PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA

1EST14 - EXPERIMENTACIÓN NUMÉRICA
Examen 1
(Segundo Semestre 2023-2)

Problema 1 (7 puntos)

Los accidentes automovilísticos fatales representan una preocupación persistente en todo el mundo, con un impacto significativo en la seguridad pública y la calidad de vida.

Para llevar a cabo este análisis, se ha recolectado la información antes descrita de 48 ciudades en el año 2022 en el archivo accidentesFatales.csv.

Se definen las siguientes variables:

variable	descripción
cantidad	Cantidad de vehículos en la ciudad (cientos de miles)
velocidad	Velocidad media a la que se transita en las carreteras de dicha ciudad (en Km/h.)
noasfaltadas	Porcentaje de carreteras no asfaltadas
malestado	Porcentaje de calles en mal estado
sobriedad	Tiene el valor de 1 en caso haya presencia de puntos de control de sobriedad y 0 en caso contrario
iluminación	Estado de la calidad de iluminación en carreteras de la ciudad. Puede ser "Baja", "Media", "Óptima"
edad	Edad media de los conductores de la ciudad
antiguedad	Media de los años de los vehículos en uso
fatales	Cantidad anual de accidentes vehiculares fatales

```
d <- read.table("D:/maya/Documentos/00  2023-2 1EST14-ExpNumerica/evaluaciones/2023-2/Ex1/accidentesFatales.csv",
  header=TRUE, stringsAsFactors=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)

d <- within(d, {
  sobriedad <- factor(sobriedad, labels=c('sin control','con control'))
})
```

a. (2 puntos) Con nivel de significancia de 5%, ¿se puede afirmar que la cantidad de accidentes depende de la instalación de puntos de control?

H0: mu(FATALES_cpc) = mu(FATALES_spc)
H1: mu(FATALES_cpc) <> mu(FATALES_spc)

Primero se verifica que las muestras son independientes

Segundo probamos normalidad del número de accidentes fatales por sobriedad

```
> normalityTest(fatales ~ sobriedad, test="shapiro.test", data=d)
 p-values adjusted by the Holm method:
          unadjusted adjusted
```

sin control 0.62100   0.62100
con control 0.27321   0.54641
Dentro de cada grupo, la variable presenta normalidad
Se acepta la normalidad

Ahora verificamos igualdad de varianzas, mediante la prueba de razón de varianzas

```
> Tapply(fatales ~ sobriedad, var, na.action=na.omit, data=d) # variances by group
sin control con control
   348512.7    145265.3
> var.test(fatales ~ sobriedad, alternative='two.sided', conf.level=.95, data=d)
F test to compare two variances
data:  fatales by sobriedad
F = 2.3991, num df = 34, denom df = 12, p-value = 0.1075
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8169584 5.6454484
sample estimates:
ratio of variances
         2.399147
```

Se acepta que las varianzas son iguales

Finalmente se puede aplicar una prueba paramétrica t para la diferencia de medias (caso 2)

```
> with(d, numSummary(fatales, groups = sobriedad, statistics=c('mean', 'sd')))
               mean       sd data:n
sin control 2877.200 590.3496     35
con control 2429.385 381.1368     13
> t.test(fatales ~ sobriedad, alternative = "two.sided", conf.level = .95, var.equal = TRUE, data = d)
                                                                 Two Sample t-test
data:  fatales by sobriedad
t = 2.5364, df = 46, p-value = 0.01466
alternative hypothesis: true difference in means between group sin control and group con control is not equal to 0
95 percent confidence interval:
  92.42248 803.20829
sample estimates:
mean in group sin control mean in group con control
```

Se rechaza que las medias son iguales, por lo que se asume que La cantidad de accidentes fatales, son diferentes. Además, se observa que disminuye significativamente en una ciudad en que hay puntos de control de sobriedad.

b. (2 puntos) Con nivel de significancia de 5%, ¿se puede afirmar que la cantidad de accidentes depende de la iluminación?

H0: Para todo i=1, 2, 3: mu(FATALES_i) = mu(FATALES)
H1: existe i tal que: mu(FATALES_i) <> mu(FATALES)

Primero se verifica que las muestras son independientes

Segundo probamos normalidad del número de accidentes fatales por sobriedad

```
> normalityTest(fatales ~ iluminacion, test="pearson.test", data=d)
p-values adjusted by the Holm method:
```

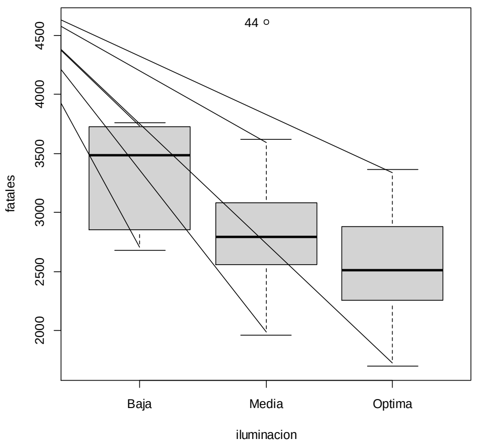
unadjusted adjusted
Baja 0.13534 0.40601
Media 0.67115 1.00000
Optima 0.82084 1.00000
Dentro de cada grupo, la variable presenta normalidad
Se acepta la normalidad

Ahora verificamos igualdad de varianzas, mediante la prueba de razón de varianzas
> Tapply(fatales ~ iluminacion, var, na.action=na.omit, data=d) # variances by group
Baja Media Optima
207791.8 388291.4 196353.0
> bartlett.test(fatales ~ iluminacion, data=d)
Bartlett test of homogeneity of variances
data: fatales by iluminacion
Bartlett's K-squared = 2.3366, df = 2, p-value = 0.3109
Se acepta que las varianzas son iguales

Finalmente se puede aplicar una prueba ANOVA para la medias
> modelo1 <- aov(fatales ~ iluminacion, data = d)
> summary(modelo1)
Df Sum Sq Mean Sq F value Pr(>F)
iluminacion 2 3529461 1764730 6.638 0.00298 **
Residuals 45 11964093 265869

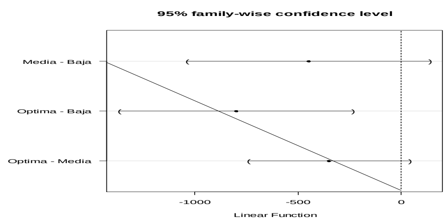
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> with(d, numSummary(fatales, groups = iluminacion, statistics=c('mean', 'sd')))
mean sd data:n
Baja 3332.833 455.8418 6
Media 2882.471 623.1303 17
Optima 2531.400 443.1174 25
Considerando un nivel de significancia alpha=0.05, se rechaza que las medias son iguales, por lo que se asume que la cantidad de accidentes fatales está influenciada por la iluminación.
En la tabla de medias encontramos que a medida que la iluminación mejor, disminuye los accidentes fatales.

> Boxplot(fatales ~ iluminacion, data=d, id=list(method="y"))
[1] "44"



Multiple Comparisons of Means: Tukey Contrasts

	Estimate	lwr	upr
Med-Baj=0	-450.36	-1040.19	139.47
Opt-Baj=0	-801.43	-1366.11	-236.76
Opt-Med=0	-351.07	-741.55	39.40



La diferencia de medias de accidentes fatales es significante si comparamos entre ciudades con iluminación optima y con iluminación baja.
La media con baja no es significante, como tampoco óptima con media

c. (2 puntos) Con nivel de significancia de 5%, ¿se puede afirmar que la velocidad depende de la existencia de los puntos de control de la sobriedad?

Hipótesis:
H0: mu(Vcpc) = mu(Vspc)
H1: mu(Vcpc) <> mu(Vspc)

Primero se verifica que las muestras son independientes
De hecho, porque las ciudades o tienen o no tienen los puntos de control, forman 2 poblaciones separadas.

Segundo probamos normalidad de la velocidad por sobriedad

```
> normalityTest(velocidad ~ sobriedad, test="shapiro.test", data=d)
p-values adjusted by the Holm method:
          unadjusted  adjusted
sin control 0.000014975 0.00002995
con control 0.75073     0.75073
```

NO se acepta la normalidad, pues para el caso de las ciudades que no poseen puntos de control de sobriedad, la velo9cidad en carretera no mostró normalidad.

Por el motivo citado arriba, ya se debe realizar una prueba NO paramétrica, en este caso la prueba Prueba de U Mann Whitney para muestras independientes.
H0: ME(Vcpc) = ME(Vspc)
H1: ME(Vcpc) <> ME(Vspc)

```
> Tapply(velocidad ~ sobriedad, median, na.action=na.omit, data=d) # medians by group
sin control con control
        91         85
> wilcox.test(velocidad ~ sobriedad, alternative="two.sided", data=d)
Wilcoxon rank sum test with continuity correction
data:  velocidad by sobriedad
W = 250.5, p-value = 0.6012
alternative hypothesis: true location shift is not equal to 0
```

Considerando un nivel de significancia alpha=0.05, se acepta que las mediana de la velocidad en las carreteras de ciudad son iguales para las ciudades con puntos de control de sobriedad y sin puntos de control de sobriedad.

d. (1 punto) El porcentaje de carreteras no asfaltadas es mayor que el porcentaje de calles en mal estado?

Hipótesis:
H0: mu(%Carr) = mu(%malestado)
H1: mu(%Carr) > mu(%malestado)

Primero se verifica que las muestras son pareadas
De hecho, porque en cada ciudad se mide de manera independiente el % de carreteras no asfaltadas y el % porcentaje de calles en mal estado.

Segundo, definimos: dCarrCall = % carreteras no asfaltadas - % calles en mal estado

```
> d$dCarrCall <- d$noasfaltadas - d$malestado
```

Luego, probamos la normalidad de la diferencia dCarrCall

> normalityTest(~dCarrCall, test="shapiro.test", data=d)
Shapiro-Wilk normality test
data: dCarrCall
W = 0.97001, p-value = 0.2535
Se acepta la normalidad de la diferencia.

Se puede realizar una prueba PARAMÉTRICA, en este caso la prueba t para una muestra de la diferencia.
H0: mu(dCarrCall) = 0
H1: mu(dCarrCall) > 0
> with(d, (t.test(dCarrCall, alternative = "greater", mu = 0.0, conf.level = .95)))
One Sample t-test
data: dCarrCall
t = 0.14179, df = 47, p-value = 0.4439
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 -6.093923   Inf
sample estimates:
mean of x
   0.5625
Considerando un nivel de significancia alpha=0.05, se acepta que la media del % de carreteras no asfaltadas es igual a la media del % porcentaje de calles en mal estado.