Statistics with SpaRRows

Lecture 11-c

Julia Schroeder

Julia.schroeder@imperial.ac.uk

Outline

- Linear models – going big
    - Multiple continuous predictors
    - Interactions between continuous predictors
    - Interactions between categorical predictors
    - …

Multiple continuous predictors

```
> summary(lm(Y0~Cont1+Cont2,data=a))

Call:
lm(formula = Y0 ~ Cont1 + Cont2, data = a)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.22886 -0.29364  0.00364  0.32803  1.25419 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.89714    0.28444    6.67 1.58e-09 ***
Cont1       -4.05761    0.04663  -87.01  < 2e-16 ***
Cont2        3.79363    0.05129   73.97  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4757 on 97 degrees of freedom
Multiple R-squared:  0.9922,	Adjusted R-squared:  0.992 
F-statistic:  6144 on 2 and 97 DF,  p-value: < 2.2e-16
```

Main effects cannot be interpreted in isolation

IF Cont2 is being held constant, then, with increasing Cont1, Y increases

If Cont1 is being held constant, then, with increasing Cont2, Y decreases

-> cannot visualize well

- helpful to account for environmental variables

Interactions between continuous predictors

- Really hard (if not impossible) to properly interpret
- Only do this if you know what it means

Interactions between categorical predictors

- Cat1: male, female
- Cat2: Orange, Green, Purple

- → Possible combinations:

- Male Orange, Male Green, Male Purple

- Female Orange, Female Green, Female Purple

Interactions between categorical predictors

```
> summary(lm(Y3~Cat1*Cat2,data=a))

Call:
lm(formula = Y3 ~ Cat1 * Cat2, data = a)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2190 -0.2519  0.0102  0.3457  1.2261 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
b0 (Intercept)        -5.7993     0.1162 -49.890  < 2e-16 ***
b1 Cat1male           16.4021     0.1567 104.646  < 2e-16 ***
b2 Cat2Orange         14.4821     0.1552  93.313  < 2e-16 ***
b3 Cat2Purple         -6.4878     0.2096 -30.960  < 2e-16 ***
b4 Cat1male:Cat2Orange  4.4325     0.2266  19.559  < 2e-16 ***
b5 Cat1male:Cat2Purple -1.9261     0.2690  -7.161 1.76e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4932 on 94 degrees of freedom
Multiple R-squared:  0.9984,	Adjusted R-squared:  0.9983 
F-statistic: 1.168e+04 on 5 and 94 DF,  p-value: < 2.2e-16
```

Female Green (Reference): b_0

Difference between reference and

Male Green: $b_0 + b_1$

Female Orange: $b_0 + b_2$

Female Purple: $b_0 + b_3$

Male Orange: $b_0 + b_4$

Male Purple: $b_0 + b_5$

Interactions between categorical predictors

- Does not test between all categories
- → Tukey test

```
> m1<-lm(Y3~Cat1*Cat2,data=a)
> an<-aov(m1)
> TK<-TukeyHSD(x=an)
> TK
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = m1)

$Cat1
                diff      lwr      upr p adj
male-female 13.91706 13.72118 14.11294     0

$Cat2
                    diff        lwr        upr p adj
Orange-Green   15.845950  15.579905  16.111995     0
Purple-Green   -7.498879  -7.810613  -7.187145     0
Purple-Orange -23.344830 -23.659461 -23.030198     0

$`Cat1:Cat2`
                                  diff        lwr        upr p adj
male:Green-female:Green      16.402149  15.946138  16.858159     0
female:Orange-female:Green   14.482141  14.030614  14.933669     0
male:Orange-female:Green     35.316741  34.815130  35.818352     0
female:Purple-female:Green   -6.487805  -7.097479  -5.878131     0
male:Purple-female:Green      7.988251   7.476961   8.499540     0
female:Orange-male:Green     -1.920007  -2.347889  -1.492126     0
male:Orange-male:Green       18.914592  18.434156  19.395029     0
female:Purple-male:Green    -22.889954 -23.482328 -22.297579     0
male:Purple-male:Green       -8.413898  -8.904431  -7.923365     0
male:Orange-female:Orange    20.834600  20.358416  21.310783     0
female:Purple-female:Orange -20.969946 -21.558877 -20.381016     0
male:Purple-female:Orange    -6.493891  -6.980259  -6.007522     0
female:Purple-male:Orange   -41.804546 -42.432699 -41.176393     0
male:Purple-male:Orange     -27.328490 -27.861680 -26.795301     0
male:Purple-female:Purple    14.476056  13.840147  15.111964     0
```

What to use when

- Depends on your question

Do you want to:

- predict
- explain
- explore
- account for variables
- …

Take home

- When running models, always be aware what is categorical and what is continuous, because the the interpretation differs
- Know your data structure!
- Do not overfit – less complex models are better