***Midterm 2***

***Part B***

***PLEASE MAKE SURE YOU READ THIS PART FIRST!***

*Support all of your results statements with* ***statistics****. Be sure* ***to interpret the biological meaning*** *of all results*. *It is your responsibility to* ***confirm that all test assumptions have been met****, but you do not need to show the probability plots, etc. you used to confirm assumptions. All datasets are available on Canvas.*Because all of these designs are balanced, you can use “**anova**” (instead of Anova) for your hypothesis testing.

1. Marisa is interested in how Red Bull affects student performance on exams. She administered three treatments to different students: a can of Red Bull, a can of Coke, or a bottle of water was given to each student just before the exam. Red Bull and Coke both contain sugar and caffeine, but Red Bull contains more caffeine (111mg) compared to Coke (34mg). The water serves as a control for fluid intake. In the class of 45 students, 15 students were each given one of the three treatments, with students assigned randomly to treatments. She recorded the score of each student on the exam.

(5pts) Use the data file “***redbull***” to perform an ANOVA to compare the effects of these treatments.

(5pts) Use Tukey’s post-hoc tests to address whether caffeine and sugar had significant effects on test performance.

(5pts) Make a graph to show differences among treatments.

I tested the effects of sugar and caffeine on student test performance using a one-way ANOVA. The data met the assumptions of normality and homoscedasticity. Ingestion of sugar and caffeine prior to an exam significantly impacted student performance (F = 5.92, df = 2 and 42, p = 0.0054). I used a Tukey HSD post-hoc test to assess the impact of additional caffeine from red bull compared to coke on student test scores. There was no difference in scores among students who drank red bull compared to those who drank coke (p > 0.98), indicating the additional caffeine had no effect. However, compared to those students who drank only water before the test, both red bull and coke significantly increased test scores (predbull = 0.010 and pcoke = 0.016).

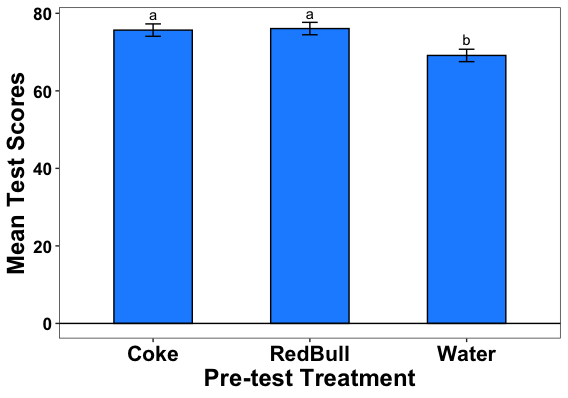


Figure 1. Bar plot showing mean test scores (+/- S.E. and Tukey HSD significance letters) of students who ingested sugar and caffeine (from either coke or red bull) before an exam compared to a control group who only drank water. Exam scores were significantly higher for those students who drank either coke or red bull compared to students who drank water (F = 5.92, df = 2 and 42, p = 0.0054).

2. Erica conducted a lab experiment on the effects of different dietary fats on the endurance of mice. Wild type lab mice were fed diets that were identical except that a portion of each diet was supplemented with a different fat: olive oil, fish oil, or lard. Mice for the experiment were selected randomly from a large colony of 12-week-old mice. Each mouse received each diet in a random order. The endurance of each mouse was tested three times on a treadmill. During each trial, the mouse was run to exhaustion and the time to exhaustion was recorded. Replicate trials on an individual were separated by one week, to allow for recovery, before each mouse was given a different diet. Use the data file “***mousefats***” to answer the following questions:

* 1. (5 pts) Did the type of dietary fat affect endurance of mice?
  2. (5 pts) How much is variance in endurance explained by different mice, by the different responses of mice to fat treatments, and how much is unexplained?

I tested the effects of different dietary fats on the endurance of mice using a repeated measures ANOVA, with treatment (fats supplemented) and trial (weeks tested) as my fixed factors and mouse as my random factor. The data met the assumptions of normality and homoscedasticity. There was no interactive effect of fats and trial (F = 0.178, df = 4 and 72, p = 0.949) nor of trial alone (F = 0.558, df = 2 and 72, p = 0.575) on mouse endurance on a treadmill. The supplemental fat provided to mice significantly impacted endurance (F = 18.3, df = 2 and 72, p = 3.63e-7), where olive oil resulted in the longest average endurance (18.7 time-units, 0.547 S.E.), and fish oil resulted in the shortest (mean = 15.3 time-units, 0.547 S.E.). The different mice used in this experiment explained 14.6% of the endurance variance, while the fat treatments explained 32.4% of endurance variance, and 52% variance was unexplained by our model. Neither trial nor the interaction of fat treatment and trial explained any detectable variation in mouse endurance.

3. Tina conducted a field experiment to test whether the amount of shelter affects recruitment of blue-banded gobies. She constructed replicate habitats with different numbers of holes drilled in them; the gobies used the holes as shelter. She used three shelter treatments: 20, 40, or 60 holes per habitat (“gobitat”). She was also interested in whether predators affect habitat use, so she manipulated the presence of predators using cages that kept predators away from the gobitats and compared these to uncaged gobitats. The experiment was conducted in an area where there was a depth gradient, and the experiment was divided into two blocks—one shallow and one deep. Because Tina was specifically interested in these different depths and whether treatment effects were different between depths, she included block as a fixed factor to test the effect of depth. Each of the shelter\*predator treatment combinations was replicated twice in each block, for a total of 24 gobitats. A month after the gobitats were installed, the total number of young bluebanded gobies that had accumulated (“recruited”) on each gobitat was recorded.

(10pts) Use the data file “***gobyshelter***” to test the effects of shelter, predators, depth, and their interactions. You may either interpret the results of your full model, or perform model selection. **Include a table** with the results from your analysis.

I tested the effects of shelter, predator presence, and depth on recruitment of bluebanded gobies using a three-way fixed factor ANOVA. The data met the assumptions of normality and homoscedasticity. A full model of our factors and interactions yielded a highly insignificant three-way interaction (p=0.627), block-predator interaction (p=0.903), and block-shelter interaction (p=0.761), with a marginally insignificant shelter-predator interaction (p=0.0587). I tested AIC values for all possible reduced models (where interactions were systematically removed), and the best-fit model (AIC = 160) included all main effects and the shelter-predator interaction as fixed factors. Results from the reduced model are displayed in Table 1. Using the reduced model, there was a 164% increase in goby recruitment from the 20-hole shelters to the 60-hole shelters, showing that goby recruitment is heavily impacted by available shelter. There were 1.3 times more young gobies at the deep sites compared to the shallow, and more than twice as many gobies at the caged sites, where no predators were present. The greatest goby recruitment occurred at sites with 60 shelter holes and no predators (mean = 37.25, 3.3 S.E.) compared to the lowest recruitment at 40-hole sites with predator presence (mean = 7.00, 3.3 S.E.).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | *F* | *p* |
| Block (B) | 240.7 | 1 | 240.7 | 6.93 | 0.017 |
| Shelter (S) | 1542.3 | 2 | 771.2 | 22.2 | 1.8e-5 |
| Predators (P) | 1350.0 | 1 | 1350.0 | 38.9 | 9.1e-6 |
| SxP | 316.0 | 2 | 158.0 | 4.55 | 0.026 |
| Residual | 590.3 | 17 | 34.73 |  |  |

Table 1. Three-way Fixed Factor ANOVA Table displaying effects all factors and interactions from a best-fit reduced model, where all remaining factors and interactions were significant predictors of goby recruitment.

4. Hannah is concerned that increased abundance of octocorals on coral reefs is decreasing the health of stony corals. She establishes 40 plots. In half of the plots, she removes all octocorals. In the other half, she leaves octocorals intact. After one month, she measures the photosynthetic rate of one randomly chosen stony coral colony in each plot. She also suspects that photosynthetic rate (Fv/Fm) might be affected by the size of the coral, so she also measures coral colony diameter (cm).

(10pts) Analyze the data in the file “***octocorals*”** and interpret the effects of octocorals and coral colony size on photosynthetic rate. Provide an ANOVA table.

(5pts) Provide a graph that summarizes your results.

I tested the effects of octocoral removal and octocoral size (colony diameter) on photosynthetic rates (Fv/Fm) of stony corals on reef plots using an ANCOVA with octocoral presence and size as my fixed covariates. I could not achieve linearity between the response variable and covariate. Model was run with this missing assumption, but still having met normality and homoscedasticity. The interaction of Treatment and Colony Diameter was a significant predictor of photosynthetic efficiency (F = 5064.3, df = 1 and 36, p < 2.2e-16).

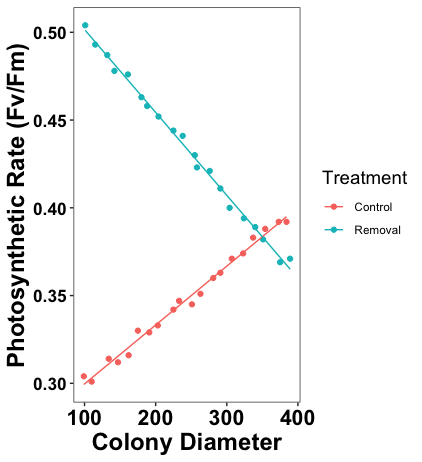


Figure 2. Plot displaying photosynthetic rate of stony corals (Fv/Fm) against coral colony diameter in control and octocoral removal plots. The removal of octocorals allowed the increase of photosynthetic rates linearly for smaller coral colonies, while plots with octocoral presence required greater colony diameters to achieve increased photosynthetic rates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | *F* | *p* |
| Treatment (T) | 0.0756 | 1 | 0.0756 | 7913.1 | <2.2e-16 |
| Diameter (D) | 0.0014 | 1 | 0.0014 | 147.8 | 2.6e-14 |
| TxD | 0.0483 | 1 | 0.0484 | 5064.3 | <2.2e-16 |
| Residual | 0.0003 | 36 | 0.0000 |  |  |

Table 2. ANCOVA Table displaying effects all factors and interactions, where the interaction of Treatment and Colony Diameter was a significant predictor of photosynthetic efficiency.

Code: (for what it’s worth, since we’re not killing trees over it!)

# Danielle Barnas Midterm 2 Part B

# October 22, 2020

####################################################

# Question 1

library(tidyverse)

library(car) # qqp() and Anova()

library(emmeans) # post-hoc test emmeans()

library(agricolae) # HSD.test()

library(lme4) # for testing random effects

library(lmerTest) #Need this to get anova results from lmer

library(MASS)

rm(list=ls())

mydata<-read\_csv("Data/Midterm2/redbull.csv")

View(mydata)

# as.factor

mydata$Treatment<-as.factor(mydata$Treatment)

# create model one-way anova

model<-aov(testscore~Treatment, mydata)

# check assumptions

plot(model)

qqp(residuals(model),"norm")

# stats

anova(model)

# post-hoc Tukey test (requires aov())

TukeyHSD(model)

# HSD.test (gives you letters for different groups which is convenient)

HSD.test(model, "Treatment", console=TRUE)

# Graph data means

graphdata<-as.data.frame(emmeans(model, "Treatment"))

graphdata

# append a new column to graph data with the Tukey results in order

graphdata$tukey<-list("a","a","b")

graphdata

ggplot(data=graphdata, aes(x=Treatment, y=emmean, fill=Treatment)) +

theme\_bw()+

theme(axis.text.x=element\_text(face="bold", color="black", size=16), axis.text.y=element\_text(face="bold", color="black", size=13), axis.title.x = element\_text(color="black", size=18, face="bold"), axis.title.y = element\_text(color="black", size=18, face="bold"),panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) +

geom\_bar(colour="black", fill="DodgerBlue", width=0.5, stat="identity") +

guides(fill=FALSE) +

ylab("Mean Test Scores") +

xlab("Pre-test Treatment") +

#scale\_x\_discrete(labels=c("Low" = "Low", "Ambient" = "Ambient", "1.5N" = "1.5 N", "2N" = "2 N")) +

geom\_errorbar(aes(ymax=emmean +SE, ymin=emmean - SE), stat="identity", position=position\_dodge(width=0.9), width=0.1)+

geom\_text(aes(label=tukey), position = position\_dodge(width=0.9), vjust=-1)+

geom\_hline(yintercept=0)

####################################################

# Question 2

rm(list=ls())

mydata<-read\_csv("Data/Midterm2/mousefats.csv")

View(mydata)

# as.factor

mydata$fat<-as.factor(mydata$fat)

mydata$mouse<-as.factor(mydata$mouse)

mydata$Trial<-as.factor(mydata$Trial)

# create model, repeated measures ANOVA

model<-lmer(endurance~fat\*Trial + (1|mouse),mydata)

# check assumptions

plot(model)

qqp(residuals(model),"norm")

# stats

anova(model)

# interaction effect is insignificant

# test reduced model

reduced<-lm(endurance~fat\*Trial, mydata)

AIC(model)

AIC(reduced)

#LRT

anova(model, reduced)

# full model is a better fit

emmeans(model, ~fat)

# variance components - random factors model

model1<-lmer(endurance~1 + (1|fat) + (1|Trial) + (1|fat:Trial) + (1|mouse), data=mydata)

# check assumptions

plot(model1)

qqp(residuals(model1),"norm")

# get variance components

summary(model1)

1.327+0+0+2.943+4.802 # total = 9.072

1.327/9.072\*100 # mouse

2.943/9.072\*100 # fat

4.802/9.072\*100 # residual

####################################################

# Question 3

rm(list=ls())

mydata<-read\_csv("Data/Midterm2/gobyshelter.csv")

View(mydata)

# fixed: Block, Shelter, Predators (no = cage, yes = uncaged)

# nested: Shelter:Gobitat #

# response: goby recruits (# of young bluebanded gobies accumulated)

mydata<-mydata%>%

rename(gobitat='Gobitat #', recruits='goby recruits')

# as.factor

mydata$Block<-as.factor(mydata$Block)

mydata$Shelter<-as.factor(mydata$Shelter)

mydata$Predators<-as.factor(mydata$Predators)

mydata$gobitat<-as.factor(mydata$gobitat) # doesn't actually matter. won't include in model because each gobitat = each data point

# create model, three-way fixed ANOVA

model<-lm(recruits~Block\*Shelter\*Predators,mydata)

# check assumptions

plot(model)

qqp(residuals(model),"norm")

# check a log transformation just in case for homoscedasticity (also tried square root and didn't help)

mydata$logrecruits<-log(mydata$recruits)

model2<-lm(logrecruits~Block\*Shelter\*Predators,mydata)

plot(model2) # nope that's worse

qqp(residuals(model2),"norm")

# stats on full model

anova(model)

# no significance of any interactions

# check reduced models

reduced1<-lm(recruits~Block + Shelter + Predators + Block:Shelter + Block:Predators + Shelter:Predators,mydata) #remove three-way interaction

reduced2<-lm(recruits~Block + Shelter + Predators + Block:Shelter + Block:Predators,mydata) #remove three-way interaction and Shelter:Predators

reduced3<-lm(recruits~Block + Shelter + Predators + Block:Shelter + Shelter:Predators,mydata) #remove three-way interaction and Block:Predators

reduced4<-lm(recruits~Block + Shelter + Predators + Block:Predators + Shelter:Predators,mydata) #remove three-way interaction and Block:Shelter

reduced5<-lm(recruits~Block + Shelter + Predators + Block:Shelter ,mydata) #remove three-way interaction and Shelter:Predators+ Block:Predators

reduced6<-lm(recruits~Block + Shelter + Predators + Block:Predators,mydata) #remove three-way interaction and Shelter:Predators+ Block:Shelter

reduced7<-lm(recruits~Block + Shelter + Predators + Shelter:Predators,mydata) #remove three-way interaction and Block:Predators+ Block:Shelter

reduced8<-lm(recruits~Block + Shelter + Predators,mydata) #remove three-way interaction and Block:Predators+ Block:Shelter + Shelter:Predators

AIC(model)

AIC(reduced1)

AIC(reduced2)

AIC(reduced3)

AIC(reduced4)

AIC(reduced5)

AIC(reduced6)

AIC(reduced7) # lowest AIC lm(recruits~Block + Shelter + Predators + Shelter:Predators,mydata)

AIC(reduced8)

# LRT

anova(model, reduced7)

# not actually different

anova(reduced7)

# table included in answer

anova(model) # to give insignificant p values

# check assumptions of new model

plot(reduced7)

qqp(residuals(reduced7),"norm")

# estimated marginal means

emmeans(model, ~Block)

emmeans(model, ~Shelter)

emmeans(model, ~Predators)

emmeans(model, ~Shelter\*Predators)

# to calculate effect of shelter:

# 60 hole mean(31.2) - 20 hole mean(11.8) = 19.4 / 20 hole mean(11.8)

# = 1.64 \* 100 = 164% increase

####################################################

# Question 4

rm(list=ls())

mydata<-read\_csv("Data/Midterm2/octocorals.csv")

View(mydata)

# fixed: Treatment, ColonyDiameter

# ignore: Plot. one plot per measurement

# as.factor

mydata$Treatment<-as.factor(mydata$Treatment)

# create model

model<-lm(Photosynthesis~Treatment\*ColonyDiameter,mydata)

#Check assumptions

plot(Photosynthesis~ColonyDiameter, data=mydata) # not linear

plot(model)

qqp(residuals(model),"norm")

#transform

mydata$logP<-log(mydata$Photosynthesis)

plot(logP~ColonyDiameter, data=mydata) # not linear

mydata$logCD<-log(mydata$ColonyDiameter)

plot(logP~ColonyDiameter, data=mydata) # not linear

plot(logP~logCD, data=mydata) # not linear

mydata$sP<-sqrt(mydata$Photosynthesis)

plot(sP~ColonyDiameter, data=mydata) # not linear

plot(sP~logCD, data=mydata) # not linear

plot(Photosynthesis~logCD, data=mydata) # not linear

mydata$sCD<-sqrt(mydata$ColonyDiameter)

plot(Photosynthesis~logCD, data=mydata) # not linear

plot(sP~logCD, data=mydata) # not linear

plot(logP~sCD, data=mydata) # not linear

plot(sP~sCD, data=mydata) # not linear

# cannot achieve linearity

#Double-check assumptions

plot(model)

qqp(residuals(model),"norm")

anova(model)

#To plot the data:

predP<-predict(model) #Gets the predicted values from the regression lines in the ANCOVA

graphdata<-cbind(mydata, predP) #attaches those predictions to the dataset

ggplot(data=graphdata, aes(x=ColonyDiameter, y=Photosynthesis, color=Treatment)) +

theme\_bw()+

theme(legend.title=element\_text(colour="black", size=14), axis.text.x=element\_text(face="bold", color="black", size=16), axis.text.y=element\_text(face="bold", color="black", size=13), axis.title.x = element\_text(color="black", size=18, face="bold"), axis.title.y = element\_text(color="black", size=18, face="bold"),panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) +

geom\_point() + geom\_line(aes(y=predP)) +

labs(x="Colony Diameter", y="Photosynthetic Rate (Fv/Fm)", fill="Treatment")