PE-Problemes-e-status-Disseny-de...

fibernauta

Probabilidad y Estadística

2º Grado en Ingeniería Informática

Facultad de Informática de Barcelona (Fib)
Universidad Politécnica de Catalunya

Probabilitat i Estadística

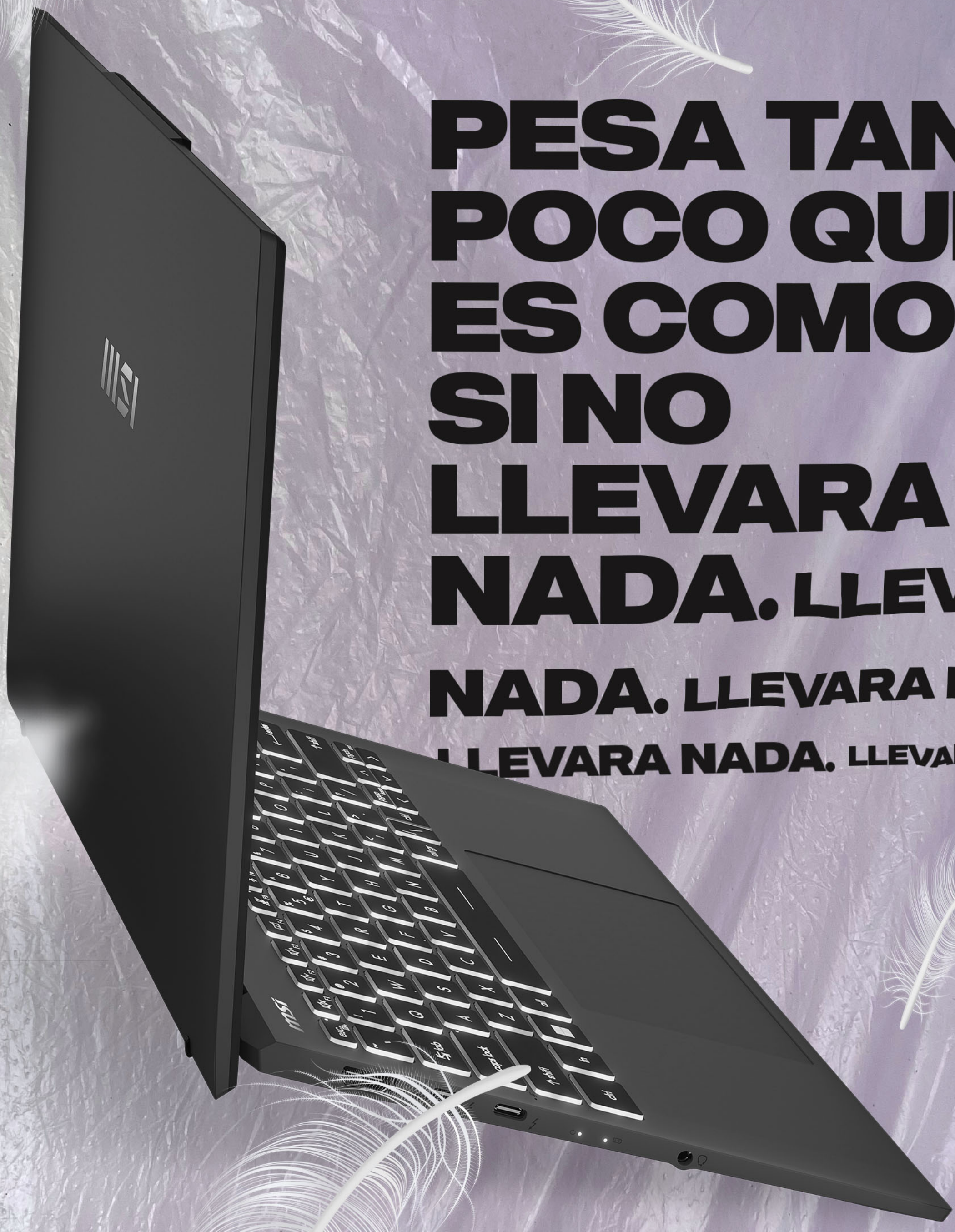
FIB-UPC

Problemes d’e-status:

B5 – Disseny d'experiments

Papilla de zombi

1. Vamos a comparar los resultados. Primero, di cuál es la diferencia promedio entre los zombies que has atropellado tú y los que ha atropellado B por partida. *Si tú has atropellado menos, el valor ha de ser negativo.*	11.85227
2. Calcula qué desviación estándar tiene los resultados del jugador A.	15.21244
3. Considerando que las características de dispersión (a lo largo de infinitas partidas) fueran iguales en los dos jugadores, calcula una estimación de la desviación tipo que tenéis en común.	16.88778
4. Sabemos que, de noche a noche, la diferencia de promedios de A y B fluctúa, con una desviación tipo que llamamos error tipo o standard error. ¿Puedes estimar este valor con los datos disponibles?	6.614518
5. La información que conocemos de la muestra de esta noche da lugar a un estadístico, utilizado para poner a prueba la hipótesis H: $\mu_A = \mu_B$, que tomaremos bilateral porque a priori ninguno de los dos tiene ventaja. ¿Cuánto vale este estadístico?	1.791857
6. Halla el valor P asociado a la prueba de hipótesis que has realizado. (4 decimales correctos)	0.08527
7. La pregunta definitiva es qué diferencia hay entre los promedios de los dos jugadores. Estima dicha diferencia con un intervalo de confianza 99% (al menos dos cifras decimales correctas).	-6.58527 30.2898
8. ¿Cuál sería la conclusión de la prueba? 1. La evidencia es que B lo hace mejor 2. No se puede descartar que A y B tengan el mismo nivel 3. Parece que los dos sois igual de buenos 4. La evidencia es que A lo hace mejor *elige un valor entre 1 y 4*	2



MSI

PESA TAN POCO QUE ES COMO SI NO LLEVARA NADA. LLEVARA NADA. LLEVARA NADA. LLEVARA NADA. LLEVARA NADA.

MSI PRESTIGE 13 AI STUDIO

¡ESTE PORTÁTIL ES LA BOMBA SI QUIERES POTENCIA Y LLEVARLO A TODOS LADOS! SOLO PESA 990 GRAMOS Y TIENE UNA PANTALLA DE 13 PULGADAS, MÁS LIGERO QUE LOS APUNTES DE MEDICINA.

ES PEQUEÑO PERO PODEROSO, CON LOS NUEVOS CHIPS INTEL CORE ULTRA 7 Y GRÁFICAS INTEL ARC.

ADEMÁS, TRAE INTELIGENCIA ARTIFICIAL PARA HACERTE LA VIDA MÁS FÁCIL. EL MSI AI ENGINE SE ADAPTA AUTOMÁTICAMENTE A LO QUE ESTÁS HACIENDO: BAJA EL RUIDO EN REUNIONES, AHORRA BATERÍA EN TAREAS SENCILLAS Y DA EL MÁXIMO EN LAS PESADAS. ¡ES COMO TENER UN ASISTENTE PERSONAL PARA TU LAPTOP!

MIRA MIRA

Script en R

```
A = c(59, 64, 82, 92, 101, 80, 108, 93, 84, 96, 73)
B = c(78, 62, 36, 90, 51, 78, 85, 79, 63, 78, 101, 62, 54,
103, 77, 69)
p1 = mean(A) - mean(B)
p2 = sd(A) # Desviación estándar del jugador B: p2 = sd(B)
n1 = length(A); n2 = length(B)
s2 = ( (n1 - 1)*(sd(A)^2) + (n2 - 1)*(sd(B)^2) ) / ( (n1 - 1)
+ (n2 - 1) )
# mA = mean(A); mB = mean(B)
# s2 = ( sum((A-mA)^2) + sum((B-mB)^2) ) / (n1 + n2 - 2)
p3 = sqrt(s2) # Estimación de la variancia en común: p3 = s2
p4 = se = sqrt(s2/n1 + s2/n2)
p5 = t = (mean(A) - mean(B)) / se
# t.test(A, B, var.equal=TRUE)$statistic
p6 = 2 * (1 - pt(abs(t), n1 + n2 - 2))
# t.test(A, B, var.equal=TRUE)$p.value
alfa = 1 - 99/100
p7 = c(p1 - qt(1 - alfa/2, n1 + n2 - 2) * se, p1 + qt(1 -
alfa/2, n1 + n2 - 2) * se)
# t.test(A, B, var.equal=TRUE, conf.level=0.99)$conf.int
p8 = 2 # p-valor > 0.05 -> No podem rebutjar H₀: μA = μB

p1; p2; p3; p4; p5; p6; p7; p8
```

Consola de R

```
> A = c(59, 64, 82, 92, 101, 80, 108, 93, 84, 96, 73)
> B = c(78, 62, 36, 90, 51, 78, 85, 79, 63, 78, 101, 62, 54,
103, 77, 69)
> p1 = mean(A) - mean(B)
> p2 = sd(A)
> n1 = length(A); n2 = length(B)
> s2 = ( (n1 - 1)*(sd(A)^2) + (n2 - 1)*(sd(B)^2) ) / ( (n1 -
1) + (n2 - 1) )
> p3 = sqrt(s2)
> p4 = se = sqrt(s2/n1 + s2/n2)
> p5 = t = (mean(A) - mean(B)) / se
> p6 = 2 * (1 - pt(abs(t), n1 + n2 - 2))
> alfa = 1 - 99/100
> p7 = c(p1 - qt(1 - alfa/2, n1 + n2 - 2) * se, p1 + qt(1 -
alfa/2, n1 + n2 - 2) * se)
> p8 = 2
> p1; p2; p3; p4; p5; p6; p7; p8

[1] 11.85227
[1] 15.21244
[1] 16.88778
[1] 6.614518
[1] 1.791857
[1] 0.08526912
[1] -6.585273 30.289818
[1] 2
```

Lentillas para el trabajo

	1. Calcula la media de la humedad medida en las lentillas de tipo New.	44.4
	2. Halla las diferencias (New - Ref) en la humedad medida entre los dos tipos de lente, y calcula el promedio de estas diferencias.	2.7875
	3. Calcula la desviación típica de las diferencias.	6.489208
	4. Calcula la desviación típica de la media de las diferencias (el error típico).	2.294282
	5. Calcula el valor del estadístico para contrastar la hipótesis nula: "no hay diferencia en humedad entre lentillas de tipo New y de tipo Ref".	1.214977
	6. ¿Cuántos grados de libertad hay que utilizar para la distribución de referencia del estadístico calculado?	7
	7. Calcula el p-valor de la prueba de hipótesis anterior.	0.2637615
	8. Con un riesgo del 5%, ¿aceptarías o rechazarías la hipótesis nula? ¿Y con un riesgo del 1%? 1. aceptar con 1 y 5; 2. aceptar con 5 y rechazar con 1; 3. rechazar con 5 y aceptar con 1; 4. rechazar con 1 y 5.	1
	9. Calcula el intervalo de confianza del 95 por ciento para el efecto que consigue el nuevo tipo de lentilla (incremento de humedad).	-2.637614 8.212614

Script en R

```
Id      Hume    Ojo     Nvo
7       64.6    1       0
3       10.6    0       0
4       1       0       1
1       62.1    1       1
8       33.8    1       0
6       43.1    0       0
6       44.7    1       1
7       60.9    0       1
5       81.8    0       1
5       77.7    1       0
2       34.9    0       0
8       36.9    0       1
4       1       1       0
2       41.6    1       1
3       26.2    1       1
1       67.2    0       0

Id      Hume    Ojo     Nvo
1       67.2    0       0
2       34.9    0       0
3       10.6    0       0
4       1       1       0
5       77.7    1       0
6       43.1    0       0
7       64.6    1       0
8       33.8    1       0
1       62.1    1       1
2       41.6    1       1
3       26.2    1       1
4       1       0       1
5       81.8    0       1
6       44.7    1       1
7       60.9    0       1
8       36.9    0       1

Ref = c(67.2, 34.9, 10.6, 1, 77.7, 43.1, 64.6, 33.8)
New = c(62.1, 41.6, 26.2, 1, 81.8, 44.7, 60.9, 36.9)
p1 = mean(New) # media de la humedad en las de tipo Ref -> mean(Ref)
Dif = New - Ref
p2 = mean(Dif)
p3 = sd(Dif)
p4 = p3/sqrt(length(Dif))
p5 = p2/p4
# t.test(New, Ref, paired = TRUE)$statistic
p6 = length(Dif) - 1
# t.test(New, Ref, paired = TRUE)$parameter
```

```
p7 = 2 * (1 - pt(abs(p5), p6))
# t.test(New, Ref, paired = TRUE)$p.value
opciones = list("aceptar con 1 y 5"=1, "aceptar con 5 y
rechazar con 1"=2, "rechazar con 5 y aceptar con 1"=3,
"rechazar con 1 y 5"=4)
if (p7 < 0.05) {
    if (p7 < 0.01) {
        p8 = opciones$"rechazar con 1 y 5";
    } else {
        p8 = opciones$"rechazar con 5 y aceptar con 1";
    }
} else if (p7 < 0.01) {
    p8 = opciones$"aceptar con 5 y rechazar con 1";
} else {
    p8 = opciones$"aceptar con 1 y 5";
}
alfa = 1 - 95/100
p9 = c(p2 - qt(1 - alfa/2, p6) * p4, p2 + qt(1 - alfa/2, p6) * p4)
# t.test(New, Ref, paired = TRUE, conf.level=0.95)$conf.int

p1; p2; p3; p4; p5; p6; p7; p8; p9
```

Consola de R

```
> Ref = c(67.2, 34.9, 10.6, 1, 77.7, 43.1, 64.6, 33.8)
> New = c(62.1, 41.6, 26.2, 1, 81.8, 44.7, 60.9, 36.9)
> p1 = mean(New)
> Dif = New - Ref
> p2 = mean(Dif)
> p3 = sd(Dif)
> p4 = p3/sqrt(length(Dif))
> p5 = p2/p4
> p6 = length(Dif) - 1
> p7 = 2 * (1 - pt(abs(p5), p6))
> opciones = list("aceptar con 1 y 5"=1, "aceptar con 5 y rechazar con 1"=2, "rechazar con 5 y aceptar con 1"=3, "rechazar con 1 y 5"=4)
> if (p7 < 0.05) {
+     if (p7 < 0.01) {
+         p8 = opciones$"rechazar con 1 y 5";
+     } else {
+         p8 = opciones$"rechazar con 5 y aceptar con 1";
+     }
+ } else if (p7 < 0.01) {
+     p8 = opciones$"aceptar con 5 y rechazar con 1";
+ } else {
+     p8 = opciones$"aceptar con 1 y 5";
+ }
> alfa = 1 - 95/100
> p9 = c(p2 - qt(1 - alfa/2, p6) * p4, p2 + qt(1 - alfa/2, p6) * p4)
> p1; p2; p3; p4; p5; p6; p7; p8; p9

[1] 44.4
[1] 2.7875
[1] 6.489208
[1] 2.294282
[1] 1.214977
```

```
[1] 7
[1] 0.2637615
[1] 1
[1] -2.637614  8.212614
```

Competencia gasolinera

✔	1. ¿Cuál es la desviación típica para la marca G?	0.04506503
✔	2. ¿Cuál es la desviación típica para la marca H?	0.03319011
✔	3. Introduzca el valor del estadístico de referencia en comparación de variancias (el cociente entre la variancia mayor y la variancia menor).	1.84358
✔	4. Bajo la hipótesis nula de que ambas variancias poblacionales son iguales, el estadístico anterior sigue una ley F de Fisher. Introduzca los parámetros de la ley que corresponde a este caso (primero, grados de libertad del numerador; después, grados de libertad del denominador).	7 3
✔	5. ¿Cuál es el valor que utilizará, con un riesgo α=5% bilateral, para decidir si el estadístico anterior permite rechazar la hipótesis nula?	14.6244
✔	6. Aunque no lo necesita, diga cuál sería el valor crítico que limita el estadístico por la parte inferior.	0.1697845
✔	7. Para hallar el p-valor de la prueba necesitará el ordenador. Tenga en cuenta que debe calcular la probabilidad de superar el valor del estadístico que ha encontrado, y multiplicar por 2, ya que se trata de una prueba bilateral.	0.6630297
✔	8. Una de estas tres frases es la conclusión correcta. ¿Cuál es? 1. no hemos hallado evidencias de que las dos marcas difieran en dispersión del precio. 2. tenemos pruebas de que las dos variancias pueden ser idénticas. 3. hemos conseguido demostrar que el coste de G y H no es el mismo.	1
✔	9. Vamos a ver si tenemos claro lo de la distribución F de Fisher-Snedecor. Pruebe a calcular aquel valor x tal que una variable que sigue la distribución F con grados de libertad {12, 18} tenga probabilidad 0.885 de ser menor que x.	1.852411

Script en R

```
G = c(1.592, 1.529, 1.531, 1.548, 1.548, 1.521, 1.478, 1.445)
H = c(1.467, 1.488, 1.523, 1.445)
p1 = sd(G)
p2 = sd(H)
if (var(G) > var(H)) { p3 = var(G)/var(H); } else { p3 =
var(H)/var(G); }
if (var(G) > var(H)) { p4 = c(length(G)-1, length(H)-1); }
else { p4 = c(length(H)-1, length(G)-1); }
p5 = qf(1 - 0.05/2, p4[1], p4[2])
p6 = qf(0.025, p4[1], p4[2])
p7 = 2 * (1 - pf(p3, p4[1], p4[2]))
if (p7 > 0.05) {
    p8 = "no hemos hallado evidencias de que las dos marcas
difieran en dispersión del precio.";
} else if (p7 < 0.05) {
    p8 = "hemos conseguido demostrar que el coste de G y H no
es el mismo.";
} else {
    p8 = "tenemos pruebas de que las dos variancias pueden
ser idénticas.";
}
p9 = qf(0.885, 12, 18) # mayor que x -> qf(1 - 0.885, 12, 18)

p1; p2; p3; p4; p5; p6; p7; p8; p9
```

Consola de R

```
> G = c(1.592, 1.529, 1.531, 1.548, 1.548, 1.521, 1.478, 1.445)
> H = c(1.467, 1.488, 1.523, 1.445)
> p1 = sd(G)
> p2 = sd(H)
> if (var(G) > var(H)) { p3 = var(G)/var(H); } else { p3 =
var(H)/var(G); }
> if (var(G) > var(H)) { p4 = c(length(G)-1, length(H)-1); }
else { p4 = c(length(H)-1, length(G)-1); }
> p5 = qf(1 - 0.05/2, p4[1], p4[2])
> p6 = qf(0.025, p4[1], p4[2])
> p7 = 2 * (1 - pf(p3, p4[1], p4[2]))
> if (p7 > 0.05) {
+     p8 = "no hemos hallado evidencias de que las dos marcas
difieran en dispersión del precio.";
+ } else if (p7 < 0.05) {
+     p8 = "hemos conseguido demostrar que el coste de G y H
no es el mismo.";
+ } else {
+     p8 = "tenemos pruebas de que las dos variancias pueden
ser idénticas.";
+ }
> p9 = qf(0.885, 12, 18)
> p1; p2; p3; p4; p5; p6; p7; p8; p9

[1] 0.04506503
[1] 0.03319011
[1] 1.84358
[1] 7 3
[1] 14.6244
[1] 0.1697845
[1] 0.6630297
[1] "no hemos hallado evidencias de que las dos marcas
difieran en dispersión del precio."
[1] 1.852411
```

El exprimidor

	1. Consideremos los datos basados en muestras independientes. Admitiendo que los dos sistemas comparten una variabilidad común en la producción de zumo, dé una estimación de la variancia común.	207.0152
	2. Resuelva la prueba de hipótesis para comparar los promedios del peso obtenido de zumo, y diga cuál es el resultado obtenido para el estadístico de la prueba.	5.36272
	3. Planteamos una prueba de variancias: H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 \neq \sigma_2^2$ Calcule el valor del estadístico de esta prueba (procurando que caiga en la zona de la derecha), y el límite establecido con un riesgo 1% que nos permitiría rechazar la hipótesis de partida.	1.665431 5.319667
	4. Consideramos ahora la segunda parte. En este caso, tendremos en cuenta que los valores esperados que se comparan corresponden al zumo de media naranja. Calcule cuánto vale ahora el estadístico de la prueba.	7.475633
	5. Estime por intervalo de confianza del 85% la diferencia de zumo promedio que nos puede dar una naranja exprimida con uno u otro sistema, de acuerdo con el procedimiento basado en muestras independientes.	22.7384 40.2616
	6. A partir del análisis según muestras independientes, elige la respuesta más acertada: 1. El exprimidor que gira solo en un sentido es mejor 2. Creemos que los dos tipos de exprimidor son equivalentes 3. El exprimidor que gira en ambos sentidos es mejor 4. No lo sabemos, no hemos hallado nada concluyente	1

Script en R

```
Antes de ejecutar el script:
- Clicar en "Copiar dades per enganxar a un altre programa"
- Seleccionar los datos
- Pulsar Ctrl+C

G       Peso    Z       Y
1       257     119     61
1       220     100     62
1       232     111     60
1       248     138     51
1       270     123     57
1       239     102     49
1       237     118     55
1       215     108     47
1       201     81      54
1       262     99      40
1       236     129     57
1       210     94      51
2       181     68      48
2       210     73      43
2       232     88      49
2       243     97      37
2       219     80      41
2       206     76      39
2       238     87      48
2       177     51      41
2       213     71      47
2       266     94      38
2       211     80      42
2       221     79      32

datos = read.table("clipboard", header=T)
datos1 = subset(datos, G==1)
datos2 = subset(datos, G==2)
n1 = length(datos1$Z)
n2 = length(datos2$Z)
p1 = (var(datos1$Z)*(n1 - 1) + var(datos2$Z)*(n2 - 1))/(n1 + n2 - 2)
# estimación de la desviación tipo común -> sqrt(p1)
p2 = (mean(datos1$Z) - mean(datos2$Z))/(sqrt(p1) * sqrt(1/n1 + 1/n2))
# valor P -> p2 = 2 * (1 - pt(abs(p2), n1 + n2 - 2))
p3 = c(var(datos1$Z)/var(datos2$Z), qf(1 - 0.01/2, n1 - 1, n2 - 1))
Dif = datos1$Y - datos2$Y
p4 = mean(Dif)/(sd(Dif)/sqrt(length(Dif)))
# valor P -> p4 = 2 * (1 - pt(abs(p4), length(Dif) - 1))
p5 = t.test(datos1$Z, datos2$Z, var.equal=TRUE, conf.level=0.85)$conf.int
```

```
# muestras apareadas -> t.test(Dif*2, var.equal=TRUE, conf.level=0.85)

p6 = 1

# p5 = [22.7384, 40.2616] -> El exprimidor que gira solo en un
sentido es mejor

# Otras ejecuciones:

# p5 = [9.438066, 24.061934] -> El exprimidor que gira solo en un
sentido es mejor

# p5 = [-21.73712, 15.73712] -> No lo sabemos, no hemos hallado
nada concluyente

# p5 = [-16.625, -4.089] -> El exprimidor que gira en ambos
sentidos es mejor


p1; p2; p3; p4; p5; p6
```

Consola de R

```
> datos = read.table("clipboard", header=T)
> datos1 = subset(datos, G==1)
> datos2 = subset(datos, G==2)
> n1 = length(datos1$Z)
> n2 = length(datos2$Z)
> p1 = (var(datos1$Z)*(n1 - 1) + var(datos2$Z)*(n2 - 1))/(n1
+ n2 - 2)
> p2 = (mean(datos1$Z) - mean(datos2$Z))/(sqrt(p1) *
sqrt(1/n1 + 1/n2))
> p3 = c(var(datos1$Z)/var(datos2$Z), qf(1 - 0.01/2, n1 - 1,
n2 - 1))
> Dif = datos1$Y - datos2$Y
> p4 = mean(Dif)/(sd(Dif)/sqrt(length(Dif)))
> p5 = t.test(datos1$Z, datos2$Z, var.equal=TRUE,
conf.level=0.85)$conf.int
> p6 = 1
> p1; p2; p3; p4; p5; p6
[1] 207.0152
[1] 5.36272
[1] 1.665431 5.319667
[1] 7.475633
[1] 22.7384 40.2616
[1] 1
```

Técnicas de optimización

<table>
<tr><td>1. Complete la tabla de frecuencias.
<table>
<tr><td></td><td>óptimo</td><td>subóptimo</td><td>imposible</td></tr>
<tr><td>Simplex</td><td></td><td></td><td></td></tr>
<tr><td>Punto Interior</td><td></td><td></td><td></td></tr>
</table>
Introduzca los valores por columnas, separados por un blanco.</td><td>90 84 51 79
63 64</td></tr>
<tr><td>2. ¿Cuál sería el número esperado de observaciones, si en cuanto al estado final los dos procedimientos respondieran sin diferencias en su distribución? Responda para el caso concreto de acabar en estado óptimo usando la formulación Simplex.</td><td>82.3573</td></tr>
<tr><td>3. La prueba de Pearson consiste en observar si los valores esperados en caso de independencia están muy alejados de los reales. Si fueran parecidos, querría decir que no hay motivos para sospechar que un método funcione globalmente de forma distinta al otro. Pero solo que haya una celda que demuestre ser muy diferente, eso hace que el estadístico de Pearson (X^2) aumente significativamente. En este caso, ¿cuánto vale el estadístico de Pearson?</td><td>5.0325</td></tr>
<tr><td>4. Escoja la opción más adecuada:
La significación estadística de la prueba de Pearson es:
1. muy escasa (1 de cada 5, o más). No hay evidencias sobre una relación
2. posible (entre 1 de cada 5, y 1 de cada 20). Hay una evidencia muy débil
3. moderada (entre 1 de cada 20, y 1 de cada 200). La evidencia sobre la relación es sustancial
4. fuerte (1 de cada 200, o menos). Es muy raro que el resultado observado se pueda deber únicamente al azar</td><td>2</td></tr>
<tr><td>5. Mirando las proporciones logradas en los estados finales, ¿son similares Simplex y Punto Interior? ¿Hay más proporción de instancias que acaban en estado óptimo con Simplex que con el otro método? Para responder, introduzca ambas proporciones, separadas por un espacio.</td><td>0.4411765
0.3700441</td></tr>
<tr><td>6. Suponiendo la hipótesis H: la proporción de resultados de tipo óptimo es la misma para ambos métodos. ¿Cuál es la estimación de la probabilidad común que presenta este tipo de resultado?</td><td>0.4037123</td></tr>
<tr><td>7. Calcule el error tipo correspondiente a la diferencia de dos proporciones, bajo el supuesto de la hipótesis H.</td><td>0.047334</td></tr>
<tr><td>8. Para responder a la prueba que compara las proporciones de ambos métodos en el estado óptimo, diga cuánto vale el estadístico asociado a la prueba (puede obviar el signo, ya que se tratará como bilateral), y el valor de P de dicha prueba.</td><td>1.50277
0.132898</td></tr>
</table>

Script en R

```
A: Simplex
B: Punto Interior
0: Óptimo
1: Subóptimo
2: Imposible

      met   stat
1     B     1
2     B     1
3     A     2
4     A     1
5     A     0
6     A     0
7     B     2
8     A     0
9     B     0
10    A     1
     ...
431   B     1

datos = read.table("clipboard")
frecuencias = list("A0"=0, "A1"=0, "A2"=0, "B0"=0, "B1"=0, "B2"=0)
for (i in 1:length(datos$met)) {
    x = paste(datos$met[i], datos$stat[i], sep="");
    frecuencias[[x]] = frecuencias[[x]] + 1;
}
p1 = c(frecuencias$"A0", frecuencias$"B0", frecuencias$"A1",
frecuencias$"B1", frecuencias$"A2", frecuencias$"B2"); p1
```

f_{ij}	óptimo	subóptimo	imposible	
Simplex	90	51	63	204
Punto Int	84	79	64	227
	174	130	127	431

```
probA = (frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2")/length(datos$met)
```

```
probB = (frecuencias$"B0" + frecuencias$"B1" + frecuencias$"B2")/length(datos$met)

prob0 = sum(p1[1:2])/length(datos$met)

prob1 = sum(p1[3:4])/length(datos$met)

prob2 = sum(p1[5:6])/length(datos$met)

frecuenciasEsperadas = list("A0" = length(datos$met) * probA * prob0,
                            "A1" = length(datos$met) * probA * prob1,
                            "A2" = length(datos$met) * probA * prob2,
                            "B0" = length(datos$met) * probB * prob0,
                            "B1" = length(datos$met) * probB * prob1,
                            "B2" = length(datos$met) * probB * prob2)

# frecuenciasEsperadas

p2 = frecuenciasEsperadas$"A0" # estado óptimo (0), formulación Simplex (A)
```

e_{ij}	óptimo	subóptimo	imposible	
Simplex	82,...	61,...	60,...	204
Punto Int	91,...	68,...	66,...	227
	174	130	127	431

```
p3 = 0

for (i in 1:length(frecuencias)) {

    x = (as.numeric(frecuencias[i]) - as.numeric(frecuenciasEsperadas[i]))^2 / as.numeric(frecuenciasEsperadas[i]);

    # print(x);

    p3 = p3 + x;

}
```

$(f_{ij} - e_{ij})^2/e_{ij}$	óptimo	subóptimo	imposible	
Simplex	0.7092355624	1.802476352	0.1388121703	2.650524085
Punto Int	0.6373746904	1.619846589	0.124747501	2.38196878
				5.032492865

$$\sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 5.032492865 \text{ (estadístic de Pearson)}$$

```
p4 = 2   # graus de llibertat = (files-1)*(columnes-1) = (2-1)*(3-1) = 2

         # P-valor = P(X^2_2 > 5.0325) = 0.08076 [R: 1-pchisq(p3,2)]
```

```
        # Punt crític = X^2_{2,0.95} = 5.991465 [R: qchisq(0.95,2)]
        # Rebutgem H_0 si P-valor < 0.05 o X^2 > Punt crític
        # En aquest cas, la resposta és 2 perquè 5.0325 < 5.991465
        # i hi ha poca diferència entre els dos valors
        # Execució 2: X^2 = 0.07983128 -> 1. muy escasa
        # Execució 3: X^2 = 6.86417    -> 3. moderada
        # Execució 4: X^2 = 18.38441   -> 4. Fuerte
p5 = c(frecuencias$"A0"/(frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2"), frecuencias$"B0"/(frecuencias$"B0" +
frecuencias$"B1" + frecuencias$"B2"))
# estado subóptimo -> numeradores: frecuencias$"A1" y frecuencias$"B1"
# estado imposible -> numeradores: frecuencias$"A2" y frecuencias$"B2"
p6 = (frecuencias$"A0" + frecuencias$"B0")/length(datos$met)
# estado subóptimo -> numerador: frecuencias$"A1" + frecuencias$"B1"
# estado imposible -> numerador: frecuencias$"A2" + frecuencias$"B2"
```

$$P_{\text{común óptimo}} = \frac{\frac{90}{204} \cdot 204 + \frac{84}{227} \cdot 227}{204 + 227} = 0.403712297$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} \quad \text{on } P_1 \text{ és } X_1/n_1 \text{ i } P_2 \text{ és } X_2/n_2$$

```
p7 = sqrt(p6*(1-p6)/(frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2") + p6*(1-p6)/(frecuencias$"B0" + frecuencias$"B1"
+ frecuencias$"B2"))
```

$$\sqrt{P(1-P)/n_1 + P(1-P)/n_2}$$

```
z = abs((p5[1]-p5[2])/p7)
p = 2*pnorm(-z) # P(|Z| > |z|) = 2*pnorm(-z)
                #             = 2*(1-pnorm(z)) = 1-pnorm(z)+pnorm(-z)
p8 = c(z,p)
```

$$\hat{z} = \frac{(P_1 - P_2)}{\sqrt{P(1-P)/n_1 + P(1-P)/n_2}}$$

```
p1; p2; p3; p4; p5; p6; p7; p8
```

Consola de R

```
> datos = read.table("clipboard")
> frecuencias = list("A0"=0, "A1"=0, "A2"=0, "B0"=0, "B1"=0, "B2"=0)
> for (i in 1:length(datos$met)) {
+     x = paste(datos$met[i], datos$stat[i], sep="");
+     frecuencias[[x]] = frecuencias[[x]] + 1;
+ }
> p1 = c(frecuencias$"A0", frecuencias$"B0", frecuencias$"A1",
frecuencias$"B1", frecuencias$"A2", frecuencias$"B2"); p1
[1] 90 84 51 79 63 64
> probA = (frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2")/length(datos$met)
> probB = (frecuencias$"B0" + frecuencias$"B1" +
frecuencias$"B2")/length(datos$met)
> prob0 = sum(p1[1:2])/length(datos$met)
> prob1 = sum(p1[3:4])/length(datos$met)
> prob2 = sum(p1[5:6])/length(datos$met)
> frecuenciasEsperadas = list("A0" = length(datos$met) * probA * prob0,
+                             "A1" = length(datos$met) * probA * prob1,
+                             "A2" = length(datos$met) * probA * prob2,
+                             "B0" = length(datos$met) * probB * prob0,
+                             "B1" = length(datos$met) * probB * prob1,
+                             "B2" = length(datos$met) * probB * prob2)
> frecuenciasEsperadas
$A0
[1] 82.35731
$A1
[1] 61.53132
$A2
[1] 60.11137
$B0
[1] 91.64269
$B1
[1] 68.46868
$B2
[1] 66.88863
```

```
> p2 = frecuenciasEsperadas$"A0"
> p3 = 0
> for (i in 1:length(frecuencias)) {
+     x = (as.numeric(frecuencias[i]) -
as.numeric(frecuenciasEsperadas[i]))^2 /
as.numeric(frecuenciasEsperadas[i]);
+     print(x);
+     p3 = p3 + x;
+ }
[1] 0.7092356
[1] 1.802476
[1] 0.1388122
[1] 0.6373747
[1] 1.619847
[1] 0.1247475
> p3; qchisq(0.95,2)
[1] 5.032493
[1] 5.991465
> p4 = 2
> p5 = c(frecuencias$"A0"/(frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2"), frecuencias$"B0"/(frecuencias$"B0" +
frecuencias$"B1" + frecuencias$"B2"))
> p6 = (frecuencias$"A0" + frecuencias$"B0")/length(datos$met)
> p7 = sqrt(p6*(1-p6)/(frecuencias$"A0" + frecuencias$"A1" +
frecuencias$"A2") + p6*(1-p6)/(frecuencias$"B0" + frecuencias$"B1"
+ frecuencias$"B2"))
> z = abs((p5[1]-p5[2])/p7)
> p = 2*pnorm(-z)
> p8 = c(z,p)
> p1; p2; p3; p4; p5; p6; p7; p8
[1] 90 84 51 79 63 64
[1] 82.35731
[1] 5.032493
[1] 2
[1] 0.4411765 0.3700441
[1] 0.4037123
[1] 0.04733415
[1] 1.5027718 0.1328979
```