Week 6 stats Interactions

Will

11/03/2021

knitr::opts\_chunk$set(echo = TRUE)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.6 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(modelr)  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(qqplotr)

##   
## Attaching package: 'qqplotr'

## The following objects are masked from 'package:ggplot2':  
##   
## stat\_qq\_line, StatQqLine

library(praise)  
library(patchwork)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(emmeans)

biomass <- read\_csv("data/biomass.csv")

## Parsed with column specification:  
## cols(  
## Fert = col\_character(),  
## Light = col\_character(),  
## FL = col\_character(),  
## Biomass.m2 = col\_double()  
## )

need to check the data for missing values and unknonw names and if it is tidy.

Question - Can you write a brief/sensible hypothesis for the effects of fertilizer and light on biomass? It is best to break down your hypotheses into sections and give direction

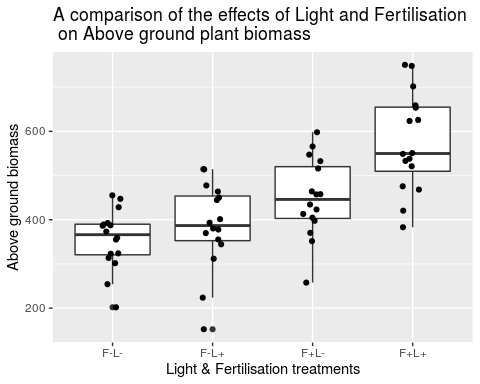
The application of fertilizer increases the biomass of the experimental plant communities

The application of light to the plant understorey increases the biomass of experimental plant communities

also there will be a postive interaction between the light and fertaliser to have a higher effect

We should also plot our data to sense-check our ideas/hypotheses

biomass %>%   
 ggplot(aes(x=FL, y=Biomass.m2))+  
 geom\_boxplot()+  
 geom\_jitter(width=0.1)+  
 labs(x="Light & Fertilisation treatments", y="Above ground biomass")+  
 ggtitle("A comparison of the effects of Light and Fertilisation \n on Above ground plant biomass")

 Now that we have both a sensible hypothesis and an observation of an apparent difference between treatments, we are justified in producing a model with one variable (Treatment) and four levels (F-L-, F-L+, F+L-, F+L+) to test this effect.

model1 <- lm(Biomass.m2~FL, data=biomass)  
summary(model1)

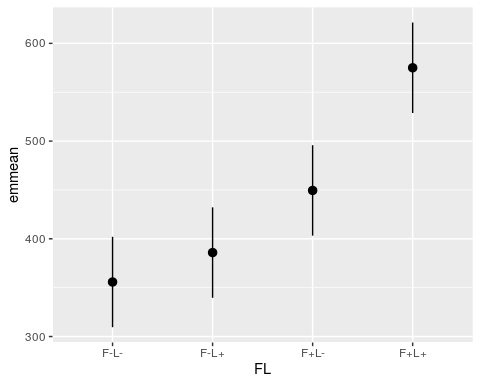
##   
## Call:  
## lm(formula = Biomass.m2 ~ FL, data = biomass)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -233.619 -42.842 1.356 67.961 175.381   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 355.79 23.14 15.376 < 2e-16 \*\*\*  
## FLF-L+ 30.12 32.72 0.921 0.36095   
## FLF+L- 93.69 32.72 2.863 0.00577 \*\*   
## FLF+L+ 219.22 32.72 6.699 8.13e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 92.56 on 60 degrees of freedom  
## Multiple R-squared: 0.4686, Adjusted R-squared: 0.442   
## F-statistic: 17.63 on 3 and 60 DF, p-value: 2.528e-08

Here the intercept is F-L- (no fertiliser or light). The overall model summary appears to support our hypotheses. There are significant mean differences between our treatments and the intercept.

If we want to plot the actual estimated means from our model, then these will be contained in broom::augment() but are probably easiest to pull out using the emmeans() function from the package of the same name.

this cannot be used directly for estimating significant differences or effect sizes.

emmeans::emmeans(model1, specs="FL") %>%   
 as\_tibble() %>%   
 ggplot(aes(x=FL,  
 y=emmean))+  
 geom\_pointrange((aes(ymin=lower.CL,  
 ymax=upper.CL)))

 ##Estimated mean differences Question - What do the three estimated differences of the mean tell us about the effect of fertiliser and light on biomass? If the combined effect of fertiliser and light was additive we would expect the F+L+ treatment to be roughly the sum of the other two mean differences (30+94) = 124.

if the light + fert mean was addidtive it would be 124 for both and not 219 showing there is a postive effect BUT the value is much greater than this FLF+L+ = 219.

This indicates an interaction - where the combined effect of Fertiliser and Light is greater than we would expect from combining the treatments.

# four possible combinations of just two treatments, not 4 independant variables

We need a model that that uses the combination of the two treatments to reflect the design and will explicitly estimate the interaction. the light:fert= start a two way interactions,

model2 <- lm(Biomass.m2~Light+Fert+Light:Fert, data=biomass)  
summary(model2)

##   
## Call:  
## lm(formula = Biomass.m2 ~ Light + Fert + Light:Fert, data = biomass)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -233.619 -42.842 1.356 67.961 175.381   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 355.79 23.14 15.376 < 2e-16 \*\*\*  
## LightL+ 30.13 32.72 0.921 0.36095   
## FertF+ 93.69 32.72 2.863 0.00577 \*\*   
## LightL+:FertF+ 95.41 46.28 2.062 0.04359 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 92.56 on 60 degrees of freedom  
## Multiple R-squared: 0.4686, Adjusted R-squared: 0.442   
## F-statistic: 17.63 on 3 and 60 DF, p-value: 2.528e-08

#(LightL+:FertF+)= estimates the effect size of the interaction.

the value is now 95 as this is the estimate that of the extra mean differene so the extra amount of grwoth showing us the interction effect

our estimated mean fro F+L+ is 219 as this was what the estimate was in model 1, if the effect of light is 30 and fertiliser is 93 then the interaction effect is what’s left over - and now our model successfully estimates this.

The intercept is the mean for the unmanipulated control F-L-

## this does what we did above step by step

coef(model2)[1]#[1]= picks out information by row and column

## (Intercept)   
## 355.7938

The estimated biomass for the Fertilised treatment (F+L-) is the intercept plus the coefficient value for fertilisation

coef(model2)[1]+coef(model2)[2]

## (Intercept)   
## 385.9188

The estimated biomass for the Light treatment (F-L+) is

coef(model2)[1]+coef(model2)[3]

## (Intercept)   
## 449.4875

estimated mean of our final full combination treatment, we take the baseline intercept and both additive main effects and the interaction term

coef(model2)[1]+coef(model2)[2]+coef(model2)[3]+coef(model2)[4]

## (Intercept)   
## 575.0188

Another way to think about the interaction term, is that if there was no absolutely no interaction between fertilizer treatments and light treatments then the effect size of the interaction term would be zero.

So the estimate would be zero and our additive effects would be the whole story. ## Main effects in the presence of a significant interaction Light appears to be non-significant P = 0.36, this is the average effect of Light on growth across both fertilised and non-fertilised treatments. We know because of our significant interaction term that whether the light has an effect depends on the fertilisation treatment - the effect of adding light is stronger on fertilised plants.

when describing a model always work from the bottom up (interactions first) If you detect a significant interaction term, then you must must include all the main effects that make up that treatment as well. Sometimes the interaction effect may be very weak, and the main effects much stronger overall in which case you definitely need to report all main and interaction effects when writing up your results.

### 3. ANOVA tables

overall summary of an effect we can use ANOVA tables.

summary(aov(model2))

## Df Sum Sq Mean Sq F value Pr(>F)   
## Light 1 96915 96915 11.31 0.00135 \*\*   
## Fert 1 319889 319889 37.34 8.02e-08 \*\*\*  
## Light:Fert 1 36409 36409 4.25 0.04359 \*   
## Residuals 60 513998 8567   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## F-values for main effects

light is sig in avo table but not in the other , whta avo does is it adds terms one so can overestimate the fist thing that gets put into the table. its good for measuring the

The main disadvantage of ANOVA tables run by the summary(aov()) route is that they work by sequentially adding effects.

This tests the main effect of factor A, followed by the main effect of factor B after the main effect of A, followed by the interaction effect AB after the main effects.= The result is that the order in which variables were specified in the model are important AND it is bad because it over-estimates the significance of main effects when there are interactions.

This is known as Type I Sums of Squares. this is ok for looking at the top-level interaction terms only we can reproduce the ANOVA for the interaction term easily by running an F-test comparing our complex model (2) with a simpler model (3).

good way to confirm that an interaction effect is significant. If it is not, you can remove it from your model (use the simple one instead).

this is just running a test to see if the results will be different if you run it with and without the interaction and it is significantly differnet as f is 0.04 which is the same as what is say at the bottom of the anova

model3 <- lm(Biomass.m2 ~ Light + Fert, data = biomass) ## simple model - interaction removed  
anova(model2,model3, test="F")

## Analysis of Variance Table  
##   
## Model 1: Biomass.m2 ~ Light + Fert + Light:Fert  
## Model 2: Biomass.m2 ~ Light + Fert  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 60 513998   
## 2 61 550407 -1 -36409 4.2501 0.04359 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

say the interaction was sig and say it has an f vlaue of 4.45 and df of 1/60 and p value= .. the main effect ###Type I,II,III sum of squares But what about producing F values for the main effects, when there are significant interactions in the model? Enter car::Anova which can specify type II and type III Sums of Squares.

car::Anova(model2, type="III")

## Anova Table (Type III tests)  
##   
## Response: Biomass.m2  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 2025427 1 236.4322 < 2.2e-16 \*\*\*  
## Light 7260 1 0.8475 0.360950   
## Fert 70228 1 8.1979 0.005769 \*\*   
## Light:Fert 36409 1 4.2501 0.043587 \*   
## Residuals 513998 60   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

this is balance of all the main effects and does not weight them differetn

the interaction effect has not hcanes but the main effects are weighted better so can report the f values say fertiliseers does and light does not E.g. Fertiliser had a sigificant effect on biomass F1,60 = 8.2, P = 0.006. #How to choose Type I,II, or III Sums of squares Type I - compare the effect of removing an interaction or main effect from a model - useful for justifying model simplification

Type II - Most accurate for describing a model with main effects only

Type III - Most accurate for describing main effects and interactions when there is an interaction term in the

### Summary

The test rejects the null hypothesis of no interaction effect of light addition and fertilization, (though only just!). And we could report this as:

There was a significant interactive effect of light addition and fertilisation (F~1,60 = 4.25, P = 0.044). And we could also report our main effects using Type III Sums of Squares.

IF our interaction was non-significant we would have failed to reject our Null Hypothesis, and we could remove the interaction term and re-do the model with main effects only - and test that as a separate hypothesis.

# 4. Confidence intervals and estimates

alternative to ANOVA reporting is to estimate effect sizes with 95% Confidence Intervals.

Providing estimates with CI allows a test of interaction effects, but also provides information on the strength of the effect (how strong is the interaction in terms of the effect on bio-mass production).

broom::tidy(model2, conf.int=T)

## # A tibble: 4 x 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 356. 23.1 15.4 1.76e-22 310. 402.   
## 2 LightL+ 30.1 32.7 0.921 3.61e- 1 -35.3 95.6  
## 3 FertF+ 93.7 32.7 2.86 5.77e- 3 28.2 159.   
## 4 LightL+:FertF+ 95.4 46.3 2.06 4.36e- 2 2.84 188.

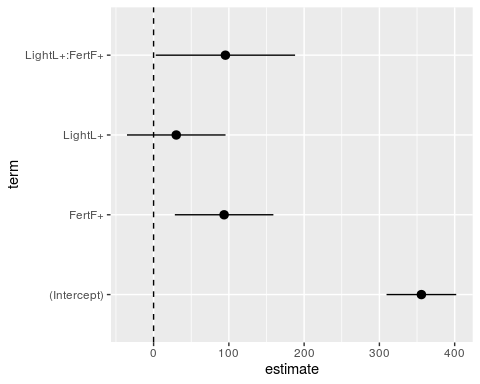
if the 95% confidence interval spans zero, then we cannot report a mean difference that is different to zero at P < 0.05. we do not have greater than 95% confidence that the estimated mean difference between the two light treatments is greater than zero.

From our confidence intervals we can also report that the addition of fertiliser increases plant biomass by a minimum of **28.23 g/m^2** and the combined effect of light and fertilisation increases biomass by at least **2.84 g/m^2** more than would expect from the additive effects of light and fertilisation. (Note I am using the lowest confidence interval margins here).

tidy\_model <- broom::tidy(model2, conf.int=T)   
  
tidy\_model

## # A tibble: 4 x 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 356. 23.1 15.4 1.76e-22 310. 402.   
## 2 LightL+ 30.1 32.7 0.921 3.61e- 1 -35.3 95.6  
## 3 FertF+ 93.7 32.7 2.86 5.77e- 3 28.2 159.   
## 4 LightL+:FertF+ 95.4 46.3 2.06 4.36e- 2 2.84 188.

tidy\_model %>%   
 ggplot(aes(x=estimate,   
 y=term))+  
 geom\_pointrange(aes(xmin=conf.low,   
 xmax=conf.high))+  
 geom\_vline(xintercept=0, ### set intercept to zero, if an interval crosses zero it means "zero difference"  
 linetype="dashed")

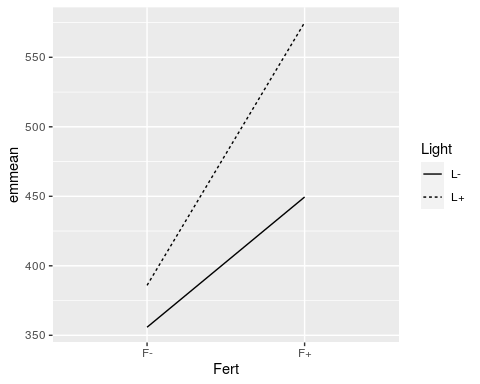
 # Emmeans

use our model to produce the estimated means and confidence intervals directly for our different categories.

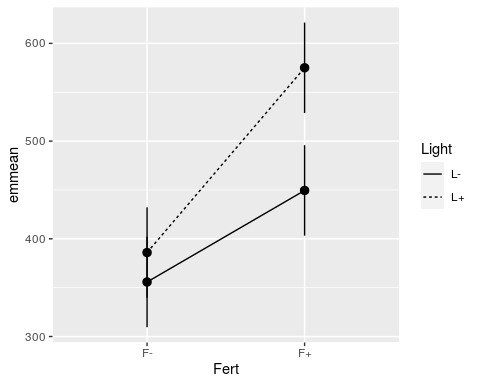
This is useful for plotting how well your model describes/explains the dataset.

The easiest way to do this is with the emmeans package, that you used earlier

means <- emmeans::emmeans(model2, specs= ~Fert:Light, type="response") %>% confint()   
 ### estimated means as predicted by the linear model, pipe to confint to add 95% Confidence intervals to estimates  
  
  
# means <- emmeans::emmeans(model2, specs= pairwise~Fert:Light, type="response") %>% confint()  
## add the argument pairwise ~ in front of the variables to produce a second table of $contrasts.   
# Contrasts allows you to estimate average and minimum effect size differences between all factor levels - this is an example of a post-hoc test - a term you should be familiar with from first year.   
  
plot1 <- means%>% ## emmeans does not output in a table format - pipe to as\_tibble() to convert it.   
 as\_tibble() %>%   
 ggplot(aes(x=Fert,  
 y=emmean,  
 group=Light))+  
 geom\_line(aes(linetype=Light))  
  
plot1

 #### geom\_pointrange() to add 95% Confidence intervals to these mean estimates

plot1+geom\_pointrange(aes(ymin=lower.CL,   
 ymax=upper.CL,  
 ))



## Write up

\*We can report the full model in a table, if we use the stargazer package it will include 95% CI for the mean differences. (Remember this is great for an instant and report worthy table when producing a markdown document - but will look like nonsense if run in R as it outputs HTML).

\*We can refer to useful figures

\*We should report the full model ANOVA (from broom::glance or the bottom line of the summary() table)

I hypothesised there would be positive interaction effect on the biomass of experimental plant communities given a combined fertiliser and artifical light application. Plants given a fertiliser treatment may experience increased competition for light as they grow, and so the application of an artificial light source to the plant understorey is expected to have a greater impact here than on the less nutrient-rich treatment.

To test this hypothesis I compared plant biomass levels in g/m^2 using a general linear model with factorial predictors of light and fertiliser treatment and an interaction term of these two predictors.

I found a strong overall effect of light and fertiliser treatment on the biomass of plants in these experimental communities (F3,60 = 17.6, P <0.001, Full model estimates Table 1). The addition of Fertiliser had a strong positive effect on biomass with an average increase in yield of 93g/m^2 (95%CI: 28.23-159.15). The application of light to the understorey of the plants did not have a significant effect on plant growth as a main effect ( estimated increase of 30.12g/m^2 (95%CI: -35.33-95.58), but there was a significant positive interaction effect of combined light and fertiliser application that increased plant biomass by an extra 95.4 g/m^2 (95% CI: 2.84-187.98) above that predicted by the additive main effects alone (Figure 1).

This suggests that biomass production may be limited by different factors in the under-storey and upper canopy of the plants. For example with the addition of fertiliser competition for light may be more intense, or more light limited.

Describe the differences observed, include values and CI where appropriate. ### Questions how do you know if an interaction effect is weak? that bit is confusing me

Philip Leftwich Philip Leftwich 13:04

Ok well there are two terms that people often trip over

13:05 It could be only marginally significant p=0.049

13:05 So is it a true effect/ how confident are we that it is real

13:05 Or there is effect size.

13:05 What’s the lowest confidence interval value for the mean difference

13:05 If it says there is HUGE biological effect then it is a very strong interaction

13:06 Or does it make more of an effect size difference than the main effects on their own?

13:06 Does that help at all?

Olivia Giffney Olivia Giffney 13:07

wait so F between L thats the interaction effect right?

Philip Leftwich Philip Leftwich 13:07

If you are on the model that specifies it then it is Light:Fert

Olivia Giffney Olivia Giffney 13:07

but L doesn’t have a significant effect on its own but it does when its FL

Philip Leftwich Philip Leftwich 13:07

Right

Olivia Giffney Olivia Giffney 13:07

okay i think i get it

Philip Leftwich Philip Leftwich 13:08

So we know the effect of light is predominantly cause by its interaction with fertiliser

13:08 For Fertiliser the conf.low value is 28.2

13:08 so adding fertilieser produces AT LEAST this increase in biomass

the interaction effect conf.low is 2.84

So we can say that adding Fertiliser is the most important effect, followed by its interaction with light