Lab 14 - Repeated Measures ANOVA

Nick Sumpter

2022-09-23

Today's Lab

This lab will teach you how to perform a one-way repeated-measures ANOVA. Repeated-measures ANOVAs test whether there is a difference in the mean value of a variable at levels of a repeated measure (i.e. where each level is measured in the same individual). An example of this would be measuring blood pressure in the same group of people at multiple times of day to see if mean blood pressure changes as a function of time.

Loading Packages and Data

We will need to use a new function for extracting residuals from the package `dae` , which you will need to install.

```
install.packages("dae")
```

```
library(car)
library(dae)
library(ez)
library(pastecs)
library(tidyverse)

theme_set(theme_bw())

setwd("~/Documents/PhD/Teaching/GRD770/R Labs 2022/Lab 14 - Repeated Measures ANOVA")

load("Repeated-Measures-ANOVA.RData")

source("../Lab 6 - Normality and Sample Properties 1/functions.R")
```

Getting to Know Your Data

The dataset we will be using is called `noise`, contained in the "Repeated-Measures-ANOVA.RData" file. This dataset describes performance on a perceptual task with varying levels of background noise. The task involved object recognition in the presence of visual noise. Higher scores indicate worse performance. `noise` has 5 columns:

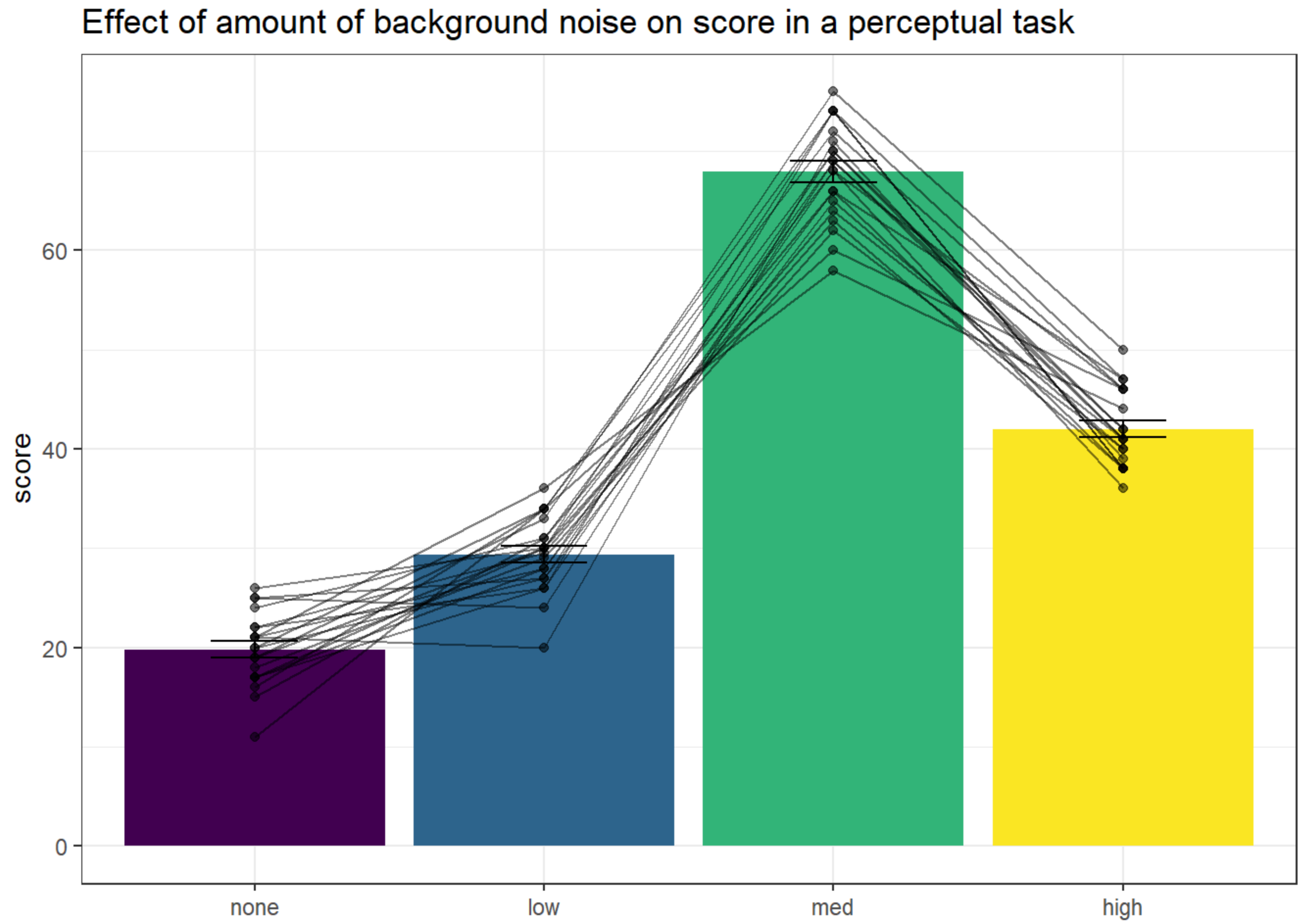
- `ID` (factor, 20 levels): ID number for each subject
- `none` (numeric): score on the task with no background noise
- `low` (numeric): score on the task with low level of background noise
- `med` (numeric): score on the task with medium level of background noise
- `high` (numeric): score on the task with high level of background noise

`none`, `low`, `med`, and `high` are all measures of the same task in the same participants under different conditions and constitute a repeated-measures design.

```
summary(noise)
```

```
##        ID          none            low             med             high      
##  1      : 1   Min.   :11.0   Min.   :20.00   Min.   :58.00   Min.   :36.00  
##  2      : 1   1st Qu.:17.0   1st Qu.:27.00   1st Qu.:64.75   1st Qu.:38.75  
##  3      : 1   Median :20.0   Median :29.75   Median :68.50   Median :41.00  
##  4      : 1   Mean   :19.8   Mean   :29.38   Mean   :67.95   Mean   :42.00  
##  5      : 1   3rd Qu.:22.0   3rd Qu.:31.50   3rd Qu.:71.25   3rd Qu.:46.00  
##  6      : 1   Max.   :26.0   Max.   :36.00   Max.   :76.00   Max.   :50.00  
##  (Other):14
```

```
noise_long <- noise %>%
  pivot_longer(cols = none:high,
               names_to = 'level',
               values_to = 'score') %>%
  mutate(level = factor(level, levels = c("none", "low", "med", "high"), ordered = TRU
E))

ggplot(data = noise_long, mapping = aes(x = level, y = score)) +
  geom_bar(mapping = aes(fill = level), stat = "summary", fun = "mean", show.legend = FA
LSE) +
  geom_errorbar(stat = "summary", fun.data = "mean_se", width = 0.3) +
  geom_point(alpha = 0.5) +
  geom_line(mapping = aes(group = ID), alpha = 0.5) +
  labs(title = "Effect of amount of background noise on score in a perceptual task") +
  theme(axis.title.x = element_blank())
```



Assumptions of Repeated Measures ANOVA

1. Normality of the model residuals at each level of the grouping variable
2. Sphericity (the repeated-measures equivalent of homogeneity of variance)

As we know that each level of noise was measured in the same individuals, we know that the original independence assumption of ANOVA is not met. This does not matter for repeated-measures ANOVA, however, it is still important that each individual (i.e. each row of the data in the `noise` dataset) is independent of each other individual (based on the experimental design).

Normality

For repeated-measures ANOVA, as for other regression based tests, the assumption of normality refers to normality of the model residuals at each level of any input categorical variables. We will therefore test this assumption after running the model.

Sphericity

Sphericity is our second assumption and refers to the condition where the variance of the differences between each level of the grouping measure is approximately equal for each pair of levels. For example, if you had 3 time points, then the variance of the difference between time point 1 and 2 for each individual should be similar to the variance of the difference between time point 1 and 3, and time point 2 and 3. Sphericity is tested automatically when running `ezANOVA`, so there is no reason to run it separately. There is also not any specific requirement that the dataset be spherical. If the dataset isn't spherical, then all that changes is the degrees of freedom at the end. This is shown later in the lab.

Running a Repeated Measures ANOVA

Using `ezANOVA`, the only difference from the one-way independent ANOVA lab is that the `between` argument has been exchanged for the `within` argument.

```
mod <- ezANOVA(data = noise_long,
               dv = score,
               within = level,
               wid = ID,
               type = 3,
               return_aov = TRUE)

mod
```

```
## $ANOVA
##   Effect DFn DFd        F            p p<.05       ges
## 2  level   3  57 482.8101 1.805257e-40     * 0.9525669
##
## $`Mauchly's Test for Sphericity`
##   Effect         W         p p<.05
## 2  level 0.7133265 0.3081504
##
## $`Sphericity Corrections`
##   Effect       GGe        p[GG] p[GG]<.05      HFe        p[HF] p[HF]<.05
## 2  level 0.8560544 5.628836e-35         * 1.001374 1.805257e-40         *
##
## $aov
##
## Call:
## aov(formula = formula(aov_formula), data = data)
##
## Grand Mean: 39.78125
##
## Stratum 1: ID
##
## Terms:
##                 Residuals
## Sum of Squares   272.7344
## Deg. of Freedom        19
##
## Residual standard error: 3.788725
##
## Stratum 2: ID:level
##
## Terms:
##                     level Residuals
## Sum of Squares  26118.834  1027.853
## Deg. of Freedom         3        57
##
## Residual standard error: 4.24647
## Estimated effects are balanced
```

The output of `ezANOVA` is a list of 4 tables (each denoted by a `$` symbol at the start of the line). The first table is called `ANOVA` and contains the main results of the ANOVA test, including the degrees of freedom, F stat, and p-value.

The second table is called `Mauchly's Test for Sphericity` and this is where the sphericity assumption is tested for you by `ezANOVA`. In this case, the p value is greater than 0.05 and thus the sphericity assumption has not been violated. This tells us that the variances of differences in the levels are not significantly different from one another. If the outcome of this test is significant, you will move to the correction factors for departure from sphericity as follows:

1. Look the at the Greenhouse-Geisser epsilon value (denoted by GGe in the `Sphericity Corrections` table).

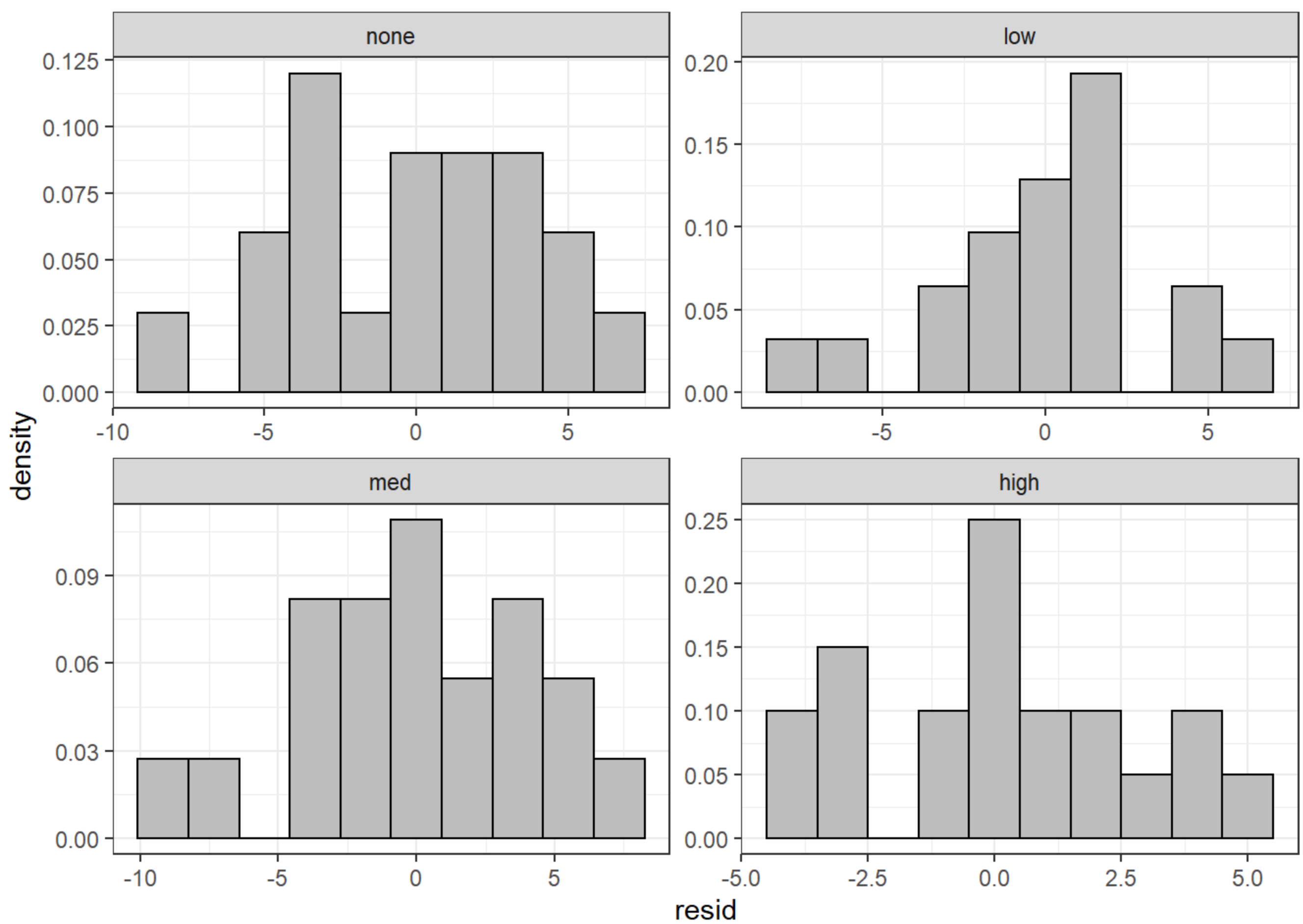
2. If the GGe value is < 0.75, multiply both the numerator and denominator degrees of freedom by the Greenhouse-Geisser epsilon value and replace the p-value with the p[GG] value from the table when you report the results.

3. If the GGe value is > 0.75, multiply the numerator and denominator degrees of freedom by the Huynh-Feldt epsilon value (denoted by HFe) and replace the p-value with the p[HF] value from the table when you report the results.

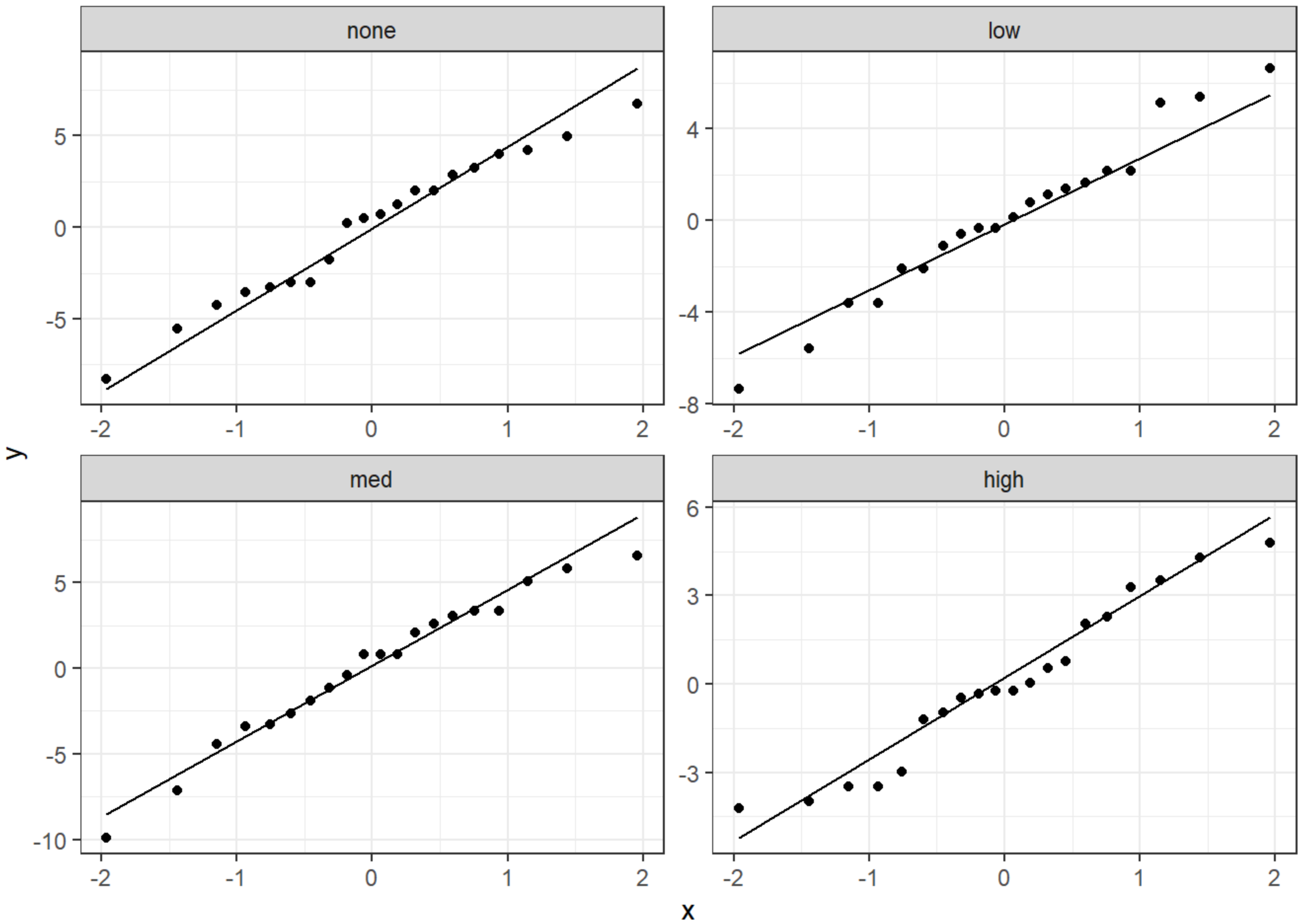
Testing Normality

We can extract residuals in pretty much exactly the same way as for the one-way independent ANOVA. The only difference is that we need to use the `residuals.aovlist` function instead of `resid`

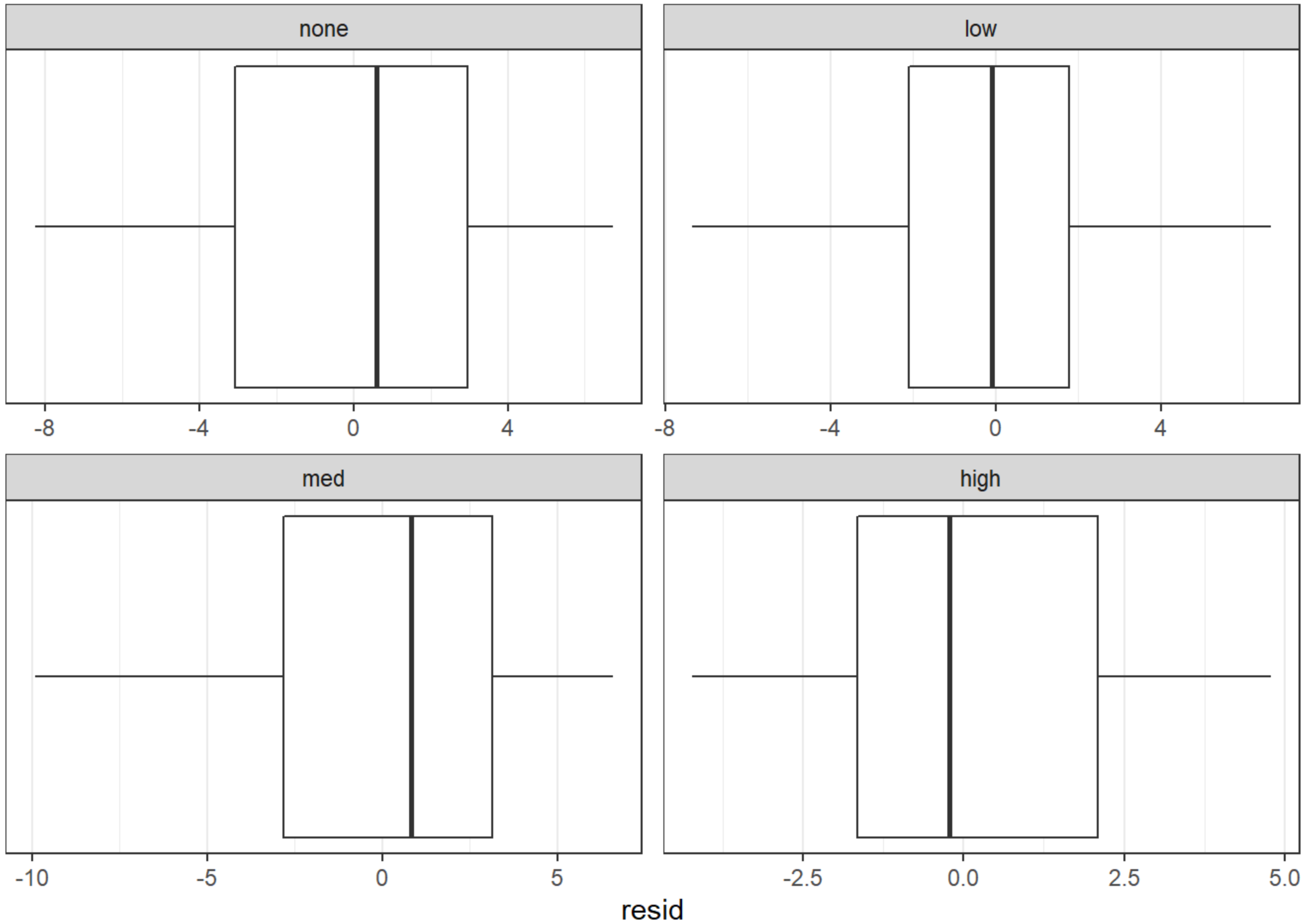
```
# Extracting residuals
residuals <- tibble(group = noise_long$level,
                    resid = residuals.aovlist(mod$aov))

# Normality within groups
ggplot(data = residuals, mapping = aes(x = resid)) +
  geom_histogram(mapping = aes(y = ..density..), bins = 10, fill = 'gray', color = 'blac
k') +
  facet_wrap(~ group, scales = "free")
```



```
ggplot(data = residuals, mapping = aes(sample = resid)) +
  geom_qq() +
  geom_qq_line() +
  facet_wrap(~ group, scales = "free")
```




```
ggplot(data = residuals, mapping = aes(x = resid)) +
  geom_boxplot() +
  theme(axis.ticks.y = element_blank(),
        axis.text.y = element_blank(),
        panel.grid.major.y = element_blank(),
        panel.grid.minor.y = element_blank()) +
  facet_wrap(~ group, scales = "free")
```




```
stat.desc.clean(dataset = residuals, variable = resid, group)
```

```
## # A tibble: 4 × 7
## # Groups:   group [4]
##   group skewness skew.2SE kurtosis kurt.2SE normtest.W normtest.p
##   <ord>    <dbl>    <dbl>    <dbl>    <dbl>      <dbl>      <dbl>
## 1 none   −0.253   −0.247    −0.950   −0.478      0.973      0.813
## 2 low    −0.0728  −0.0711   −0.495   −0.249      0.979      0.914
## 3 med    −0.494   −0.482    −0.507   −0.255      0.967      0.698
## 4 high    0.0909   0.0887   −1.16    −0.585      0.950      0.370
```

Based on this information, the normality assumption was met and thus the model was valid.

Reporting Results from a One-Way Repeated-Measures ANOVA

If the sphericity assumption is met:

"Mauchly's test for sphericity indicated that the assumption of sphericity had been preserved for the main effect of noise level, W = 0.7133, p = 0.308. There was a significant main effect of noise level on score, F(3,57) = 482.81. Post-hoc tests revealed that….."

If the sphericity assumption is violated, you will need to multiply the numerator and denominator degrees of freedom by either the GG or HF ϵ value in your report, and replace the p-value with the corrected p-value next to the test you ran:

“Mauchly’s test for sphericity indicated that the assumption of sphericity had been violated for the main effect of noise level, W = , p = . There was a significant main effect of noise level on score, F(corrected df, corrected df) = F value, p = GG or HF p value. Post-hoc tests revealed that…..”

Independent Practice

For these exercises, we are going to be using the `Loblolly` dataset that is stored in R (load it in with `Loblolly <- Loblolly`). We hypothesize that tree height will significantly differ based on the age of the tree. Investigate this using a one-way repeated measures ANOVA.

Note: you will need to convert the `age` variable to an ordered factor prior to running ezANOVA (see previous labs for how to do this).

1. Get to know your data: describe your variables and their distribution
2. Run the repeated-measures ANOVA
3. Assumptions
    a. Determine whether assumptions are met
    b. If assumptions are not met, describe what you will do to account for this. If possible, modify your model to meet the assumptions
4. Report your findings as you would describe them in the results section