

Estadística (Química) - Primer Cuatrimestre- 2020 - Coronavirus
Práctica 4 - Estadística Descriptiva

1. Una muestra estándar de suero sanguíneo contiene 42.0 g de albúmina por litro. Cuatro laboratorios (A – D) realizan seis determinaciones cada uno (en el mismo día) de la concentración de albúmina, con los siguientes resultados (en g/l):

A	42.5	41.6	42.1	41.9	41.1	42.2
B	39.8	43.6	42.1	40.1	43.9	41.9
C	43.5	42.8	43.8	43.1	42.7	43.3
D	39.4	41.9	40.7	40.3	42.6	39.0

(para resolver este ejercicio, mire las instrucciones en el archivo `instruccionesej1.R` cuyo link figura cerquita de dónde bajó este archivo).

- (a) Grafique en cada caso los datos en un Diagrama de Punto, todos en la misma escala.
 (b) Comente la insesgadez (o exactitud) y la precisión de cada una de estas series de resultados. ¿Qué estadísticos utilizó para responder?
2. En un experimento se midió la temperatura de sublimación del iridio y del rodio. Los datos figuran más adelante. La siguiente instrucción puede ser útil para leer el archivo en R (ya que las longitudes de los dos conjuntos de datos son diferentes).

`read.table(file.choose(), head=TRUE, fill=TRUE)`

- (a) Compare los dos conjuntos de datos mediante histogramas y box plots. Grafique los dos boxplots en el mismo gráfico.
 (b) Halle las medias, las medianas y las medias podadas al 10% y 20%. Compare.
 (c) Halle el desvío estándar, la distancia intercuartil y la MAD como medidas de dispersión.
 (d) Halle los percentiles 0.90, 0.75, 0.50, 0.25 y 0.10.
 (e) Se cuenta con la siguiente información adicional sobre este conjunto de datos: se listaron en el orden en que fueron tomadas. Grafique las observaciones en el orden en que fueron tomadas. ¿Es razonable el modelo de errores independientes e idénticamente distribuidos? ¿Sería correcto analizar todo el conjunto de datos? ¿A partir de cuál observación considera que el proceso se ha estabilizado?
 (f) Sobre el conjunto de datos correspondiente al proceso estabilizado, resuelva nuevamente los ítems a) – d).

IRIDIO								
136.6	145.2	151.5	162.7	159.1	159.8	160.8	173.9	160.1
160.4	161.1	160.6	160.2	159.5	160.3	159.2	159.3	159.6
160.0	160.2	160.1	160.0	159.7	159.5	159.5	159.6	159.5

RODIO							
126.4	135.7	132.9	131.5	131.1	131.1	131.9	132.7
133.3	132.5	133.0	133.0	132.4	131.6	132.6	132.2
131.3	131.2	132.1	131.1	131.4	131.2	131.1	131.1

3. En un estudio nutricional se consideran las calorías y el contenido de sodio de tres tipos de salchichas y se obtuvieron los datos que se muestran en la tabla final.
- (a) Realice un histograma para las calorías de cada tipo de salchichas. ¿Observa grupos en algún gráfico? ¿Cuántos grupos observa? ¿Observa algún candidato a outlier?

- (b) Repita con la cantidad de sodio.
- (c) Realice los box plots paralelos para las calorías. ¿Observa la misma cantidad de grupos que antes? ¿A qué conclusión llega? De acuerdo con los box plots hallados, ¿cómo caracterizaría la diferencia entre los tres tipos de salchichas desde el punto de vista de las calorías?
- (d) ¿Le parece que los datos sustentan la igualdad del contenido medio calórico de las tres variedades?

A		B		C	
Calorías	Sodio	Calorías	Sodio	Calorías	Sodio
186	495	173	458	129	430
181	477	191	506	132	375
176	425	182	473	102	396
149	322	190	545	106	383
184	482	172	496	94	387
190	587	147	360	102	442
158	370	146	387	87	359
139	322	139	386	99	357
175	479	175	507	170	528
148	375	136	393	113	513
152	330	179	405	135	426
111	300	153	372	142	513
141	386	107	144	86	358
153	401	195	511	143	581
190	645	135	405	152	588
157	440	140	428	146	522
131	317	138	339	144	545
149	319				
135	298				
132	253				

4. Los siguientes datos corresponden a 100 determinaciones repetidas de la concentración de ion nitrato (en $\mu\text{g/l}$), 50 de ellas corresponden a un grupo de estudiantes (Grupo A) y las restantes 50 a otro grupo (Grupo B):

Grupo A				Grupo B			
0.51	0.51	0.50	0.50	0.61	0.67	0.70	0.56
0.51	0.50	0.48	0.50	0.61	0.67	0.63	0.75
0.49	0.50	0.47	0.50	0.55	0.49	0.75	0.55
0.51	0.48	0.52	0.49	0.65	0.77	0.64	0.58
0.51	0.49	0.52	0.51	0.73	0.51	0.66	0.65
0.51	0.48	0.52	0.49	0.65	0.57	0.61	0.53
0.52	0.53	0.49	0.51	0.53	0.64	0.52	0.69
0.48	0.51	0.50	0.47	0.54	0.57	0.63	0.62
0.51	0.49	0.50	0.50	0.68	0.56	0.58	0.58
0.50	0.49	0.53	0.47	0.59	0.62	0.60	0.66
0.51	0.50	0.49	0.48	0.61	0.54	0.63	0.61
0.53	0.52	0.51	0.51	0.69	0.63	0.53	0.48
0.46	0.49			0.65	0.61		

- (a) Estudie si la distribución de los conjuntos de datos para ambos grupos es normal, realizando los correspondientes histogramas y superponiendo la curva normal. Además dibuje los qqplots para cada conjunto de datos.

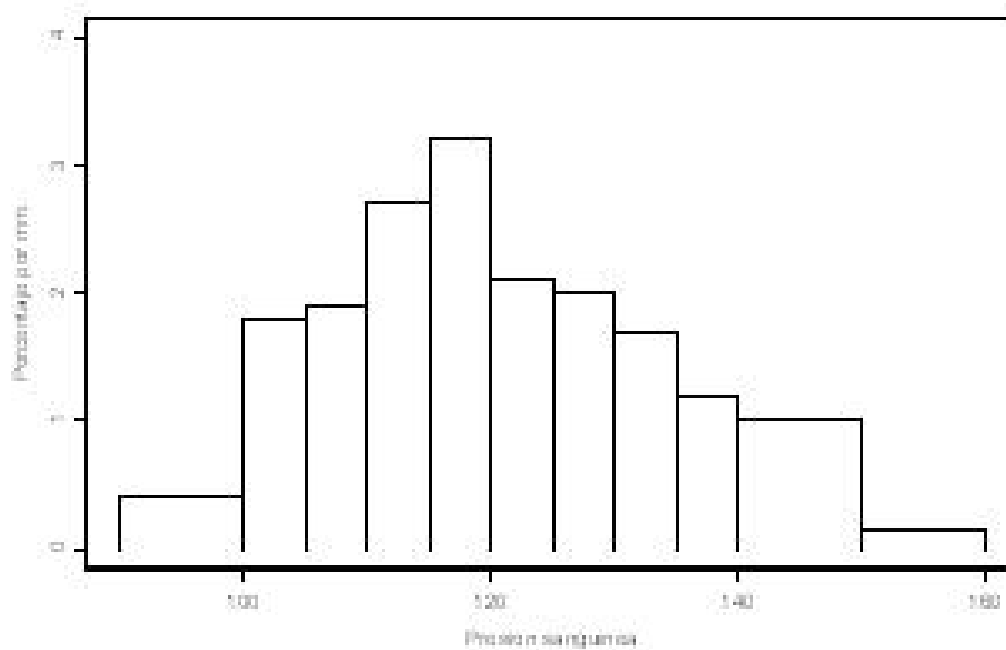
- (b) ¿Le parece a partir de estos datos que ambos grupos están midiendo lo mismo? Responda comparando medidas de centralidad y de dispersión de los datos. Hacer un box-plot de ambos conjuntos de datos en el mismo gráfico.
5. En una clase grande, hay 150 estudiantes hombres y 100 mujeres, los estudiantes promediaban en altura 178cm, con un desvío standard de 25cm y las estudiantes promediaban 162cm, con un desvío standard de 25cm. Tomando conjuntamente los hombres y las mujeres:
- (a) La altura promedio es _____.
- (b) El desvío estándar de las alturas es ¿un poco mayor que 25cm, exactamente 25cm, ó un poco inferior a 25cm?
6. Con la finalidad de incrementar las lluvias en zonas desérticas, se desarrolló un método que consiste en el bombardeo de la nube con átomos. Para evaluar la efectividad del método se realizó el siguiente experimento:
- Para cada nube que se podía bombardear se decidió al azar si se la trataba o no.
 - Las nubes no tratadas fueron denominadas nubes controles.

Las tablas presentan la cantidad de agua caída de 26 nubes tratadas y 26 nubes controles.

NUBES TRATADAS								
129.6	31.4	2745.6	489.1	430.0	302.8	119.0	4.1	92.4
17.5	200.7	274.7	274.7	7.7	1656.0	978.0	198.6	703.4
1697.8	334.1	118.3	255.0	115.3	242.5	32.7	40.6	

NUBES CONTROLES								
26.1	26.3	87.0	95.0	372.4	0.01	17.3	24.4	11.5
321.2	68.5	81.5	47.3	28.6	830.1	345.5	1202.6	36.6
4.9	4.9	41.1	29.9	163.0	244.3	147.8	21.7	

- (a) Realice box plots para ambos grupos. ¿Le parece que el método produce algún efecto?
- (b) Analice la normalidad realizando qqplots e histogramas (de densidad) para ambos conjuntos de datos y superponiendo la curva normal. Para hacerlo, considere la instrucción (donde en el vector `ww` guardamos los datos de controles o tratadas):
- ```
hist(ww,probability=T)
equis<-seq(min(ww),max(ww),by=0.01)
lines(equis,dnorm(equis, mean = mean(ww),sd = sqrt(var(ww))),col=2,lwd=2)
```
- (c) Realice la transformación log a los datos y repita b) para los datos transformados.
7. El histograma muestra la distribución de presión sanguínea de 15000 mujeres en estudio. Utilice el histograma para responder las siguientes preguntas:
- (a) ¿El porcentaje de mujeres con presión superior a los 130mm es más cercano a 25%, 50% o 75%?
- (b) ¿El porcentaje de mujeres con presión entre 90 y 160mm es más cercano al 1%, 50% o 99%?
- (c) ¿Cuál intervalo tiene más mujeres: 130-140mm o 140-150mm?
- (d) ¿Cuál intervalo tiene más mujeres: 130-135mm o 140-150mm?



- (e) En el intervalo 125-130mm la altura del histograma es de alrededor de 2.2% por mm. ¿Qué porcentaje de mujeres tuvo presión en ese intervalo?
8. Considere  $x_1, \dots, x_n$  una muestra de una población cualquiera. Sean  $\bar{x}$  y  $\tilde{x}$  la media y la mediana muestral respectivamente.
- Si se suma una constante  $a$  a cada uno de los  $x_i$  de la muestra, obteniéndose  $y_i = x_i + a$ , ¿cómo se relacionan  $\bar{x}$  con  $\bar{y}$  y  $\tilde{x}$  con  $\tilde{y}$ ?
  - Si cada uno de los  $x_i$  es multiplicado por una constante  $b$ , obteniéndose  $y_i = bx_i$ , ¿cómo se relacionan  $\bar{x}$  con  $\bar{y}$  y  $\tilde{x}$  con  $\tilde{y}$ ?
9. Sea  $s_X^2$  la varianza muestral correspondiente a los datos  $x_1, \dots, x_n$  observados. Demuestre que
- $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2$ .
  - Probar que si  $y_i = x_i + a$  con  $a$  constante, entonces  $s_X = s_Y$ .
  - Probar que si  $y_i = bx_i$  con  $b$  constante, entonces  $s_X = |b| s_Y$ .