Linear Combinations and Multiple Comparisons

CONTRASTS

(NOTE: SKIP THE ”LINEAR TRENDS” SUBSECTION)

Overview

- ANOVA provides a test for the equality of several means

- The main weaknesses are
  - it doesn't tell us **which** means are different
  - It doesn't account for any **structure** in the groups

    (Example: Is the average treatment effect across 3 levels of treatments different than the placebo?)

- The downside to this more refined analysis is that we need to control for the number of comparisons we end up making

- We need to develop additional techniques for making more specific comparisons

Example:
Handicap & Capability Study

- **Goal:** How do physical handicaps affect perception of employment qualification?
- The researchers prepared 5 recorded job interviews with same actors
- The tapes differed only in the handicap of the applicant:
  - No handicap
  - One leg amputated
  - Crutches
  - Hearing Impaired
  - Wheelchair

14 people were randomly assigned to each tape to rate applicants: 0-10 pts

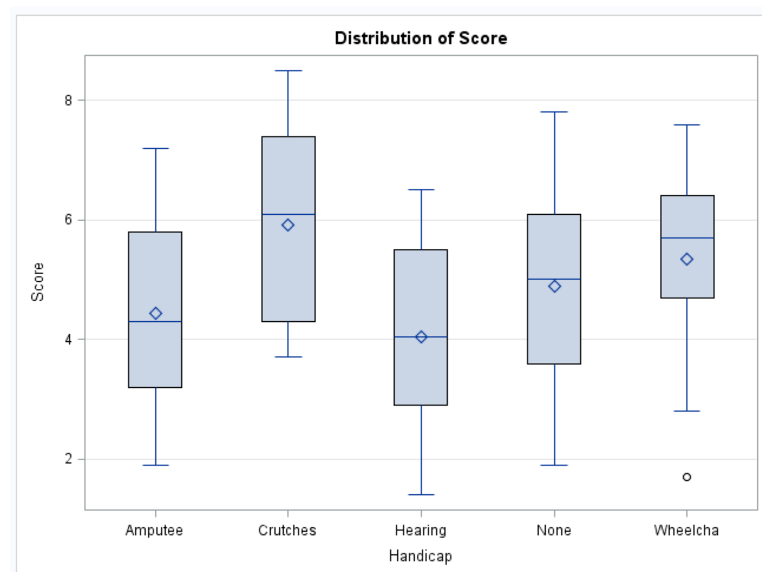
Example: Handicap & Capability Study

Do subjects systematically evaluate qualifications differently according to handicap?

If so, which handicaps are evaluated differently?

	None	Amputee	Crutches	Hearing	Wheelchair
0					
1	9	9		4	7
2	5	56		149	8
3	06	268	7	479	5
4	129	06	033	237	78
5	149	3589	18	589	03
6	17	1	0234	5	1124
7	48	2	445		246
8			5		
9					

Legend: 7 | 4 represents a score of 7.4 on the Applicant Qualification Scale.



Is There Any Difference at All?

We should begin any several group analysis by examining whether there is any evidence that at least any two groups are different

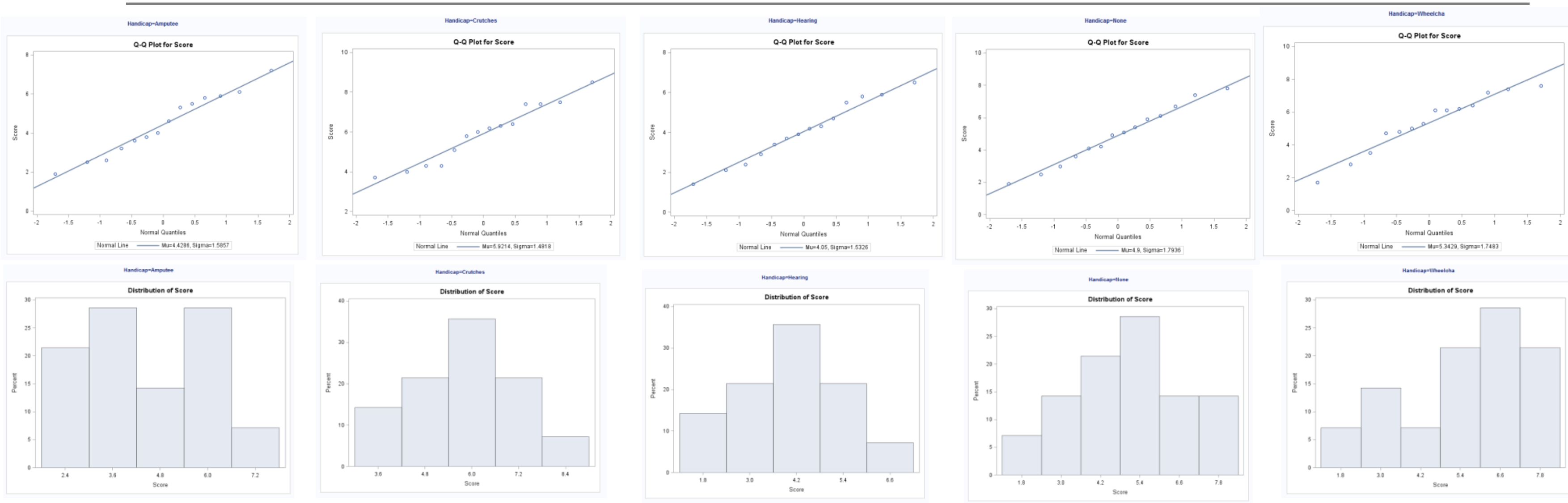
If there isn't any (statistically) significant difference in the groups, then there is no reason to address more refined questions

- The tapes differed only in the handicap of the applicant:
  - No handicap
  - One leg amputated
  - Crutches
  - Hearing Impaired
  - Wheelchair

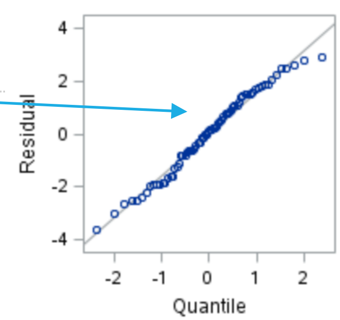
```
DATA handicap;
      INPUT score handicap $ @@;
      DATALINES;
1.9 None 2.5 None 3.0 None 3.6 None 4.1 None 4.2 None 4.9 None
5.1 None 5.4 None 5.9 None 6.1 None 6.7 None 7.4 None 7.8 None
1.9 Amp  2.5 Amp  2.6 Amp  3.2 Amp  3.6 Amp  3.8 Amp  4.0 Amp
4.6 Amp  5.3 Amp  5.5 Amp  5.8 Amp  5.9 Amp  6.1 Amp  7.2 Amp
3.7 Crut 4.0 Crut 4.3 Crut 4.3 Crut 5.1 Crut 5.8 Crut 6.0 Crut
6.2 Crut 6.3 Crut 6.4 Crut 7.4 Crut 7.4 Crut 7.5 Crut 8.5 Crut
1.4 Hear 2.1 Hear 2.4 Hear 2.9 Hear 3.4 Hear 3.7 Hear 3.9 Hear
4.2 Hear 4.3 Hear 4.7 Hear 5.5 Hear 5.8 Hear 5.9 Hear 6.5 Hear
1.7 Whee 2.8 Whee 3.5 Whee 4.7 Whee 4.8 Whee 5.0 Whee 5.3 Whee
6.1 Whee 6.1 Whee 6.2 Whee 6.4 Whee 7.2 Whee 7.4 Whee 7.6 Whee
;
```

ANOVA is a natural test for this task → need to check the assumptions

Handicap & Capability Study: Normality Assumption

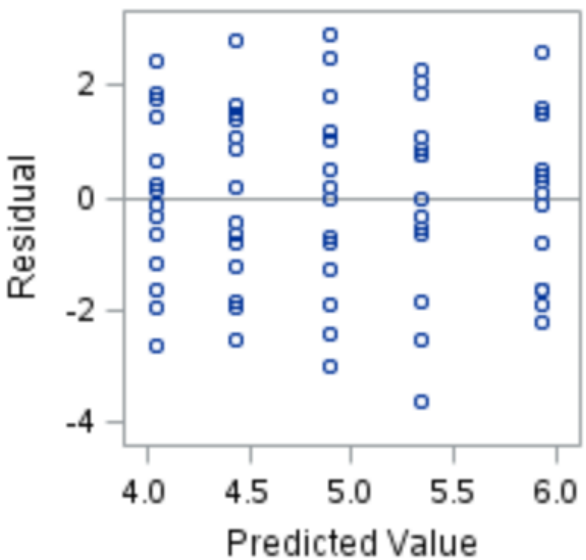



```
ods graphics on;
PROC GLM DATA = handicap ORDER=DATA PLOTS=diagnostics;
      CLASS handicap;
      MODEL score = handicap;
RUN;
ods graphics off;
```



There is no visual evidence to suggest that the data are not normally distributed

Handicap & Capability Study: Equal Variances Assumption




```
ods graphics on;
PROC GLM DATA = handicap ORDER=DATA PLOTS=diagnostics;
      CLASS handicap;
      MODEL score = handicap;
RUN;
ods graphics off;
```

Formal ANOVA model:

$i = 1, 2, \ldots, I$
$j = 1, 2, \ldots, n_i$

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$\overset{iid}{\sim}$ means independent, identically distributed

There is no evidence to suggest variances are unequal

Is There Any Difference at All?

Now that we have checked assumptions, let's run an ANOVA

If there isn't any (statistically) significant difference in the population means, then there is no reason to address more refined questions

- The tapes differed only in the handicap of the applicant:
  - No handicap (μ_{None})
  - One leg amputated (μ_{Amp})
  - Crutches (μ_{Crutch})
  - Hearing Impaired (μ_{Hear})
  - Wheelchair (μ_{Wheel})

$H_0: \mu_{None} = \mu_{Amp} = \mu_{Crutch} = \mu_{Hear} = \mu_{Wheel} = \mu$

$H_A: \mu_j \neq \mu_k$ for some j, k

Handicap & Capability Study: ANOVA results

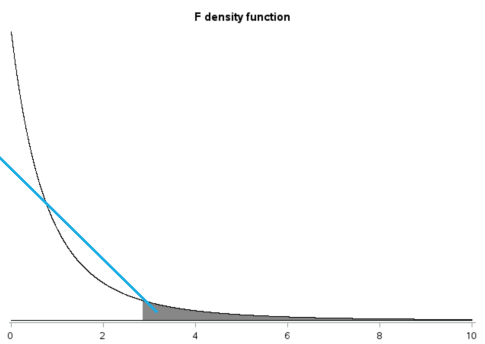
$H_0: \mu_{None} = \mu_{Amp} = \mu_{Crutch} = \mu_{Hear} = \mu_{Wheel} = \mu$

$H_A: \mu_j \neq \mu_k$ for some j, k

```
PROC GLM DATA = handicap;
    CLASS handicap;
    MODEL score = handicap;
RUN;
```

Extra Sums of Squares (ESS)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	30.5214286	7.6303571	2.86	0.0301
Error	65	173.3214286	2.6664835		
Corrected Total	69	203.8428571			



There is evidence to that there are at least two population means different (p-value of 0.0301 from a 1-way ANOVA).

Handicap & Capability Study: More Specific Questions

$$H_0: \frac{\mu_{Amp} + \mu_{Hear}}{2} = \frac{\mu_{Crutch} + \mu_{Wheel}}{2}$$
$$H_A: \frac{\mu_{Amp} + \mu_{Hear}}{2} \neq \frac{\mu_{Crutch} + \mu_{Wheel}}{2}$$

$$\updownarrow$$

$$H_0: \frac{\mu_{Amp} + \mu_{Hear}}{2} - \frac{\mu_{Crutch} + \mu_{Wheel}}{2} = 0$$
$$H_A: \frac{\mu_{Amp} + \mu_{Hear}}{2} - \frac{\mu_{Crutch} + \mu_{Wheel}}{2} \neq 0$$

$$\updownarrow$$

$$H_0: \mu_{Amp} + \mu_{Hear} - \mu_{Crutch} - \mu_{Wheel} = 0$$
$$H_A: \mu_{Amp} + \mu_{Hear} - \mu_{Crutch} - \mu_{Wheel} \neq 0$$

$$\updownarrow$$

(CONTRAST)

$$\gamma = 1\mu_{amp} - 1\mu_{Crutch} + 1\mu_{Hear} + 0\mu_{None} - 1\mu_{Wheel} \leftrightarrow$$

“gamma”

$$H_0: \gamma = 0$$
$$H_A: \gamma \neq 0$$

Level of Handicap	N	Score	
		Mean	Std Dev
Amputee	14	4.42857143	1.58571924
Crutche	14	5.92142857	1.48177574
Hearing	14	4.05000000	1.53259458
None	14	4.90000000	1.79357829
Wheelch	14	5.34285714	1.74828016

Linear Combinations & Contrasts

$$\gamma = C_1\mu_1 + C_2\mu_2 + \cdots + C_I\,\mu_I$$

(Constraint: $C_1 + C_2 + \cdots + C_I = 0$)

$$g = C_1\overline{Y}_1 + C_2\overline{Y}_2 + \cdots + C_I\overline{Y}_I.$$

$$\mathrm{SE}(g) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \cdots + \frac{C_I^2}{n_I}}.$$

(this requires independence)

Example: $\gamma = 1\mu_{Amp} - 1\mu_{Crutch} + 1\mu_{Hear} + 0\mu_{None} - 1\mu_{Wheel}$

The test statistic is t = (g- γ_0)/SE(g)

(γ_0 is the H_0 value, γ_0 = 0 in the current example)

(This has an approximate t-distribution w/ df = $n - I$)

(In this case, $n - I$ = 65)

Linear Combinations & Contrasts

$$\gamma = C_1\mu_1 + C_2\mu_2 + \cdots + C_I\mu_I$$

(Constraint: $C_1 + C_2 + \cdots + C_I = 0$)

$$g = C_1\overline{Y}_1 + C_2\overline{Y}_2 + \cdots + C_I\overline{Y}_I.$$

$$\mathrm{SE}(g) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \cdots + \frac{C_I^2}{n_I}}.$$

(this requires independence)

How do we get the pooled variance estimator?

```
PROC GLM DATA = handicap;
      CLASS handicap;
      MODEL score = handicap;
RUN;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	30.5214286	7.6303571	2.86	0.0301
Error	65	173.3214286	$\boxed{2.6664835}$		
Corrected Total	69	203.8428571			

Handicap & Capability Study: By Hand

H_0: $\mu_{Amp} + \mu_{Hear} = \mu_{Crutch} + \mu_{Wheel}$
H_A: $\mu_{Amp} + \mu_{Hear} \neq \mu_{Crutch} + \mu_{Wheel}$

$\gamma = 1\mu_{Amp} - 1\mu_{Crutch} + 1\mu_{Hear} + 0\mu_{None} - 1\mu_{Wheel}$
$g = 1\bar{Y}_{Amp} - 1\bar{Y}_{Crutch} + 1\bar{Y}_{Hear} + 0\bar{Y}_{None} - 1\bar{Y}_{Wheel}$
$g = (1)4.4 - (1)5.9 + (1)4.1 + (0)4.9 - (1)5.3 = -2.8$

$$SE(g) = \sqrt{2.666}\sqrt{\frac{1}{14} + \frac{1}{14} + \frac{1}{14} + \frac{0}{14} + \frac{1}{14}} = .873$$

Level of Handicap	N	Score Mean	Score Std Dev
Amputee	14	4.42857143	1.58571924
Crutche	14	5.92142857	1.48177574
Hearing	14	4.05000000	1.53259458
None	14	4.90000000	1.79357829
Wheelch	14	5.34285714	1.74828016

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	30.5214286	7.6303571	2.86	0.0301
Error	65	173.3214286	2.6664835		
Corrected Total	69	203.8428571			

R-Square	Coeff Var	Root MSE	Score Mean
0.149730	33.13206	$\boxed{1.632937}$	4.928571

$t_{65}(0.975) = 1.9971$

There is evidence that the sum of points assigned to Amp & Hear handicaps is smaller than the sum of points assigned to Crutch & Wheel handicaps

(95% t-tools CI of $-2.78577 \pm (1.9971)(0.87286)$ = (-4.529, -1.043)

Handicap & Capability Study: In SAS

```
DATA handicap;
      INPUT score handicap $ @@;
      DATALINES;
1.9 None 2.5 None 3.0 None 3.6 None 4.
5.1 None 5.4 None 5.9 None 6.1 None 6.
1.9 Amp  2.5 Amp  2.6 Amp  3.2 Amp  3.
4.6 Amp  5.3 Amp  5.5 Amp  5.8 Amp  5.
3.7 Crut 4.0 Crut 4.3 Crut 4.3 Crut 5.
6.2 Crut 6.3 Crut 6.4 Crut 7.4 Crut 7.
1.4 Hear 2.1 Hear 2.4 Hear 2.9 Hear 3.
4.2 Hear 4.3 Hear 4.7 Hear 5.5 Hear 5.
1.7 Whee 2.8 Whee 3.5 Whee 4.7 Whee 4.
6.1 Whee 6.1 Whee 6.2 Whee 6.4 Whee 7.
;
```

In SAS, “contrast” is only a test while “estimate” also gives an estimate

```
PROC GLM DATA = handicap ORDER=DATA;
     CLASS handicap;
     MODEL score = handicap;
     MEANS handicap;
     CONTRAST 'Avg. Amp & Hear vs Avg Crutch & Wheel' handicap 0 1 -1 1 -1;
     ESTIMATE 'Avg. Amp & Hear vs Avg Crutch & Wheel' handicap 0 1 -1 1 -1 / DIVISOR = 2;
     ESTIMATE 'Sum Amp & Hear vs Sum Crutch & Wheel' handicap 0 1 -1 1 -1;
RUN;
```

Handicap & Capability Study: In SAS

```
PROC GLM DATA = handicap ORDER=DATA;
     CLASS handicap;
     MODEL score = handicap;
     MEANS handicap;
     CONTRAST 'Avg. Amp & Hear vs Avg Crutch & Wheel' handicap 0 1 -1 1 -1;
     ESTIMATE 'Avg. Amp & Hear vs Avg Crutch & Wheel' handicap 0 1 -1 1 -1 / DIVISOR = 2;
     ESTIMATE 'Sum Amp & Hear vs Sum Crutch & Wheel' handicap 0 1 -1 1 -1;
RUN;
```

$$\gamma = 1\mu_{Amp} - 1\mu_{Crutch} + 1\mu_{Hear} + 0\mu_{None} - 1\mu_{Wheel}$$

$$\gamma = 0.5\mu_{Amp} - 0.5\mu_{Crutch} + 0.5\mu_{Hear} + 0\mu_{None} - 0.5\mu_{Wheel}$$

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Avg. Amp & Hear vs Avg Crutch & Wheel	1	27.16071429	27.16071429	10.19	0.0022

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Avg. Amp & Hear vs Avg Crutch & Wheel	-1.39285714	0.43642079	-3.19	0.0022
Sum Amp & Hear vs Sum Crutch & Wheel	-2.78571429	0.87284159	-3.19	0.0022

Handicap & Capability Study: In SAS

```
MODEL score = handicap / ALPHA=.01 CLPARM;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Avg. Amp & Hear vs Avg Crutch & Wheel	-1.39285714	0.43642079	-3.19	0.0022	-2.55094531	-0.23476898
Sum Amp & Hear vs Sum Crutch & Wheel	-2.78571429	0.87284159	-3.19	0.0022	-5.10189062	-0.46953795

There is evidence that the average points assigned to Amp & Hear handicaps is smaller than the average points assigned to Crutch & Wheel handicaps (pooled t-tools linear contrast p-value of 0.0022). We estimate that this difference is -1.39 pts with an associated 99% confidence interval of….

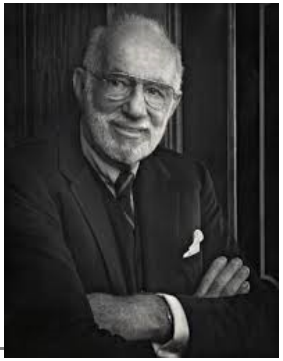
```
DATA quantile;
    quant = QUANTILE('t',0.995,70-5);
RUN;

PROC PRINT DATA = quantile;
RUN;
```

Obs	quant
1	2.65360

(-1.39-(2.65)*(0.436), -1.39+(2.65)*(0.436)) = (-2.5454 -0.2346)

Return to Spock Example

Spock Trial (Reminder)



- To test the claim, the Spock Judge's (which we will call S) recent venires are compared with 6 other Judge's recent venires (which we notate A to F)

- There are two key questions
  1. **Is there evidence that women are unrepresented on S's venire relative to A to F's?**
  2. Is there evidence of a difference in women's representation on A to F's venire

Framing the Question Statistically

Is there evidence that women are unrepresented on S's venire relative to A-F?

This question can be quantified in a few ways:

1. Does the mean % of females of the Spock Judge's venires equal the mean % for the A-F judges?

$$H_0: \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$$
$$H_A: \mu_S\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$$

2. Does the mean % of females of the Spock Judge's venires equal the average of the mean % for the A-F judges?

$$H_0:\ \mu_S - \frac{\mu_A + \mu_B + \mu_C + \mu_D + \mu_E + \mu_F}{6} = 0$$
$$H_A:\ \mu_S - \frac{\mu_A + \mu_B + \mu_C + \mu_D + \mu_E + \mu_F}{6} \neq 0$$

Framing the Question Statistically

If we proceed under the assumption that the mean % of women for judges A-F are equal, we can test whether the Spock judge has a mean % different than the other judges by testing:

$H_0: \mu \mu \mu \mu \mu \mu \mu$

$H_A: \mu_S \mu \mu \mu \mu \mu \mu$

```
PROC GLM DATA = spock2;
    CLASS OthersModel;
    MODEL percFemale = OthersModel;
RUN;
```

Obs	percFemale	judge	OthersModel
1	6.4	S	S
2	8.7	S	S
3	13.3	S	S
4	13.6	S	S
5	15.0	S	S
6	15.2	S	S
7	17.7	S	S
8	[illegible]	S	S
9	23.1	S	S
[illegible]	[illegible]	[illegible]	Others
[illegible]	[illegible]	[illegible]	Others
[illegible]	[illegible]	[illegible]	Others
[illegible]	[illegible]	[illegible]	Others
14	[illegible]	A	Others
15	27.0	B	Others
16	28.9	B	Others
[illegible]	[illegible]	[illegible]	Others
[illegible]	[illegible]	B	Others
[illegible]	[illegible]	[illegible]	Others
20	45.6	B	Others
21	21.0	C	Others
22	23.4	C	Others
[illegible]	[illegible]	[illegible]	Others
[illegible]	[illegible]	[illegible]	[illegible]

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1600.622964	1600.622964	32.15	<.0001
Error	44	2190.903123	49.793253		
Corrected Total	45	3791.526087			

There is strong evidence to support the claim that the mean % of women in the Spock judge's venires does not equal the mean % for the other 6 judges (one-way ANOVA p-value <0.0001) if the other judges are assumed to have the same mean

Framing the Question Statistically

$$H_0\colon \mu_S - \frac{\mu_A + \mu_B + \mu_C + \mu_D + \mu_E + \mu_F}{6} = 0$$

$$H_A\colon \mu_S - \frac{\mu_A + \mu_B + \mu_C + \mu_D + \mu_E + \mu_F}{6} \neq 0$$

```
PROC GLM DATA = spock ORDER=DATA;
      CLASS judge;
      MODEL percFemale = judge;
      CONTRAST 'Contrast Spock judge to A-F judge' judge -6 1 1 1 1 1 1;
      ESTIMATE 'Estimate Spock judge to A-F judge' judge -6 1 1 1 1 1 1 / DIVISOR = 6;
RUN;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Contrast Spock judge to A-F judge	1	1536.777000	1536.777000	32.15	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Estimate Spock judge to A-F judge	14.9783333	2.64180389	5.67	<.0001	9.6347806	20.3218861

Compare to:

$H_0\colon \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$

$H_A\colon \mu_S\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1600.622964	1600.622964	32.15	<.0001
Error	44	2190.903123	49.793253		
Corrected Total	45	3791.526087			

The Pooled Standard Deviation Estimator

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1600.622964	1600.622964	32.15	<.0001
Error	44	2190.903123	49.793253		
Corrected Total	45	3791.526087			

$H_0: \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$
$H_A: \mu_S\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1927.080865	321.180144	6.72	<.0001
Error	39	1864.445222	47.806288		
Corrected Total	45	3791.526087			

$H_0: \mu\ \mu\ \mu\ \mu\ \mu\ \mu\ \mu$
$H_A: \mu_S\ \mu_A\ \mu_B\ \mu_C\ \mu_D\ \mu_E\ \mu_F$

The pooled standard deviation estimator is from the largest possible full model:

$$s_p=\sqrt{47.806288} = 6.91421$$

Spock Contrast: 95% Confidence Interval

Level of Judge	N	Percent	
		Mean	Std Dev
A	5	34.1200000	11.9418173
B	6	33.6166667	6.5822235
C	9	29.1000000	4.5929293
D	2	27.0000000	3.8183766
E	6	26.9666667	9.0101424
F	9	26.8000000	5.9688776
Spock's	9	14.6222222	5.0387939

$$\gamma = \mu_S - \frac{\mu_A + \mu_B + \mu_C + \mu_D + \mu_E + \mu_F}{6}$$

$$g = \bar{Y}_S - \frac{\bar{Y}_A + \bar{Y}_B + \bar{Y}_C + \bar{Y}_D + \bar{Y}_E + \bar{Y}_F}{6}$$

$$= 14.6222 – (34.12 + 33.6167 + 29.1 + 27.0 + 26.967 + 26.8)/6 = \text{-}14.98$$

$$\text{SE(g)} = 6.91421 * \sqrt{\frac{1}{9} + \left(\frac{1}{36}\right)\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{9} + \frac{1}{2} + \frac{1}{6} + \frac{1}{9}\right)} = 2.641804$$

Margin of Error: 5.275947 $= 1.9971 * 2.641804$ $\color{red}{\rightarrow}$ (-14.98 - 5.28, -14.98 + 5.28)

$(t_{0.025,65} = 1.9971)$

There is strong evidence that the mean % of women from the Spock judges' venires is less that that of the average of the A-F judges (95% CI of (-20.25, -9.7) percentage points).

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Estimate Spock judge to A-F judge	14.9783333	2.64180389	5.67	<.0001	9.6347806	20.3218861