Biostatistics: Theory and Applications in R (Virtual)

Week5_Session2_R_training5

Contents

Prof Dr Mohammed Abu Sayed Arfin Khan
Department of Forestry and Environmental Science
Shahjalal University of Science and Technology, Sylhet
+8801917174537, khan-for@sust.edu, nobelarfin@yahoo.com
Homepage | Google scholar | Researchgate | ORCID | Publons | BayCEER

#Task week 4

```
###############
# Use the study_1 data set to create the following box plot:
# title="Write your name",
# subtitle="Nursery experiment",
# y axis level ="Species Biomass (gm)",
# x axis level ="Species number",
# caption = "Task: Week 4 R training course")
# plot title text (size=18, color = "green"),
# plot subtitle text(size =16, color = "red"),
# x axis title text(size=14),
# y axis.title text(size=14),
# x axis text(size=14),
# y axis text(size=14),
# all text (color="black"),
# plot.background=element_rect(fill="grey"),
# add plot margin = unit(c(1, 1, 1, 1), "cm"))
# legend position="right",
# y axix limits (ylim=c(0, 350))
# export the plot to a JPEG file


#Set the working directory- getwd()/ setwd("Y:/")
getwd()
setwd("C:/Users/Fahmida Sultana/Desktop/R training/R_training_Class_05")


#install openxlsx package or xlsx package
library(openxlsx)
library(readxl)


#####import data set from xlsx file
study_1 <- read.xlsx("Tree_height.xlsx")
str(study_1)

study_2 <- read_excel("Tree_height.xlsx", sheet = "study_2")
str(study_2)
```

#Task Week 4: box plot script

```
library(ggplot2)
ggplot(study_1, aes(x = plant, y = biomass, fill = plant)) +
  stat_boxplot(geom = "errorbar", width = 0.25) +
  geom_boxplot() +
  guides(fill = guide_legend(title = "Plant"))+
  labs(title="Write your name",
       subtitle="Nursery experiment",
       y="Species Biomass (gm)",
       x="Species number",
       caption = "Task: Week 4 R training course")+
  theme(plot.title=element_text(size=20, color = "green"),
        plot.subtitle = element_text(size = 16, color = "red"),
        axis.title.x=element_text(size =14),
        axis.title.y=element_text(size=14),
        axis.text.x=element_text(size=14),
        axis.text.y=element_text(size=14),
        text=element_text(color="black"),
        plot.background=element_rect(fill="grey"), #background
        plot.margin = unit(c(1, 1, 1, 1), "cm"),      #plot margin: top, right, bottom, left
        legend.position="right",                      # legend at right
        legend.text = element_text(size=12))+
  coord_cartesian(ylim=c(0, 350))
```

#Significance Brackets for box plot in 'ggplot2'

```
#geom_signif()
install.packages("ggsignif")

library(ggplot2)
library(ggsignif)
```

#For single comparison

```
p1 <-ggplot(study_1, aes(x = plant, y = biomass, fill = plant)) +
  stat_boxplot(geom = "errorbar", width = 0.25) +
  geom_boxplot() +
  guides(fill = guide_legend(title = "plant"))+
  labs(title="Plant responses",
       subtitle="Nursery experiment",
       y="Species Biomass (gm)",
       x="Species number",
       caption = "Task: Week 5 R training course")+
```

```
 theme(plot.title=element_text(size=20, color = "blue"),
      plot.subtitle = element_text(size = 16, color = "red"),
      axis.title.x=element_text(size =14),
      axis.title.y=element_text(size=14),
      axis.text.x=element_text(size=14),
      axis.text.y=element_text(size=14),
      text=element_text(color="black"),
      plot.background=element_rect(fill="grey"), #background
      plot.margin = unit(c(1, 1, 1, 1), "cm"),     #plot margin: top, right, bottom, left
      legend.position="right",                     # legend at right
      legend.text = element_text(size=12))+
 coord_cartesian(ylim=c(0, 400))+
 geom_signif(comparisons = list(c("p2", "p4")),
           map_signif_level = TRUE,
           textsize = 5)

p1
```

#For multiple comparisons

```
#margin_top = 0.1, step_increase = 0.2, tip_length = 0.03

p2 <-ggplot(study_1, aes(x = treatment, y = biomass, fill = treatment)) +
 stat_boxplot(geom = "errorbar", width = 0.25) +
 geom_boxplot() +
 guides(fill = guide_legend(title = "Treatment"))+
 labs(title="Plant responses under climate",
     subtitle="Nursery experiment",
     y="Species Biomass (gm)",
     x="Species number",
     caption = "Task: Week 5 R training course")+
 theme(plot.title=element_text(size=20, color = "blue"),
      plot.subtitle = element_text(size = 16, color = "red"),
      axis.title.x=element_text(size =14),
      axis.title.y=element_text(size=14),
      axis.text.x=element_text(size=14),
      axis.text.y=element_text(size=14),
      text=element_text(color="black"),
      plot.background=element_rect(fill="grey"), #background
      plot.margin = unit(c(1, 1, 1, 1), "cm"),     #plot margin: top, right, bottom, left
      legend.position="right",                     # legend at right
      legend.text = element_text(size=12))+
 coord_cartesian(ylim=c(0, 500))+
 geom_signif(comparisons = list(c("Control", "Drought"), c("Rain", "Control")),
      map_signif_level = TRUE, margin_top = 0.1,step_increase = 0.2,
      tip_length = 0.03,
      textsize = 5)

p2
```

#Control the direction (either x or y) via orientation

```
#coord_flip()

p2 + coord_flip()
```

#Bar graph in ggplot2

```
head(study_1)
ggplot(study_1, aes(x=treatment, y=biomass, fill=treatment)) +
  geom_bar(position="dodge",stat="identity")
```

#group bar by the function fill

```
ggplot(study_1, aes(x=treatment, y=biomass, fill=plant)) +
  geom_bar(position="dodge", stat="identity")
```

#add a title, change color palette and axis titles

```
#title and axis title: labs()
#color palette: scale_fill_viridis(discrete = T)

install.packages("viridis")
install.packages("viridisLite")

# library
library(ggplot2)
library(viridis)
library(viridisLite)


ggplot(study_1, aes(x=treatment, y=biomass, fill=plant)) +
  geom_bar(position="dodge", stat="identity") +
  guides(fill = guide_legend(title = "Treatment"))+
  scale_fill_viridis(discrete = T) +
  labs(title="Plant responses under climate",
       subtitle="Nursery experiment",
       y="Species Biomass (gm)",
       x="Treatment",
       caption = "Task: Week 5 R training course")

#class task
#give a name t1 to the above mention bar graph and plot it
```

#Use some built in theme for ggplot2

```
install.packages("ggthemes")
install.packages("hrbrthemes")

library(ggplot2)
library(ggthemes)
library(hrbrthemes)

t1<- ggplot(study_1, aes(x=treatment, y=biomass, fill=plant)) +
  geom_bar(position="dodge", stat="identity") +
  guides(fill = guide_legend(title = "Treatment"))+
  scale_fill_viridis(discrete = T) +
  labs(title="Plant responses under climate",
       subtitle="Nursery experiment",
       y="Species Biomass (gm)",
       x="Treatment",
       caption = "Task: Week 5 R training course")

t1+theme_classic()
```

#Class Task
#use the following theme to the bar graph t1

```
theme_ipsum()
theme_ipsum_ps()
theme_bw()
theme_test()
theme_light()
theme_gray()
```

#Separate bar graph with the function facet_wrap()

```
str(study_1)
ggplot(study_1, aes(fill=plant, y=biomass, x=treatment)) +
  geom_bar(position="dodge", stat="identity") +
  guides(fill = guide_legend(title = "Treatment"))+
  scale_fill_viridis(discrete = T) +
  labs(title="Plant responses under climate",
       subtitle="Nursery experiment",
       y="Species Biomass (gm)",
       x="Treatment",
       caption = "Task: Week 5 R training course") +
  theme(legend.position="none")+
  theme_gray()+
  facet_wrap(~plant)
```

#create data set and row bind in R

```
a <- data.frame(Study_Area=c("KNP","LNP","SNP"), Species=c(3,2,4), frame=rep('a',3))
b <- data.frame(Study_Area=c("KNP","LNP","SNP"), Species=c(5,3,7), frame=rep('b',3))
study_3 <- rbind(a,b)
head(study_3)
```

#Animated barplot transition with R

#build an animated plot with gganimate

```
install.packages("gganimate")
library(ggplot2)
library(gganimate)

ggplot(study_3, aes(x=Study_Area, y=Species, fill=Study_Area)) +
  geom_bar(stat='identity') +
  theme_bw() +
  # gganimate specific bits:
  transition_states(
    frame,
    transition_length = 2,
    state_length = 1) +
  ease_aes('sine-in-out')

# Save at gif:
anim_save("288-animated-barplot-transition.gif")
```

#Linear model and confidence interval in ggplot2

```
head(study_1)
```

 # normal plot

```
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  theme_gray()
```

 # with linear trend

```
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm , color="red", se=FALSE) +
  theme_gray()
```

#Add linear trend line + confidence interval

```
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm , color="red", se=TRUE, fill="yellowgreen") +
  theme_gray()
```

#Customize confidence interval level

```
#level=0.90 or 0.95 or 0.99; confidence level
#lwd = 1; line width, increase the width by number
#labs(title = "90% Confidence Interval")

#90%
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm , color="red", lwd=1, se=TRUE, level = 0.90, fill="yellow") +
  theme_gray()+
  labs(title = "90% Confidence Interval")

#95%
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm , color="red", lwd=1, se=TRUE,level = 0.95, fill="yellowgreen") +
  theme_gray()+
  labs(title = "95% Confidence Interval")


#99%
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm , color="red",lwd=1, se=TRUE,level = 0.99, fill="green") +
  theme_gray()+
  labs(title = "99% Confidence Interval")
```

#Scatter plot with ellipses in ggplot2

```
#Adding ellipses with stat_ellipse

head(study_1)
ggplot(study_1, aes(x=year, y=biomass)) +
  geom_point() +
  stat_ellipse()

head(study_2)
ggplot(study_2, aes(x=Elevation_m, y=Height_m)) +
```

```
 geom_point() +
 stat_ellipse()
```

#Customization color, line type and line width

```
#color = 2,
#linetype = 2, try 1-6
#lwd = 2

ggplot(study_2, aes(x=Elevation_m, y=Height_m)) +
 geom_point() +
 stat_ellipse(color = 2,
              linetype = 2,
              lwd = 2)
```

#Add 95% Confidence levels with ellipses

```
ggplot(study_2, aes(x=Elevation_m, y=Height_m)) +
 geom_point() +
 stat_ellipse(level = 0.95, color = "red")


#Task
#Add 90% Confidence levels with ellipses
#Add 99% Confidence levels with ellipses

ggplot(study_2, aes(x=Elevation_m, y=Height_m)) +
 geom_point() +
 stat_ellipse(level = 0.9, color= "yellowgreen") +
 stat_ellipse(level = 0.95, color = "red") +
 stat_ellipse(level = 0.99, color = "blue")
```

#Ellipses by group

```
#color= by treatment or study_area
#When you create a scatter plot by group, the ellipses are created for each group.

ggplot(study_1, aes(x=year, y=biomass, color=treatment)) +
 geom_point()+
 stat_ellipse()

ggplot(study_2, aes(x=Elevation_m, y=Height_m, color=study_area)) +
 geom_point()+
 stat_ellipse()
```

#Linetype by group in Ellipses

```
ggplot(study_1, aes(x=year, y=biomass, color=treatment, linetype =treatment)) +
 geom_point()+
 stat_ellipse()

ggplot(study_2, aes(x=Elevation_m, y=Height_m, color=study_area, linetype =study_area)) +
 geom_point()+
 stat_ellipse()
```

#Fill the area with ellipses

```
#geom = "polygon"

ggplot(study_1, aes(x=year, y=biomass, color=treatment)) +
 geom_point() +
 stat_ellipse(geom = "polygon",
          aes(fill = treatment))

ggplot(study_2, aes(x=Elevation_m, y=Height_m, color=study_area)) +
 geom_point() +
 stat_ellipse(geom = "polygon",
          aes(fill = study_area))
```

#Polygon by group with transparency

```
#alpha=

ggplot(study_1, aes(x=year, y=biomass, color=treatment)) +
 geom_point() +
 stat_ellipse(geom = "polygon",
          aes(fill = treatment),
          alpha = 0.25)

ggplot(study_2, aes(x=Elevation_m, y=Height_m, color=study_area)) +
 geom_point() +
 stat_ellipse(geom = "polygon",
          aes(fill = study_area),
          alpha = 0.25)
```

#how ellipses looks like with a big data set
 # with a Data simulation

```
set.seed(2)
```

```
x <- runif(500)
y <- 4 * x ^ 2 + rnorm(length(x), sd = 4)
group <- ifelse(x < 0.4, "A",
                ifelse(x > 0.8, "C", "B"))
x <- x + runif(length(x), -0.15, 0.15)

# Data frame
df <- data.frame(x = x, y = y, group = group)
str(df)

ggplot(df, aes(x = x, y = y, color = group)) +
  geom_point() +
  stat_ellipse()

#Polygon by group with transparency
ggplot(df, aes(x = x, y = y, color = group)) +
  geom_point() +
  stat_ellipse(geom = "polygon",
               aes(fill = group),
               alpha = 0.25)

###################################
```

#Task week 5#

```
#use the data set study_2 to create a bar graph
#plot study_area in the x axis and species_richness in the y axis
#use position="dodge" and stat="identity"
#Use legend title "Study Area"
#Use plot title="Write your name",
#Use plot subtitle="Bar plot",
#use y axis name ="Species Richness",
#Use x axis name ="Study Area",
#Use caption as "Task: Week 5 R training course"
#give a name t2 to the bar graph
#add the following theme to the bar graph t2
  #theme_excel()
  #theme_excel_new()
  #theme_few()
  #theme_foundation()
  #theme_clean()
  #theme_igray()
#Create a pdf file with all the bar graph and submit the task using the following Google class link
https://classroom.google.com/c/NTgwNjY0OTY2MDA1?cjc=tlqfgnj
Class code: tlqfgnj
####################
```