Workshop 7.4b: Single factor ANOVA (Bayesian)

Murray Logan

19 Jul 2017

Section 1

Revision

Estimation

Y	X
3	0
2.5	1
6	2
5.5	3
9	4
8.6	5
12	6

$$
\begin{array}{ccccccc}
3.0 & = & \beta_0 \times 1 & + & \beta_1 \times 0 & + & \varepsilon_1 \\
2.5 & = & \beta_0 \times 1 & + & \beta_1 \times 1 & + & \varepsilon_1 \\
6.0 & = & \beta_0 \times 1 & + & \beta_1 \times 2 & + & \varepsilon_2 \\
5.5 & = & \beta_0 \times 1 & + & \beta_1 \times 3 & + & \varepsilon_3
\end{array}
$$

Estimation

Y	X
3	0
2.5	1
6	2
5.5	3
9	4
8.6	5
12	6

$$
\begin{aligned}
3.0 &= \beta_0 \times 1 + \beta_1 \times 0 + \varepsilon_1 \\
2.5 &= \beta_0 \times 1 + \beta_1 \times 1 + \varepsilon_1 \\
6.0 &= \beta_0 \times 1 + \beta_1 \times 2 + \varepsilon_2 \\
5.5 &= \beta_0 \times 1 + \beta_1 \times 3 + \varepsilon_3
\end{aligned}
$$

Matrix algebra

$$\underbrace{\begin{pmatrix} 3.0 \\ 2.5 \\ 6.0 \\ 5.5 \\ 9.0 \\ 8.6 \\ 12.0 \end{pmatrix}}_{\text{Response values}} = \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{pmatrix}}_{\text{Model matrix}} \times \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\text{Parameter vector}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}}_{\text{Residual vector}}$$

$$Y = X\beta + \epsilon$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Section 2

Anova Param-eterization

Simple ANOVA

Three treatments (One factor - 3 levels), three replicates

Simple ANOVA

Two treatments, three replicates

Categorical predictor

Y	A	dummy1	dummy2	dummy3
2	G1	1	0	0
3	G1	1	0	0
4	G1	1	0	0
6	G2	0	1	0
7	G2	0	1	0
8	G2	0	1	0
10	G3	0	0	1
11	G3	0	0	1
12	G3	0	0	1

$$\mathtt{y}_{\mathtt{ij}} = \mu + \beta_1(\mathtt{dummy}_1)_{\mathtt{ij}} + \beta_2(\mathtt{dummy}_2)_{\mathtt{ij}} + \beta_3(\mathtt{dummy}_3)_{\mathtt{ij}} + \varepsilon_{\mathtt{ij}}$$

Overparameterized

$$y_{ij} = \mu + \beta_1(\mathtt{dummy}_1)_{ij} + \beta_2(\mathtt{dummy}_2)_{ij} + \beta_3(\mathtt{dummy}_3)_{ij} + \varepsilon_{ij}$$

Y	A	Intercept	dummy1	dummy2	dummy3
2	G1	1	1	0	0
3	G1	1	1	0	0
4	G1	1	1	0	0
6	G2	1	0	1	0
7	G2	1	0	1	0
8	G2	1	0	1	0
10	G3	1	0	0	1
11	G3	1	0	0	1
12	G3	1	0	0	1

Overparameterized

$$\mathtt{y}_{\mathtt{ij}} = \mu + \beta_1(\mathtt{dummy}_1)_{\mathtt{ij}} + \beta_2(\mathtt{dummy}_2)_{\mathtt{ij}} + \beta_3(\mathtt{dummy}_3)_{\mathtt{ij}} + \varepsilon_{\mathtt{ij}}$$

- three treatment groups
- four parameters to estimate
- need to re-parameterize

Categorical predictor

$$y_i = \mu + \beta_1(\mathtt{dummy}_1)_i + \beta_2(\mathtt{dummy}_2)_i + \beta_3(\mathtt{dummy}_3)_i + \varepsilon_i$$

EFFECTS PARAMETERIZATION

$$y_{ij} = \mu + \beta_2(\mathtt{dummy}_2)_{ij} + \beta_3(\mathtt{dummy}_3)_{ij} + \varepsilon_{ij}$$

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = p - 1$$

Categorical predictor

Means parameterization

$$y_i = \alpha + \beta_2(\mathtt{dummy}_2)_i + \beta_3(\mathtt{dummy}_3)_i + \varepsilon_i$$

Y	A	alpha	dummy2	dummy3
2	G1	1	0	0
3	G1	1	0	0
4	G1	1	0	0
6	G2	1	1	0
7	G2	1	1	0
8	G2	1	1	0
10	G3	1	0	1
11	G3	1	0	1
12	G3	1	0	1

Categorical predictor

MEANS PARAMETERIZATION

$$y_i = \alpha + \beta_2 D_{2i} + \beta_3 D_{3i} + \varepsilon_i$$

	Y	A
1	2.00	G1
2	3.00	G1
3	4.00	G1
4	6.00	G2
5	7.00	G2
6	8.00	G2
7	10.00	G3
8	11.00	G3
9	12.00	G3

$y_i = \alpha_p + \varepsilon_i$,

where p= number of levels of the factor minus 1

and D = dummy variables

$$\begin{pmatrix} 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 10 \\ 11 \\ 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix}$$

Categorical predictor

TREATMENT CONTRASTS

Parameter	Estimates	Null Hypothesis
Intercept	mean of control group	H_0: $\mu = \mu_1 = 0$
α_2^*	mean of group 2 minus mean of control group	H_0: $\alpha_2 = \alpha_2 = 0$
α_3^*	mean of group 3 minus mean of control group	H_0: $\alpha_3 = \alpha_3 = 0$

```
> contrasts(A) <-contr.treatment
> contrasts(A)
```

Categorical predictor

TREATMENT CONTRASTS

Parameter	Estimates	Null Hypothesis
Intercept	mean of control group	H$_0$: $\mu = \mu_1 = 0$
α_2^*	mean of group 2 minus mean of control group	H$_0$: $\alpha_2 = \alpha_2 = 0$
α_3^*	mean of group 3 minus mean of control group	H$_0$: $\alpha_3 = \alpha_3 = 0$

```
> summary(lm(Y~A))$coef
```

Assumptions

- Normality
- Homogeneity of variance
- Independence

- As for regression

The model

Y	A	dummy1	dummy2	dummy3
2	G1	1	0	0
3	G1	1	0	0
4	G1	1	0	0
6	G2	0	1	0
7	G2	0	1	0
8	G2	0	1	0
10	G3	0	0	1
11	G3	0	0	1
12	G3	0	0	1

$$y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

Section 3

Worked Examples

Worked Examples

```
> day <- read.csv('../data/day.csv', strip.white=T)
> head(day)
```

```
  TREAT BARNACLE
1  ALG1       27
2  ALG1       19
3  ALG1       18
4  ALG1       23
5  ALG1       25
6  ALG2       24
```

Worked Examples

Question: what effects do different substrate types have on barnacle recruitment

Linear model:

$$
\begin{aligned}
\texttt{Barnacle}_{\texttt{ij}} &\sim \texttt{N}(\mu_{\texttt{ij}}, \sigma^2) \\
\mu_{\texttt{ij}} &= \beta_0 + \beta_{\texttt{j}}\texttt{Treat}_{\texttt{ij}}
\end{aligned}
$$

$$
\begin{aligned}
\beta_0 &\sim \texttt{N}(0, \sigma^2_{\texttt{Int}}) \\
\beta_{\texttt{j}} &\sim \texttt{N}(0, \sigma^2_{\texttt{Treat}}) \\
\sigma^2 &\sim \texttt{cauchy}(0, 4)
\end{aligned}
$$