HW 5

NRG

2025-10-02

```r
library("readxl")
library('car')
```

Libraries

```
## Loading required package: carData
```

```r
library("ggplot2")
library("emmeans")
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

Question 1

Import data

```r
hemp <- read_excel("/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datasets/H

temp <- factor(hemp$Temperature)
fert <- factor(hemp$Fertilizer)
grow <- hemp$`Growth Rate (cm/week)`

df <- data.frame(grow, fert, grow)
```

Research question: What is the effect of temperature and fertilizer on plant growth?

Statistical method: two way anova

Hypothesis statements H0: There is no effect of temperature and fertilizer on plant growth and there is no interaction between temperature and fertilizer on plant growth. H1: Temperature affects plant growth or fertilizer affects plant growth or the effect of fertilizer on plant growth depends on temperature.

Assumptions

```r
model <- aov(grow ~ temp * fert, data = df)

# Independent within and between all factor level combinations
plot((aov(grow ~ temp * fert, data = df)), which=1)
```

Residuals vs Fitted

Fitted values
aov(grow ~ temp * fert)
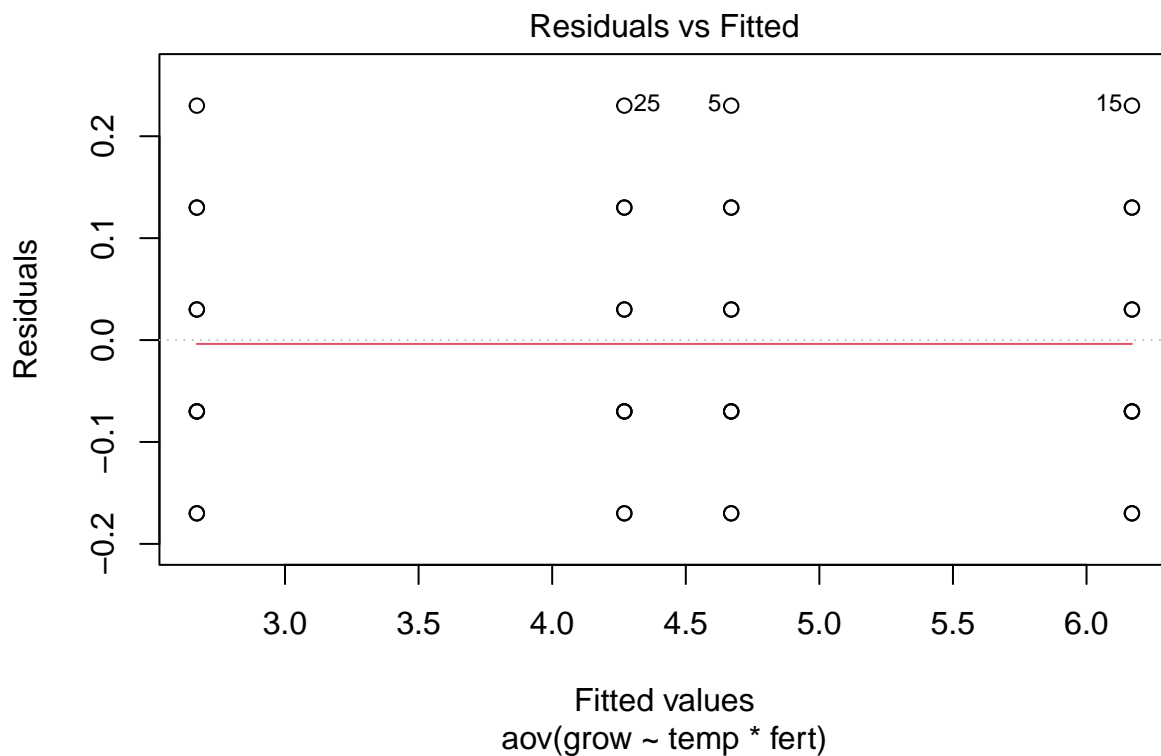
```r
# Pass: data passes as residuals look weird but it is fabricated data so we assume they are okay

# Residuals are normally distributed within each cell (combination of the two factors)
model <- aov(grow ~ temp * fert, data = df)
mf <- model.frame(model)
r  <- residuals(model)
tapply(r, list(mf$temp, mf$fert),
       function(x) if (length(x) >= 3) shapiro.test(x)$p.value else NA)
```

```
##          Added Not Added
## High 0.4663507 0.4663507
## Low  0.4663507 0.4663507
```

```r
# Pass: residuals look normal

# Homogeneity of variance of the dependent variable are equal across all cells
leveneTest(grow ~ interaction(temp, fert), data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##       Df F value Pr(>F)
## group  3       0      1
##       36
```

```
# Passes test

# Count of data and interactive plot to determine ANOVA type
with(df, table(temp, fert))
```

```
##       fert
## temp   Added Not Added
##   High    10        10
##   Low     10        10
```

```r
with(hemp, interaction.plot(
  x.factor = Temperature,
  trace.factor = Fertilizer,
  response = `Growth Rate (cm/week)`,
  fun = mean,
  type = "b",
  ylab = "Growth Rate Mean",
  xlab = "Temperature",
  pch = c(1, 19),
  col = c("blue", "red"),
  trace.label = "Fertilizer"
  )
)
```

Growth Rate Mean vs Temperature (High, Low) with Fertilizer (Added, Not Added)
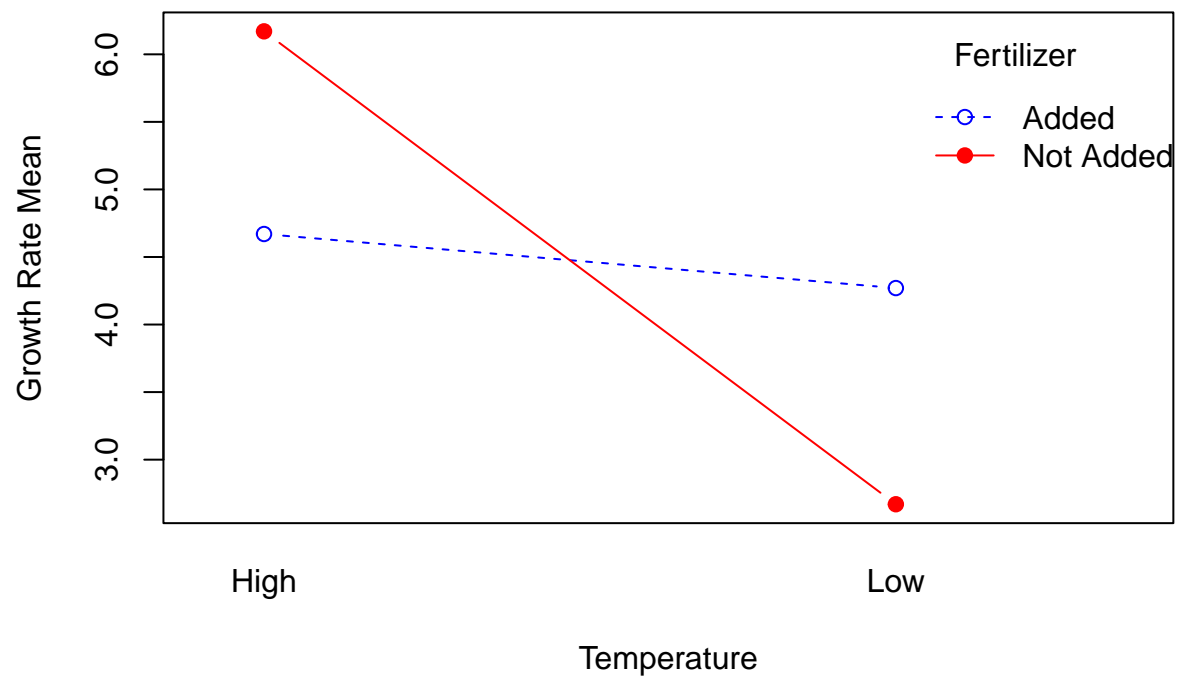
```r
# The crossing lines in the figure suggest that the effect of fertilizer depends on that of temp
# We should use type III.
```

Test

```r
model <- aov(grow ~ temp * fert, data = df)
summary(model)
```

```
##             Df Sum Sq Mean Sq  F value Pr(>F)
## temp         1  38.02   38.02 2125.621 <2e-16 ***
## fert         1   0.02    0.02    1.398  0.245
## temp:fert    1  24.03   24.03 1343.012 <2e-16 ***
## Residuals   36   0.64    0.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Posthoc

```r
emmeans(model, pairwise ~ temp * fert)
```

```
## $emmeans
##  temp fert      emmean     SE df lower.CL upper.CL
##  High Added       4.67 0.0423 36     4.58     4.76
##  Low  Added       4.27 0.0423 36     4.18     4.36
```

```
##  High Not Added   6.17 0.0423 36     6.08    6.26
##  Low  Not Added   2.67 0.0423 36     2.58    2.76
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast                          estimate      SE df t.ratio p.value
##  High Added - Low Added                 0.4 0.0598 36   6.687  <.0001
##  High Added - High Not Added           -1.5 0.0598 36 -25.078  <.0001
##  High Added - Low Not Added             2.0 0.0598 36  33.437  <.0001
##  Low Added - High Not Added            -1.9 0.0598 36 -31.765  <.0001
##  Low Added - Low Not Added              1.6 0.0598 36  26.749  <.0001
##  High Not Added - Low Not Added         3.5 0.0598 36  58.514  <.0001
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

Results Fertilizer has a conditional effect on growth rate given certain temperatures. Under low temperature fertilizer increases growth rate (4.27 vs 2.67) and given high temperature fertilizer decreases growth rate (4.67 vs 6.17). A two-way ANOVA showed significant effects of temperature ($F(1,36) = 44.72$, $p < .0001$), fertilizer ($F(1,36) = 628.88$, $p < .0001$), and their interaction ($F(1,36) = 1343.01$, $p < .0001$). The model explained 98% of the variance ($R^2 = 0.98$). Growth was highest at high temperature without fertilizer (6.17) and lowest at low temperature without fertilizer (2.67); all pairwise differences were significant ($p < .0001$).

Question 2

Import data

```
fruit <- read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datasets/

# Independent variable
acti <- factor(fruit$activity)

# Dependent variable
long <- fruit$longevity

# Covariate
thor <- fruit$thorax

df <- data.frame(long, acti, thor)
```

Research question: What is the effect of female partner treatment on male longevity when accounting for thorax length?

Statistical method: ANCOVA

Hypothesis statements H0: There is no effect of female partners treatment on male longevity, when accounting for thorax length, and there is no interaction between female partner treatment and thorax length. H1: There is an effect of female partners treatment on male longevity, when accounting for thorax length, or there is an interaction between female partner treatment and thorax length.

Assumptions

```
# Observations are independent within and between groups (study design)
# Pass: assumed to be good

# Residuals are normally distributed (residuals vs fitted plot)
plot((aov(long ~ thor + acti, data = df)), which=1)
```

Residuals vs Fitted

Fitted values
aov(long ~ thor + acti)
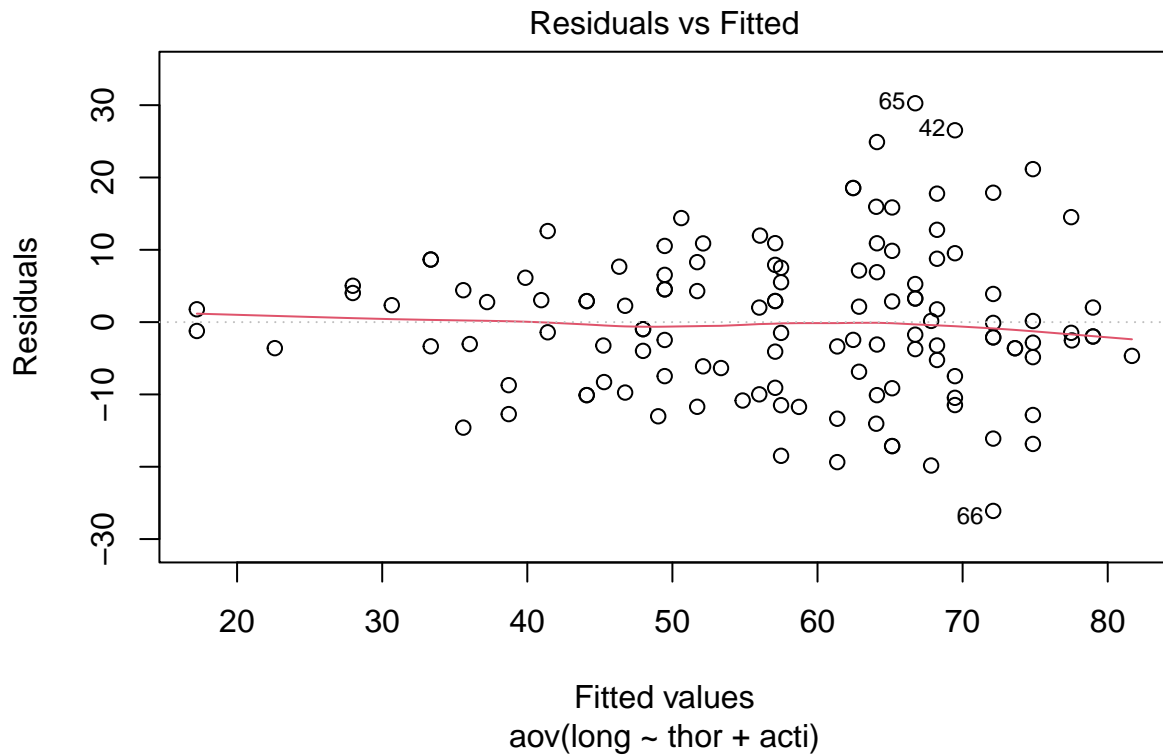
```
# Fail: residuals look screwed to the right

# Attempt log() to fix
plot((aov(log(long) ~ thor + acti, data = df)), which=1)
```

Residuals vs Fitted

Fitted values
aov(log(long) ~ thor + acti)
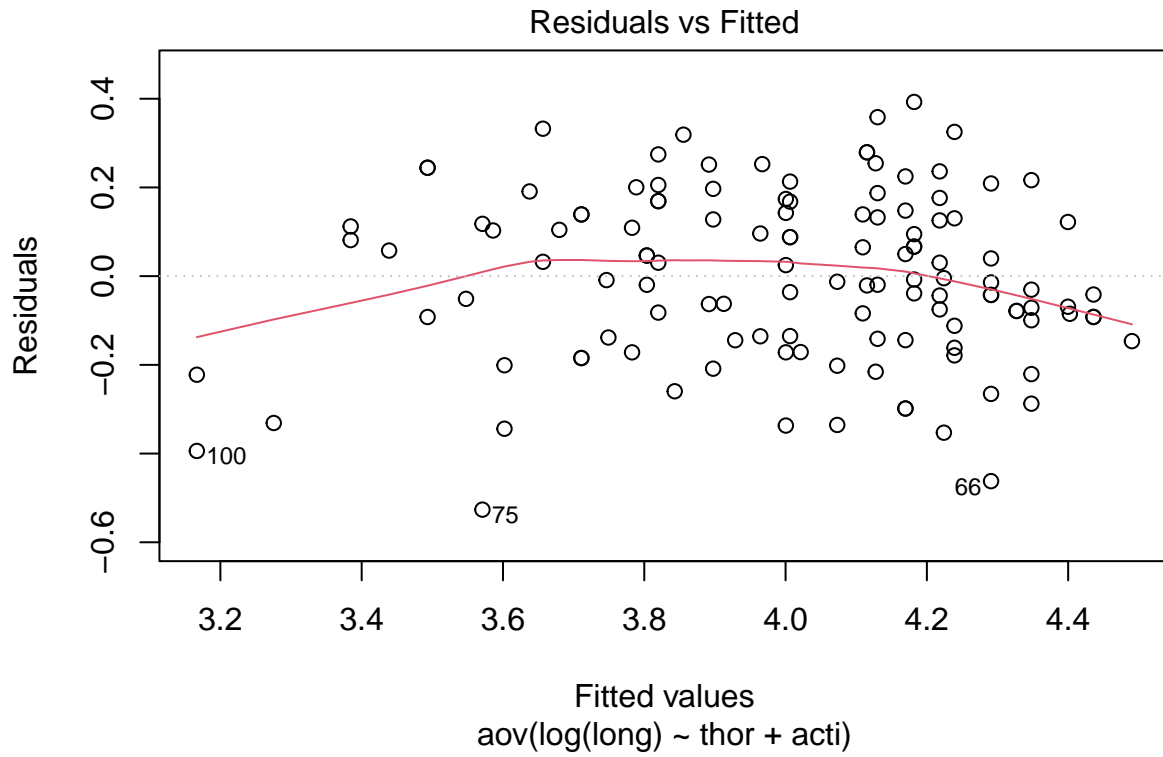
```
# Pass: log() looks better

# Homogeneity of variance across dependent variable groups
leveneTest(log(long) ~ acti, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  1.1247 0.3482
##       119
```
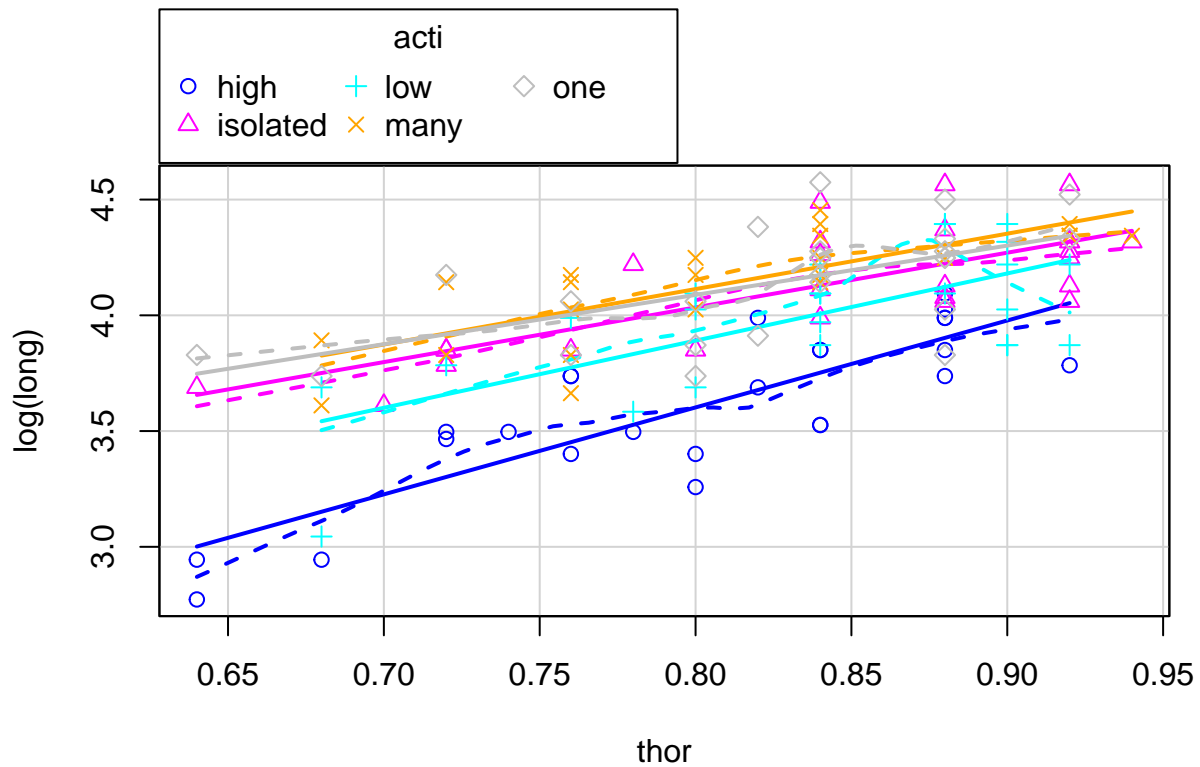
```
# Pass

# Linear relationship between the covariates and the dependent variable
scatterplot(log(long) ~ thor | acti, df)
```

```
# Pass

# Slope between the dependent variable and the covariates remains consistent across all groups
anova(aov(log(long) ~ thor * acti, df))
```

```
## Analysis of Variance Table
##
## Response: log(long)
##             Df Sum Sq Mean Sq  F value Pr(>F)
## thor         1 6.2378  6.2378 170.8335 <2e-16 ***
## acti         4 4.1542  1.0386  28.4428 <2e-16 ***
## thor:acti    4 0.2386  0.0596   1.6334 0.1706
## Residuals  114 4.1626  0.0365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Pass but no interaction

# Balance of data to determine ANOVA type
with(df, table(acti))
```

```
## acti
##     high isolated      low     many      one
##       25       25       25       24       25
```

```
# Not the same number of measurements, so use type III

# Passes all assumptions with log transform
```
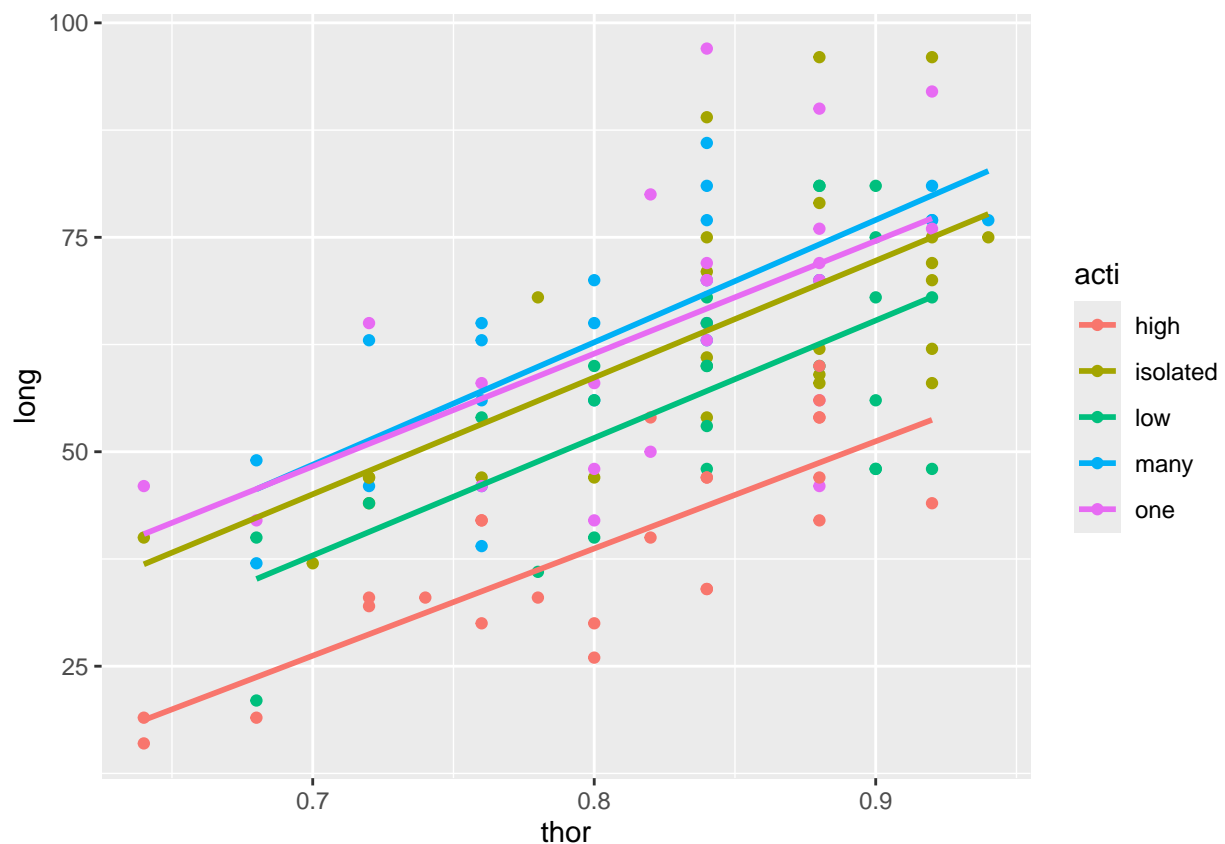
Test

```
summary(aov(long ~ thor * acti, df))
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## thor           1  15003   15003 130.733 < 2e-16 ***
## acti           4   9635    2409  20.988 5.5e-13 ***
## thor:acti      4     24       6   0.053   0.995
## Residuals    114  13083     115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(df, aes(x = thor, y = long, color = acti)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
# Pass: significant p value so we would continue to posthoc but because we only have two levels we dont
```

Results There were significant effects of thorax length ($F(1, 114) = 130.73$, $p < 0.001$) and female partner treatment ($F(4, 114) = 20.99$, $p < 0.001$) on male longevity. The interaction between thorax length and female partner treatment was not significant ($F(4, 114) = 0.05$, $p = 0.99$). Given these results, we accept the alternative hypothesis that male longevity differs among female partner treatments when accounting for thorax length, but reject the part of the hypothesis predicting an interaction. The relationship between thorax length and longevity is positive and consistent across all treatment groups.

Questions 3

Import data

```
cricket <- read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datasets

species <- factor(cricket$Species)
temp <- cricket$Temp_C
pulse <- cricket$pulses

df <- data.frame(species, temp, pulse)
```

Research question: Do pulse rates differ between species of crickets (O. exclamationis and O. niveus) at different temperatures?

Variables Independent variable: temperature Dependent variable: pulse Levels: species

Stats: ANCOVA

Hypothesis statements: H0: Pulse rates do not differ between O. exclamationis and O. niveus when accounting for temperature, and there is no interaction between species and temperature. H1: Pulse rates differ between O. exclamationis and O. niveus when accounting for temperature, and there is an interaction between species and temperature.

Assumptions

```
# Observations are independent within and between groups
# Pass: assumed to be good

# Residuals are normally distributed
plot((aov(pulse ~ temp + species, data = df)), which=1)
```

Residuals vs Fitted

Fitted values
aov(pulse ~ temp + species)
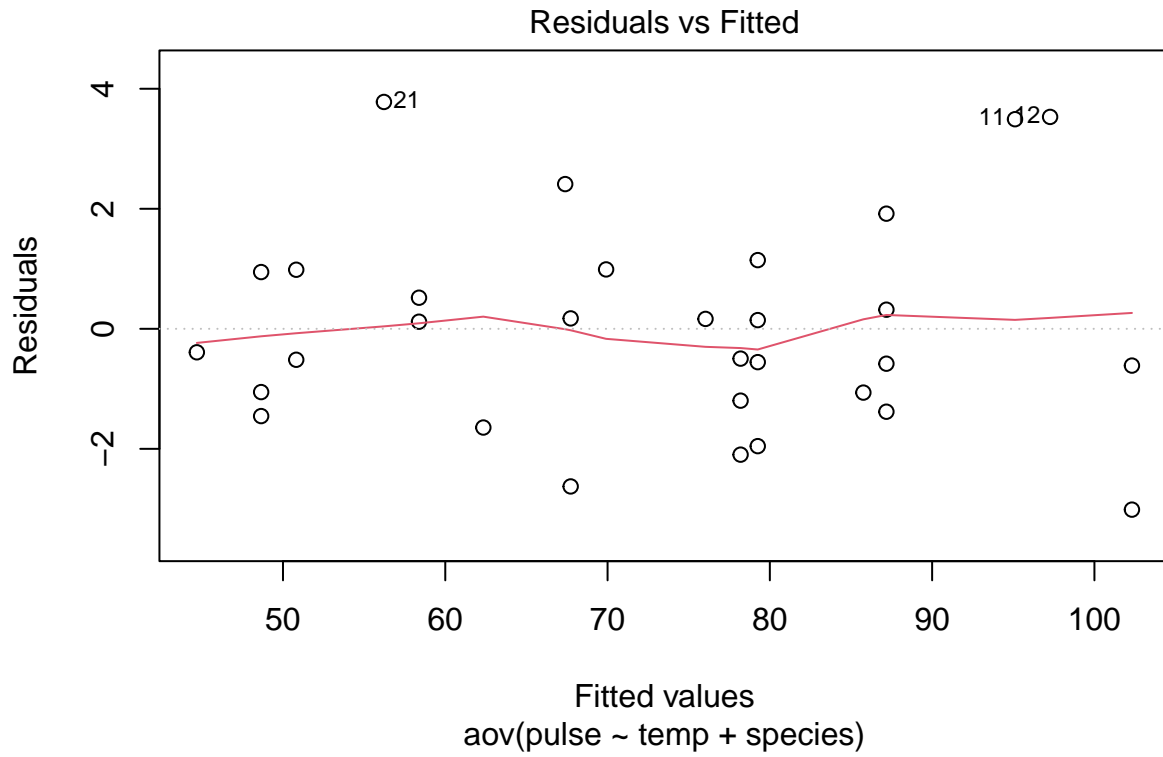
```
# Pass: residuals look randomly distributed

# Homogeneity of variance across dependent variable groups
leveneTest(pulse ~ species, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.4486 0.5083
##       29
```
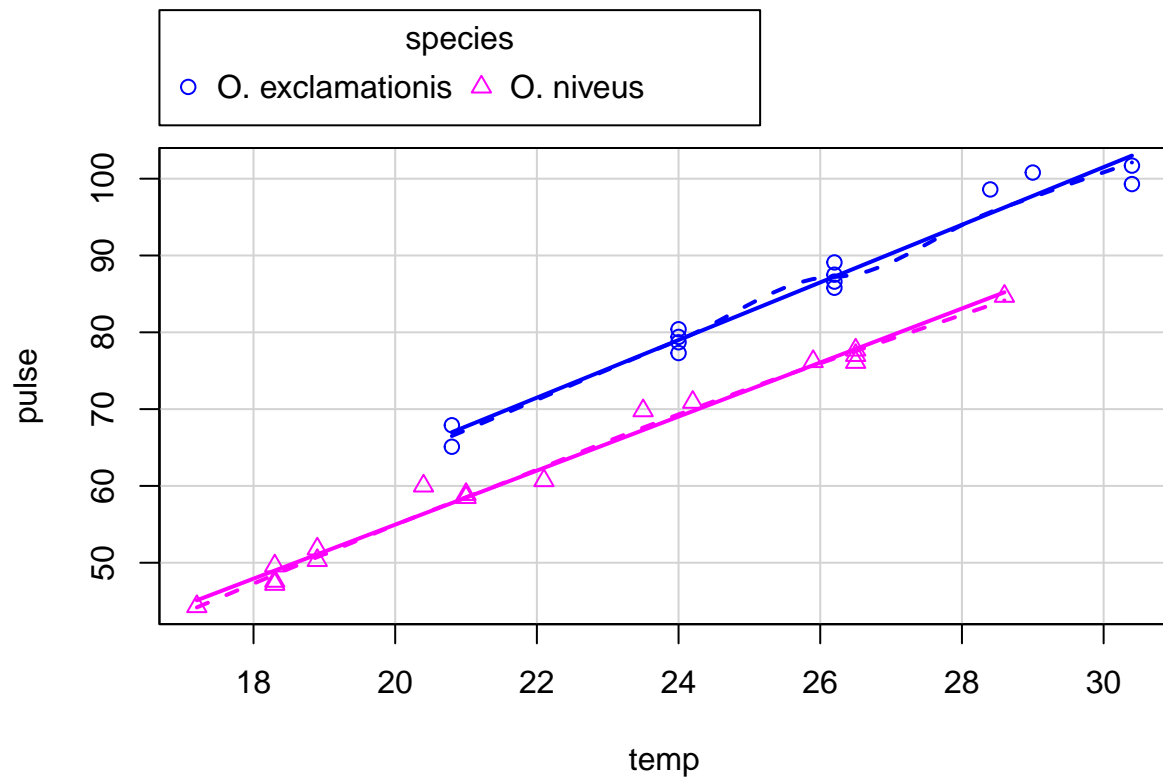
```
# Pass: failed to reject null no evidence of unequal variance

# Linear relationship between the covariates and the dependent variable
scatterplot(pulse ~ temp | species, df)
```

```
# Pass: relationship look linear

# Slope between the dependent variable and the covariates remains consistent across all groups
anova(aov(pulse ~ temp * species, df))
```

```
## Analysis of Variance Table
##
## Response: pulse
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## temp          1 7894.8  7894.8 2505.583 < 2.2e-16 ***
## species       1  598.0   598.0  189.789 9.907e-14 ***
## temp:species  1    4.3     4.3    1.357    0.2542
## Residuals    27   85.1     3.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Pass: we fail to reject null so we assume equal response

# No transformation needed as we pass assumptions

# Balance of data to determine ANOVA type
with(df, table(species))
```

```
## species
## O. exclamationis       O. niveus
```

```
##                    14                17
```

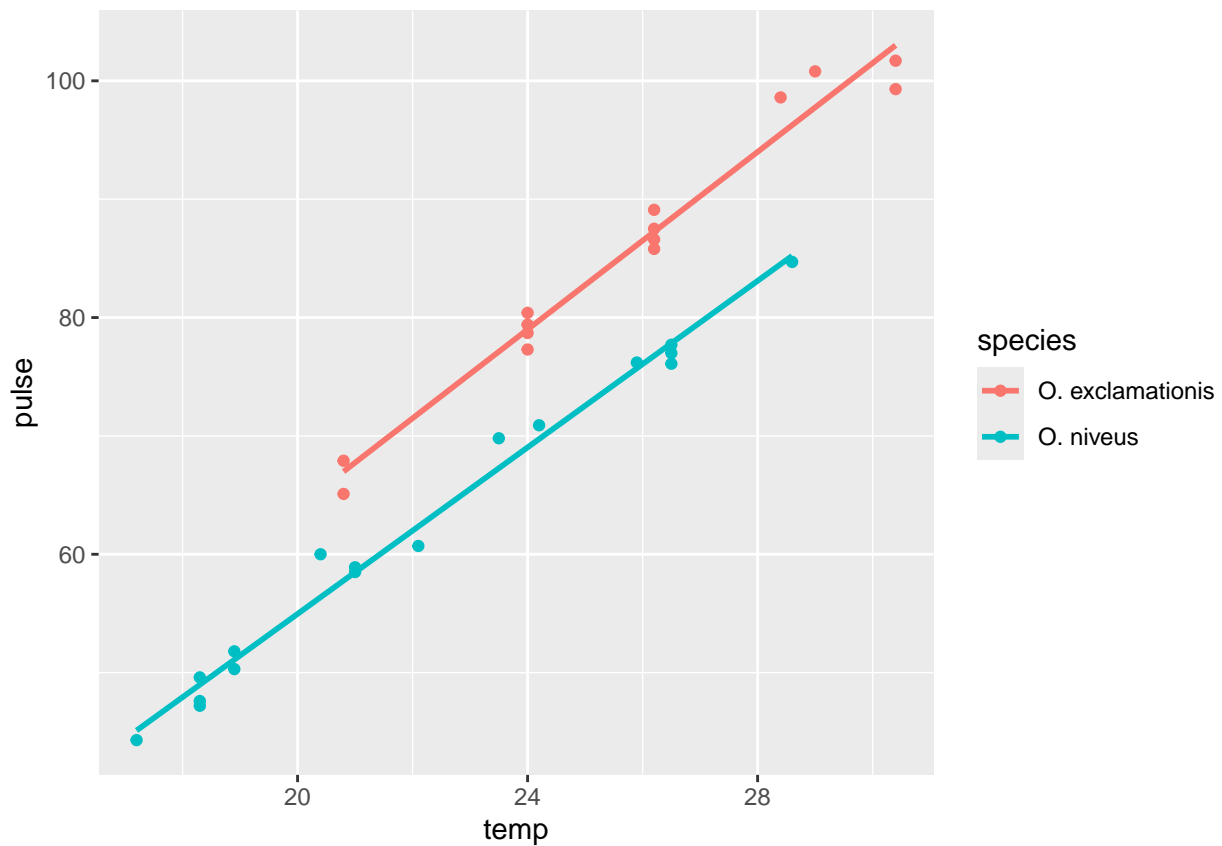Run ANCOVA

```r
summary(aov(pulse ~ temp * species, df))
```

```
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## temp          1   7895    7895 2505.583  < 2e-16 ***
## species       1    598     598  189.789 9.91e-14 ***
## temp:species  1      4       4    1.357    0.254
## Residuals    27     85       3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ggplot(df, aes(x = temp, y = pulse, color = species)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Result: There were significant effects of temperature (F(1, 27) = 2505.58, p < 0.001) and species (F(1, 27) = 189.79, p < 0.001). The interaction between temperature and species was not significant (F(1, 27) = 1.36, p = 0.25). Given these results we can accept our alternative hypothesis that the pulse rates differ between O. exclamationis and O. niveus when accounting for temperature, and there is an interaction between species and temperature. The interaction plot shows us that for each species, as temperature rises, so does pulse rate.