KEY

Bill Kristan

Fri Feb 4 13:13:32 2022

Import the Daphnia data:

library(readxl)  
read\_excel("daphnia.xlsx") -> daphnia

Put the levels of cyandensity in order (but NOT as an ordered factor):

daphnia$cyandensity <- factor(daphnia$cyandensity, levels = c("low","med","high"))

Summarize the data with summarySE:

library(Rmisc)

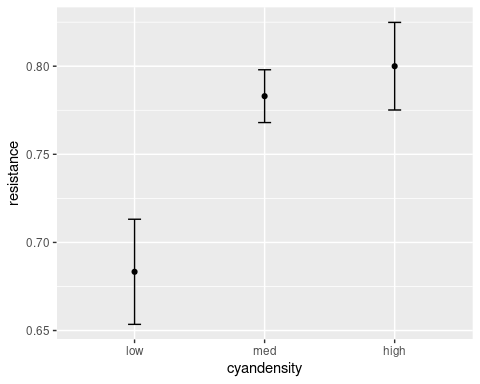
## Loading required package: lattice

## Loading required package: plyr

summarySE(daphnia, "resistance", "cyandensity") -> daphnia.summ

Plot the means +- one standard error:

library(ggplot2)  
ggplot(daphnia.summ, aes(x = cyandensity, y = resistance)) + geom\_point() + geom\_errorbar(aes(ymin = resistance - se, ymax = resistance + se), width = 0.1)



Subset the data so that we can use just the high and low groups first.

subset(daphnia, subset = cyandensity != "med") -> daphnia.high.low

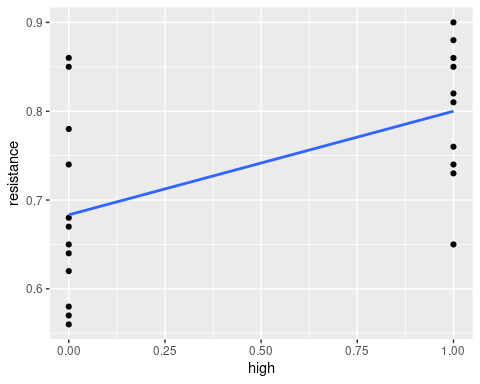
Make a dummy coded variable - low will be the baseline because it now comes first, so high is the dummy coded variable:

as.numeric(daphnia.high.low$cyandensity == "high") -> daphnia.high.low$high

Plot the two group data, using high as the x-axis variable:

ggplot(daphnia.high.low, aes(x = high, y = resistance)) + geom\_point() + geom\_smooth(method = "lm", se = F)

## `geom\_smooth()` using formula 'y ~ x'



Run the analysis as a regression, using high as the predictor variable and resistance as the response:

lm(resistance ~ high, data = daphnia.high.low) -> daphnia.hl.dummy.lm

Now run the analysis as an ANOVA - that is, use cyandensity as a predictor, which is a categorical variable:

lm(resistance ~ cyandensity, data = daphnia.high.low) -> daphnia.hl.categorical.lm

Get the ANOVA tables for both of the models:

cat("Dummy coded version\n\n")

## Dummy coded version

anova(daphnia.hl.dummy.lm)

## Analysis of Variance Table  
##   
## Response: resistance  
## Df Sum Sq Mean Sq F value Pr(>F)   
## high 1 0.074242 0.074242 8.5796 0.008296 \*\*  
## Residuals 20 0.173067 0.008653   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

cat("\n\nCategorical version\n\n")

##   
##   
## Categorical version

anova(daphnia.hl.categorical.lm)

## Analysis of Variance Table  
##   
## Response: resistance  
## Df Sum Sq Mean Sq F value Pr(>F)   
## cyandensity 1 0.074242 0.074242 8.5796 0.008296 \*\*  
## Residuals 20 0.173067 0.008653   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Next get the model summary for each model:

cat("Dummy coded version\n")

## Dummy coded version

summary(daphnia.hl.dummy.lm)

##   
## Call:  
## lm(formula = resistance ~ high, data = daphnia.high.low)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.150000 -0.062500 -0.008333 0.059167 0.176667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.68333 0.02685 25.447 <2e-16 \*\*\*  
## high 0.11667 0.03983 2.929 0.0083 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09302 on 20 degrees of freedom  
## Multiple R-squared: 0.3002, Adjusted R-squared: 0.2652   
## F-statistic: 8.58 on 1 and 20 DF, p-value: 0.008296

cat("\n\nCategorical version\n")

##   
##   
## Categorical version

summary(daphnia.hl.categorical.lm)

##   
## Call:  
## lm(formula = resistance ~ cyandensity, data = daphnia.high.low)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.150000 -0.062500 -0.008333 0.059167 0.176667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.68333 0.02685 25.447 <2e-16 \*\*\*  
## cyandensityhigh 0.11667 0.03983 2.929 0.0083 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09302 on 20 degrees of freedom  
## Multiple R-squared: 0.3002, Adjusted R-squared: 0.2652   
## F-statistic: 8.58 on 1 and 20 DF, p-value: 0.008296

Now repeat the analysis with all three levels. First, dummy code the daphnia data set to have one dummy coded column for med, and a second for high:

daphnia$med <- as.numeric(daphnia$cyandensity == 'med')  
daphnia$high <- as.numeric(daphnia$cyandensity == 'high')

Use the dummy codes in a multiple regression:

lm(resistance ~ med + high, data = daphnia) -> daphnia.dummy.lm

Now use cyandensity as a categorical predictor:

lm(resistance ~ cyandensity, data = daphnia) -> daphnia.categorical.lm

Get the ANOVA tables:

cat("Dummy coded multiple regression\n\n")

## Dummy coded multiple regression

anova(daphnia.dummy.lm)

## Analysis of Variance Table  
##   
## Response: resistance  
## Df Sum Sq Mean Sq F value Pr(>F)   
## med 1 0.014953 0.014953 2.2436 0.144977   
## high 1 0.074242 0.074242 11.1396 0.002329 \*\*  
## Residuals 29 0.193277 0.006665   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

cat("\n\nCategorical predictor\n\n")

##   
##   
## Categorical predictor

anova(daphnia.categorical.lm)

## Analysis of Variance Table  
##   
## Response: resistance  
## Df Sum Sq Mean Sq F value Pr(>F)   
## cyandensity 2 0.089195 0.044598 6.6916 0.004078 \*\*  
## Residuals 29 0.193277 0.006665   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Confirm that summing the SS for high and med gives the SS for cyandensity:

sum(anova(daphnia.dummy.lm)$`Sum Sq`[1:2])

## [1] 0.08919521

Get the model summaries for both models:

cat("Dummy coded\n\n")

## Dummy coded

summary(daphnia.dummy.lm)

##   
## Call:  
## lm(formula = resistance ~ med + high, data = daphnia)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.150000 -0.047500 -0.003167 0.051667 0.176667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.68333 0.02357 28.996 < 2e-16 \*\*\*  
## med 0.09967 0.03496 2.851 0.00794 \*\*   
## high 0.11667 0.03496 3.338 0.00233 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08164 on 29 degrees of freedom  
## Multiple R-squared: 0.3158, Adjusted R-squared: 0.2686   
## F-statistic: 6.692 on 2 and 29 DF, p-value: 0.004078

cat("\n\nCategorical predictor\n\n")

##   
##   
## Categorical predictor

summary(daphnia.categorical.lm)

##   
## Call:  
## lm(formula = resistance ~ cyandensity, data = daphnia)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.150000 -0.047500 -0.003167 0.051667 0.176667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.68333 0.02357 28.996 < 2e-16 \*\*\*  
## cyandensitymed 0.09967 0.03496 2.851 0.00794 \*\*   
## cyandensityhigh 0.11667 0.03496 3.338 0.00233 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08164 on 29 degrees of freedom  
## Multiple R-squared: 0.3158, Adjusted R-squared: 0.2686   
## F-statistic: 6.692 on 2 and 29 DF, p-value: 0.004078