

Set R07 - Contrasts

STAT 401 (Engineering) - Iowa State University

April 5, 2017

Simple hypothesis

Consider the one-way ANOVA model: $Y_{ij} \sim N(\mu_j, \sigma^2)$ where $j = 1, \dots, J$.

Here are a few simple alternative hypotheses:

1. Mean lifetimes for N/R50 and R/R50 diet are different.
2. Mean lifetimes for N/R40 is different than for N/R50 and R/R50 combined.
3. Mean lifetimes for high calorie (NP and N/N85) diets is different than for low calorie diets combined.

$$H_0 : \gamma = 0 \quad H_1 : \gamma \neq 0 :$$

$$\gamma_1 = \mu_{R/R50} - \mu_{N/R50}$$

$$\gamma_2 = \mu_{N/R40} - \frac{1}{2}(\mu_{N/R50} + \mu_{R/R50})$$

$$\gamma_3 = \frac{1}{4}(\mu_{N/R50} + \mu_{R/R50} + \mu_{N/R40} + \mu_{lopro}) - \frac{1}{2}(\mu_{NP} + \mu_{N/N85})$$

Contrasts

Definition

A **linear combination** of group means has the form

$$\gamma = C_1\mu_1 + C_2\mu_2 + \dots + C_J\mu_J$$

where C_j are known coefficients and μ_j are the unknown population means.

Definition

A linear combination with $C_1 + C_2 + \dots + C_J = 0$ is a **contrast**.

Remark Contrast interpretation is usually best if

$|C_1| + |C_2| + \dots + |C_J| = 2$, i.e. the positive coefficients sum to 1 and the negative coefficients sum to -1.

Inference on contrasts

$$\gamma = C_1\mu_1 + C_2\mu_2 + \cdots + C_J\mu_J$$

Estimated by

$$g = C_1\bar{Y}_1 + C_2\bar{Y}_2 + \cdots + C_J\bar{Y}_J$$

with standard error

$$SE(g) = s_p \sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \cdots + \frac{C_J^2}{n_J}}$$

t-statistic (compare to t_{n-J}) and CI:

$$t = \frac{g}{SE(g)} \quad g \pm t_{n-J, 1-\alpha/2} SE(g)$$

Contrasts for mice lifetime dataset

For these contrasts:

1. Mean lifetimes for N/R50 and R/R50 diet are different.
2. Mean lifetimes for N/R40 is different than for N/R50 and R/R50 combined.
3. Mean lifetimes for high calorie (NP and N/N85) diets is different than for low calorie diets combined.

$$H_0 : \gamma = 0 \quad H_1 : \gamma \neq 0 :$$

$$\gamma_1 = \mu_{R/R50} - \mu_{N/R50}$$

$$\gamma_2 = \mu_{N/R40} - \frac{1}{2}(\mu_{N/R50} + \mu_{R/R50})$$

$$\gamma_3 = \frac{1}{4}(\mu_{N/R50} + \mu_{R/R50} + \mu_{N/R40} + \mu_{lopro}) - \frac{1}{2}(\mu_{NP} + \mu_{N/N85})$$

	N/N85	N/R40	N/R50	NP	R/R50	lopro
early rest - none @ 50kcal	0.00	0.00	-1.00	0.00	1.00	0.00
40kcal/week - 50kcal/week	0.00	1.00	-0.50	0.00	-0.50	0.00
lo cal - hi cal	-0.50	0.25	0.25	-0.50	0.25	0.25

Mice lifetime examples

	Diet	n	mean	sd
1	N/N85	57	32.69	5.13
2	N/R40	60	45.12	6.70
3	N/R50	71	42.30	7.77
4	NP	49	27.40	6.13
5	R/R50	56	42.89	6.68
6	lopro	56	39.69	6.99

Contrasts:

	g	SE(g)	t	p	L	U
early rest - none @ 50kcal	0.59	1.19	0.49	0.62	-1.76	2.94
40kcal/week - 50kcal/week	2.53	1.05	2.41	0.02	0.46	4.59
lo cal - hi cal	12.45	0.78	15.96	0.00	10.92	13.98

```
library(multcomp)
m = lm(Lifetime~Diet-1, case0501) # The -1 indicates no intercept (see Ch 7)
summary(m)
```

Call:

```
lm(formula = Lifetime ~ Diet - 1, data = case0501)
```

Residuals:

Min	1Q	Median	3Q	Max
-25.5167	-3.3857	0.8143	5.1833	10.0143

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
DietN/N85	32.6912	0.8846	36.96	<2e-16 ***
DietN/R40	45.1167	0.8622	52.33	<2e-16 ***
DietN/R50	42.2972	0.7926	53.37	<2e-16 ***
DietNP	27.4020	0.9540	28.72	<2e-16 ***
DietR/R50	42.8857	0.8924	48.06	<2e-16 ***
Dietlopro	39.6857	0.8924	44.47	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.678 on 343 degrees of freedom

Multiple R-squared: 0.9724, Adjusted R-squared: 0.9719

F-statistic: 2011 on 6 and 343 DF, p-value: < 2.2e-16

K

	N/N85	N/R40	N/R50	NP	R/R50	lopro
early rest - none @ 50kcal	0.0	0.00	-1.00	0.0	1.00	0.00
40kcal/week - 50kcal/week	0.0	1.00	-0.50	0.0	-0.50	0.00
lo cal - hi cal	-0.5	0.25	0.25	-0.5	0.25	0.25

```
t = glht(m, linfct=K)
summary(t)
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lm(formula = Lifetime ~ Diet - 1, data = case0501)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
early rest - none @ 50kcal == 0	0.5885	1.1936	0.493	0.9457
40kcal/week - 50kcal/week == 0	2.5252	1.0485	2.408	0.0488 *
lo cal - hi cal == 0	12.4497	0.7800	15.961	<1e-04 ***

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```
confint(t, calpha=univariate_calpha())
```

Simultaneous Confidence Intervals

```
Fit: lm(formula = Lifetime ~ Diet - 1, data = case0501)
```

```
Quantile = 1.9669
```

```
95% confidence level
```

Linear Hypotheses:

	Estimate	lwr	upr
early rest - none @ 50kcal == 0	0.5885	-1.7591	2.9361
40kcal/week - 50kcal/week == 0	2.5252	0.4628	4.5876
lo cal - hi cal == 0	12.4497	10.9155	13.9839

Summary

- Contrasts are linear combinations of means where the coefficients sum to zero
- t-test tools are used to calculate pvalues and confidence intervals

Sulfur effect on scab disease in potatoes

The experiment was conducted to investigate the effect of sulfur on controlling scab disease in potatoes. There were seven treatments: control, plus spring and fall application of 300, 600, 1200 lbs/acre of sulfur. The response variable was percentage of the potato surface area covered with scab averaged over 100 random selected potatoes. A completely randomized design was used with 8 replications of the control and 4 replications of the other treatments.

Cochran and Cox. (1957) Experimental Design (2nd ed). pg96 and Agron. J. 80:712-718 (1988)

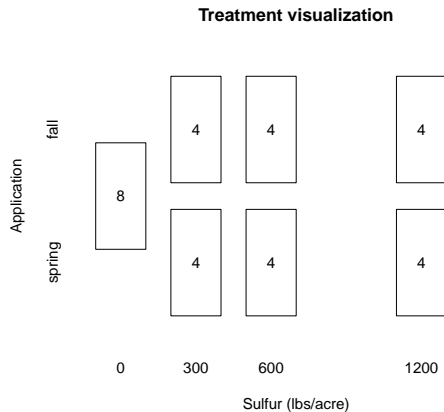
Scientific question:

- Does sulfur have any impact at all?
- Is there a difference between spring and fall?
- Is there an effect of increased sulfur (expect more sulfur causes less scab)?

Data

	inf	trt	row	col
1	9	F3	4	1
2	12	0	4	2
3	18	S6	4	3
4	10	F12	4	4
5	24	S6	4	5
6	17	S12	4	6
7	30	S3	4	7
8	16	F6	4	8
9	10	0	3	1
10	7	S3	3	2
11	4	F12	3	3
12	10	F6	3	4
13	21	S3	3	5
14	24	0	3	6
15	29	0	3	7
16	12	S6	3	8
17	9	F3	2	1
18	7	S12	2	2
19	18	F6	2	3
20	30	0	2	4
21	18	F6	2	5
22	16	S12	2	6
23	16	F3	2	7
24	4	F12	2	8
25	9	S3	1	1
26	18	0	1	2
27	17	S12	1	3
28	19	S6	1	4
29	32	0	1	5
30	5	F12	1	6

Design



Design

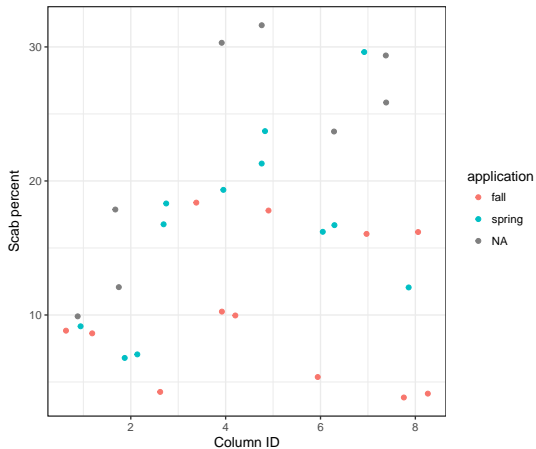
**Completely randomized design
potato scab experiment**

row	4	F3	O	S6	F12	S6	S12	S3	F6
	3	O	S3	F12	F6	S3	O	O	S6
	2	F3	S12	F6	O	F6	S12	F3	F12
	1	S3	O	S12	S6	O	F12	O	F3
		1	2	3	4	5	6	7	8
		col							

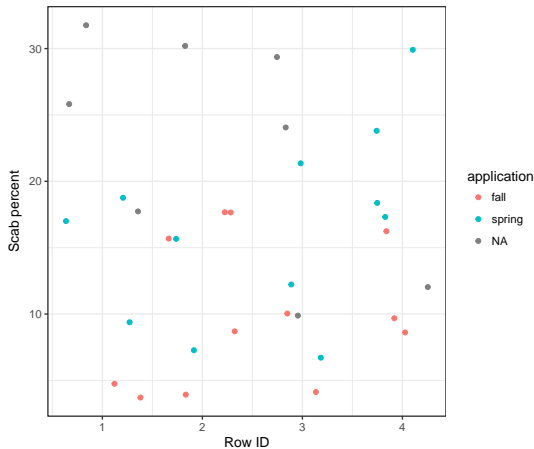
Data

```
Error in title(xlab = "Sulfur", ylab = "Scab percent"): plot.new has not been called yet
```

Data



Data



Model

Y_{ij} : avg % of surface area covered with scab for plot i in treatment j for $j = 1, \dots, 7$.

Assume $Y_{ij} \stackrel{ind}{\sim} N(\mu_j, \sigma^2)$.

Hypotheses:

- Difference amongst any means: One-way ANOVA F-test
- *Any effect*: Control vs sulfur
- *Fall vs spring*: Contrast comparing fall vs spring applications
- *Sulfur level*: Linear trend contrast

Control vs sulfur

$$\begin{aligned}\gamma &= \frac{1}{6}(\mu_{F12} + \mu_{F6} + \mu_{F3} + \mu_{S3} + \mu_{S6} + \mu_{S12}) - \mu_O \\ &= \frac{1}{6}(\mu_{F12} + \mu_{F6} + \mu_{F3} + \mu_{S3} + \mu_{S6} + \mu_{S12} - 6\mu_O)\end{aligned}$$

Fall vs spring contrast

- *Fall vs spring*: Contrast comparing fall vs spring applications

$$\begin{aligned}\gamma &= \frac{1}{3}(\mu_{F12} + \mu_{F6} + \mu_{F3}) + 0\mu_O - \frac{1}{3}(\mu_{S3} + \mu_{S6} + \mu_{S12}) \\ &= \frac{1}{3}\mu_{F12} + \frac{1}{3}\mu_{F6} + \frac{1}{3}\mu_{F3} + 0\mu_O - \frac{1}{3}\mu_{S3} - \frac{1}{3}\mu_{S6} - \frac{1}{3}\mu_{S12} \\ &= \frac{1}{3} [\mu_{F12} + \mu_{F6} + \mu_{F3} + 0\mu_O - 1\mu_{S3} - 1\mu_{S6} - 1\mu_{S12}]\end{aligned}$$

Sulfur level: linear trend contrasts

- The group sulfur levels (X_i) are 12, 6, 3, 0, 3, 6, and 12 (100lbs/acre).
- So the linear trend contrast ($X_i - \bar{X}$) is

X_i	12	6	3	0	3	6	12
$X_i - \bar{X}$	6	0	-3	-6	-3	0	6

$$\gamma = 6\mu_{F12} + 0\mu_{F6} - 3\mu_{F3} - 6\mu_O - 3\mu_{S3} + 0\mu_{S6} + 6\mu_{S12}$$

Contrasts

Trt	F12	F6	F3	O	S3	S6	S12	Div
Sulfur v control	1	1	1	-6	1	1	1	6
Fall v Spring	1	1	1	0	-1	-1	-1	3
Linear Trend	-6	0	-3	-6	-3	0	6	1

```
library(multcomp)
K = rbind("sulfur - control" = c(1, 1, 1,-6, 1, 1, 1)/6,
          "fall - spring"    = c(1, 1, 1, 0,-1,-1,-1)/3,
          "linear trend"     = c(6, 0,-3,-6,-3, 0, 6)/1)
m = lm(inf~trt,d)
anova(m)
```

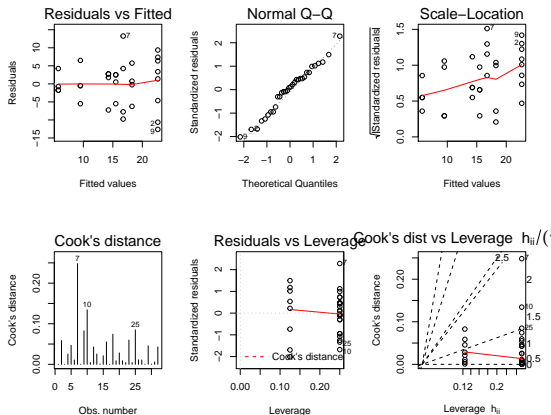
Analysis of Variance Table

Response: inf

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt	6	972.34	162.057	3.6081	0.01026 *
Residuals	25	1122.88	44.915		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
par(mfrow=c(2,3))
plot(m,1:6)
```



```
g = glht(lm(inf~trt-1,d), linfct=K) # notice the -1 in the model
summary(g, test=adjusted(type="none")) # unadjusted pvalues
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lm(formula = inf ~ trt - 1, data = d)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
sulfur - control == 0	-9.292	2.736	-3.396	0.00229	**
fall - spring == 0	-6.167	2.736	-2.254	0.03322	*
linear trend == 0	-94.500	34.824	-2.714	0.01188	*

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- none method)
```

```
confint(g, calpha=univariate_calpha()) # unadjusted confidence intervals
```

Simultaneous Confidence Intervals

```
Fit: lm(formula = inf ~ trt - 1, data = d)
```

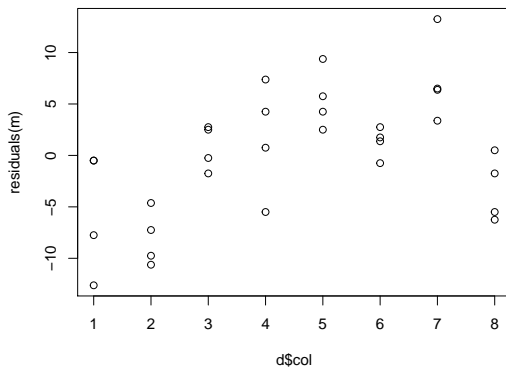
```
Quantile = 2.0595
```

```
95% confidence level
```

Linear Hypotheses:

	Estimate	lwr	upr
sulfur - control == 0	-9.2917	-14.9266	-3.6567
fall - spring == 0	-6.1667	-11.8016	-0.5317
linear trend == 0	-94.5000	-166.2212	-22.7788


```
plot(d$col,residuals(m))
```



Summary

For this particular data analysis

- Significant differences in means between the groups (ANOVA $F_{6,25} = 3.61$ $p=0.01$)
- Having sulfur was associated with a reduced scab % of 9 (4,15) compared to no sulfur
- Fall application reduced scab % by 6 (0.5,12) compared to spring application
- Linear trend in sulfur was significant ($p=0.01$)
- Concerned about spatial correlation among columns
- Consider a transformation of the response
 - CI for F12 (-1.2, 12.7) (not shown)
 - Non-constant variance (residuals vs predicted, sulfur, application)