lab9.R

stonehuang

2022-11-09

birds = read.csv("/Users/stonehuang/Documents/environmental\_data/data/bird.sta.csv")  
hab = read.csv("/Users/stonehuang/Documents/environmental\_data/data/hab.sta.csv")  
  
birdhab = merge(  
 birds,  
 hab, by=c("basin", "sub", "sta"))  
  
# Create a contingency table for edge/interior and brown creeper presence/absence  
table(  
 birdhab$s.edge,  
 birdhab$BRCR > 0)

##   
## FALSE TRUE  
## E 144 29  
## I 559 314

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

##   
## TRUE FALSE  
## E 29 144  
## I 314 559

#Q1. The presence/absence of Brown Creepers does not vary between the interior and edge of forest stands.  
#Q2. Brown Creepers show a significant habitat preference because the p-value is lower than 0.05.  
chisq\_BRCR = chisq.test(br\_creeper\_table)  
chisq\_BRCR

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: br\_creeper\_table  
## X-squared = 23.3, df = 1, p-value = 1.386e-06

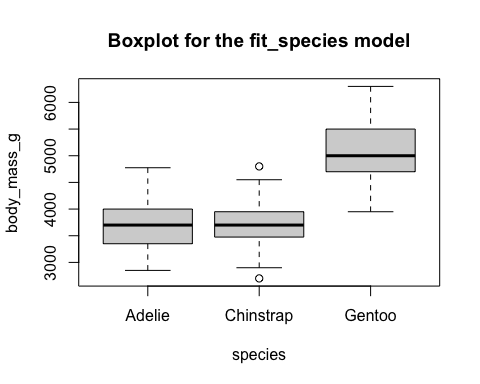
require(palmerpenguins)

## Loading required package: palmerpenguins

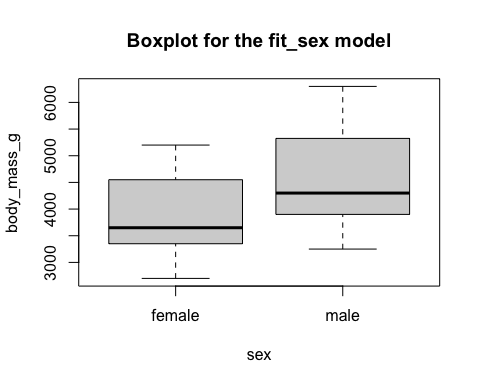
head(penguins)

## # A tibble: 6 × 8  
## species island bill\_length\_mm bill\_depth\_mm flipper\_length\_… body\_mass\_g sex   
## <fct> <fct> <dbl> <dbl> <int> <int> <fct>  
## 1 Adelie Torge… 39.1 18.7 181 3750 male   
## 2 Adelie Torge… 39.5 17.4 186 3800 fema…  
## 3 Adelie Torge… 40.3 18 195 3250 fema…  
## 4 Adelie Torge… NA NA NA NA <NA>   
## 5 Adelie Torge… 36.7 19.3 193 3450 fema…  
## 6 Adelie Torge… 39.3 20.6 190 3650 male   
## # … with 1 more variable: year <int>

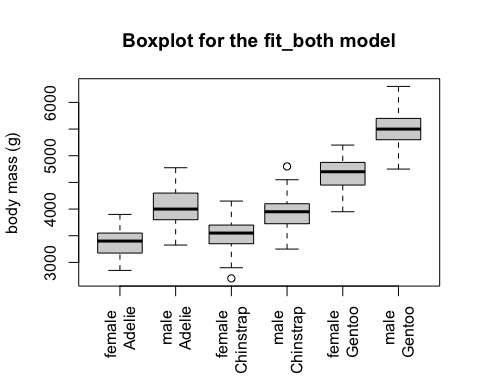
#Q3  
fit\_species = lm(formula = body\_mass\_g ~ species, data = penguins)  
#Q4  
fit\_sex = lm(formula = body\_mass\_g ~ sex, data = penguins)  
#Q5  
fit\_both = lm(formula = body\_mass\_g ~ sex\*species, data = penguins)  
  
#Q6  
boxplot(body\_mass\_g ~ species, data = penguins, main = "Boxplot for the fit\_species model")



#Q7  
boxplot(body\_mass\_g ~ sex, data = penguins, main = "Boxplot for the fit\_sex model")



#Q8  
boxplot(body\_mass\_g ~ sex\*species, data = penguins, las = 3, names = c("female \n Adelie", "male \n Adelie", "female \n Chinstrap", "male \n Chinstrap", "female \n Gentoo", "male \n Gentoo"), ylab = "body mass (g)", xlab = "", main = "Boxplot for the fit\_both model")



#Q9  
#fit\_species model may have problems fulfilling the homogeneity assumption because the boxes have very different width.  
  
#Q10. The null hypothesis of the Bartlett test is that the variances are the same among groups.  
#Q11. 0.0501  
bartlett.test(body\_mass\_g ~ species, data = penguins)

##   
## Bartlett test of homogeneity of variances  
##   
## data: body\_mass\_g by species  
## Bartlett's K-squared = 5.9895, df = 2, p-value = 0.05005

#Q12. 0.0319  
bartlett.test(body\_mass\_g ~ sex, data = penguins)

##   
## Bartlett test of homogeneity of variances  
##   
## data: body\_mass\_g by sex  
## Bartlett's K-squared = 4.6017, df = 1, p-value = 0.03194

#Q13. 0.1741  
dat\_groups = aggregate(  
 body\_mass\_g ~ sex\*species,  
 data = penguins,  
 FUN = c)  
str(dat\_groups)

## 'data.frame': 6 obs. of 3 variables:  
## $ sex : Factor w/ 2 levels "female","male": 1 2 1 2 1 2  
## $ species : Factor w/ 3 levels "Adelie","Chinstrap",..: 1 1 2 2 3 3  
## $ body\_mass\_g:List of 6  
## ..$ : int 3800 3250 3450 3625 3200 3700 3450 3325 3400 3800 ...  
## ..$ : int 3750 3650 4675 3800 4400 4500 4200 3600 3950 3800 ...  
## ..$ : int 3500 3525 3950 3250 4150 3800 3700 3575 3700 3450 ...  
## ..$ : int 3900 3650 3725 3750 3700 3775 4050 4050 3300 4400 ...  
## ..$ : int 4500 4450 4550 4800 4400 4650 4650 4200 4150 4800 ...  
## ..$ : int 5700 5700 5400 5200 5150 5550 5850 5850 6300 5350 ...

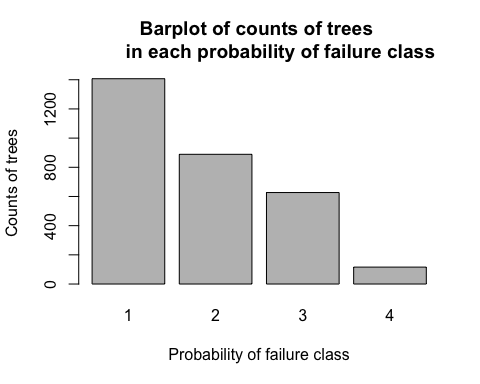
bartlett.test(dat\_groups$body\_mass\_g)

##   
## Bartlett test of homogeneity of variances  
##   
## data: dat\_groups$body\_mass\_g  
## Bartlett's K-squared = 7.6908, df = 5, p-value = 0.1741

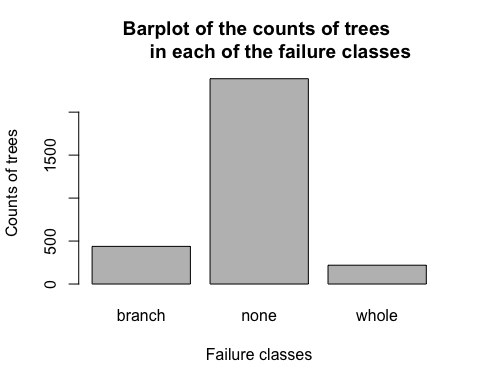
#Q14  
#The fit\_sex model has issues with heterogeneity. The p-value is lower than 0.05, so we reject the null hypothesis that the variances are the same among groups.  
  
#Q15  
dat\_fl = read.csv("/Users/stonehuang/Documents/environmental\_data/data/trees\_FL.csv")  
head(dat\_fl)

## DBH\_in HeighttoTop\_ft Species CanopyWidth\_ft MostSignificantDefect  
## 1 8.0 12.5 Chinese elm 13 Dead  
## 2 8.0 16.0 Chinese elm 20 Decay  
## 3 5.5 15.5 Crapemyrtle 15 Poor Tree Architecture  
## 4 6.2 14.0 Crapemyrtle 24 Poor Tree Architecture  
## 5 16.0 17.0 Crapemyrtle 27 Poor Tree Architecture  
## 6 15.2 20.0 Crapemyrtle 19 Poor Tree Architecture  
## Failure\_Standardized ProbabilityofFailure  
## 1 whole 4  
## 2 whole 3  
## 3 whole 1  
## 4 whole 1  
## 5 none 2  
## 6 none 2

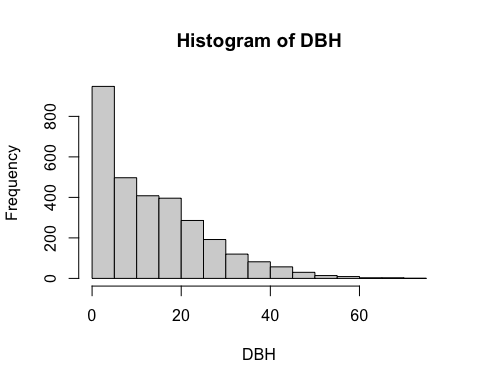
barplot(table(dat\_fl$ProbabilityofFailure), ylab = "Counts of trees", xlab = "Probability of failure class", main = "Barplot of counts of trees   
 in each probability of failure class")



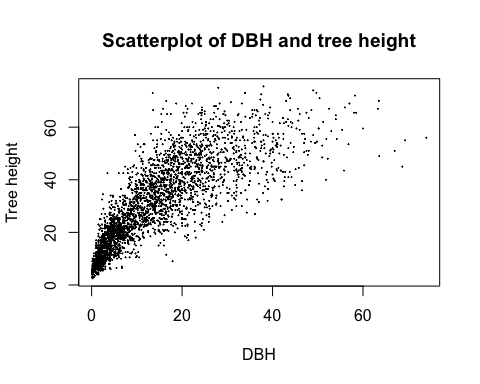
barplot(table(dat\_fl$Failure\_Standardized), ylab = "Counts of trees", xlab = "Failure classes", main = "Barplot of the counts of trees   
 in each of the failure classes")



hist(dat\_fl$DBH\_in, xlab = "DBH", main = "Histogram of DBH")



plot(dat\_fl$HeighttoTop\_ft ~ dat\_fl$DBH\_in, ylab = "Tree height", xlab = "DBH", main = "Scatterplot of DBH and tree height", cex = 0.1)



#Q16. The null hypothesis for the Kolmogorov-Smirnov test is that there is no difference in DBH between whole-tree failures and intact trees.  
#Q17. 0.02125. The distribution of DBH is not the same for the two groups. The p-value is lower than 0.05 so we reject the null hypothesis.  
whole = subset(dat\_fl, Failure\_Standardized == "whole")  
none = subset(dat\_fl, Failure\_Standardized == "none")  
ks.test(whole$DBH\_in, none$DBH\_in)

## Warning in ks.test(whole$DBH\_in, none$DBH\_in): p-value will be approximate in  
## the presence of ties

##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: whole$DBH\_in and none$DBH\_in  
## D = 0.10643, p-value = 0.02125  
## alternative hypothesis: two-sided

#Q18. Tree height increases with increasing DBH. The relationship is curved and monotonic.   
#Q19. Spearman  
cor.test(  
 dat\_fl$DBH\_in,  
 dat\_fl$HeighttoTop\_ft,  
 method='spearman',  
 exact=FALSE)

##   
## Spearman's rank correlation rho  
##   
## data: dat\_fl$DBH\_in and dat\_fl$HeighttoTop\_ft  
## S = 541454081, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.8850462

#Q20. p-value < 2.2e-16.I conclude that the two variables are significantly correlated.  
  
#Q21. X-squared = 202.65, p-value < 2.2e-16  
dat\_fl$fail = factor(dat\_fl$Failure\_Standardized != "none")  
levels(dat\_fl$fail) = c("No Fail", "Fail")  
fl\_table\_2 = table(  
 dat\_fl$ProbabilityofFailure,  
 dat\_fl$fail)  
fl\_table\_2

##   
## No Fail Fail  
## 1 1239 168  
## 2 670 219  
## 3 421 206  
## 4 53 63

chisq\_trees = chisq.test(fl\_table\_2)  
chisq\_trees

##   
## Pearson's Chi-squared test  
##   
## data: fl\_table\_2  
## X-squared = 202.65, df = 3, p-value < 2.2e-16

#Q22. -136  
round(  
 chisq\_trees$observed - chisq\_trees$expected,  
 digits = 0)

##   
## No Fail Fail  
## 1 136 -136  
## 2 -27 27  
## 3 -71 71  
## 4 -38 38

#Q23. There were fewer tree failures than expected by chance in failure probability category #1.  
#Q24. There more tree failures than expected by chance in failure probability category #4.  
#Q25. I conclude that the probability of failure rating system is effective! There were more tree failures than expected in higher failure probability categories.