Universitat Rovira i Virgili

Escola Tècnica Superior d’Enginyeria

ESTADÍSTICA

PRÀCTICA 7

FASE 2

NOM: IVAN MORILLAS GOMEZ

3/12/2021

2021-2022

Plantejament inicial:

Iniciar la resolución d'aquest problema requereix d'operacions amb les dades lliurades, per tant, el primer que s'ha de fer és carregar-les per tal de poder començar a treballar. El desenvolupament de la práctica ha estar realitzat amb el llenguatge de programación R.

Primer de tot, he creat un arxiu .txt amb les dades de la pràctica i després l'importaré dins del programa on les comes són el separador i els punts són els caràcters decimals.

```
> Dades <- read.table("/home/dades.txt", header=FALSE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

```
> Dades
      V1    V2
1   2.76  7.02
2   5.18  3.10
3   2.68  5.44
4   3.05  3.99
5   4.10  5.21
6   7.05 10.26
7   6.60 13.91
8   4.79 18.53
9   7.39  7.91
10  7.30  4.85
11 11.78 11.10
12  3.90  3.74
13 26.00 94.03
14 67.48 94.03
15 17.04 41.70
```

On V1 és el conjunt de dades al moment de la injecció i V2 és el conjunt de dades després de 30 minuts de la injecció

I ara, amb una comanda del llenguatge R, puc posar les dades horitzontalment per a una millor visualització i poder començar a treballar.

```
> V1 = as.numeric(as.character(Dades$V1))

> print(V1)
 [1]  2.76  5.18  2.68  3.05  4.10  7.05  6.60  4.79  7.39  7.30 11.78  3.90 26.00 67.48 17.04

> V2 = as.numeric(as.character(Dades$V2))

> print(V2)
 [1]  7.02  3.10  5.44  3.99  5.21 10.26 13.91 18.53  7.91  4.85 11.10  3.74 94.03 94.03 41.70
```

Doneu una estimació puntual del valor en mitjana i de la desviació poblacional de la variable X_M:

La primera tasca demanada és el calcul de l'estimació puntual del valor en mitjana i de la desviació poblaciones del conjunt de dades V1.

Per a calcular la estimació puntual de la mitjana, sencillament amb la comanda 'mean' ens calcula la mitjana.

```
> mitjanaV1 = mean(V1)
```

```
> cat("L'estimació puntual de la mitjana de V1 és de: ", mitjanaV1, "\n")
L'estimació puntual de la mitjana de V1 és de:  11.80667
```

Per a calcular la estimació puntual de la desviació poblacional, sencillament amb la comanda 'sd' ens calcula la desviació poblacional.

```
> desviacioV1 = sd(V1)
> cat("L'estimació puntual de la desviació poblacional de V1 és de: ", desviacioV1, "\n")
L'estimació puntual de la desviació poblacional de V1 és de:  16.63641
```

Calculeu un interval de confiança al 99% per al valor de la diferència de concentracions:

Per calcular intervals de confiança, R facilita funcions que desenvolupen automàticament els càlculs per arribar a diferents resultats sobre l'estudi de mostres, en concret, un dels resultats és l'interval de confiança.

Per a començar, restem els valors dels dos conjunts de concentracions i ens quedaría un conjunt així:

```
> Vresta = (V1-V2)
```

```
> print(Vresta)
 [1]  -4.26   2.08  -2.76  -0.94  -1.11  -3.21  -7.31 -13.74  -0.52   2.45   0.68   0.16 -68.03 -26.55 -24.66
```

I ara, per tal d'estudiar aquest conjunt, s'utilitzará la funció de R 't.test()', una prova t-Student. Agafarem el conjunt de dades i un nivell de confiança del 99%=0.99 (el demanat) i ens surt això:

```
> t.test(Vresta, conf.level = 0.99)

        One Sample t-test

data:  Vresta
t = -2.0646, df = 14, p-value = 0.058
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 -24.047153   4.351153
sample estimates:
mean of x
   -9.848
```

A la cinquena línea dels resultats es mostren dos números decimals separats per un espai, aquest és l'interval que buscavem:

{-24.047153 , 4.351153}

Amb un risc de 0.05, pot assegurar-se que les concentracions d'andrògens augmenten després de 30 minuts d'inhibició?:

Per comprovar si les concentracions d'andrògens augmenten, faré dos tests d'hipotesis, un per a cada conjunt de dades amb la comanda t.test() amb un nivell de confiança del 95%=0.95

```
        One Sample t-test

data:  V1
t = 2.7486, df = 14, p-value = 0.01569
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  2.593728 21.019605
sample estimates:
mean of x
 11.80667
```

```
        One Sample t-test

data:  V2
t = 2.7124, df = 14, p-value = 0.01684
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  4.531354 38.777980
sample estimates:
mean of x
 21.65467
```

Com es pot veure, si comparem les dos mitjanes obtingudes amb les hipótesis, la diferencia de mitjanes entre el conjunt de dades de V2, o sigui, després de 30 minuts d'inhibició i el conjunt de dades de V1, o sigui, al moment de la injecció, és d'aproximadamente 10. Llavors podem concluïr que es pot assegurar que les concentracions augmenten després de 30 minuts d'inhibició.

Podíem fer la suposició que la distribución dels dos conjunts de dades és normal? I que el primer conjunt té distribució exponencial?:

La següent demostració que es demana es per la confirmació de normalitat en les dades. Com no es tenen gaires dades ($n < 30$) s'aplicarà el test de Kolmogorov-Smirnov per verificar si les dades es poden considerar distribucions normals. En aquest test s'haurà d'avaluar el resultat de l'estadistic de prova i comparar-lo amb el corresponent valor de la taula KSL, en cas que es compleixi EP < valorTaula s'acceptará la hipótesis nul·la H_0, per tant, es podrá considerar normalitat en les dades. En cas contrari, no es por confirmar normalitat. El nivell de significació que es seguirá per al test será de 0.05. La hipótesis será bilateral, ja que es vol comprobar si es pot considerar o no una distribució.

```
> ks.test(V1, pnorm, mitjanaV1, desviacioV1)

        One-sample Kolmogorov-Smirnov test

data:  V1
D = 0.33801, p-value = 0.04955
alternative hypothesis: two-sided


> ks.test(V2, pnorm, mitjanaV2, desviacioV2)

        One-sample Kolmogorov-Smirnov test

data:  V2
D = 0.34025, p-value = 0.06204
alternative hypothesis: two-sided
```

En ambdues mostres no es poden considerar una distribució normal en les dades, ja que els valors de la D, que es l'estadistic de prova, són superiors, en ambdós casos, al valor de les taules (per a α=0.05).

```
> ks.test(V1, pexp, mitjanaV1, desviacioV1)

	One-sample Kolmogorov-Smirnov test

data:  V1
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

```
> ks.test(V2, pexp, mitjanaV2, desviacioV2)

	One-sample Kolmogorov-Smirnov test

data:  V2
D = 1, p-value = 1.872e-13
alternative hypothesis: two-sided
```

Oblidant ara tots els resultats obtinguts anteriormente, hi ha evidències significatives que els dos conjunts de dades tenen una diferencia de mitjanes de, com a mínim, 3 nanograms?:

En el següent apartat ens demana cercar si existeix evidencia suficient per considerar que hi ha diferència de com a mínim 3 nanograms entre els dos conjunts de dades.

Es planteja com a hipótesis nul·la H_0: μ_0 >= 3 i com a hipótesis alternativa H_1: μ_0 < 3.

```
> t.test(V1, V2, "less", mu = 3, conf.level = 0.99)

	Welch Two Sample t-test

data:  V1 and V2
t = -1.4172, df = 21.479, p-value = 0.08539
alternative hypothesis: true difference in means is less than 3
99 percent confidence interval:
     -Inf 12.93524
sample estimates:
mean of x mean of y
 11.80667  21.65467
```

Per a aquest cas, es compleix la hipótesis nul·la, la diferencia entre les mitjanes tenen una diferència de com a mínim 3 nanograms.

Mitjançant un procés d'analisis de la variància, contrasteu si hi ha diferencia significativa entre les mitjanes de les tres mostres:

Per tal de generar les dades necessàries amb una distribución normal es farà ús de la funció de R 'rnorm()'.

```
> V3 = rnorm(15, mean(mitjanaV1, mitjanaV2), mean(desviacioV1, desviacioV2))

> print(V3)
 [1]  28.848212  -4.179103  10.459258 -15.480646   8.840779  26.391388   9.809877  28.895268  -3.138826  44.230319
[11] -14.636376   4.484494  21.350797 -10.360441   6.295155
```

Posteriorment, s'han d'unir les dades donades amb les noves generades i, per preparar les dades per a l'ús de la funció de test ANOVA, es juntaran les dades en un mateix vector.

```
> V4 = c(V1, V2, V3)

> print(V4)
 [1]   2.760000   5.180000   2.680000   3.050000   4.100000   7.050000   6.600000   4.790000   7.390000   7.300000  11.780000
[12]   3.900000  26.000000  67.480000  17.040000   7.020000   3.100000   5.440000   3.990000   5.210000  10.260000  13.910000
[23]  18.530000   7.910000   4.850000  11.100000   3.740000  94.030000  94.030000  41.700000  28.848212  -4.179103  10.459258
[34] -15.480646   8.840779  26.391388   9.809877  28.895268  -3.138826  44.230319 -14.636376   4.484494  21.350797 -10.360441
[45]   6.295155
```

Per discernir entre les classificacions de les dades del vector V4, es crearan vectors que classifiquin la informació. Aquest vector classificadors s'apliquen amb la funció 'factor()'.

```
> tipus = rep(1:3, each = 15)

> tipus = factor(tipus, labels = c("Al moment", "30 minuts", "Noves dades"))
```

L'últim pas que queda és fer el test ANOVA, en R la funció 'aov()' realitzará aquesta feina.

```
> p.aov = aov(V4 ~ tipus)

> model.tables(p.aov)
Tables of effects

 tipus
tipus
  Al moment   30 minuts Noves dades
     -2.498       7.350      -4.851

> model.tables(p.aov, type = "mean")
Tables of means
Grand mean

14.30511

 tipus
tipus
  Al moment   30 minuts Noves dades
     11.807      21.655       9.454
```

Segons les dades resultants, com que el p-valor es molt petit, es pot concluïr que les diferències de les mitjanes entre els tres conjunts és significatiu.