

Estadística No Paramétrica

Tarea 2

Prof. H Vaquera H

2022-05-24

Instrucciones: Algunos ejercicios se pueden realizar manualmente y verificar usando R o SAS.

Problema 1

La siguiente tabla contiene información sobre la longitud de las hojas de plantas de la misma especie en cada uno de cuatro lugares pantanosos no urbanizados. En cada uno de éstos, se seleccionaron aleatoriamente seis plantas. Se seleccionaron al azar diez hojas de cada una de las plantas y la media de las diez mediciones (en centímetros) se registró para cada planta de cada uno de los lugares. Use la prueba *H de Kruskal-Wallis* para determinar si hay suficiente evidencia para afirmar que la distribución de las longitudes medias de las hojas difiere en localizaciones para al menos dos de los sitios. Use $\alpha = .05$. encuentre el p -value aproximado.

Localización	Longitud media de la hoja (cm)					
1	5.7	6.3	6.1	6.0	5.8	6.2
2	6.2	5.3	5.7	6.0	5.2	5.5
3	5.4	5.0	6.0	5.6	4.0	5.2
4	3.7	3.2	3.9	4.0	3.5	3.6

- Establezca el Juego de hipótesis para probar si los niveles de contaminación por Hg son similares entre grupos.
- realice una grafica de caja (Boxplot)
- Utilice la prueba del Kruskal Wallis para probar la hipotesis en a) usando un $\alpha = 0,05$. Concluya.
- Utilice la prueba de Nemenyi y Dunn para investigar cual es el la localidad con mayor longitud usando un $\alpha = 0,05$. Concluya.
- Utilice la el ANOVA de una via de clasificación para probar la la direrencia entre localidades usando un $\alpha = 0,05$. verifique supuesto de normalidad, y homocedasticidad Concluya. Anexe un grafica de medias.

Problema 2

Stelzer et al (2012) realizaron ensayos de reacción en cadena de la polimerasa cuantitativa (qPCR) con indicadores fecales para determinar si tres laboratorios estaban proporcionando resultados similares. Se enviaron 21 muestras de fuente fecal a cada uno de los 3 laboratorios. Con las muestras como bloques. Se desea decidir si al menos un laboratorio proporcionó resultados más altos/más bajos que los demás.

Stelzer, E.A., Strickler, K.M., and Schill, W.B., 2012, Interlaboratory comparison of three microbial source tracking quantitative polymerase chain reaction (qPCR) assays from fecal-source and

environmental samples: U.S. Geological Survey Scientific Investigations Report 2012-5087, 10 p.,
<https://doi.org/10.3133/sir20125087>

- a). Indique la hipótesis nula y alterna para la prueba de Friedman en este problema.
- b). Obtenga el estadístico de Friedman y pruebe $H : 0$ usando $\alpha = 0.05$.
- c). Use una prueba de comparación múltiple para determinar qué laboratorios difieren de otros. Apoye su decisión usando gráficos.

ver tabla 1

Table 1: Tabla de datos problema2

Bacterias Totales	Muestra	Laboratorio
44	1	Lab 1
42	2	Lab 1
180	3	Lab 1
200	4	Lab 1
190	5	Lab 1
200	6	Lab 1
190	7	Lab 1
48	8	Lab 1
54	9	Lab 1
33	10	Lab 1
57	11	Lab 1
55	12	Lab 1
44	13	Lab 1
53	14	Lab 1
60	15	Lab 1
127	16	Lab 1
3000	17	Lab 1
620	18	Lab 1
1300	19	Lab 1
170	20	Lab 1
170	21	Lab 1
45	1	Lab 2
40	2	Lab 2
190	3	Lab 2
210	4	Lab 2
180	5	Lab 2
190	6	Lab 2
190	7	Lab 2
49	8	Lab 2
54	9	Lab 2
32	10	Lab 2
56	11	Lab 2
53	12	Lab 2
43	13	Lab 2
51	14	Lab 2
59	15	Lab 2
120	16	Lab 2
3000	17	Lab 2
620	18	Lab 2
1200	19	Lab 2
170	20	Lab 2
170	21	Lab 2
120	1	Lab 3
99	2	Lab 3
480	3	Lab 3
480	4	Lab 3
470	5	Lab 3
390	6	Lab 3
410	7	Lab 3
120	8	Lab 3
130	9	Lab 3
70	10	Lab 3
140	11	Lab 3
130	12	Lab 3
98	13	Lab 3
120	14	Lab 3
140	15	Lab 3
320	16	Lab 3
7200	17	Lab 3
1500	18	Lab 3
2900	19	Lab 3
450	20	Lab 3
360	21	Lab 3

1

Problema 3

A series of experiments was designed to test a hypothesis that massive silver iodide seeding can, under specified conditions, lead to increased precipitation. The data from these experiments were reported in the article “A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification” [Technometrics (1975) 17:161–166]. The rain volume falling from the cloud after seeding with silver iodide is reported here

Tabla 2

Table 2: Tabla de datos de siembra de nubes

Rainfall (acre-feet) Unseeded Clouds							
129.6	31.4	2745.6	489.1	430	302.8	119	4.1
92.4	17.5	200.7	274.7	274.7	7.7	1656	978
198.6	703.4	1697.8	334.1	118.3	255	115.3	242.5
32.7	40.6						
Rainfall (acre-feet) Seeded Clouds							
26.1	26.3	87	95	372.4	0	17.3	24.4
11.5	321.2	68.5	81.2	47.3	28.6	830.1	345.5
1202.6	36.6	4.9	4.9	41.1	29	163	244.3
147.8	21.7						

- use la prueba de Ansari-Bradley para probar si la variabilidad en ambos tratamientos es igual, establezca su juego de hipótesis, graficos y conclusiones con $\alpha = 0.05$.
- Use la prueba de Kolmogorov-Smirnov para probar si las distribuciones de precipitación son iguales en ambos tratamientos, use $\alpha = 0.05$.

Problema 4

Nappi (E12) investigated the changes occurring in the haemocytes of larvae of *Drosophila algonquin* during parasitization by the hymenopterous parasite (parasitoid) *Pseudeucoila bochei*. Twenty-seven hours after parasitization of *Drosophila algonquin* larvae, differential counts (%) of plasmatocytes were made on three groups: host larvae in which reaction was successful (S), those in which the reaction was unsuccessful (U), and those in which there was no visible host reaction (N). The results are shown in Table 6.16. We wish to test the null hypothesis of no difference among the three groups against the alternative that the differential counts of plasmatocytes (%) decrease in the three groups from group N to group S.

Differential plasmatocyte counts, percent from larvae of *Drosophila algonquin* 27 hours after parasitization by *Pseudeucoila bochei* (host age 91 hours when parasitized)

Successful host reactions (S)	Unsuccessful host reactions (U)	No visible host reactions (N)
54.0	79.8	98.6
67.0	82.0	99.5
47.2	88.8	95.8
71.1	79.6	93.3
62.7	85.7	98.9
44.8	81.7	91.1
67.4	88.5	94.5
80.2		

Source: A. J. Nappi, "Cellular Immune Reactions of Larvae of *Drosophila algonquin*," *Parasitology*, 70 (1975), 189–194: published by Cambridge University Press.

- a) Use prueba de the **Jonckheere-Terpstra** escriba el juego de hipótesis y concluya ($\alpha = 0.05$).

Problema 5: Parcelas divididas con ARTools

Del artículo: Edward Durner, 2019. Effective Analysis of Interactive Effects with Non Normal Data Using the Aligned Rank Transform, ARTool and SAS University Edition, *Horticulturae* 2019, 5, 57; doi:10.3390/horticulturae5030057

Escribir un reporte: con tablas graficos interpretacion y conclusiones.

Del experimento: consider an experiment where six rates of three nitrogen sources were evaluated for strawberry yield ($\text{g} \cdot \text{plant}^{-1}$). The experimental design was a split plot with nitrogen source as the main plot and nitrogen rate as the sub-plot. There were five replicates of the main plot and the main plots were set in a randomized complete block design.

- a) Analice los datos para un modelo de parcelas divididas en forma paramétrica.

$$Y_{ijk} = \mu + S_k + \alpha_i + \delta_{ik} + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

donde: - Y_{ijk} es la observación del nivel i del factor A (tratamiento de parcela grande) y el nivel j del factor B (tratamiento de subparcela) para el nivel k de S (bloque), - μ es la media de la población, - S_k es el efecto aleatorio del bloque k (bkl) - α_i es el efecto fijo del nivel i del factor A (nit), - δ_{ik} es el efecto de error aleatorio de la parcela grande. - β_j es el efecto fijo del nivel j del factor B (rate), - $\alpha\beta_{ij}$ el efecto de interacción entre los factores A y B, - ϵ_{ijk} es el efecto de error aleatorio de la subparcela. $\sim n(0, \sigma^2)$

- c). Utilizando el modelo de parcelas divididas (parametrico)

En sas

```
proc anova;
classes blk nit rate;
model yield = blk nit blk*nit rate nit*rate;
test h = nit e = blk*nit;
run;
```

En R

```
library(agricolae)
anova_par_div = aov(yield ~ blk+ nit*rate + Error(blk:nit), data = datos3)
summary(anova_par_div)
```

- a) pruebe los efectos de tratamientos y la interacción tratamiento vs tiempo, use $\alpha = 0.05$. Verifique supuestos.
- b). utilice el método de Tukey para investigar cual es el mejor tratamiento.
- c) Realizar la gráfica de interacción nit*rate e interpretar
- d) Realice el análisis no paramétrico de utilizando rangos alineados con ARTool Código en R:

```
library(ARTool)
anova_rangos<- art(Yield ~ Nitrogen*Rate+ (1|blk),data=datos3)
anova(anova_rangos)
```

```
library(emmeans)
library(dplyr)
library(tidyr)
library(phia)
library(ggplot2)
media=artlm(anova_rangos,"Nitrogen:Rate")
marginal=emmeans(media, ~Nitrogen*Rate)
marginal
pares=pairs(marginal,adjust = "tukey")
pares
plot(marginal)
plot(pares)
```

Guia: Resolver los incisos y prepare un reporte.

blk	Nitrogen	Rate	Yield
1	Urea	1	430
2	Urea	1	343
3	Urea	1	337
4	Urea	1	322
5	Urea	1	228
1	Urea	2	318
2	Urea	2	337
3	Urea	2	341
4	Urea	2	300
5	Urea	2	418
1	Urea	3	328
2	Urea	3	311
3	Urea	3	349
4	Urea	3	301
5	Urea	3	240
1	Urea	4	335
2	Urea	4	303
3	Urea	4	309
4	Urea	4	310
5	Urea	4	402
1	Urea	5	314
2	Urea	5	333
3	Urea	5	346
4	Urea	5	403
5	Urea	5	350
1	Urea	6	301
2	Urea	6	341
3	Urea	6	442
4	Urea	6	310
5	Urea	6	328
1	Calcium nitrate	1	353
2	Calcium nitrate	1	381
3	Calcium nitrate	1	355
4	Calcium nitrate	1	358
5	Calcium nitrate	1	485
1	Calcium nitrate	2	381
2	Calcium nitrate	2	405
3	Calcium nitrate	2	377
4	Calcium nitrate	2	382
5	Calcium nitrate	2	380
1	Calcium nitrate	3	420
2	Calcium nitrate	3	427
3	Calcium nitrate	3	512
4	Calcium nitrate	3	434
5	Calcium nitrate	3	408
1	Calcium nitrate	4	453
2	Calcium nitrate	4	438
3	Calcium nitrate	4	540
4	Calcium nitrate	4	453
5	Calcium nitrate	4	470
1	Calcium nitrate	5	467
2	Calcium nitrate	5	467
3	Calcium nitrate	5	434
4	Calcium nitrate	5	425
5	Calcium nitrate	5	450
1	Calcium nitrate	6	426
2	Calcium nitrate	6	437
3	Calcium nitrate	6	452
4	Calcium nitrate	6	444
5	Calcium nitrate	6	368
1	Potassium nitrate	1	325
2	Potassium nitrate	1	343
3	Potassium nitrate	1	372
4	Potassium nitrate	1	445
5	Potassium nitrate	1	353
1	Potassium nitrate	2	334
2	Potassium nitrate	2	365
3	Potassium nitrate	2	342
4	Potassium nitrate	2	348
5	Potassium nitrate	2	372
1	Potassium nitrate	3	325
2	Potassium nitrate	3	341
3	Potassium nitrate	3	349
4	Potassium nitrate	3	362
5	Potassium nitrate	3	374
1	Potassium nitrate	4	369
2	Potassium nitrate	4	351
3	Potassium nitrate	4	371
4	Potassium nitrate	4	356
5	Potassium nitrate	4	439
1	Potassium nitrate	5	334
2	Potassium nitrate	5	365
3	Potassium nitrate	5	345
4	Potassium nitrate	5	365
5	Potassium nitrate	5	358
1	Potassium nitrate	6	326
2	Potassium nitrate	6	333
3	Potassium nitrate	6	337
4	Potassium nitrate	6	350
5	Potassium nitrate	6	250

Problema 6

Los machos de la magnífica fragata (*Fregata magnificens*) tienen una gran bolsa roja en la garganta. Muestran visualmente esta bolsa y la usan para emitir un sonido de tambor cuando buscan pareja. Madsen et al. (2004) querían saber si las hembras, que presumiblemente eligen pareja en función del tamaño de su bolsa, podían utilizar el tono del sonido del tambor como indicador del tamaño de su bolsa. Los autores estimaron el volumen de la bolsa y la frecuencia fundamental del sonido del tambor en 18 machos.

Volume (cm ³)	Frequency (Hz)
1760	529
2040	566
2440	473
2550	461
2730	465
2740	532
3010	484
3080	527
3370	488
3740	485
4910	478
5090	434
5090	468
5380	449
5850	425
6730	389
6990	421
7960	416

- b) Realice una grafica X vs Y.
- c) Obtenga el coeficiente de correlación de Spearman, kendall, y el de Pearson. Interprete.

Problema 7

Del trabajo de:

KOKLU, M. and OZKAN, I.A., (2020), Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques. Computers and Electronics in Agriculture, 174, 105507.
<https://doi.org/10.1016/j.compag.2020.105507>

En esta investigación se utilizaron siete diferentes variedades de frijoles secos, teniendo en cuenta las características como forma, aspecto, tipo y estructura según la situación del mercado. Se desarrolló un sistema de visión por computadora para distinguir siete variedades diferentes registradas de frijol seco con características similares para obtener una clasificación uniforme de semillas. Para el modelo de clasificación se tomaron imágenes de 13,611 granos de siete variedades de granos de frijol secos diferentes registrados con una cámara de alta resolución. Las imágenes de frijoles obtenidas por el sistema de visión por computadora

se sometieron a etapas de segmentación y extracción de características, y un total de 16 características; De los granos se obtuvieron 12 dimensiones y 4 formas.

Información de atributos:

- 1.) **Area (A)**: The area of a bean zone and the number of pixels within its boundaries.
- 2.) **Perimeter (P)**: Bean circumference is defined as the length of its border.
- 3.) **Major axis length (L)**: The distance between the ends of the longest line that can be drawn from a bean.
- 4.) **Minor axis length (l)**: The longest line that can be drawn from the bean while standing perpendicular to the main axis.
- 5.) **Aspect ratio (K)**: Defines the relationship between L and l.
- 6.) **Eccentricity (Ec)**: Eccentricity of the ellipse having the same moments as the region.
- 7.) **Convex area (C)**: Number of pixels in the smallest convex polygon that can contain the area of a bean seed.
- 8.) **Equivalent diameter (Ed)**: The diameter of a circle having the same area as a bean seed area.
- 9.) **Extent (Ex)**: The ratio of the pixels in the bounding box to the bean area.
- 10.) **Solidity (S)**: Also known as convexity. The ratio of the pixels in the convex shell to those found in beans.
- 11.) **Roundness (R)**: Calculated with the following formula: $(4\pi A)/(P^2)$
- 12.) **Compactness (CO)**: Measures the roundness of an object: Ed/L
- 13.) **ShapeFactor1 (SF1)**
- 14.) **ShapeFactor2 (SF2)**
- 15.) **ShapeFactor3 (SF3)**
- 16.) **ShapeFactor4 (SF4)**
- 17.) **Variety** (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira)

obtener datos:

<https://archive.ics.uci.edu/ml/machine-learning-databases/00602/DryBeanDataset.zip>

- a) Realice un análisis de correlación para investigar la relación entre atributos de las semillas de frijol utilizando los coeficientes de Pearson, Sperman y Kendall, interprete sus resultados. Los atributos presentan normalidad.
- b) Investigue que atributos pueden detectar mejor la diferencia entra variedades de frijol, use Kruskall Wallis o tranformacion a rangos de conover $\alpha = .05$
- c) Realice las gráficas que apoyen su interpretación,