Bioestadística – Laboratorio

**MBIO2401: Laboratorio (1 crédito)**

*Profesores del Laboratorio:* Luisa A. Castellanos.

*Profesor del Curso:* Andrew J. Crawford.

*Monitora:* Valentina Muñoz

Departamento de Ciencias Biológicas, Universidad de los Andes

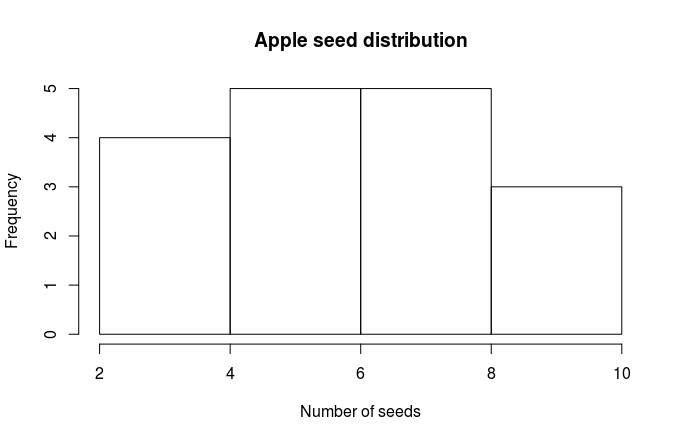
**Lab 7 – Pruebas de normalidad, transformaciones y ANOVA**

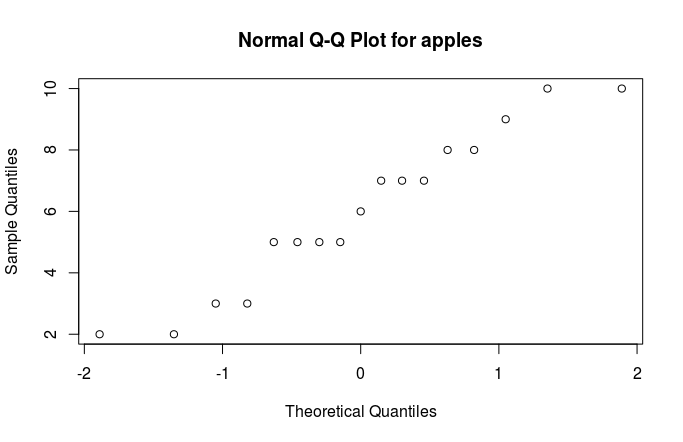
**La asistencia al laboratorio es obligatoria y por lo tanto un requisito para presentar este taller**

Para cada uno de los siguientes planteamientos debe probar la normalidad de las variables con los tres métodos vistos (dos gráficos y uno cuantitativo). En el caso de que estas variables resulten no normales debe hacer las transformaciones pertinentes y probar la normalidad nuevamente.

De acuerdo con los resultados deberá hacer la prueba estadística que le permita probar su hipótesis. Recuerden plantear siempre las hipótesis tanto para las pruebas de normalidad como para las pruebas estadísticas que decidan realizar.

1. **[1 Punto]** En la naturaleza las cosas tienden a distribuirse de manera normal: altura en una población humana, numero de rocas en un rio, numero de escamas en una especie de pez etc. Ustedes cortaran su manzana al medio y contarán el numero de semillas que tiene, este numero lo colocarán en la tabla grupal que está en SICUA. Realice los análisis gráficos correspondientes y la prueba estadística para determinar si estos datos se ajustan a una distribución normal.

A partir del histograma, podemos ver como “al ojo”, los datos siguen una distribución normal.

Por otro lado, en el q-q plot se aprecia como los datos siguen aproximadamente una linea vertical.

> shapiro.test(manzanas$semillas)

Shapiro-Wilk normality test

data: manzanas$semillas

W = 0.94455, p-value = 0.3762

Encima de todo, el test the Shapiro-Wilk tiene un p-value de 0.3762, luego no rechazamos H0 y afirmamos que los datos siguen una distriución normal.

Así pues, la cantidad de semillas en una manzana sigue una distribución normal.

1. **[2 Puntos]** Los datos que se encuentran en Sicua+ con el nombre “Punto1\_ANOVA” representan el tamaño de cono (masa) de una especie de pino en 16 sitios de estudio en tres tipos de ambientes en el occidente de USA (Edelaar & Benkman, 2006). Los tres ambientes hacían referencia a:

Ambiente 1: Isla con ardillas ausentes

Ambiente 2: Isla con ardillas presentes

Ambiente 3: Continente con ardillas presentes.

* 1. Existen diferencias entre los tratamientos?

> modelo1<-aov(data=data1,Masa\_cono~Tipo\_ambiente)

> summary(modelo1)

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

Tipo\_ambiente 2 29.404 14.702 50.09 7.79e-07 \*\*\*

Residuals 13 3.816 0.294

---

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

Tenemos un p-value<<005 luego rechazamos H0 y afirmamos que por lo menos una de las medias es diferente a las otras. Es decir, sí existen diferencias entre los tratamientos.

* 1. ¿Qué tratamiento afectó en mayor medida la sobrevivencia? (Utilice las medias obtenidas para cada tratamiento para responder a esta pregunta)

> mean(I$Masa\_cono)

[1] 8.9

> mean(II$Masa\_cono)

[1] 6.08

> mean(III$Masa\_cono)

[1] 6.12

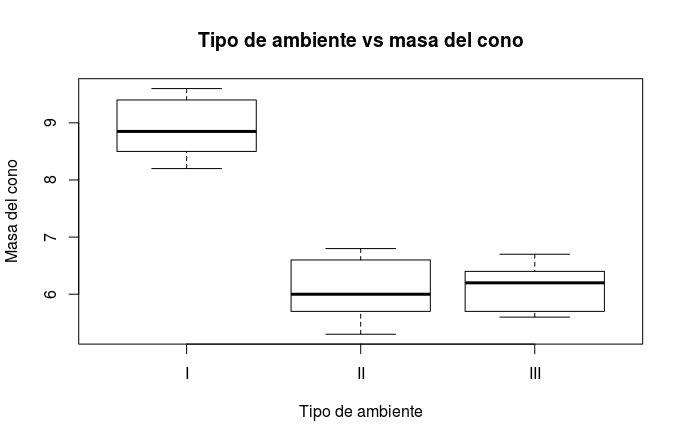
El tratamiento que afecto en mayor medida la sobrevivencia fue el tratamiento II, la isla con ardillas presentes. Esto tiene sentido pues las ardillas se alimentan de los pinos y les reducen la taza de supervivencia.

* 1. ¿Qué tratamiento afectó en menor medida la sobrevivencia?

Mirando unicamente las meias, el tratamiento I fue el que afecto en menor medida la supervivencia pues

sin ardillas, los pinos no tienen quien los depreden. Así pues, su supervivencia no se ve afectada.

* 1. Realice una gráfica para ilustrar los resultados.



* 1. Conclusión biológica

Las presencia de las ardillas reduce significativamente la masa del cono de los pinos.

1. **[2 Puntos]** Investigadores desean saber si es posible concluir que dos poblaciones de juveniles de cierta especie difieren en la edad promedio en la cual llegan a la madurez sexual. Los datos son los siguientes: (Asuma que el muestreo fue aleatorio e independiente)

Muestra población A:

9.5, 10.5, 9, 10, 13, 10, 13.5, 10, 9.5, 10, 9.75

Muestra población B:

12.5, 9.5, 13.5, 13.75, 12, 13.75, 12.5, 9.5, 12, 13.5, 12, 12

¿Qué pueden concluir los investigadores? Utilice un alfa de 0.05.

> var.test(A,B)

F test to compare two variances

data: A and B

F = 1.0004, num df = 10, denom df = 11, p-value = 0.9916

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2837362 3.6662358

sample estimates:

ratio of variances

1.000361

> shapiro.test(A)

Shapiro-Wilk normality test

data: A

W = 0.73964, p-value = 0.001523

> shapiro.test(B)

Shapiro-Wilk normality test

data: B

W = 0.83472, p-value = 0.0239

Aunque ambas muestras tengan varianzas similares, no podemos usar ANOVA pues ni A ni B siguen una distribución normal bajo un alpha = 0.05.

Así pues, usaremos una prueba no parametrica:

|  |
| --- |
| > wilcox.test(A,B,alternative = "two.sided")  Wilcoxon rank sum test with continuity correction  data: A and B  W = 31, p-value = 0.03216  alternative hypothesis: true location shift is not equal to 0  Warning message:  In wilcox.test.default(A, B, alternative = "two.sided") :  cannot compute exact p-value with ties |
| |  | | --- | | > | |
|  |

Vemos que hay un problema por tener “ties” (valores repetidos dentro de las muestras). Sin embargo, el internet parece concordad en que si hay pocos, el efecto no es significativo. Así pues, seguiremos el análisis.

Obtuvimos un p-value de 0.03216<0.05 luego rechazamos H0 y establecemos que las dos poblaciones de juveniles difieren en la edad promedio en la cual llegan a la madurez sexual.

1. **[2.5 puntos]** Los datos que se encuentran en Sicua+ con el nombre “Punto2\_ANOVA” corresponden a los resultados después del siguiente experimento: Se midió el tiempo de supervivencia de una serie de animales a los cuales se les asignó aleatoriamente un tipo de veneno (I, II y III) y un tipo de tratamiento (A, B, C, D).
2. ¿Existen diferencias entre los tratamientos?

|  |
| --- |
| > modelo2Trat <- aov(data=data2,Supervivencia~Tratamiento)  > summary(modelo2Trat)  Df Sum Sq Mean Sq F value Pr(>F)  Tratamiento 3 0.9283 0.30942 6.622 0.000865 \*\*\*  Residuals 44 2.0559 0.04673  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
| |  | | --- | | > | |
|  |

Tenemos un p-value de 0.000865<<0.05 luego rechazamos H0 y confirmamos que hay una diferencia entre los tratamientos.

1. ¿Qué tratamiento afectó en mayor medida la sobrevivencia? (Utilice las medias obtenidas para cada tratamiento para responder a esta pregunta)

|  |
| --- |
| > TA<-subset(data2,Tratamiento=="A")  > TB<-subset(data2,Tratamiento=="B")  > TC<-subset(data2,Tratamiento=="C")  > TD<-subset(data2,Tratamiento=="D")  > mean(TA$Supervivencia)  [1] 0.3141667  > mean(TB$Supervivencia)  [1] 0.6766667  > mean(TC$Supervivencia)  [1] 0.3891667  > mean(TD$Supervivencia)  [1] 0.5341667 |
| |  | | --- | | > | |
|  |

El tratamiento que afecto en mayor medida la superviveincia fue el tratamiento A. Es el que menor taza de supervivencia tiene.

1. ¿Hay diferencias entre los venenos? ¿Qué veneno afectó en menor medida la sobrevivencia?

|  |
| --- |
| > modelo2Ven<-aov(data=data2,Supervivencia~Veneno)  > summary(modelo2Ven)  Df Sum Sq Mean Sq F value Pr(>F)  Veneno 2 1.022 0.5110 11.72 8e-05 \*\*\*  Residuals 45 1.962 0.0436  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
| |  | | --- | | > | |
|  |

El p-value es <<0.05 luego rechazamos H0 y confirmamos que si hay una diferencia entre los venenos.

|  |
| --- |
| > VI <-subset(data2,Veneno=="I")  > VII <-subset(data2,Veneno=="II")  > VIII <-subset(data2,Veneno=="III")  > mean(VI$Supervivencia)  [1] 0.615  > mean(VII$Supervivencia)  [1] 0.544375  > mean(VIII$Supervivencia)  [1] 0.27625 |
| |  | | --- | | > | |
|  |

El veneno que menos afecto la supervivencia fue el 1ro. Es el que mas alta tasa de supervivencia tiene.

1. ¿Existe interacción? Incluya el Valor-P y el gráfico.

> modelo3<- aov(data=data2,Supervivencia~Veneno\*Tratamiento)

> summary(modelo3)

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

Veneno 2 1.0220 0.5110 23.390 3.10e-07 \*\*\*

Tratamiento 3 0.9283 0.3094 14.162 2.97e-06 \*\*\*

Veneno:Tratamiento 6 0.2474 0.0412 1.887 0.11

Residuals 36 0.7865 0.0218

---

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

El valor p correspondiente a la interaccion es de 0.11>>0.05 luego no podemos rechazar H0. Es decir, no se puede afirmar que la interaccion tiene un efecto significativo sobre la supervivencia.

1. Conclusión biológica

El veneno y los tratamientos por separado tienen un efecto considerable sobre la supervivencia pero no se observa ningun efecto significativo en la interacción de veneno-tratamiento.