bonferroni Procedure

Pairwise comparison $\alpha = 0.05$: not good enough!

Goal: when doing many comparisons, we want the **overall error rate** to be α , meaning that the probability of making **at least one type I error** after performing **all** the comparisons is α .

$$1 - (1 - \alpha^*)^c = \alpha, \text{ where } c = \binom{k}{2}$$

Note, when α^* is small: $(1 - \alpha^*)^c \approx 1 - c\alpha^*$. We have:

$$1 - (1 - c\alpha^*) \approx \alpha \Rightarrow c\alpha^* \approx \alpha \Rightarrow \alpha^* \approx \frac{\alpha}{c} = \frac{\alpha}{\binom{k}{2}}$$

Bonferroni correction

Named after Carlo Emilio Bonferroni

The Bonferroni Procedure

To control the experiment-wise error rate to be α , we need to let the significance level α^* in each of the pairwise comparison to be α/c , where c is the # of comparison.

For each comparison,
if the $p < \alpha^*$,
then H_0 is rejected.



If $p < \alpha/c$, then
 H_0 is rejected.



If $p \times c < \alpha$, then
 H_0 is rejected.

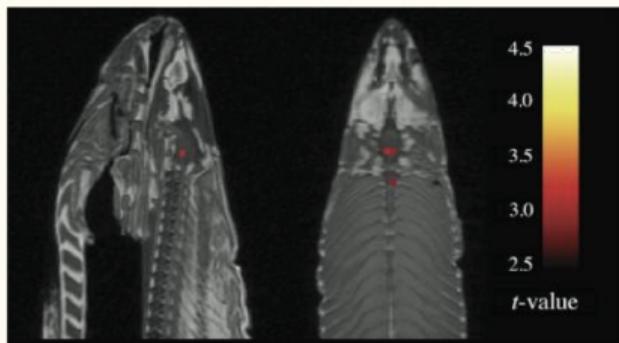
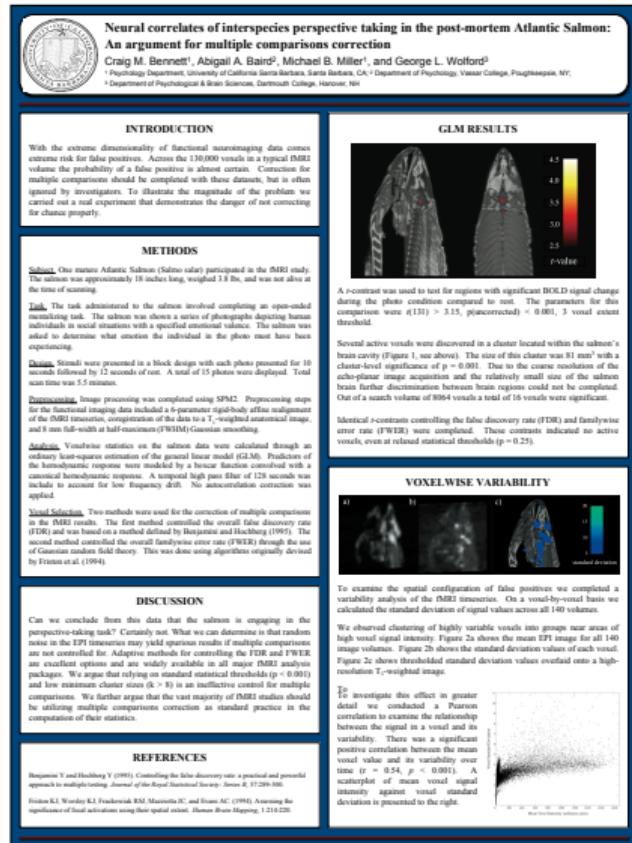
Corrected p -value

p_{adj}

$p.adj$

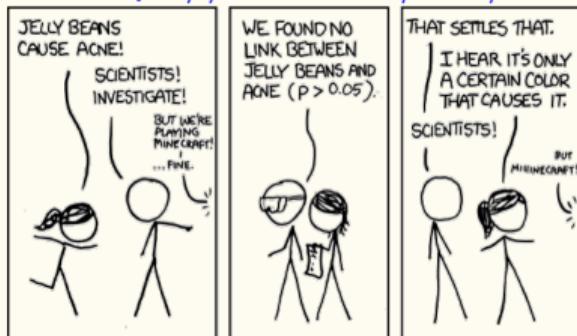
$p.adj = \min \left[p \times \binom{k}{2}, 1 \right]$, if $p.adj < \alpha$, then H_0 is rejected.

Multiple Comparisons - The Salmon Test



Multiple Comparisons - Significant

<https://xkcd.com/882/>



WE FOUND NO
LINK BETWEEN
JELLY BEANS AND
ACNE ($P > 0.05$).

THAT SETTLES THAT.
I HEAR IT'S ONLY
A CERTAIN COLOR
THAT CAUSES IT.

SCIENTISTS!

BUT MINECRAFT!

