Contrasts

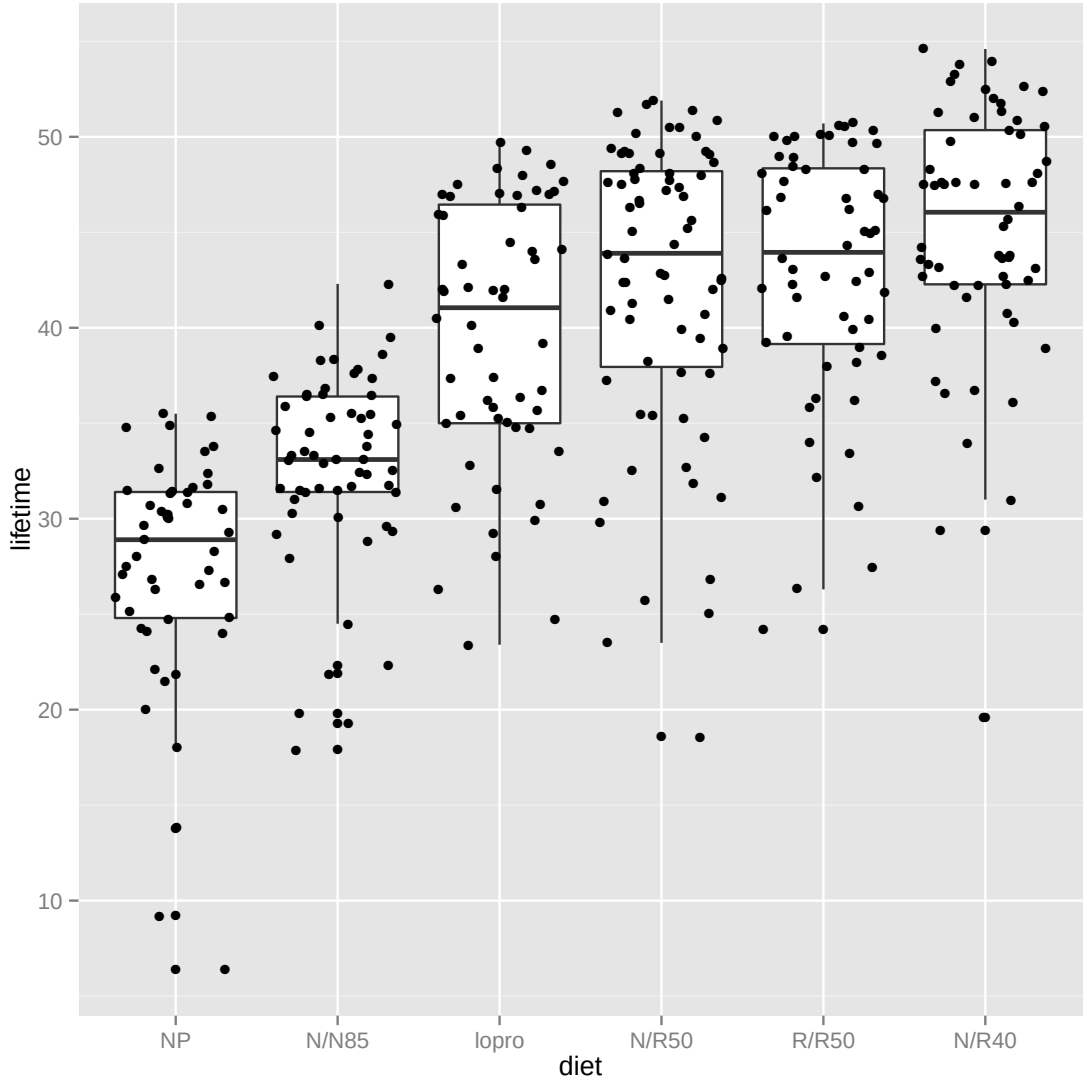
STAT 401 - Statistical Methods for Research Workers

Jarad Niemi

Iowa State University

21 September 2013

Mice diet



Simple hypothesis

Consider the one-way ANOVA model: $Y_{ij} \sim N(\mu_i, \sigma^2)$ where $i = 1, \ldots, \mathrm{I}$.

Here are a few simple alternative hypotheses:

- Mean lifetime for N/R40 and N/R50 diet is different.
- Mean lifetime for N/R40 is different than for N/R50 and R/R50 combined.
- Mean lifetime for NP and N/N85 combined is different than for low calorie diets combined.

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

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

Contrasts

i	NP	N/N85	lopro	N/R50	R/R50	N/R40
1	0	0	0	-1	0	1
2	0	0	0	$-1/2$	$-1/2$	1
3	$1/2$	$1/2$	$-1/4$	$-1/4$	$-1/4$	$-1/4$
	C_1	C_2	C_3	C_4	C_5	C_6

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

Estimated by

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

with standard error

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

t-statistic and CI

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

Examples

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

$s_p = 6.68$ with 343 degrees of freedom

	g	SEg	t	p	L	U
1	2.82	1.17	2.41	0.02	0.52	5.12
2	2.53	1.05	2.41	0.02	0.46	4.59
3	-12.45	0.78	-15.96	0.00	-13.98	-10.92

```
DATA case0501;
  INFILE 'case0501.csv' DSD FIRSTOBS=2;
  INPUT lifetime diet $ ;

PROC MEANS DATA=case0501;
  CLASS diet;
  VAR lifetime;
  RUN;
```

diet	Obs	N	Mean	Std Dev	Minimum	Maximum
N/N85	57	57	32.6912281	5.1252972	17.9000000	42.3000000
N/R40	60	60	45.1166667	6.7034058	19.6000000	54.6000000
N/R50	71	71	42.2971831	7.7681947	18.6000000	51.9000000
NP	49	49	27.4020408	6.1337010	6.4000000	35.5000000
R/R50	56	56	42.8857143	6.6831519	24.2000000	50.7000000
lopro	56	56	39.6857143	6.9916945	23.4000000	49.7000000

```
PROC GLM;
  CLASS diet;
  MODEL lifetime = diet / CLPARM;
  ESTIMATE 'N/R40 - N/R50'
    diet 0 1 -1 0 0 0 ;
  ESTIMATE 'N/R40 - (N/R50+R/R50)/2'
    diet 0 1 -.5 0 -.5 0 ;
  ESTIMATE '(NP+N/N85)/2 - (N/R50+R/R50+N/R40+lopro)/4'
    diet 2 -1 -1 2 -1 -1 / DIVISOR = 4 ;
  RUN;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
N/R40 - N/R50	2.8194836	1.17109686	2.41	0.0166
N/R40 - (N/R50+R/R50)/2	2.5252180	1.04854904	2.41	0.0166
(NP+N/N85)/2 - (N/R50+R/R50+N/R40+lopro)/4	-12.4496851	0.78001425	-15.96	<.0001

Parameter	95% Confidence Limits	
N/R40 - N/R50	0.5160481	5.1229190
N/R40 - (N/R50+R/R50)/2	0.4628224	4.5876136
(NP+N/N85)/2 - (N/R50+R/R50+N/R40+lopro)/4	-13.9838985	-10.9154718

Summary

- Contrasts are linear combinations that sum to zero
- t-test tools are used to calculate pvalues and confidence intervals