Set R07 - Contrasts

STAT 401 (Engineering) - Iowa State University

April 21, 2017

Simple hypothesis

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

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.

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

$$
\begin{aligned}
\gamma_1 &= \mu_{R/R50} - \mu_{N/R50} \\
\gamma_2 &= \mu_{N/R40} - \tfrac{1}{2}(\mu_{N/R50} + \mu_{R/R50}) \\
\gamma_3 &= \tfrac{1}{4}(\mu_{N/R50} + \mu_{R/R50} + \mu_{N/R40} + \mu_{lopro}) - \tfrac{1}{2}(\mu_{NP} + \mu_{N/N85})
\end{aligned}
$$

Contrasts

Definition

A linear combination of group means has the form

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

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

Definition

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

Remark Contrast interpretation is usually best if $|C_1| + |C_2| + \cdots + |C_J| = 2$, 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

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

with standard error

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

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

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

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.

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

$$\begin{aligned}
\gamma_1 &= \mu_{R/R50} - \mu_{N/R50} \\
\gamma_2 &= \mu_{N/R40} - \frac{1}{2}(\mu_{N/R50} + \mu_{R/R50}) \\
\gamma_3 &= \frac{1}{4}(\mu_{N/R50} + \mu_{R/R50} + \mu_{N/R40} + \mu_{lopro}) - \frac{1}{2}(\mu_{NP} + \mu_{N/N85})
\end{aligned}$$

	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

```
m = lm(Lifetime~Diet, data = Sleuth3::case0501)
summary(m)


Call:
lm(formula = Lifetime ~ Diet, data = Sleuth3::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|)
(Intercept)  32.6912     0.8846  36.958  < 2e-16 ***
DietN/R40    12.4254     1.2352  10.059  < 2e-16 ***
DietN/R50     9.6060     1.1877   8.088 1.06e-14 ***
DietNP       -5.2892     1.3010  -4.065 5.95e-05 ***
DietR/R50    10.1945     1.2565   8.113 8.88e-15 ***
Dietlopro     6.9945     1.2565   5.567 5.25e-08 ***
---
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:  0.4543,Adjusted R-squared:  0.4463
F-statistic:  57.1 on 5 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
```

```
library("lsmeans")
ls = lsmeans(m, ~ Diet)
ls


 Diet     lsmean        SE  df lower.CL upper.CL
 N/N85 32.69123 0.8845544 343 30.95139 34.43106
 N/R40 45.11667 0.8621570 343 43.42089 46.81245
 N/R50 42.29718 0.7925612 343 40.73829 43.85608
 NP    27.40204 0.9540342 343 25.52555 29.27853
 R/R50 42.88571 0.8924172 343 41.13041 44.64101
 lopro 39.68571 0.8924172 343 37.93041 41.44101

Confidence level used: 0.95


co = contrast(ls,
              #                                    N/N85 N/R40 N/R50  NP R/R50 lopro
              list("early rest - none @ 50kcal"=c(    0,    0,   -1,  0,    1,    0),
                   "40kcal/week - 50kcal/week" =c(    0,    2,   -1,  0,   -1,    0) / 2,
                   "lo cal - hi cal"           =c(   -2,    1,    1, -2,    1,    1) / 4))
confint(co)


 contrast                     estimate        SE  df   lower.CL  upper.CL
 early rest - none @ 50kcal  0.5885312 1.1935501 343 -1.7590676  2.936130
 40kcal/week - 50kcal/week   2.5252180 1.0485490 343  0.4628224  4.587614
 lo cal - hi cal            12.4496851 0.7800142 343 10.9154718 13.983899

Confidence level used: 0.95
```

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)

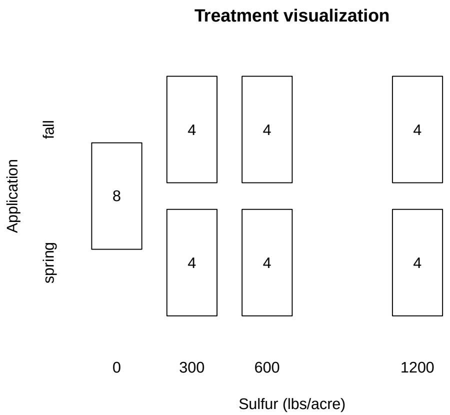
Scientific question:

- 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   O   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   O   3   1
10   7  S3   3   2
11   4 F12   3   3
12  10  F6   3   4
13  21  S3   3   5
14  24   O   3   6
15  29   O   3   7
16  12  S6   3   8
17   9  F3   2   1
18   7 S12   2   2
19  18  F6   2   3
20  30   O   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   O   1   2
27  17 S12   1   3
28  19  S6   1   4
29  32   O   1   5
30   5 F12   1   6
```

Design

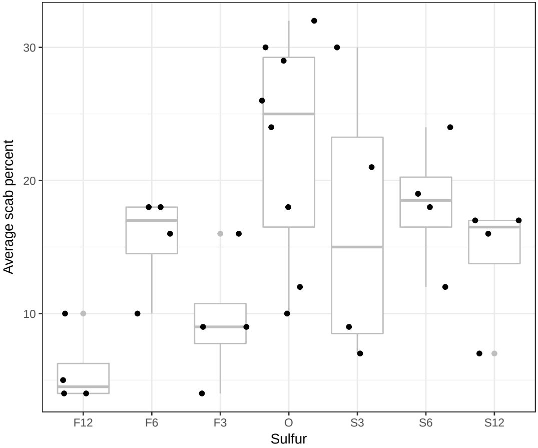

Treatment visualization
Application
fall
spring
8
4
4
4
4
4
4
0
300
600
1200
Sulfur (lbs/acre)

Design

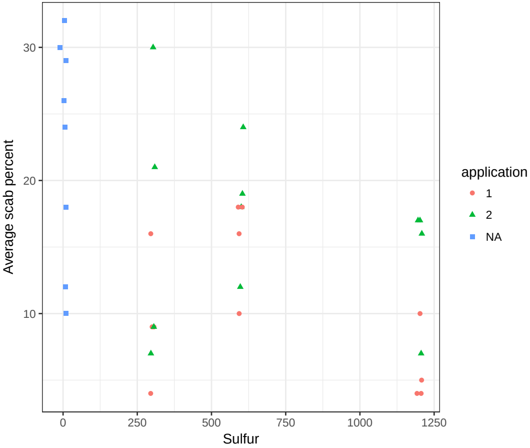
Completely randomized design
potato scab experiment

row \ col	1	2	3	4	5	6	7	8
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

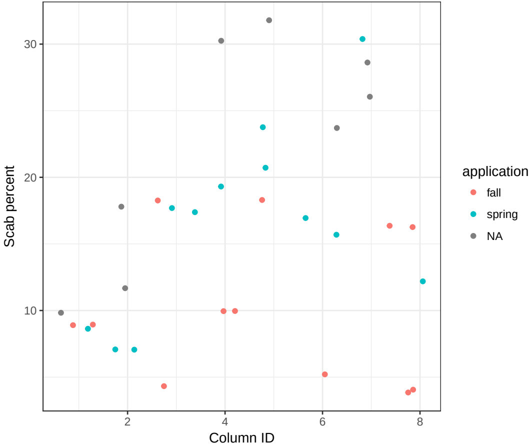
Data



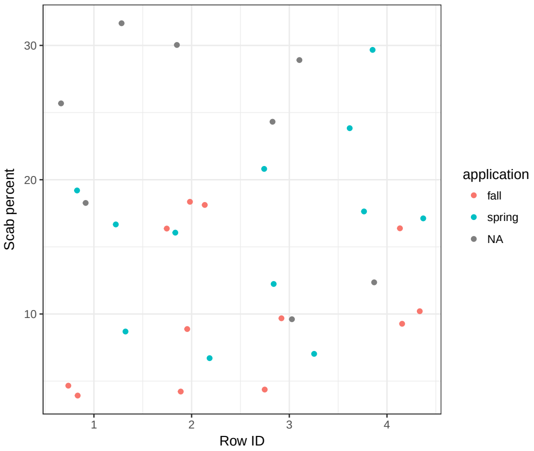
Data


Average scab percent
30
20
10
0
250
500
750
1000
1250
Sulfur
application
1
2
NA

Data


Scab percent
30
20
10
2
4
6
8
Column ID
application
fall
spring
NA

Data



Model

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

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

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

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

Sulfur level: linear trend contrasts

- The group sulfur levels (X_j) are 12, 6, 3, 0, 3, 6, and 12 (100 lbs/acre)
- and a linear trend contrast is $X_j - \overline{X}$

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

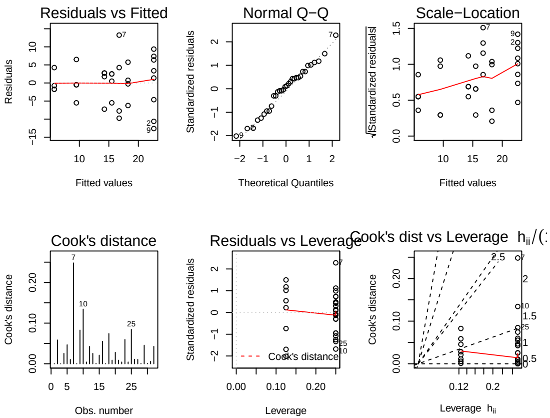
$$\gamma \quad = 6\mu_{F12} + 0\mu_{F6} - 3\mu_{F3} - 6\mu_{O} - 3\mu_{S3} + 0\mu_{S6} + 6\mu_{S12}$$

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

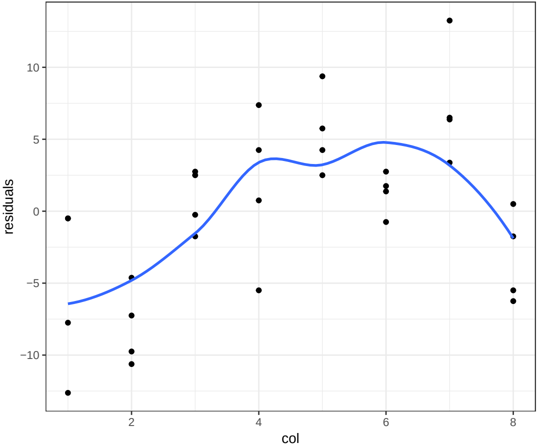
```
K = rbind("sulfur - control" = c(1, 1, 1,-6, 1, 1, 1)/6,
          "fall - spring"    = c(1, 1, 1, 0,-1,-1,-1)/3,
          "linear trend"     = c(6, 0,-3,-6,-3, 0, 6)/1)
m = lm(inf~trt,d)
anova(m)


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)
```



```
ls <- lsmeans(m, ~trt)
ls


 trt lsmean       SE df  lower.CL upper.CL
 F12  5.750 3.350933 25 -1.151375 12.65138
 F6  15.500 3.350933 25  8.598625 22.40138
 F3   9.500 3.350933 25  2.598625 16.40138
 0   22.625 2.369467 25 17.744991 27.50501
 S3  16.750 3.350933 25  9.848625 23.65138
 S6  18.250 3.350933 25 11.348625 25.15138
 S12 14.250 3.350933 25  7.348625 21.15138

Confidence level used: 0.95


co <- contrast(ls,
#                                        F12 F6 F3  0 S3 S6 S12
               list("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))
confint(co)


 contrast           estimate        SE df   lower.CL    upper.CL
 sulfur - control  -9.291667  2.736025 25  -14.92662  -3.6567175
 fall - spring     -6.166667  2.736025 25  -11.80162  -0.5317175
 linear trend     -94.500000 34.823914 25 -166.22119 -22.7788062

Confidence level used: 0.95
```

```r
d$residuals <- residuals(m)
ggplot(d, aes(col, residuals)) + geom_point() + stat_smooth(se=FALSE) + theme_bw()
```



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 reducted 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)