Module 9: Interactions

Emily Simmonds

3/7/2022

Instructions

This document contains information, questions, R code, and plots.

Hints and reminders are in bold

Questions appear in blue.

```
# R code looks like this.
# Comments are in grey and explain the code

2+2 # add two and two
```

```
## [1] 4
```

outputs from R appear like the above ^^^^

Contents of module

- Part A = Choosing a model
- Part B = Two categorical explanatory variables
- Part C = Interactions
- Part D = Mixed continuous and categorical variables
- Part E = Two continuous variables
- Part F = Detective skills

R skills

- Using `lm()` and interpretting output
- Plotting `lm()` results

Begin with the intro video. click here if it does not work.

Part A: Choosing a model

Video that recaps last week and gets ready for this week. click here if it does not work.

In this first part, you will be presented with some example data. You will then have to think about what sort of question you can ask with the model and therefore what you expect the `lm()` output to represent.

A. If you fit a linear model to following datasets what would the model be asking each time? i.e. what is the biological question that you would be using the model to answer?

B. What would you expect the `lm()` estimates for the intercept (α) and β to represent in each example?

Dataset 1: Data collected on wing length of butterflies from three different populations; one at sea level, one half-way up a mountain, and one at the top. Wing length is in mm, the populations are "low", "mid", and "high" altitude. There are 100 measurements per population.

Dataset 2: Data collected on body weight and length of frogs. Body weight is in grams and length in mm. There are 200 measurements.

Dataset 3: Data collected on bacteria growth rate under two different treatments. The first treatment controls food and is "low" and "high". The second concerns the presence of a fungus and is "present" or "not". There are 70 observations per treatment.

Answer video

Answer videos: Video1 Video2 Video3

Extra explanation of dataset 3. In this dataset you have two categorical variables, just like the example last week with Treatment and Time. The model I discuss in the video is an additive model (Food and Fungus are included in the `lm()` using a +). This is why you would only get 3 coefficients, an intercept, which = estimate of mean of the group Food = 0 and Fungus = 0, then a single effect of Food = 1 and a single effect of Fungus = 1. To get to the combined effect of Food and Fungus both = 1 and an estimate of the mean of that group, you would need to calculate yourself by adding the individual effects for each treatment to the estimated intercept group mean (Food = 0 and Fungus = 0). (Adding all 3 coefficients together.)

e.g.

- Group mean for Food = 0 and Fungus = 0 (intercept)
- Group mean for Food = 0 and Fungus = 1 (intercept + coef for fungus)
- Group mean for Food = 1 and Fungus = 0 (intercept + coef for food)
- Group mean for Food = 1 and Fungus = 1 (intercept + coef for food + coef for fungus)

OR you could use an interaction model. Then you could get estimates for the effects of all treatments and calculate the group means:

- Group mean for Food = 0 and Fungus = 0 (intercept)
- Group mean for Food = 0 and Fungus = 1 (intercept + coef of fungus)
- Group mean for Food = 1 and Fungus = 0 (intercept + coef of food)
- Group mean for Food = 1 and Fungus = 1 (intercept + coef of food + coef of fungus + coef of interaction)

This model would give 4 coefficients in R (the extra is the interaction estimate). But, as mentioned below in the written answers, you can't know if the data was collected in a way to allow this analysis or not.

There is a lot more on this in the rest of the module, if it doesn't make sense yet.

Show me the answer as words

Dataset 1: You would be asking if there is a difference in mean wing length between different populations. This is because we would assume that altitude might explain wing length (you could maybe argue the other way, but I think that would better be addressed looking at wing length and ability to fly up or something similar) and altitude here is represented by the variable population, which is categorical. This means you should expect the output of the `lm()` to give the estimate of a group mean and some differences between that group mean and other group means.

Dataset 2: You would be asking if there is a relationship between body weight and length. This is because we would assume that length would influence weight and these are both continuous variables. Therefore you should expect the output of the `lm()` to give an intercept and a slope, which represent a continuous straight line.

Dataset 3: You would be asking if there is a difference in mean bacteria growth rate between different food or fungus treatments. Growth rate is the response variable and the treatments are the explanatory variables, which are both categorical. This means you should expect the output of the `lm()` to give an estimate of a

group mean and some differences from that estimate to the other group means. It is not completely clear from the description if this experiment had a full factorial design i.e. if all combinations of both treatments were tried. So you could test an interaction or not, both are correct.

Write tips to your future self on how to decide which model to use when.

Show me Emily's example

If I were to do this, I would emphasise checking what kind of data your variables are made from e.g. categorical or continuous and thinking about which is a response and which is explanatory. You could also try and draw it out like in the examples.

Part B: Two categorical explanatory variables

Last week you had data from an experiment carried out at Rothamsted research lab in the UK. **Notice I have renamed one of the variables. After1970 is now called Time. I thought this was a bit less to type.**

```r
# Import the data
Yields <- read.csv("https://www.math.ntnu.no/emner/ST2304/2021v/Week09/Hoosfield_Yields_Renamed.csv")

# first check how the data are stored
str(Yields)
```

```
## 'data.frame':    72 obs. of  6 variables:
##  $ Years    : chr  "1852-61" "1862-71" "1872-81" "1882-91" ...
##  $ StartYear: int  1852 1862 1872 1882 1892 1902 1912 1922 1932 1942 ...
##  $ EndYear  : int  1861 1871 1881 1891 1901 1911 1921 1931 1941 1951 ...
##  $ Treatment: chr  "Control" "Control" "Control" "Control" ...
##  $ yield    : num  1.43 1.1 0.85 0.78 0.67 0.65 0.82 0.46 0.87 1.17 ...
##  $ Time     : chr  "Before" "Before" "Before" "Before" ...
```

```r
# IF Time is chr not fct run this line
Yields$Time <- as.factor(Yields$Time)

# Make sure to relevel the time variable
# to make 'before' the intercept
Yields$Time <- relevel(Yields$Time, ref="Before")
```


The data had four fertiliser treatments:

- Control
- Manure
- Fertilised,
- Stopped

You analysed the effect of these different treatments last week using t-tests, linear models, and one-way ANOVA ideas.

Towards the end of last week Bob told you that in 1970 something changed in the experiment. We don't know exactly what changed, but we want to account for it. So, we introduced an extra variable:

Time: with levels "Before" and "After", i.e. before and after 1970. And you fit a model with Time and Treatment included and looked at how this changed your results.

Load your model (the `mod.2way` model, i.e. `yield ~ Treatment + After1970`) from last week or run it again. Remember that After 1970 is now called Time.

Calculate the estimated mean of each group using the model coefficients e.g. calculate the mean for the manure treatment before and after 1970 etc

Now you have an estimated mean for each group, check this against the actual mean of each group.

Code hint

The code below shows you how you can easily (easy in terms of less typing) take the mean of each group in the data. Remember that `[]` are index brackets, they tell R, "I want something inside the object I just mentioned" e.g. `object[1]` takes the first entry in `object`.

The `==` means matching and selects only entries that match what follows the `==`. So, in words `mean(Yields$yield[Yields$Treatment == "Control" & Yields$Time == "Before"])` tells R to take the mean of the column `yields` inside data frame `Yields` but only for rows where column `Treatment` = Control and column `Time` = Before.

You can then check if these match your estimates from the `lm()` (some of these you will need to calculate).

Why does it not add up?

Show me the answer

Make sure to have at least a guess first! Think about what the model could be missing?

Answer video: Link

Part C: Interactions

Make sure to watch the video in the answers for Part B.

Often in nature, effects are not separate but interact. This means that the effect of one variable is different for different values of another, e.g. plants might not be able to grow more in response to fertiliser, unless they also have enough light. So, the effect of fertiliser would be different in low and high light.

Often, we want to include interactions in our models. We do this by changing the `+` in `lm()` to `*` or `:`. If you use `:` need to use `+` too, `*` does everything. So we recommend using `*`! It's much easier. e.g. `lm(Yield ~ Treatment * Time, data = Yields)`

By adding in an interaction, we now change what R calculates from the model. In the previous example, without an interaction, the output of the `lm()` had no row for the Fertilised, Manure, or Stopped treatments for the time period after 1970. To work out what the estimated means of these groups you would have had to add the effect of the fertiliser treatment to the control and then add the effect of time after 1970. This assumes that the effect of time is the same for each fertiliser treatment AND that the effect of each fertiliser treatment is the same regardless of time.

But in reality, the effect of time might not be the same for each fertiliser treatment and vice versa, **this is an interaction**. So, when you fit a model with an interaction in R it fits a separate effect for each combination of groups within the interacting variables e.g. one for fertiliser after 1970 and a separate effect for fertiliser before 1970 and a separate one for control after 1970 etc.

When you run a model with an interaction, you should now get a row for each combination of time and fertiliser treatment.

Run the interactive model.

What do these coefficients mean? Which parameters do they relate to?

Check your answer!

Answer video

Link

Answer in writing

The (Intercept) row is still the estimate of the mean of the control group before 1970. Each row that is labelled with 'Treatment' followed by a group name e.g. 'TreatmentFertilised' (but without a :) represent the estimated difference in mean between the (Intercept) group and the named group (all before 1970). E.g. Fertilised before 1970 group has a mean 1.6225 more than the Control group before 1970.

The row that is labelled with TimeAfter gives the estimated difference in mean between the (Intercept) group and the Control group after 1970.

The last few numbers are the interactions. They all include a : which tells you they are an interaction. They represent the estimated difference between the mean of the Treatment group before 1970 and the mean of the Treatment group after 1970. For example, the row labelled TreatmentManure:TimeAfter represents the difference between the mean for the Manure treatment before 1970 (0.89750000 + 2.14083333) and the mean for the Manure treatment after 1970. After 1970, the Manure Treatment had a mean yield approximately 2.6 higher than before 1970.

```
# Run the model
modelBothI <- lm(yield ~ Treatment*Time, data = Yields)

# Look at coefficient estimates
coef(modelBothI)
```

```
##                  (Intercept)            TreatmentFertilised
##                   0.89750000                     1.62250000
##              TreatmentManure                TreatmentStopped
##                   2.14083333                     0.87416667
##                    TimeAfter  TreatmentFertilised:TimeAfter
##                   0.04416667                     1.01750000
##     TreatmentManure:TimeAfter      TreatmentStopped:TimeAfter
##                   2.61416667                    -0.19416667
```

```
# Look at uncertainty
confint(modelBothI)
```

```
##                                  2.5 %     97.5 %
## (Intercept)                   0.6204900  1.1745100
## TreatmentFertilised           1.2307487  2.0142513
## TreatmentManure               1.7490820  2.5325847
## TreatmentStopped              0.4824153  1.2659180
## TimeAfter                    -0.4356288  0.5239621
## TreatmentFertilised:TimeAfter 0.3389668  1.6960332
## TreatmentManure:TimeAfter     1.9356334  3.2926999
## TreatmentStopped:TimeAfter   -0.8726999  0.4843666
```

Do things add up now? (we are not aiming for perfection)

Show me the answer

Mostly. Not perfect, but close.

What can you conclude from these results?

Hint

Try and think about what you want to know and what information tells you about this.

I.e. do you want to know how yields changed by treatment? by time? both? will it be the same for all treatments? how much change is happening?

Show me the answer

Need to use confidence intervals to get the conclusions. You can conclude that in the period before 1970 all fertiliser treatments seem to have a positive effect on yield compared to the control. All have estimates of positive mean difference, manure is highest. The effect of time is different for each treatment. For control, no clear direction of pattern (CI span 0. For Fertilised there was an interaction and you see an extra increase in mean after 1970 (0.34 to 1.7 increase) same for manure, but even bigger extra increase (1.94 to 3.29) but for the stopped treatment, you do not see an interaction (effect spans 0). So control and stopped show no difference with time. Makes me wonder if they changed the type of manure/fertiliser in 1970? But we don't know! **Remember, to find the mean of the Treatment groups e.g. fertilised, stopped, and manure, for after 1970 you must add the intercept the effect of time, the effect of the treatment AND the interaction effect.**

Part D: Mixed continuous and categorical variables

What about continuous data? So far, we have treated both categorical and continuous explanatory variables separately but there are many scenarios where we would want to have both types of explanatory variable in a single analysis.

Example 1: **Body length and temperature and water treatment**. Scientists measured the body length of insects in micrometers at different temperatures either with or without rainfall.

Example 2: **Lay date and insect biomass and habitat**. Scientists (including Emily) measured the date in the year when birds lay their eggs (it can take any value even 1.5 or -2). These birds live in different habitats in a woodland: good, bad, ok, which impacts when they lay eggs. You can also measure the biomass of insects close to each nest box. Biomass is measured in grams and there are several nest boxes in each habitat.

Example 3: **Tail length and dinosaur family and age of fossil**. Paleontologists measured the tail length of fossils in cm and recorded the family the dinosaur belonged to and the age of the fossil in million years.

In the examples, which variable is categorical and which is continuous?

Hint

If you are not sure, have a go at drawing out a graph for the data or designing a data collection table. For each variable, would you expect its values to be numbers e.g. a graph with an axis 0-100 (continuous) or would you expect categories e.g. A, B, C or species1, species2, species3 (categorical)?

(like the graphs from the videos in Part A)

Show me the answer

Example 1 = categorical (water) and continuous (temperature).

Example 2 = categorical (habitat) and continuous (biomass).

Example 3 = categorical (family) and continuous (fossil age).

The data for example 1 (made up) can be found here: https://www.math.ntnu.no/emner/ST2304/2020v/Week09/BodyLengthExperiment.csv

It is a csv file and has column headers.

Run an interactive model for example 1.

Code help

```
# Import data

BodyLength <- read.csv("https://www.math.ntnu.no/emner/ST2304/2020v/Week09/BodyLengthExperiment.csv",
                       header=T)

# create the lm object
BodyLengthModel <- lm(length ~ temperature*water, data = BodyLength)

# extract the coefficient estimates
coef(BodyLengthModel)
```

```
##          (Intercept)          temperature             waterYes
##            46.365831             5.191970            25.267954
## temperature:waterYes
##            -3.643074
```

```
# extract confidence intervals
confint(BodyLengthModel)
```

```
##                           2.5 %     97.5 %
## (Intercept)           31.804175 60.927487
## temperature            4.239380  6.144560
## waterYes               4.674663 45.861245
## temperature:waterYes -4.990240 -2.295909
```

Link

What do these coefficients mean? Which parameters do they relate to?

Try and draw out the results on a piece of paper, or an app. i.e. draw what they would look like on a graph

When drawing think about what data you have and how you can best represent it in a graph. Think about what the response variable is, what the explanatory variables are and how you would like to model them e.g. a straight line or several? Or differences in means etc. Don't worry, we will go through this afterwards so it doesn't matter if you are wrong, **just have a go!**

ANSWERS at the end of this section

Interpret the output in biological terms.

Hint

This is where you translate the raw coefficients into something meaningful.

EXAMPLE (not on this data just to show the sort of things to consider):

IF the slope estimate was 7, is not very meaningful on its own. This number should have units for a start.

TO INTERPRET: variable Y is estimated to increase by between 5 and 8 cms for every 1 mm change in X. This suggests that Y gets bigger as X does but much faster. So, if Y is a wing length and X is a body length, you would have a bigger wing to body size ratio as the body gets longer.

Was an interaction term needed in this model?

Hint

This is basically asking, was the interaction term clearly different to 0.

Show me the answers

Link

The (Intercept) is the intercept of the line for temperature when water = No. The next row is the slope of the effect of temperature when water = No. Then the next is the effect of water = Yes, in this case a change in intercept of the line. The final number is the interaction so the change in the effect of temperature when water = Yes, it is a difference in slope.

Plot in R (so a bit neater below too).
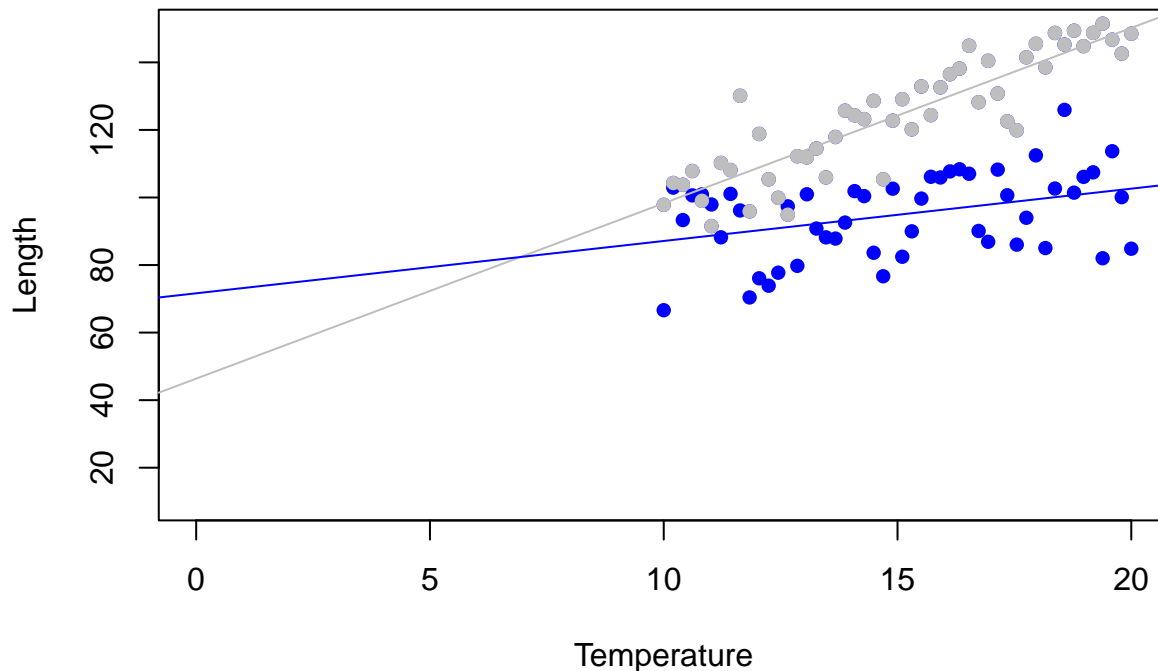
Part E: Continuous and continuous

It is not only categorical variables that can have interactions.

Two continuous variables can also have interactions. So that we give you a complete set of information, we also have a small amount on this here.

There can be many cases in the real world, where two continuous variables influence each others effect such as the amount of rainfall (in mm) might change the relationship between light (in PAR) and plant growth. It turns out - a lot of Emily's work has this in! In the example below, we will use the data from one of the papers from Emily's PhD.

You have already included two continuous variables in a linear model in R during week 7. To add an interaction you do exactly what you did for two categorical variables and use * e.g.:
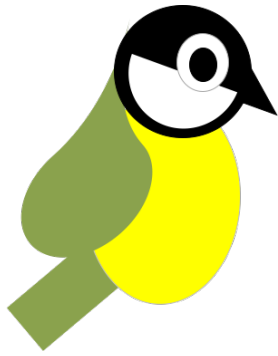
```
lm(y ~ x1 * x2)
```

Example

The example we will look at for continuous interactions is from this paper:

https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.3446

It looks at the timing of incubation onset in great tits (kjøttmeis på norsk). What I wanted to look at is whether birds start incubating their eggs earlier when it is warmer in spring.

To look at this, I put thermometers into nests to see when incubation was started. In one of the analyses, I looked at the relative start of incubation (whether the female great tit started incubating the eggs before she had finished laying or after) and whether this was influenced by when she started laying the eggs (clutch initiation date) or mean temperature.

You can find the data for this analysis here

How to import the data   The link above will download the excel file. This is different to all other data you have used so far, but is good practice for your own analyses. This file is a `.xls` NOT a `.csv` and therefore you can open in Excel or Numbers or another spreadsheet program. You will notice there are A LOT of columns.

What you will need to do is create a `.csv` file that we can use in R. To do this, save as this file and change the file type to **comma separated values** (`.csv`). You might get a warning about losing information or sheets, that is ok, it is what we want. You may also want to change the file name to something more simple.

Remember to save into the same folder R is working from. To check this use `getwd()`.

Back to the analysis

Once you have the data in R, we can work on building an interaction model.

The variables you will want are:

- relative_incu_onset - this is the timing that incubation started relative to when the female finished laying. 0 = same day as last egg, negative = before last egg and positive = after. **This is the response**. It is what I wanted to explain.

- Clutch_initiation_date - this is the date since 1st April when the first egg in the nest was laid.

- mean_temp8 - this is mean temperature at the time of laying in Celsius

Run a linear model to look at whether clutch initiation date and mean temperature impact incubation onset AND whether they have an interaction.

Hint

Remember the `*`

Have a look at the coefficients, how easy are these to interpret? Have a go if you can!

Hopefully you can interpret some of the results. But the interaction might be a bit challenging. Can we look at this in another way? Yes, we can plot the results instead. There are a couple of ways to do this.

Option 1: As a surface (complicated but fun)

The code for this is quite complex (I only learnt it today). If you are interested, this is the page I used https://rpubs.com/aagarwal29/179912

You will need to run the code yourself and the plot pops up in another window. Install the package `rgl` to do this.

This plots the interaction between clutch initiation date and mean temperature as a 3D surface. You can move this plot if you click on it and rotate to see the whole surface. You can see how the relationship between incubation onset and either of the explanatory variables changes with the other.

```r
library('rgl') # you will need to install this package

# import data (you will have saved it under a different name)
BirdData <- read.csv('BirdData.csv', header = TRUE)

# run a 3d plot of the data
plot3d(BirdData$Clutch_initiation_date,
       BirdData$mean_temp8,
       BirdData$relative_incu_start,
       type="s", size=1, lit=FALSE)

# run the interaction model
ContinuousInteractionModel <- lm(relative_incu_start ~ Clutch_initiation_date *
                                     mean_temp8, data = BirdData)

# predict values of the response variable
# this is done for ALL values in the sequence of both explanatory variables in ALL
# combinations. This is what expand.grid does - it makes a big object.

newdata <-  expand.grid(Clutch_initiation_date = seq(min(BirdData$Clutch_initiation_date),
                              max(BirdData$Clutch_initiation_date),
                              length.out = 100),
                        mean_temp8 = seq(min(BirdData$mean_temp8),
                              max(BirdData$mean_temp8),
                              length.out = 100))

# then predict the response
predictions <- predict(ContinuousInteractionModel, newdata, type = NULL)

# Add the mesh of predicted values to the plot

# first save out the unique values of both explanatory variables
x1 <- unique(newdata$Clutch_initiation_date)
x2 <- unique(newdata$mean_temp8)

# save the predictions in a matrix
y <- matrix(predictions, nrow = length(x1), ncol = length(x2))

# then add a 3d surface to the plot
surface3d(x1,
          x2,
          y, alpha=0.4, front="lines", back="lines")
```

Option 2: hold one variable at fixed values

This way is a bit more straightforward and I find it a bit easier to read (that's why I used it in the paper). Here you predict new values of Y (incubation onset) for a sequence of X1 values (mean temperature) but for only a few fixed values of X2 (clutch initiation date).

The steps needed are written in the code and are in the summary video.

```r
# first, run the model
ContinuousInteractionModel <- lm(relative_incu_start ~ Clutch_initiation_date *
                                     mean_temp8, data = BirdData)


# make 3 sets of 'new data' to use for prediction
# the values of mean temperature will all be the same but the value of
# clutch initiation will change

# vector of mean temperature values

mean_temp <- seq(min(BirdData$mean_temp8),
                 max(BirdData$mean_temp8),
                 length.out = 50)

# the first quartile of the clutch initiation date is 8
newdata1 <- data.frame(Clutch_initiation_date = 8,
                       mean_temp8 = mean_temp)

# now clutch initiation is at the median
newdata2 <- data.frame(Clutch_initiation_date = 14,
                       mean_temp8 = mean_temp)

# the third quartile of the clutch initiation date is 16
newdata3 <- data.frame(Clutch_initiation_date = 16,
                       mean_temp8 = mean_temp)

# now we want to make the predictions
predictions1 <- predict(ContinuousInteractionModel, newdata1)
predictions2 <- predict(ContinuousInteractionModel, newdata2)
predictions3 <- predict(ContinuousInteractionModel, newdata3)


### NOW PLOT

# first the basic plot of the data (incubation onset and temperature)

plot(y = BirdData$relative_incu_start,
     x = BirdData$mean_temp8, pch = 16, col = "grey",
     ylab = "Incubation onset", xlab = "Temperature")

# now add the lines

lines(y = predictions1, x = mean_temp, col = "blue")
lines(y = predictions2, x = mean_temp, col = "orange")
lines(y = predictions3, x = mean_temp, col = "purple")

# and a legend

legend("topright", c("clutch initiation = 8",
       "clutch initiation = 14",
       "clutch initiation = 16"),
     col = c("blue", "orange", "purple"), lty = 1,
```
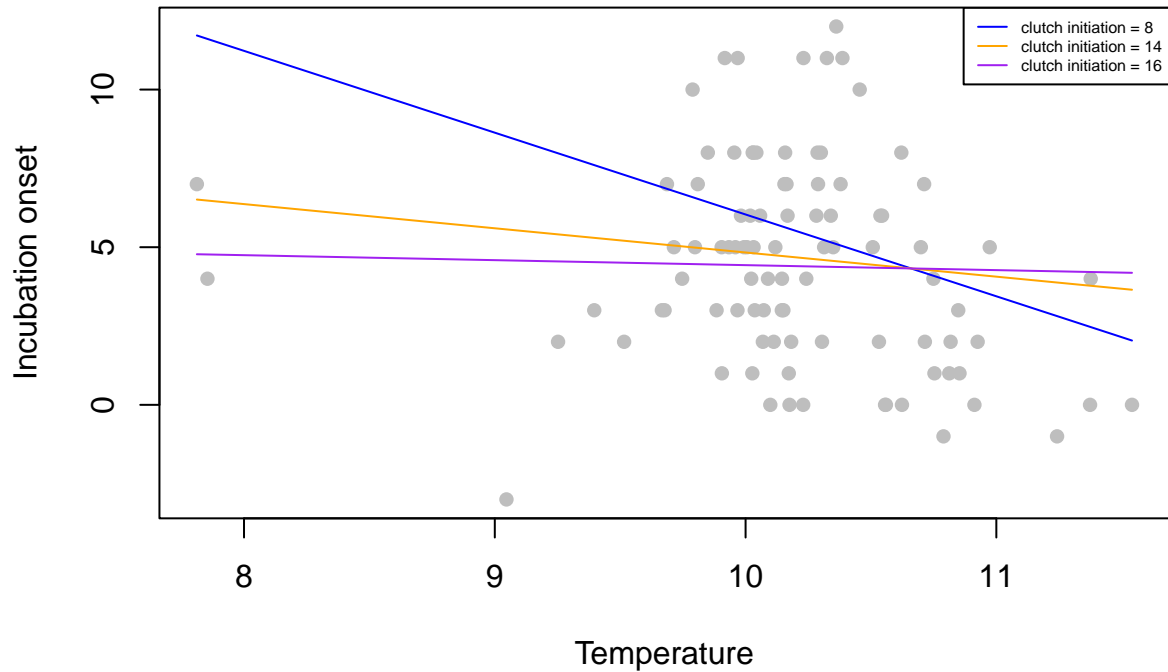
```
cex = 0.5)
```


Using these plots, can you now interpret the interaction?

Summary video on continuous interactions: Link

Part F: Detective skills

Link

Look at the model outputs above, match the following model types to the output:

They are all in different styles and I have deliberately given no information on the type of data. You have to work it out from the output.

Example 1

```
lm( log(Yield)~Ozone*Water)

coef()

(Intercept)           Ozone           WaterStressed Ozone:WaterStressed
8.7544603           -7.8144403           -0.2642307           1.3472822

confint()

(Intercept)              8.5633724  8.945548201
Ozone                  -10.5611355 -5.067745192
WaterStressed           -0.5344698  0.006008381
Ozone:WaterStressed     -2.5371313  5.231695737
```

Example 2

```
Call:
lm(formula = Lay_date ~ Long_term, data = occupancy)

Residuals:
    Min      1Q  Median      3Q     Max
-7.7843 -3.0140 -0.2233  4.2360  6.1953

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.9064     4.2683   4.898 0.000235 ***
Long_term    -0.2035     5.3802  -0.038 0.970368
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.217 on 14 degrees of freedom
  (71 observations deleted due to missingness)
Multiple R-squared:  0.0001021, Adjusted R-squared:  -0.07132
F-statistic: 0.00143 on 1 and 14 DF,  p-value: 0.9704
```

Example 3

```
##
## Call:
## lm(formula = altitude ~ Occupied, data = occupancy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.316  -8.304  -1.985   5.201  30.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   86.003      1.336  64.382   <2e-16 ***
## OccupiedY     -4.158      3.022  -1.376    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.18 on 85 degrees of freedom
## Multiple R-squared:  0.02179,    Adjusted R-squared:  0.01028
## F-statistic: 1.893 on 1 and 85 DF,  p-value: 0.1724
```

Example 4: lm(time ~ mean_max_temp1*April_lay)

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	22.9933197	0.3171830	72.49
mean_max_temp1	-0.5247670	0.0200886	-26.12
April_lay	-0.1569613	0.0114586	-13.70
mean_max_temp1:April_lay	0.0071048	0.0006913	10.28

- One categorical explanatory variable
- One categorical and one continuous explanatory variable
- One continuous explanatory variables
- Two continuous explanatory variables
- Model with an interaction

Show me the answer

- One categorical explanatory variable   Example3

- One categorical and one continuous explanatory variable   Example1

- One continuous explanatory variables   Example2

- Two continuous explanatory variables   Example4

- Model with an interaction   Examples1 and 4

How would interpretation be different for a) a model with an interaction and the one with one categorical and one continuous explanatory variable b) the same data but no interaction?

Show me the answer

With no interaction there will be the same slope for each group but with interaction there is a difference slope for the continuous variable within each group of the categorical.

Final Comment

Interactions are an important part of modelling, because the world is not simple, so effects can act together. Calculating the effects of interactions can be complicated, especially if we get more complex models (e.g. interactions between 3 or 4 covariates). The calculations follow what we do here, so to work out what is happening you just need to take a very deep breath and work out what levels you are using and what is going into the calculation. Then you can let R do the calculations: your job is to understand them.