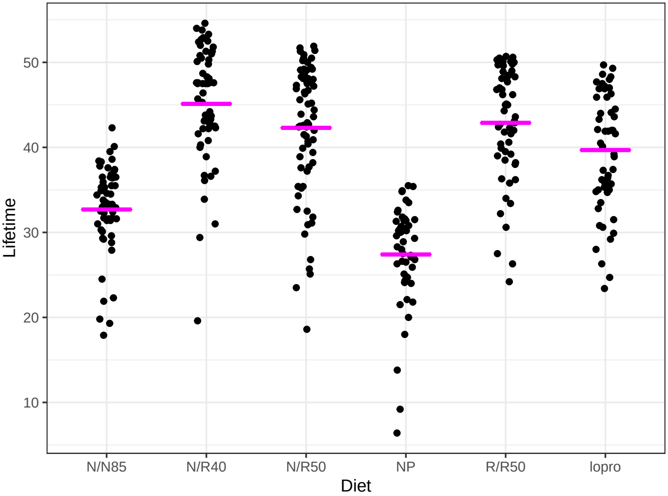
R07 - Contrasts

STAT 587 (Engineering)
Iowa State University

November 9, 2020

Diet Effect on Mice Lifetimes



ANOVA and Regression Models

ANOVA model:

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

with Y_{ij} being the lifetime for the ith mouse on the jth diet for $j = 0, 1, 2, 3, 4, 5$.

Regression model:

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_{i,1} + \ldots + \beta_p X_{i,p}, \sigma^2)$$

where Y_i is the lifetime for the ith mouse and $X_{i,j}$ is an indicator for the ith mouse being on the jth diet.

Reparameterized model since

$$\mu_0 = \beta_0 \quad \text{and} \quad \mu_j = \beta_0 + \beta_j$$

for $j > 0$.

Scientific questions

Here are a few example scientific questions:

1. What is the effect of pre-wean calorie restriction on mean lifetimes?
2. What is the difference in mean lifetimes between mice on a 40 kcal diet compared to those on a 50 kcal diet?
3. What is the effect of high calorie vs low calorie diets on mean lifetimes?

We can compute contrasts:

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

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

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

Contrasts

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.

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

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

Contrast

$$\gamma = C_1\mu_1 + C_2\mu_2 + \cdots + C_J\mu_J \quad \text{with} \quad \hat{\gamma} = C_1\overline{Y}_1 + C_2\overline{Y}_2 + \cdots + C_J\overline{Y}_J$$

with standard error

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

p-values for $H_0 : \gamma = g_0$ vs $H_A : \gamma \neq g_0$ and posterior probabilities (i.e. $2P(\gamma > 0|y)$ or $2P(\gamma < 0|y)$):

$$t = \frac{g - g_0}{SE(g)}, \quad p = 2P(T_{n-J} < -|t|).$$

Two-sided equal-tail $100(1-\alpha)\%$ confidence/credible intervals:

$$g \pm t_{n-J,1-\alpha/2} SE(g).$$

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.

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

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

	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

Fit the Multiple Regression Model

```
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
```

Estimate Group Means

```
library("emmeans")
em = emmeans(m, ~ Diet)
em

 Diet   emmean    SE  df lower.CL upper.CL
 N/N85    32.7 0.885 343     31.0     34.4
 N/R40    45.1 0.862 343     43.4     46.8
 N/R50    42.3 0.793 343     40.7     43.9
 NP       27.4 0.954 343     25.5     29.3
 R/R50    42.9 0.892 343     41.1     44.6
 lopro    39.7 0.892 343     37.9     41.4

Confidence level used: 0.95
```

```
K_list

$`early rest - none @ 50kcal`
[1]  0  0 -1  0  1  0

$`40kcal/week - 50kcal/week`
[1]  0.0  1.0 -0.5  0.0 -0.5  0.0

$`lo cal - hi cal`
[1] -0.50  0.25  0.25 -0.50  0.25  0.25
```

```
co = contrast(em, K_list)

# p-values (and posterior tail probabilities)
co

 contrast                   estimate   SE  df t.ratio p.value
 early rest - none @ 50kcal    0.589 1.19 343   0.493  0.6223
 40kcal/week - 50kcal/week     2.525 1.05 343   2.408  0.0166
 lo cal - hi cal              12.450 0.78 343  15.961  <.0001

# confidence/credible intervals
confint(co)

 contrast                   estimate   SE  df lower.CL upper.CL
 early rest - none @ 50kcal    0.589 1.19 343   -1.759     2.94
 40kcal/week - 50kcal/week     2.525 1.05 343    0.463     4.59
 lo cal - hi cal              12.450 0.78 343   10.915    13.98

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 questions:

- Does sulfur have any impact at all?
- What is the difference between spring and fall application of sulfur?
- What is the effect of increased sulfur application?

Data

```
   inf trt row col sulfur application treatment
1    9  F3   4   1    300        fall        F3
2   12   0   4   2      0   (Missing)         0
3   18  S6   4   3    600      spring        S6
4   10 F12   4   4   1200        fall       F12
5   24  S6   4   5    600      spring        S6
6   17 S12   4   6   1200      spring       S12
7   30  S3   4   7    300      spring        S3
8   16  F6   4   8    600        fall        F6
9   10   0   3   1      0   (Missing)         0
10   7  S3   3   2    300      spring        S3
11   4 F12   3   3   1200        fall       F12
12  10  F6   3   4    600        fall        F6
13  21  S3   3   5    300      spring        S3
14  24   0   3   6      0   (Missing)         0
15  29   0   3   7      0   (Missing)         0
16  12  S6   3   8    600      spring        S6
17   9  F3   2   1    300        fall        F3
18   7 S12   2   2   1200      spring       S12
19  18  F6   2   3    600        fall        F6
20  30   0   2   4      0   (Missing)         0
21  18  F6   2   5    600        fall        F6
22  16 S12   2   6   1200      spring       S12
23  16  F3   2   7    300        fall        F3
24   4 F12   2   8   1200        fall       F12
25   9  S3   1   1    300      spring        S3
26  18   0   1   2      0   (Missing)         0
27  17 S12   1   3   1200      spring       S12
28  19  S6   1   4    600      spring        S6
29  32   0   1   5      0   (Missing)         0
30   5 F12   1   6   1200        fall       F12
```

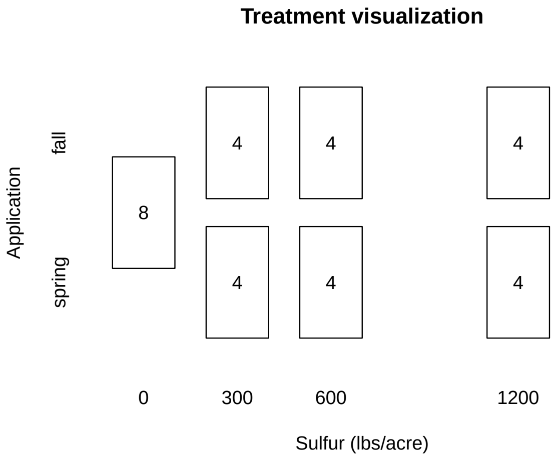
Design

Completely randomized design
potato scab experiment

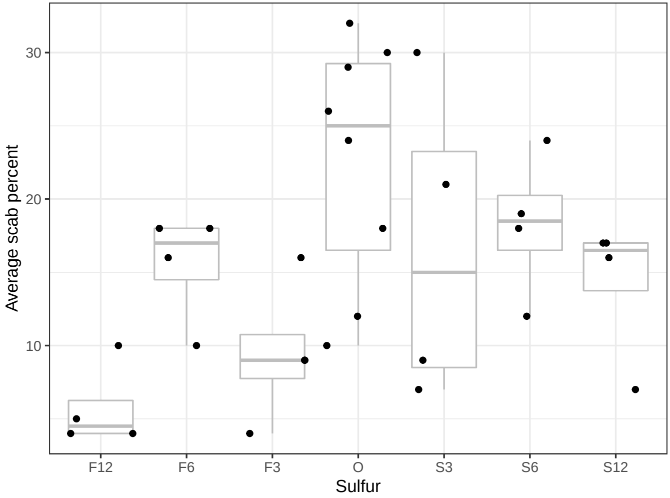
4	F3	O	S6	F12	S6	S12	S3	F6
3	O	S3	F12	F6	S3	O	O	S6
2	F3	S12	F6	O	F6	S12	F3	F12
1	S3	O	S12	S6	O	F12	O	F3
row	1	2	3	4	5	6	7	8

col

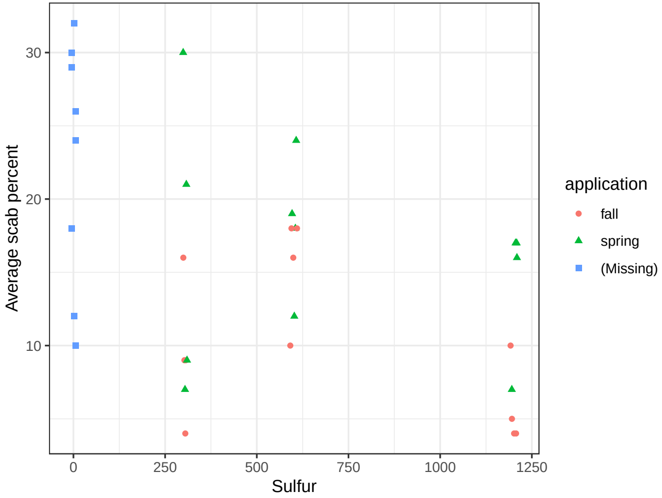
Design


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

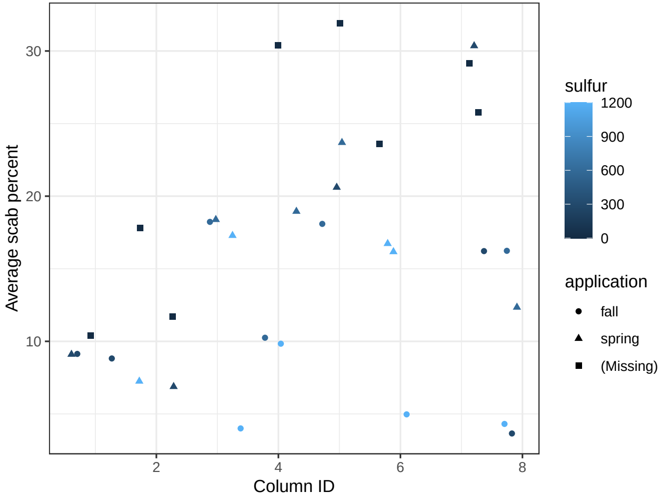
Data



Data

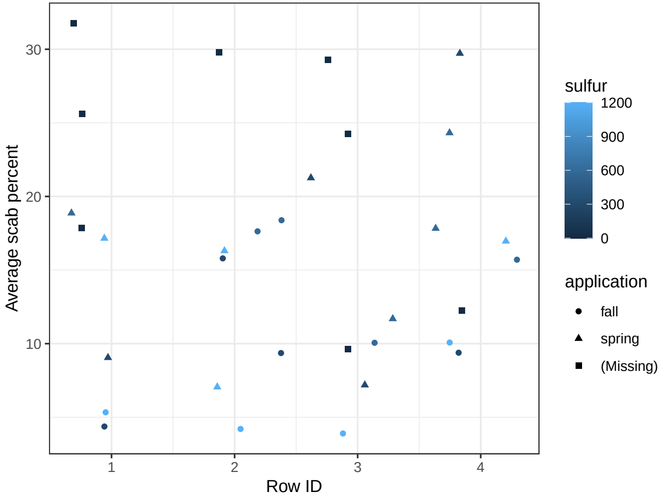

Average scab percent
30
20
10
0
250
500
750
1000
1250
Sulfur
application
fall
spring
(Missing)

Data



Average scab percent
30
20
10
2
4
6
8
Column ID
sulfur
1200
900
600
300
0
application
fall
spring
(Missing)

Data


Average scab percent
30
20
10
1
2
3
4
Row ID
sulfur
1200
900
600
300
0
application
fall
spring
(Missing)

Model

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

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

Hypotheses:

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

Contrasts

- *Sulfur effect*: Any sulfur vs none

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

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

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

Contrasts (cont.)

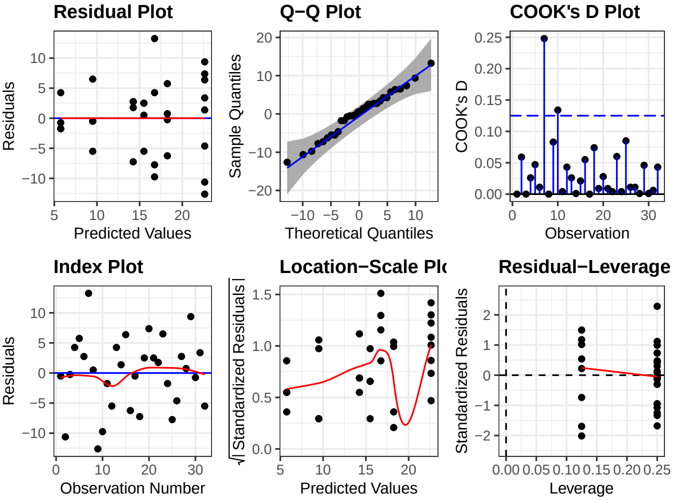
- Sulfur linear trend
  - The group sulfur levels (X_j) are 12, 6, 3, 0, 3, 6, and 12 (100 lbs/acre)
  - and a linear trend contrast is $X_j - \overline{X}$

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

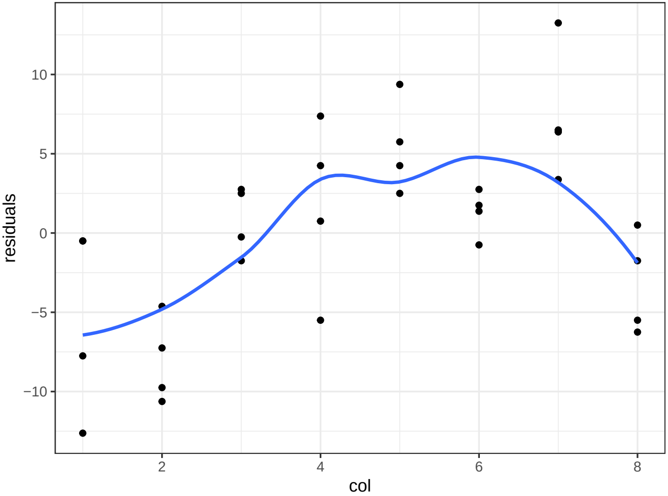
$$\gamma = 6\mu_{F12} + 0\mu_{F6} - 3\mu_{F3} - 6\mu_{O} - 3\mu_{S3} + 0\mu_{S6} + 6\mu_{S12}$$

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 =
#                                       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)

m = lm(inf ~ treatment, data = d)
anova(m)

Analysis of Variance Table

Response: inf
          Df  Sum Sq Mean Sq F value  Pr(>F)
treatment  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
```


Residual Plot
Residuals
10
5
0
−5
−10
5 10 15 20
Predicted Values
Q−Q Plot
Sample Quantiles
20
10
0
−10
−20
−10 −5 0 5 10
Theoretical Quantiles
COOK's D Plot
COOK's D
0.25
0.20
0.15
0.10
0.05
0.00
0 10 20 30
Observation
Index Plot
Residuals
10
5
0
−5
−10
0 10 20 30
Observation Number
Location−Scale Plo
√|Standardized Residuals|
1.5
1.0
0.5
0.0
5 10 15 20
Predicted Values
Residual−Leverage
Standardized Residuals
2
1
0
−1
−2
0.00 0.05 0.10 0.15 0.20 0.25
Leverage

```
em <- emmeans(m, ~treatment); em

 treatment emmean   SE df lower.CL upper.CL
 F12         5.75 3.35 25    -1.15     12.7
 F6         15.50 3.35 25     8.60     22.4
 F3          9.50 3.35 25     2.60     16.4
 O          22.62 2.37 25    17.74     27.5
 S3         16.75 3.35 25     9.85     23.7
 S6         18.25 3.35 25    11.35     25.2
 S12        14.25 3.35 25     7.35     21.2

Confidence level used: 0.95

co <- contrast(em, K)
confint(co)

 contrast          estimate    SE df lower.CL upper.CL
 sulfur - control     -9.29  2.74 25    -14.9   -3.657
 fall - spring        -6.17  2.74 25    -11.8   -0.532
 linear trend        -94.50 34.82 25   -166.2  -22.779

Confidence level used: 0.95
```


residuals
col
10
5
0
−5
−10
2
4
6
8

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 reduction in 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 logarithm of the response
  - CI for F12 (-1.2, 12.7)
  - Non-constant variance (residuals vs predicted, sulfur, application)