Exam 3

Izzi Grasso Due: 5/2/2020

library(tidyverse)

## ── Attaching packages ───────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.2.1 ✔ purrr 0.3.2  
## ✔ tibble 2.1.3 ✔ dplyr 0.7.5  
## ✔ tidyr 0.8.1 ✔ stringr 1.4.0  
## ✔ readr 1.1.1 ✔ forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.5.2

## Warning: package 'tibble' was built under R version 3.5.2

## Warning: package 'purrr' was built under R version 3.5.2

## Warning: package 'stringr' was built under R version 3.5.2

## Warning: package 'forcats' was built under R version 3.5.2

## ── Conflicts ──────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(ggplot2)

## Problem 1

#### Data

df1 <- data.frame(  
 Depletion = c(85, 55, 40, 77),  
 ModifiedHess = c(75, 45, 35, 67),  
 Surber = c(31, 20, 9, 37),  
 SubstrateRemoval = c(43, 21, 15, 27),  
 Kicknet = c(17, 10, 8, 15)  
)

#### a. Plot

df2 <- df1 %>%  
 summarise\_each(funs(mean)) %>%  
 gather(Var, Val)

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over all variables, use `summarise\_all()`

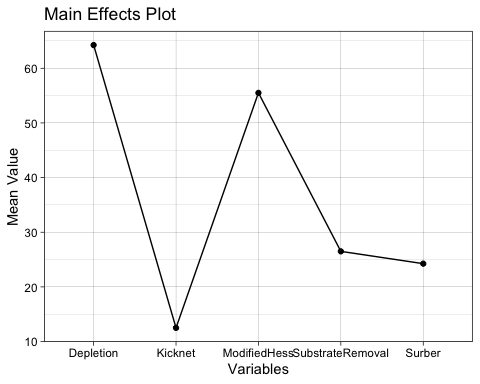
## Warning: `is\_lang()` is deprecated as of rlang 0.2.0.  
## Please use `is\_call()` instead.  
## This warning is displayed once per session.

## Warning: `lang()` is deprecated as of rlang 0.2.0.  
## Please use `call2()` instead.  
## This warning is displayed once per session.

## Warning: `mut\_node\_car()` is deprecated as of rlang 0.2.0.  
## This warning is displayed once per session.

## Warning: The `printer` argument is deprecated as of rlang 0.3.0.  
## This warning is displayed once per session.

ggplot(df2, aes(x = as.factor(Var), y = Val)) +   
 geom\_line(aes(group = 1)) +  
 geom\_point() +   
 theme\_linedraw() +   
 labs( title = "Main Effects Plot",  
 x = "Variables",  
 y = "Mean Value")



Interpretation: Since the line is not at all flat, I would expect to reject the null hypothesis. Kicknet seems to be much lower than Modified Hess and Depletion.

#### b.

Null Hypothesis: The distributions are the same location Alternative Hyp: Not null.

df1 <- df1 %>%   
 gather(Var, Val)  
  
kruskal.test(Val ~ as.factor(Var), data = df1)

##   
## Kruskal-Wallis rank sum test  
##   
## data: Val by as.factor(Var)  
## Kruskal-Wallis chi-squared = 13.986, df = 4, p-value = 0.007341

At = 0.05 we reject the null hypothesis. There is significant evidence that the sampling procedures yield different species counts.

#### c.

pairwise.wilcox.test(df1$Val, df1$Var, p.adjust.method = "bonferroni")

## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot  
## compute exact p-value with ties

##   
## Pairwise comparisons using Wilcoxon rank sum test   
##   
## data: df1$Val and df1$Var   
##   
## Depletion Kicknet ModifiedHess SubstrateRemoval  
## Kicknet 0.29 - - -   
## ModifiedHess 1.00 0.29 - -   
## SubstrateRemoval 0.57 0.81 0.57 -   
## Surber 0.29 1.00 0.57 1.00   
##   
## P value adjustment method: bonferroni

At = 0.05, we fail to reject the null hypothesis.

## Problem 2

#### Data

TarContent <- c(14, 17, 28, 17, 16, 13, 24, 25, 18, 31)  
NicotineContent <- c(0.9, 1.1, 1.6, 1.3, 1.0, 0.8, 1.5, 1.4, 1.2, 2.0)

#### a & b

Null hypothesis: = 0 Alternative hyp: 0

cor.test(TarContent, NicotineContent, method = "spearman")

## Warning in cor.test.default(TarContent, NicotineContent, method =  
## "spearman"): Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: TarContent and NicotineContent  
## S = 5.516, p-value = 5.248e-06  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.9665698

At = 0.05 we reject the null hypothesis. There is a linear correlation between the amount of tar and the amount of nicotine found in cigarettes.

## Problem 3

#### Data

x = 261  
n = 302  
y = 401  
m = 454  
p\_A = x/n  
p\_B = y/m

#### a.

z\_ = qnorm(0.005, mean = 0, sd = 1, lower.tail = FALSE)  
  
LCI = (p\_A - p\_B) - z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
UCI = (p\_A - p\_B) + z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
  
LCI

## [1] -0.08293269

UCI

## [1] 0.04488969

#### b.

z\_ = qnorm(0.05, mean = 0, sd = 1, lower.tail = FALSE)  
  
LCI = (p\_A - p\_B) - z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
UCI = (p\_A - p\_B) + z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
  
LCI

## [1] -0.05983343

UCI

## [1] 0.02179042

#### c.

z\_ = qnorm(0.025, mean = 0, sd = 1, lower.tail = FALSE)  
  
LCI = (p\_A - p\_B) - z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
UCI = (p\_A - p\_B) + z\_\*sqrt(x\*(n-x)/(n^3)+y\*(m-y)/(m^3))  
  
LCI

## [1] -0.06765191

UCI

## [1] 0.02960891

#### d.

Null Hypothesis: = Alternative Hyp:

p\_hat = (x+y)/(n+m)  
z\_star = (p\_A-p\_B)/sqrt(p\_hat\*(1-p\_hat)\*(1/n+1/m))  
2\*pnorm(z\_star, mean = 0, sd = 1)

## [1] 0.4375565

At = 0.05 we fail to reject the null hypothesis.

## Problem 4

Null Hypothesis: Alternative hyp: >

SSquared = 0.0002  
SigmaSquared = 0.0003  
alpha = 0.05  
n = 10  
  
TestStatistic <- ((n-1)\*SSquared)/SigmaSquared  
  
ChiSq <- qchisq((1-alpha), df = (n-1))  
  
TestStatistic > ChiSq

## [1] FALSE

## Problem 5

#### Data

df3 <- data.frame(  
 Value = c(12.8, 11.7, 11.5, 12.6,  
 10.6, 14.2, 14.7, 16.5,  
 11.7, 11.8, 13.6, 15.4,  
 10.7, 9.9, 10.7, 9.6,  
 11.0, 13.8, 15.9, 17.1),  
 Treatment = rep(1:4, times = 5),  
 Block = rep(1:5, times = 1, each = 4)  
)

#### Friedman Test, RBD

Null hypothesis: Distributions are identical Alternative hypothesis: Not null hypothesis.

friedman.test(df3$Value, df3$Treatment, df3$Block)

##   
## Friedman rank sum test  
##   
## data: df3$Value, df3$Treatment and df3$Block  
## Friedman chi-squared = 2.8776, df = 3, p-value = 0.4109

At = 0.05, we fail to reject the null hypothesis. There is not significant evidence to show that there is a difference between treatment means.