Categorical predictor grouping factor

Group means would be fit line

Equal variance

Since have three or more levels, we can’t specify direct alterative hypothesis

H0: all groups are the same

Alternative: at least one is different

Is grouping factor effective at reducing the variance?

When group the data, within groups we manage to reduce the variance > explaining the means of our groups data

Detour:

Quantify some sort of uncertainty > by squaring difference from mean we get positive values

+ Penalty for values that are far away from mean

Define model error and certainty, statistical significance

Grand mean > square > total sum of squares > total variability of data

Within groups > grouped according to group factor > sum of squares within groups and add them up > how much is explained by grouping factor

* Look at differences between group means

To calculate the total sum of squares, calculate group mean

Within group > differences withing groups in sd

And differences between groups > multiply by number of observations

I observations

G groups

Difference between group mean and grand mean > between groups sq

And multiply by observations of each group and then sum

From grand mean to group mean and multiply by number of observations

Degrees of freedom for F statistic of each effect

Total sum of squares df > number of total observations -1 : grand mean we calculates

Within > calculate group mean > number of levels/group we have > subtract one from the within groups

Between > number of group means

Mean squares >

Degrees of freedom of within groups/ between groups

By normalizing allows us to make direct comparison

Table has group and within groups values

Between > how different are In mean from overall

Within > how difference are withing

Whether grouping factor is significant F value

F distribution > ratio of normal distribution or ration of variances

Two parameters > df for numerator and denominator

Between groups nominator is within groups

Denominator > residuals

Nominator > difference between groups

Did our grouping help to understand the pattern?

Large btw group differences > larger numerator

High/significant f value

* Large btw group

All small within group variability

Residuals similar to between group variation > low F value

Grouping factor does not help to understand noise in the data

About 1 F value > same variance in groups and total

Btw group variance high > low p value

Anov ais about grouping the data > understanding what factor causes variance

Model coefficient

* Means of groups

Anova table > which is helping to reduce noise

Mean squares, corresponding p value and F test for each level

Does including a predictor is helping to explain model significantly? Anova table

Coefficient

Group means > adding values to base case

What body mass of different species

Doesn’t tell us how well did grouping by species do

ANIVA table

Overall significance/importance of species factor/of a variable > how well does explain variance in data

**Probability distribution functions**

Use distribution functions to answer questions

How is sample space related to some likelihood

Density>

Cumulative > smaller or larger

Cumulative > are under curve

Probability density > height of a curve

Total are under curve is 1 > probability always between zero and one

Discrete > mass > read event from height of mass curve > relative measure of likelihood

Continuous > density > underneath curve

Length of fish linked to relative frequency of observing this > height of curve

Cumulative > integrals/underneath curve > greater or less than …. , in range of 4-8

Can do 1- to calculate other part

Quantile function

What will 90% of all fish be shorter than?

* Can read from height/graph

P/d/q

* Name
* For normal > norm
* Binomial distribution > binom

Creates a map of events to likelihoods

Depending on which distribution we are working with

* Have to know the parameters

Histograms > check if residuals are normally distributed

Likelihood

* How likely has a data set occurred?
* Height of probability mass or density curves

Take log function when doing likelihood > monotonic increasing, bur rate of increase decreases!!!

Maximize likelihood > sliding distribution left and right, maximize likelihoods so that likelihood is the highest