Chloe Lang

Professor Nelson

Lab 9 Modeling 2 Report

November 14, 2021

Q1: Chi-square test **null** hypothesis: There would be no relationship between Brown Creeper presence/absence in edge and interior habitats.

Q2: **Results**: Based on the results of the chi-square test (**1.386e-06**), it appears that Brown Creepers prefer interior habitats as opposed to exterior habitats. This can also be observed in the contingency table, as the number of sightings was significantly higher in interior habitats.

Q3: **Code** for a model fit of penguin body mass as predicted by penguin species:

fit\_species =

lm(

formula = body\_mass\_g ~ species,

data = penguins

)

Q4: **Code** for a model fit of penguin body mass as predicted by sex:

fit\_sex =

lm(

formula = body\_mass\_g ~ sex,

data = penguins

)

Q5: **Code** for model fit of penguin body mass as predicted by species & sex:

fit\_both =

lm(

formula = body\_mass\_g ~ species \* sex,

data = penguins

)

Q6: Conditional **boxplot** of fit\_species model:

Chart, box and whisker chart

Description automatically generated

Q7: Conditional **boxplot** of fit\_sex model:

Chart, box and whisker chart

Description automatically generated

Q8: Conditional **boxplot** of fit\_both model:

Chart, box and whisker chart

Description automatically generated

Q9: Based on the shapes of the boxes, the model I think may have problems fulfilling the homogeneity assumption would be the fit**\_both** model because you are comparing variances with models using different numbers of variables. You are comparing three species while only comparing two types of sexes, which may lead to some issues with fulfilling the assumption as they are heterogenous and not homogenous.

Q10: Bartlett test **null** hypothesis: There will be no differences in variance between the two groups.

Q11: P-value when grouped by species: **0.05005**

Q12: P-value when grouped by sex: **0.03194**

Q13: P-value when grouped by both factors: **0.1741**

Q14: Based on the results of the Bartlett tests, I anticipate there to be issues in homogeneity for the fit\_both model. This is because both the models were grouped by sex OR species, the p-value fell below 0.05, indicating a rejection of the null hypothesis or a rejection in there being no differences in variance. However, when grouped by sex AND species, the p-value was above 0.05, indicating a failure to reject the hypothesis and thereby failing to reject a difference in variance between the two groups.

**CODE USED FOR LAB:**

## Walkthrough

head(catrate)

#Binomial test for Proportions

#Reproductive Success & Failure

n\_success = sum(catrate$success)

n\_years = sum(catrate$years)

binom.test(n\_success, n\_years)

#Reproductive Cat. & Late Filling

late\_fill\_rate = 2/7

normal\_fill\_rate = 1 - late\_fill\_rate

binom.test(

n\_success,

n\_years,

p = normal\_fill\_rate)

binom.test(

n\_success,

n\_years,

p = normal\_fill\_rate,

alternative ='less')

#Two Sample Tests

#Comparing 2 variances

#F-distn EX: Veg data

head(veg)

boxplot(pine ~ treatment, data = veg)

#Variance test

var.test(

pine ~ treatment,

data = veg,

subset = treatment %in% c('control','clipped'))

#F-tests assumes normality

shapiro.test(veg$pine[veg$treatment=="control"])

shapiro.test(veg$pine[veg$treatment=="clipped"])

#Non-parametric Variance Test

fligner.test(

pine ~ treatment,

data = veg,

subset = treatment %in% c('control','clipped'))

bartlett.test(pine ~ treatment, data=veg)

fligner.test(pine ~ treatment, data = veg)

#Comparing 2 sample means

#t-tests

t.test(

pine ~ treatment,

data = veg,

subset = treatment %in% c('control','clipped'))

#Wilcox test

wilcox.test(

pine ~ treatment,

data = veg,

subset = treatment %in% c('control','clipped'))

#Tests for paired samples

control = veg$pine[veg$treatment=='control']

clipped = veg$pine[veg$treatment=='clipped']

t.test(control, clipped, paired=TRUE)

wilcox.test(control, clipped, paired=TRUE)

#Correlation

#Marbled Salamander

disp

plot(disp$disp.rate.ftb, disp$disp.rate.eb)

cor.test(

disp$disp.rate.ftb,

disp$disp.rate.eb,

use='complete.obs')

cor.test(

disp$disp.rate.ftb,

disp$disp.rate.eb,

use='complete.obs',

method='spearman')

#Comparing 2 distn's

plot(

ecdf(disp$disp.rate.ftb),

verticals=TRUE)

plot(

ecdf(disp$disp.rate.eb),

verticals=TRUE,

lty=3,

add=TRUE)

ks.test(disp$disp.rate.ftb,disp$disp.rate.eb)

#Comparing 2+ proportions

#sex-linked killing

prop.test(c(4,16),c(40,250))

#Dependence of variables in a contingency table

#contingency: Chi-square test

owls = matrix(c(16, 9, 4, 11), nrow=2)

rownames(owls) = c("present", "absent")

colnames(owls) = c("old", "young")

chisq.test(owls)

fisher.test(owls)

#Bird habitat data

birdhab = merge(birds, hab, by=c("basin", "sub", "sta"))

table(birdhab$s.edge, birdhab$BRCR > 0)

br\_creeper\_table = table(birdhab$s.edge, birdhab$BRCR > 0)[, 2:1]

#Lab Questions

#Chi-square tests

#Q2 chi-square test on contingency table of brown creeper presence/absence in edge & interior habitats

chisq.test(br\_creeper\_table)

#Building models for ANOVA

require(palmerpenguins)

fit\_fl\_sp =

lm(

formula = flipper\_length\_mm ~ species,

data = penguins)

#Q3 model fit of penguin body mass as predicted by penguin species

fit\_species =

lm(

formula = body\_mass\_g ~ species,

data = penguins

)

#Q4 model fit of penguin body mass as predicted by sex

fit\_sex =

lm(

formula = body\_mass\_g ~ sex,

data = penguins

)

#Q5 model fit for both sex and species

fit\_both =

lm(

formula = body\_mass\_g ~ species \* sex,

data = penguins

)

#Homogeneity Assumption: Graphical

#Q6 Boxplot of fit\_species model

boxplot(body\_mass\_g ~ species, data = penguins,

main = "Penguin Body Mass

by Species",

xlab = "Penguin Species",

ylab = "Penguin Body Mass (g)",

col = "lightblue1")

#Q7 Boxplot of fit\_sex model

boxplot(body\_mass\_g ~ sex, data = penguins,

main = "Penguin Body Mass

by Sex",

xlab = "Sex",

ylab = "Penguin Body Mass (g)",

col ="plum2" )

#Q8 Boxplot of fit\_both model

boxplot(body\_mass\_g ~ species:sex, data = penguins,

main = "Plot of Penguin Body Mass

by Species and Sex",

xlab = " ",

ylab = "Body Mass (g)",

names = c("Female\nAdelie", "Male\nAdelie", "Female\nChinstrap", "Male\nChinstrap", "Female\nGentoo", "Male\nGentoo"),

col = "slategray1")

#Homogeneity Assumption: Barlett Test 1

#Q11 p-value for species

bartlett.test(body\_mass\_g ~ species, data = penguins)

#Q12 p-value for sex

bartlett.test(body\_mass\_g ~ sex, data = penguins)

#Homogeneity Assumption: Bartlett Test 2

#Q13 p-value when grouped by both

dat\_groups = aggregate(

body\_mass\_g ~ sex \* species,

data = penguins,

FUN = c)

str(dat\_groups)

bartlett.test(dat\_groups$body\_mass\_g)