Racine,

Sophia

40095905

**Assignment 1**

**Questions/Answers:**

2. **What is the mean root length below the membrane barrier for each genotype of beet plant root?**

Answer 1.A

R code

LineA=c(490,390,520,400,420,580,620,530,350,500,690,590,550,490,440)

round(mean(LineA),digits = 2)

[1] 504

LineB=c(760,370,570,540,590,600,480,70,540,520,60,650,400,50,540)

round(mean(LineB),digits = 2)

[1] 449.33

M82=c(540,490,450,50,470,580,420,330,480,480,410,540,700,680,600)

round(mean(M82), digits = 2)

[1] 481.33

1. **What is the standard deviation of penetrating root dry weight for each genotype of beet plant root?**

Answer 1.B

R code

sdLineA=c(0.3373,0.3198,0.5463,0.3715,0.5145,0.4708,0.3751,0.4357,0.487,0.3071,0.4564,0.5768,0.5403,0.4496,0.5334)

round(sd(sdLineA),digits = 2)

[1] 0.09

sdLineB=c(0.0768, 0.3641, 0.4222, 0.3118, 0.1960, 0.4698, 0.1894, 0.2726, 0.1950, 0.3537, 0.1091, 0.3235, 0.4448,0.2226, 0.2826)

round(sd(sdLineB),digits = 2)

[1] 0.12

sdM82=c(0.3204,0.3677,0.1459,0.2154,0.3288,0.2519,0.3024,0.4247,0.3459,0.3431,0.3388,0.3765,0.6458,0.3835,0.509)

round(sd(sdM82),digits = 2)

[1] 0.12

1. Chart, box and whisker chart

   Description automatically generated**Compare the distributions of penetrating root dry weights using three boxplots on a single plot. Explain what this plot is illustrating.**

Answer 1.C

The multiple boxplots are demonstrating the different rooting properties of each beet plants ’genotype; Line A, Line B, and M82 in terms of their penetrating root dry weight in grams. We can observe three outliers in genotype M82; two which are above maximum values and one below the lowest value. The plot also indicates that genotype Line A has a higher penetrating root dry weight in grams than Line B and M82, with Line B having the lowest values in the three genotypes.

1. **Is the penetrating dry root weight the same for the three genotypes of beet plant root? Perform a statistical test and interpret the results.**

Answer 1.D

We are looking at similarities between the three genotypes, therefore we will use ANOVA 1 test.

H0 : µM82 = µLB = µLA

H1: not H0

R code

Shapiro-Wilk normality test

data: sdLineA

W = 0.94766, p-value = 0.4883

Shapiro-Wilk normality test

data: sdLineB

W = 0.97075, p-value = 0.8691

Shapiro-Wilk normality test

data: sdM82

W = 0.92985, p-value = 0.2714

freq=c(sdM82,sdLineA,sdLineB)

resi=c(rep("sdM82",15),rep("sdLineA",15),rep("sdLineB",15))

PenetratingRDW=data.frame(freq,resi)

fresi=factor(resi)

fligner.test(freq~fresi)

Fligner-Killeen test of homogeneity of variances

data: freq by fresi

Fligner-Killeen:med chi-squared = 1.2188, df = 2, p-value = 0.543

\*\* variances similar

summary(aov(freq~fresi))

Df Sum Sq Mean Sq F value Pr(>F)

fresi 2 0.1960 0.09801 8.298 0.000919 \*\*\*

Residuals 42 0.4961 0.01181

TukeyHSD(aov(freq~fresi))

Fit: aov(formula = freq ~ fresi)

$fresi

diff lwr upr p adj

sdLineB-sdLineA -0.16128667 -0.2577008 -0.06487253 0.0005947

sdM82-sdLineA -0.09023333 -0.1866475 0.00618080 0.0707179

sdM82-sdLineB 0.07105333 -0.0253608 0.16746747 0.1851312

P value for ANOVA testing is smaller than 0.05 therefore we reject the null, we continue the test with Tukey to see where the difference lies, and we see significance difference between Line B and Line A genotype beet plants.

1. **Is canopy dry weight dependent on penetrating root dry weight? Illustrate this relationship if possible. How much variation in penetrating root dry weight is explained by this relationship?**

Answer 1.E

Using a scatterplot, we can identify the relationship between canopy dry weight and penetrating root dry weight. We also look at the Pearson r and the regression line associated to the two quantitative variables.

Chart, scatter chart

Description automatically generated

R code

cor(CanopyDwt,RootDwt)

[1] -0.4948796

summary(lm(CanopyDwt~RootDwt))

Residuals:

Min 1Q Median 3Q Max

-7.5998 -2.3996 0.1285 2.4788 7.8934

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 61.011 1.828 33.367 < 2e-16 \*\*\*

RootDwt -17.947 4.806 -3.735 0.000549 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.998 on 43 degrees of freedom

Multiple R-squared: 0.2449, Adjusted R-squared: 0.2273

F-statistic: 13.95 on 1 and 43 DF, p-value: 0.0005486

Very low Pearson’s r, there’s no linear relationship between the two variables. The variation (R^2) for RootDwt is 0.24 in relation to CanopyDwt relationship.

1. **Create a time series plot and decomposed time series plot to describe a trend over time.**

Graphical user interface

Description automatically generatedAnswer 2

R code

attach(Aviation\_Accident)

Accident<-ts(Aviation.accident.and.incident...Canada.,start = c(2004,1),end=c(2022,4),frequency = 12)

plot(Accident,main="Canada's aviation accidents from 2004 to 2022 ",xlab="Years",ylab="Accidents")

We have our first plot demonstrating accidents over time from year 2004 to 2022. There is clearly a big increase in the year 2020 for some reasons, it could also be identified as an outlier.

R code

newts<-log(Accident)

fit<-stl(newts,s.window = "period")

Histogram

Description automatically generated plot(fit,main="Canada's aviation accidents from 2004 to 2022",col="4")

I converted the data (Accident) value into its logarithmic value with log(), it gave me a clearer graphical analysis. We can see that Canada’s aviation accidents have a big peak in 2020 than dropping to a lower amount in the year 2021 and 2022. The residuals are also interesting to look at, demonstrating high variability throughout the time series, but especially in around the year 2015. The seasonal component does not seem to change very fast over the years, in other words similar pattern. The grey bars on the side also confirm that the remainder and data show similar variation compared to seasonal and trend components having less variation.

2. **As part of their research, the scientist would like to know if the species, site and winter months GDD for December (SA.Dec) influence the DPO. Build a model to explain this relationship and interpret the coefficients.**

Answer 3

Since we are looking at the strength of association for three independent variables (X1=sites,X2=species,X3=SA.Dec) on the dependent variable DPO (y) we will use a multiple regression model.

R code

> attach(ElmP)

> summary(lm(DPO~Species+Site+SA.Dec,data = ElmP))

Residuals:

Min 1Q Median 3Q Max

-31.4892 -8.1936 0.5566 8.8613 30.3010

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 70.017081 5.586099 12.534 < 2e-16 \*\*\*

Species 27.675994 5.879446 4.707 1.07e-05 \*\*\*

Site -0.967579 1.602946 -0.604 0.548

SA.Dec -0.004929 0.014714 -0.335 0.739

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.05 on 78 degrees of freedom

Multiple R-squared: 0.4876, Adjusted R-squared: 0.4679

F-statistic: 24.74 on 3 and 78 DF, p-value: 2.38e-11

> y'=70.017081+27.675994(X1)-0.967579(X2)-0.004929(X3)

From our coefficients we can see that the variable species has a greater influence on the DPO, with a P-value of 1.07e-05 < than our significance level 0.05 , meaning significance. Both variables X2 site and X3 SA.Dec has negative magnitudes coefficient and a P-value greater than 0.05, therefore not showing significancy in relation to the DPO.

1. **Does there seem to be a difference between species 1 and species 2 GDD in September (SA.Sept)?**

Answer 3.B

Since we are looking if there is a difference between specie 1 and 2 in the data of September SA.Sept , and the data sample size is greater than 30 and we have no information that it is a normal distributed population, we use Rank-Sum (WRS). Unequal sample size for Specie 1 and Species 2 in SA.Sept.

H0: median 1 = median 2

H1: median 1 does not equal median 2

R code

data.frame(Species,SA.Sept)

Elmp1=c(395.97823,1753.06683, 1434.94530,1664.55078, 11.09545,864.36069, 1735.74459, 2323.18266, -735.06505, -628.48284,1623.25395, -773.32749, 609.79322, 1816.27809, 2622.85628, -209.71205,-597.73388,1226.87819,70.54945,895.32337,497.89831,1287.26870,826.76754,-298.38202,1424.32982,912.40749,1961.19089, -552.71085, -37.03944,636.70233,3452.58393,1719.59154,907.33301,222.42819,2554.56024,1280.29466,1322.92943,1839.13745,1753.05343, 2362.55131)

Elmp2=c( 1119.37920,1211.35832,973.51177,167.24104,346.20602,1947.79330,418.66163, 624.30772,1465.60019,2392.37085,1836.70335,727.47285,-954.30526,655.16858,1903.91049,840.00509,2435.64908,726.27383,1223.56220,2406.10039,2205.99498, -1491.79307,591.04989,1471.06040,2099.19931,549.86365,1062.03078, 809.59152,2544.19441,2472.35922,2134.11298, 874.47780, 1045.11479,1068.58367,889.91552,872.50631,2159.35556, 774.61570,2612.62157, -32.04914,-674.63726,535.32159)

wilcox.test(Elmp1,Elmp2,exact=F,paired=F)

Wilcoxon rank sum test with continuity correction

data: Elmp1 and Elmp2

W = 786, p-value = 0.6197

alternative hypothesis: true location shift is not equal to 0

P-value of 0.62 greater than alpha .05 we fail to reject the null , there is similarity in the median for Specie 1 and 2 GDD in September.

1. **Additionally, for the species variable, does the DPO exceed 103 for all years recorded?**

Answer 3. C

y'=70.017081+27.675994(X1)-0.967579(X2)-0.004929(X3)

R code