

Estadística

Práctica 11

Informe de estadística inferencial

Contenido

1. Introducción	4
2. Preparar los datos	5
3. Obtener muestras	6
3.1 Obtener muestras de tamaño grande	6
3.2 Obtener muestras de tamaño pequeño	6
3.3 Guardar las muestras	7
4. Calcular intervalos de confianza	8
4.1 Intervalos de confianza para la media	8
4.1.1 Muestras grandes	8
4.1.2 Muestras pequeñas	8
4.1.3 Diagrama con los intervalos	9
4.2 Intervalos de confianza para proporciones	10
4.2.1 Proporción de alumnos con calificación de Aprobado	10
4.2.2 Proporción de alumnos con calificación de Notable	13
4.2.3 Proporción de alumnos con calificación de Sobresaliente	16
5. Realizar contrastes de hipótesis	19
5.1 Contraste de hipótesis sobre la diferencia de medias	19
5.1.1 Muestras grandes	19
5.1.2 Muestras pequeñas	20
5.2 Contraste de hipótesis para la diferencia de proporciones	20
5.2.1 Diferencia de proporciones de alumnos con calificación de Aprobado	20
5.2.2 Diferencia de proporciones de alumnos con calificación de Notable	22
5.2.3 Diferencia de proporciones de alumnos con calificación de Sobresaliente	25
6. Escribir el informe	27
6.1 Título	27
6.2 Introducción	27
6.3 Metodología	28
6.4 Resultados	29
6.5 Análisis de resultados	32
6.6 Conclusiones	33

6.7 Referencias 34

1. Introducción

Con esta práctica se trata de aprender a realizar un informe de estadística inferencial sobre dos poblaciones a partir de una muestra de cada una de ellas, para comprobar si hay diferencias significativas entre ambas poblaciones.

Se utilizarán los mismos datos que en la práctica de elaboración de un informe estadístico descriptivo, es decir los disponibles en el archivo "[encuesta.csv](#)".

Para elaborar un informe de estadística inferencial se seguirán los siguientes pasos:

- Plantear el objetivo de la investigación mediante la formulación de hipótesis sobre las poblaciones a estudiar.
- Obtener una muestra de cada población.
- Realizar los cálculos estadísticos para obtener intervalos de confianza y para calcular los p-valores que permitan aceptar o rechazar las hipótesis
- Generar diagramas que ayuden a entender el informe.
- Escribir el informe. Se trata de recoger en un documento los resultados de la investigación.

2. Preparar los datos

Se van a utilizar los mismos datos que en la práctica 3, en la que, a partir del archivo encuesta.csv se guardaron los datos de todas las notas de los alumnos en el vector `nota`.

```
> encuesta = read.csv2("encuesta.csv")
> (nota=encuesta$NOTA)
 [1]  8.500  7.100  8.635  8.624  8.200
 [6]  8.700  7.210  7.630  8.400  8.300
[11]  8.200  9.100  9.789 10.107  8.015
[16]  7.310  7.500  8.710  8.336  9.210
[21]  7.800 10.300  7.990  6.900  7.800
[26] 10.000  8.590  7.000  8.050 10.799
[31]  7.994  8.550  7.340  6.750  9.560
[36]  7.417  6.936  7.210  7.680 10.277
[41]  7.860 10.260  7.270  5.800  7.300
[46]  7.140  8.600  7.500  8.000  7.540
[51]  7.292  7.830  6.750  9.806  6.800
[56]  6.445  6.650  7.804 10.270  7.600
[61]  7.870  7.000  7.085  7.480  8.070
[66]  5.820  6.500  9.900  7.500  6.500
[71]  9.456  8.000  7.800  7.654)
```

Y también tenemos separadas las notas de los alumnos del turno de mañana en el vector `notaM`, y las del turno de tarde en el vector `notaT`.

```
> encuesta.mañana=encuesta[((encuesta$GRUPO=="A1") | (encuesta$GRUPO=="A2")), ]
> encuesta.tarde=encuesta[((encuesta$GRUPO=="B1") | (encuesta$GRUPO=="B2")), ]

> (notaM=encuesta.mañana$NOTA)
 [1]  8.500  7.100  8.635  8.624  8.200
 [6]  8.700  7.210  7.630  8.400  8.300
[11]  8.200  9.100  9.789 10.107  8.015
[16]  7.310  7.500  8.710  8.336  9.210
[21]  7.800 10.300  7.990  6.900  7.800
[26] 10.000  8.590  7.000  8.050 10.799
[31]  7.994  8.550  7.340  6.750  9.560
[36]  7.417  6.936  7.210  7.680 10.277
[41]  7.860 10.260

> (notaT=encuesta.tarde$NOTA)
 [1]  7.270  5.800  7.300  7.140  8.600
 [6]  7.500  8.000  7.540  7.292  7.830
[11]  6.750  9.806  6.800  6.445  6.650
[16]  7.804 10.270  7.600  7.870  7.000
[21]  7.085  7.480  8.070  5.820  6.500
[26]  9.900  7.500  6.500  9.456  8.000
[31]  7.800  7.654)
```

El número de notas de cada turno se puede calcular con la función `length`.

```
> length(notaM)
[1] 42
> length(notaT)
[1] 32
```

3. Obtener muestras

3.1 Obtener muestras de tamaño grande

En los vectores `notaM` y `notaT` ya tenemos dos muestras de los alumnos de cada turno, son los datos de los que respondieron a la encuesta. La población de cada turno estaría formada por todos los alumnos matriculados en cada turno, cuyo valor se desconoce.

Para que quede más claro que son dos muestras, vamos a utilizar dos variables con la palabra nombre en su nombre.

```
> (notaM.muestra.grande=notaM)
 [1]  8.500  7.100  8.635  8.624  8.200  8.700  7.210  7.630  8.400  8.300  8.200
[12]  9.100  9.789 10.107  8.015  7.310  7.500  8.710  8.336  9.210  7.800 10.300
[23]  7.990  6.900  7.800 10.000  8.590  7.000  8.050 10.799  7.994  8.550  7.340
[34]  6.750  9.560  7.417  6.936  7.210  7.680 10.277  7.860 10.260

> (notaT.muestra.grande=notaT)
 [1]  7.270  5.800  7.300  7.140  8.600  7.500  8.000  7.540  7.292  7.830  6.750
[12]  9.806  6.800  6.445  6.650  7.804 10.270  7.600  7.870  7.000  7.085  7.480
[23]  8.070  5.820  6.500  9.900  7.500  6.500  9.456  8.000  7.800  7.654
```

3.2 Obtener muestras de tamaño pequeño

Para poder comparar resultados utilizando diferentes tamaños de muestras, vamos a seleccionar aleatoriamente, usando la función `sample()` de R, una muestra pequeña de notas en cada turno, por ejemplo, dos muestras de tamaño 20.

```
> (notaM.muestra.pequeña=sample(notaM, 20))
 [1] 10.000  7.630  9.100  7.417  6.750  8.050  7.500  8.700  9.210  6.936 10.260
[12]  8.200  7.860  9.789  7.000  9.560 10.277  8.015  8.500  8.300

> (notaT.muestra.pequeña=sample(notaT, 20))
 [1] 8.000 9.900 6.650 8.070 7.804 7.500 9.456 6.445 7.140 7.870 7.300 7.270
7.292
[14] 7.500 7.800 6.800 8.600 9.806 6.500 5.800
```

IMPORTANTE: Las muestras pequeñas serán diferentes para cada persona que realice esta práctica, pues la función `sample` habrá seleccionado unas notas diferentes. Por tanto, los resultados para muestras pequeñas en los siguientes apartados de esta práctica no coincidirán.

3.3 Guardar las muestras

Las muestras deben guardarse en archivos csv. Las muestras grandes ya están en el archivo "encuesta.scv".

Las muestras pequeñas se van a guardar en fichero diferentes usando los siguientes comandos:

```
> write.csv2(notaM.muestra.pequeña,"20notasmañana.csv")
> write.csv2(notaT.muestra.pequeña,"20notastarde.csv")
```

De esta forma, si en algún momento borramos las variables del proyecto, podemos recuperar los datos con los siguientes comandos:

```
> x=read.csv2("20notasmañana.csv")
> (notaM.muestra.pequeña=x$x)
 [1] 10.000  7.630  9.100  7.417  6.750  8.050  7.500  8.700  9.210  6.936 10.260
[12]  8.200  7.860  9.789  7.000  9.560 10.277  8.015  8.500  8.300

> x=read.csv2("20notastarde.csv")
> (notaT.muestra.pequeña=x$x)
 [1] 8.000 9.900 6.650 8.070 7.804 7.500 9.456 6.445 7.140 7.870 7.300 7.270
7.292
[14] 7.500 7.800 6.800 8.600 9.806 6.500 5.800
```

4. Calcular intervalos de confianza

Se calcularán intervalos de confianza para la media y para proporciones con un nivel de confianza del 95% ($\alpha=0.05$).

4.1 Intervalos de confianza para la media

4.1.1 Muestras grandes

Se usarán las muestras `notaM.muestra.grande` y `notaT.muestra.grande`, con un tamaño de 42 y 32 respectivamente.

Al ser muestras grandes se utilizará como estimador una variable con distribución Normal.

Lo más sencillo es calcular el intervalo usando la función `z.test` del paquete `BSDA`, que hay que instalar previamente.

```
> install.packages("BSDA")
> library(BSDA)

> z.test(notaM.muestra.grande, sigma.x=sd(notaM.muestra.grande), conf.level=0.95)

95 percent confidence interval:
 8.024751 8.672345
```

El intervalo de confianza para la media poblacional de la nota de los alumnos del turno de mañana, a partir de una muestra grande, es [8.02, 8.67].

Se haría de igual forma con los del turno de tarde.

```
> z.test(notaT.muestra.grande, sigma.x=sd(notaT.muestra.grande), conf.level=0.95)

95 percent confidence interval:
 7.22098 7.96852
```

El intervalo de confianza para la media poblacional de la nota de los alumnos del turno de tarde, a partir de una muestra grande, es [7.22, 7.97].

4.1.2 Muestras pequeñas

Cuando las muestras son pequeñas (<30) el estimador que se utiliza tiene distribución T-Student, por lo que se debe usar la función `t.test` de R.

```
> t.test(notaM.muestra.pequeña, conf.level=0.95)

95 percent confidence interval:
 7.928723 8.976677
```

El intervalo de confianza para la media poblacional de la nota de los alumnos del turno de mañana, a partir de una muestra pequeña, es [7.93, 8.98].

Se haría de igual forma con los del turno de tarde.

```
> t.test(notaT.muestra.pequeña, conf.level=0.95)
```

```
95 percent confidence interval:  
7.161617 8.188683
```

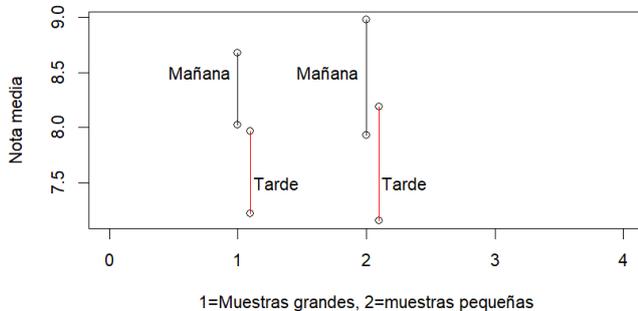
El intervalo de confianza para la media poblacional de la nota de los alumnos del turno de tarde, a partir de una muestra pequeña, es [7.16, 8.19].

4.1.3 Diagrama con los intervalos

```
test.M.grande=z.test(notaM.muestra.grande, sigma.x=sd(notaM.muestra.grande),  
conf.level=0.95)  
Li.M.grande=test.M.grande$conf.int[1]  
Ls.M.grande=test.M.grande$conf.int[2]  
test.T.grande=z.test(notaT.muestra.grande, sigma.x=sd(notaT.muestra.grande),  
conf.level=0.95)  
Li.T.grande=test.T.grande$conf.int[1]  
Ls.T.grande=test.T.grande$conf.int[2]
```

```
test.M.pequeña=t.test(notaM.muestra.pequeña, conf.level=0.95)  
Li.M.pequeña=test.M.pequeña$conf.int[1]  
Ls.M.pequeña=test.M.pequeña$conf.int[2]  
test.T.pequeña=t.test(notaT.muestra.pequeña, conf.level=0.95)  
Li.T.pequeña=test.T.pequeña$conf.int[1]  
Ls.T.pequeña=test.T.pequeña$conf.int[2]
```

```
plot(c(1,1,1.1,1.1,2,2,2.1,2.1),c(Li.M.grande, Ls.M.grande, Li.T.grande,  
Ls.T.grande, Li.M.pequeña, Ls.M.pequeña, Li.T.pequeña, Ls.T.pequeña), xlim =  
c(0,4), xlab = "1=Muestras grandes, 2=muestras pequeñas", ylab = "Nota")  
lines(c(1,1),c(Li.M.grande,Ls.M.grande))  
lines(c(1.1,1.1),c(Li.T.grande,Ls.T.grande), col="red")  
lines(c(2,2),c(Li.M.pequeña,Ls.M.pequeña))  
lines(c(2.1,2.1),c(Li.T.pequeña,Ls.T.pequeña), col="red")  
text(0.7, 8.5, label = "Mañana")  
text(1.3, 7.5, label = "Tarde")  
text(1.7, 8.5, label = "Mañana")  
text(2.3, 7.5, label = "Tarde")
```



Este diagrama puede servir para comprobar visualmente si se solapan los intervalos de los dos grupos, ya que, si no hay solapamiento, sí habrá diferencia entre las medias de ambos grupos, como ocurre en este caso para las muestras grandes; mientras que, si hay solapamiento, y éste es amplio, es muy posible que no haya diferencia significativa entre las medias de los dos grupos, pero esta apreciación visual, deberá comprobarse formalmente mediante un contraste de hipótesis. En este caso, para muestras pequeñas se aprecia un pequeño solapamiento, pero no podremos asegurar si es suficiente para afirmar que no hay diferencia, hasta hacer el contraste de hipótesis.

4.2 Intervalos de confianza para proporciones

Vamos a considerar los grupos que se crearon en la práctica 3:

- Aprobado: Notas en el rango [5, 7)
- Notable: Notas en el rango [7,9)
- Sobresaliente: Notas en el rango [9,11]

Para los cálculos necesitaremos el tamaño de las muestras.

```
> (n.notaM.muestra.grande=length(notaM.muestra.grande))
[1] 42
> (n.notaT.muestra.grande=length(notaT.muestra.grande))
[1] 32
> (n.notaM.muestra.pequeña=length(notaM.muestra.pequeña))
[1] 20
> (n.notaT.muestra.pequeña=length(notaT.muestra.pequeña))
[1] 20
```

4.2.1 Proporción de alumnos con calificación de Aprobado

4.2.1.1 Muestras grandes

Primero calculamos el número de alumnos con calificación de Aprobado en cada turno.

```
> (aprobadoM.muestra.grande =
notaM.muestra.grande[(notaM.muestra.grande>=5) & (notaM.muestra.grande<7)])
[1] 6.900 6.750 6.936
```

```
> (aprobadoT.muestra.grande =  
notaT.muestra.grande[(notaT.muestra.grande>=5) & (notaT.muestra.grande<7)])  
[1] 5.800 6.750 6.800 6.445 6.650 5.820 6.500 6.500  
> (ng.aprobadoM.muestra.grande=length(aprobadoM.muestra.grande))  
[1] 3  
> (ng.aprobadoT.muestra.grande=length(aprobadoT.muestra.grande))  
[1] 8
```

Entonces, las proporciones muestrales son:

```
> (p.aprobadoM.muestra.grande=ng.aprobadoM.muestra.grande/n.notaM.muestra.grande)  
[1] 0.07142857  
> (p.aprobadoT.muestra.grande=ng.aprobadoT.muestra.grande/n.notaT.muestra.grande)  
[1] 0.25
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> prop.test(ng.aprobadoM.muestra.grande, n.notaM.muestra.grande, conf.level=0.95,  
correct=FALSE)  
95 percent confidence interval:  
 0.02459041 0.19009427
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Aprobado en el turno de mañana, a partir de una muestra grande, es [0.0246, 0.1901].

Se haría de igual forma con los del turno de tarde.

```
> prop.test(ng.aprobadoT.muestra.grande, n.notaT.muestra.grande, conf.level=0.95,  
correct=FALSE)  
95 percent confidence interval:  
 0.1325240 0.4210656
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Aprobado en el turno de tarde, a partir de una muestra grande, es [0.1325, 0.4211].

4.2.1.2 Muestras pequeñas

Primero calculamos el número de alumnos con calificación de Aprobado en cada turno.

```
> (aprobadoM.muestra.pequeña =  
notaM.muestra.pequeña[(notaM.muestra.pequeña>=5) & (notaM.muestra.pequeña<7)])  
[1] 6.750 6.936  
> (aprobadoT.muestra.pequeña =  
notaT.muestra.pequeña[(notaT.muestra.pequeña>=5) & (notaT.muestra.pequeña<7)])  
[1] 6.650 6.445 6.800 6.500 5.800  
> (ng.aprobadoM.muestra.pequeña=length(aprobadoM.muestra.pequeña))  
[1] 2  
> (ng.aprobadoT.muestra.pequeña=length(aprobadoT.muestra.pequeña))  
[1] 5
```

Entonces, las proporciones muestrales son:

```
>
(p.aprobadoM.muestra.pequeña=ng.aprobadoM.muestra.pequeña/n.notaM.muestra.pequeña
)
[1] 0.1
>
(p.aprobadoT.muestra.pequeña=ng.aprobadoT.muestra.pequeña/n.notaT.muestra.pequeña
)
[1] 0.25
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> prop.test(ng.aprobadoM.muestra.pequeña, n.notaM.muestra.pequeña,
conf.level=0.95, correct=FALSE)
95 percent confidence interval:
 0.02786648 0.30103365
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Aprobado en el turno de mañana, a partir de una muestra pequeña, es [0.0279, 0.3010].

Se haría de igual forma con los del turno de tarde.

```
> prop.test(ng.aprobadoT.muestra.pequeña, n.notaT.muestra.pequeña,
conf.level=0.95, correct=FALSE)
95 percent confidence interval:
 0.1118617 0.4687009
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Aprobado en el turno de tarde, a partir de una muestra pequeña, es [0.1119, 0.4687].

4.2.1.3 Diagrama

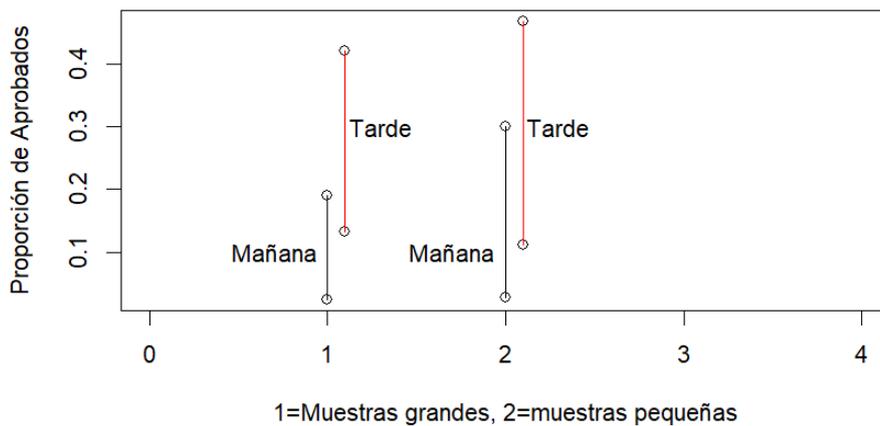
```
test.ApM.grande=prop.test(ng.aprobadoM.muestra.grande, n.notaM.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.ApM.grande=test.ApM.grande$conf.int[1]
Ls.ApM.grande=test.ApM.grande$conf.int[2]
test.ApT.grande=prop.test(ng.aprobadoT.muestra.grande, n.notaT.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.ApT.grande=test.ApT.grande$conf.int[1]
Ls.ApT.grande=test.ApT.grande$conf.int[2]

test.ApM.pequeña=prop.test(ng.aprobadoM.muestra.pequeña, n.notaM.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.ApM.pequeña=test.ApM.pequeña$conf.int[1]
Ls.ApM.pequeña=test.ApM.pequeña$conf.int[2]
test.ApT.pequeña=prop.test(ng.aprobadoT.muestra.pequeña, n.notaT.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.ApT.pequeña=test.ApT.pequeña$conf.int[1]
Ls.ApT.pequeña=test.ApT.pequeña$conf.int[2]

plot(c(1,1,1.1,1.1,2,2,2.1,2.1),c(Li.ApM.grande, Ls.ApM.grande, Li.ApT.grande,
Ls.ApT.grande, Li.ApM.pequeña, Ls.ApM.pequeña, Li.ApT.pequeña, Ls.ApT.pequeña),
xlim = c(0,4), xlab = "1=Muestras grandes, 2=muestras pequeñas", ylab =
"Proporción de Aprobados")
lines(c(1,1),c(Li.ApM.grande,Ls.ApM.grande))
lines(c(1.1,1.1),c(Li.ApT.grande,Ls.ApT.grande),col="red")
```



```
lines(c(2,2),c(Li.ApM.pequeña,Ls.ApM.pequeña))
lines(c(2.1,2.1),c(Li.ApT.pequeña,Ls.ApT.pequeña),col="red")
text(0.7, 0.1, label = "Mañana")
text(1.3, 0.3, label = "Tarde")
text(1.7, 0.1, label = "Mañana")
text(2.3, 0.3, label = "Tarde")
```



4.2.2 Proporción de alumnos con calificación de Notable

4.2.2.1 Muestras grandes

Primero calculamos el número de alumnos con calificación de Notable en cada turno.

```
> (notableM.muestra.grande =
notaM.muestra.grande[(notaM.muestra.grande>=7) & (notaM.muestra.grande<9)])
 [1] 8.500 7.100 8.635 8.624 8.200 8.700 7.210 7.630 8.400 8.300 8.200 8.015
 7.310
 [14] 7.500 8.710 8.336 7.800 7.990 7.800 8.590 7.000 8.050 7.994 8.550 7.340
 7.417
 [27] 7.210 7.680 7.860
> (notableT.muestra.grande =
notaT.muestra.grande[(notaT.muestra.grande>=7) & (notaT.muestra.grande<9)])
 [1] 7.270 7.300 7.140 8.600 7.500 8.000 7.540 7.292 7.830 7.804 7.600 7.870
 7.000
 [14] 7.085 7.480 8.070 7.500 8.000 7.800 7.654
> (ng.notableM.muestra.grande=length(notableM.muestra.grande))
 [1] 29
> (ng.notableT.muestra.grande=length(notableT.muestra.grande))
 [1] 20
```

Entonces, las proporciones muestrales son:

```
> (p.notableM.muestra.grande=ng.notableM.muestra.grande/n.notaM.muestra.grande)
 [1] 0.6904762
> (p.notableT.muestra.grande=ng.notableT.muestra.grande/n.notaT.muestra.grande)
```

```
[1] 0.625
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> prop.test(ng.notableM.muestra.grande, n.notaM.muestra.grande, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
 0.5397401 0.8092890
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Notable en el turno de mañana, a partir de una muestra grande, es [0.5397, 0.8093].

Se haría de igual forma con los del turno de tarde.

```
> prop.test(ng.notableT.muestra.grande, n.notaT.muestra.grande, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
 0.4525441 0.7706611
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Notable en el turno de tarde, a partir de una muestra grande, es [0.4525, 0.7707].

4.2.2.2 Muestras pequeñas

Primero calculamos el número de alumnos con calificación de Notable en cada turno.

```
> (notableM.muestra.pequeña =
notaM.muestra.pequeña[(notaM.muestra.pequeña>=7) & (notaM.muestra.pequeña<9)])
 [1] 7.630 7.417 8.050 7.500 8.700 8.200 7.860 7.000 8.015 8.500 8.300
> (notableT.muestra.pequeña =
notaT.muestra.pequeña[(notaT.muestra.pequeña>=7) & (notaT.muestra.pequeña<9)])
 [1] 8.000 8.070 7.804 7.500 7.140 7.870 7.300 7.270 7.292 7.500 7.800 8.600
> (ng.notableM.muestra.pequeña=length(notableM.muestra.pequeña))
 [1] 11
> (ng.notableT.muestra.pequeña=length(notableT.muestra.pequeña))
 [1] 12
```

Entonces, las proporciones muestrales son:

```
>
(p.notableM.muestra.pequeña=ng.notableM.muestra.pequeña/n.notaM.muestra.pequeña)
 [1] 0.55
>
(p.notableT.muestra.pequeña=ng.notableT.muestra.pequeña/n.notaT.muestra.pequeña)
 [1] 0.6
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> (prop.test(ng.notableM.muestra.pequeña, n.notaM.muestra.pequeña,
conf.level=0.95, correct=FALSE))
95 percent confidence interval:
 0.3420853 0.7418021
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Notable en el turno de mañana, a partir de una muestra pequeña, es [0.3421, 0.7418].

Se haría de igual forma con los del turno de tarde.

```
> (prop.test(ng.notableT.muestra.pequeña, n.notaT.muestra.pequeña,
conf.level=0.95, correct=FALSE))
95 percent confidence interval:
 0.3865815 0.7811935
```

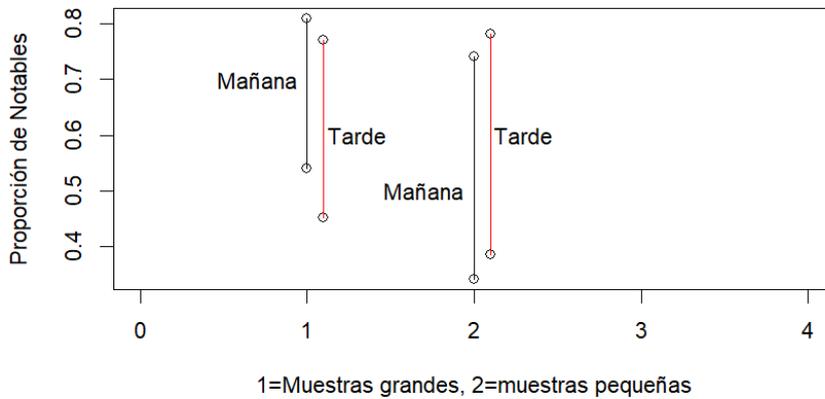
El intervalo de confianza para la proporción poblacional de alumnos con calificación de Notable en el turno de tarde, a partir de una muestra pequeña, es [0.3866, 0.7812].

4.2.2.3 Diagrama

```
test.NotM.grande=prop.test(ng.notableM.muestra.grande, n.notaM.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.NotM.grande=test.NotM.grande$conf.int[1]
Ls.NotM.grande=test.NotM.grande$conf.int[2]
test.NotT.grande=prop.test(ng.notableT.muestra.grande, n.notaT.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.NotT.grande=test.NotT.grande$conf.int[1]
Ls.NotT.grande=test.NotT.grande$conf.int[2]

test.NotM.pequeña=prop.test(ng.notableM.muestra.pequeña, n.notaM.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.NotM.pequeña=test.NotM.pequeña$conf.int[1]
Ls.NotM.pequeña=test.NotM.pequeña$conf.int[2]
test.NotT.pequeña=prop.test(ng.notableT.muestra.pequeña, n.notaT.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.NotT.pequeña=test.NotT.pequeña$conf.int[1]
Ls.NotT.pequeña=test.NotT.pequeña$conf.int[2]

plot(c(1,1,1.1,1.1,2,2,2.1,2.1),c(Li.NotM.grande, Ls.NotM.grande, Li.NotT.grande,
Ls.NotT.grande, Li.NotM.pequeña, Ls.NotM.pequeña, Li.NotT.pequeña,
Ls.NotT.pequeña), xlim = c(0,4), xlab = "1=Muestras grandes, 2=muestras
pequeñas", ylab = "Proporción de Notables")
lines(c(1,1),c(Li.NotM.grande,Ls.NotM.grande))
lines(c(1.1,1.1),c(Li.NotT.grande,Ls.NotT.grande),col="red")
lines(c(2,2),c(Li.NotM.pequeña,Ls.NotM.pequeña))
lines(c(2.1,2.1),c(Li.NotT.pequeña,Ls.NotT.pequeña),col="red")
text(0.7, 0.7, label = "Mañana")
text(1.3, 0.6, label = "Tarde")
text(1.7, 0.5, label = "Mañana")
text(2.3, 0.6, label = "Tarde")
```



4.2.3 Proporción de alumnos con calificación de Sobresaliente

4.2.3.1 Muestras grandes

Primero calculamos el número de alumnos con calificación de Sobresaliente en cada turno.

```
> (sobM.muestra.grande = notaM.muestra.grande[notaM.muestra.grande>=9])
[1] 9.100 9.789 10.107 9.210 10.300 10.000 10.799 9.560 10.277 10.260
> (sobT.muestra.grande = notaT.muestra.grande[notaT.muestra.grande>=9])
[1] 9.806 10.270 9.900 9.456
> (ng.sobM.muestra.grande=length(sobM.muestra.grande))
[1] 10
> (ng.sobT.muestra.grande=length(sobT.muestra.grande))
[1] 4
```

Entonces, las proporciones muestrales son:

```
> (p.sobM.muestra.grande=ng.sobM.muestra.grande/n.notaM.muestra.grande)
[1] 0.2380952
> (p.sobT.muestra.grande=ng.sobT.muestra.grande/n.notaT.muestra.grande)
[1] 0.125
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> prop.test(ng.sobM.muestra.grande, n.notaM.muestra.grande, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
0.1348096 0.3852754
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Sobresaliente en el turno de mañana, a partir de una muestra grande, es [0.1348, 0.3853].

Se haría de igual forma con los del turno de tarde.

```
> prop.test(ng.sobT.muestra.grande, n.notaT.muestra.grande, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
 0.04970134 0.28068305
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Sobresaliente en el turno de tarde, a partir de una muestra grande, es [0.0497, 0.2807].

4.2.3.2 Muestras pequeñas

Primero calculamos el número de alumnos con calificación de Sobresaliente en cada turno.

```
> (sobM.muestra.pequeña = notaM.muestra.pequeña[notaM.muestra.pequeña>=9])
[1] 10.000  9.100  9.210 10.260  9.789  9.560 10.277
> (sobT.muestra.pequeña = notaT.muestra.pequeña[notaT.muestra.pequeña>=9])
[1] 9.900 9.456 9.806
> (ng.sobM.muestra.pequeña=length(sobM.muestra.pequeña))
[1] 7
> (ng.sobT.muestra.pequeña=length(sobT.muestra.pequeña))
[1] 3
```

Entonces, las proporciones muestrales son:

```
> (p.sobM.muestra.pequeña=ng.sobM.muestra.pequeña/n.notaM.muestra.pequeña)
[1] 0.35
> (p.sobT.muestra.pequeña=ng.sobT.muestra.pequeña/n.notaT.muestra.pequeña)
[1] 0.15
```

Y el intervalo de confianza se calcula con la función `prop.test`.

```
> prop.test(ng.sobM.muestra.pequeña, n.notaM.muestra.pequeña, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
 0.1811918 0.5671457
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Sobresaliente en el turno de mañana, a partir de una muestra pequeña, es [0.1812, 0.5671].

Se haría de igual forma con los del turno de tarde.

```
> prop.test(ng.sobT.muestra.pequeña, n.notaT.muestra.pequeña, conf.level=0.95,
correct=FALSE)
95 percent confidence interval:
 0.05236875 0.36041886
```

El intervalo de confianza para la proporción poblacional de alumnos con calificación de Sobresaliente en el turno de tarde, a partir de una muestra pequeña, es [0.0524, 0.3604].

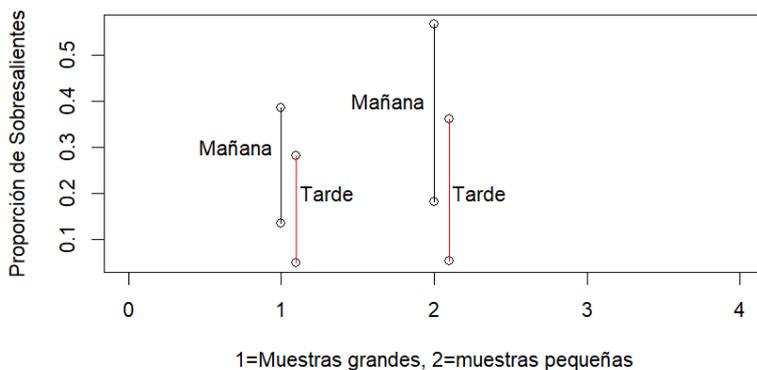
4.2.3.3 Diagrama

```
test.sobM.grande=prop.test(ng.sobM.muestra.grande, n.notaM.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.sobM.grande=test.sobM.grande$conf.int[1]
Ls.sobM.grande=test.sobM.grande$conf.int[2]
```

```
test.sobT.grande=prop.test(ng.sobT.muestra.grande, n.notaT.muestra.grande,
conf.level=0.95, correct=FALSE)
Li.sobT.grande=test.sobT.grande$conf.int[1]
Ls.sobT.grande=test.sobT.grande$conf.int[2]
```

```
test.sobM.pequeña=prop.test(ng.sobM.muestra.pequeña, n.notaM.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.sobM.pequeña=test.sobM.pequeña$conf.int[1]
Ls.sobM.pequeña=test.sobM.pequeña$conf.int[2]
test.sobT.pequeña=prop.test(ng.sobT.muestra.pequeña, n.notaT.muestra.pequeña,
conf.level=0.95, correct=FALSE)
Li.sobT.pequeña=test.sobT.pequeña$conf.int[1]
Ls.sobT.pequeña=test.sobT.pequeña$conf.int[2]
```

```
plot(c(1,1,1.1,1.1,2,2,2.1,2.1),c(Li.sobM.grande, Ls.sobM.grande, Li.sobT.grande,
Ls.sobT.grande, Li.sobM.pequeña, Ls.sobM.pequeña, Li.sobT.pequeña,
Ls.sobT.pequeña), xlim = c(0,4), xlab = "1=Muestras grandes, 2=muestras
pequeñas", ylab = "Proporción de Sobresalientes")
lines(c(1,1),c(Li.sobM.grande,Ls.sobM.grande))
lines(c(1.1,1.1),c(Li.sobT.grande,Ls.sobT.grande),col="red")
lines(c(2,2),c(Li.sobM.pequeña,Ls.sobM.pequeña))
lines(c(2.1,2.1),c(Li.sobT.pequeña,Ls.sobT.pequeña),col="red")
text(0.7, 0.3, label = "Mañana")
text(1.3, 0.2, label = "Tarde")
text(1.7, 0.4, label = "Mañana")
text(2.3, 0.2, label = "Tarde")
```



5. Realizar contrastes de hipótesis

5.1 Contraste de hipótesis sobre la diferencia de medias

5.1.1 Muestras grandes

Las medias muestrales son:

```
> (mean(notaM.muestra.grande))  
[1] 8.348548  
> (mean(notaT.muestra.grande))  
[1] 7.59475
```

Aunque la nota media muestra del turno de mañana es mayor que la del turno de tarde, no sabemos si la media de poblacional también será mayor, para ello hay que formular un contraste de hipótesis.

Podemos plantear primero un contraste para determinar si hay diferencias entre las medias.

$$H_0: \mu_{\text{Mañana}} = \mu_{\text{Tarde}} \text{ VS } H_A: \mu_{\text{Mañana}} \neq \mu_{\text{Tarde}}$$

```
> z.test(notaM.muestra.grande, notaT.muestra.grande, alternative="two.sided",  
sigma.x=sd(notaM.muestra.grande), sigma.y=sd(notaT.muestra.grande),  
conf.level=0.95)
```

Two-sample z-Test

```
data: notaM.muestra.grande and notaT.muestra.grande  
z = 2.9876, p-value = 0.002812  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 0.2592794 1.2483159  
sample estimates:  
mean of x mean of y  
 8.348548  7.594750
```

Como $p\text{-valor} < 0.05$, se puede aceptar la hipótesis alternativa de que hay diferencia significativa entre las medias, y entonces podemos probar la hipótesis alternativa de que la media poblacional del turno de mañana es mayor, como ocurre con las medias muestrales:

$$H_0: \mu_{\text{Mañana}} = \mu_{\text{Tarde}} \text{ VS } H_A: \mu_{\text{Mañana}} > \mu_{\text{Tarde}}$$

```
> z.test(notaM.muestra.grande, notaT.muestra.grande, alternative="greater",  
sigma.x=sd(notaM.muestra.grande), sigma.y=sd(notaT.muestra.grande),  
conf.level=0.95)
```

Two-sample z-Test

```
data: notaM.muestra.grande and notaT.muestra.grande  
z = 2.9876, p-value = 0.001406
```

```
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3387848      NA
sample estimates:
mean of x mean of y
 8.348548  7.594750
```

Como p -valor < 0.05 , se puede aceptar la hipótesis alternativa de que la media poblacional del turno de mañana es mayor que la del turno de tarde.

5.1.2 Muestras pequeñas

Las medias muestrales son:

```
> (mean(notaM.muestra.pequeña))
[1] 8.4527
> (mean(notaT.muestra.pequeña))
[1] 7.67515
```

Como la media muestral del turno de mañana es bastante superior a la del turno de tarde, planteamos directamente la hipótesis alternativa de que la media poblacional también es mayor:

$$H_0: \mu_{Mañana} = \mu_{Tarde} \text{ vs } H_A: \mu_{Mañana} > \mu_{Tarde}$$

```
> t.test(notaM.muestra.pequeña, notaT.muestra.pequeña, alternative="greater",
sigma.x=sd(notaM.muestra.pequeña), sigma.y=sd(notaT.muestra.pequeña),
conf.level=0.95)
```

Welch Two Sample t-test

```
data: notaM.muestra.pequeña and notaT.muestra.pequeña
t = 2.2182, df = 37.985, p-value = 0.0163
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1865667      Inf
sample estimates:
mean of x mean of y
 8.45270  7.67515
```

Como p -valor < 0.05 , se puede aceptar la hipótesis alternativa también en el caso de muestras pequeñas.

5.2 Contraste de hipótesis para la diferencia de proporciones

5.2.1 Diferencia de proporciones de alumnos con calificación de Aprobado

5.2.1.1 Muestras grandes

Las proporciones para las muestras grandes son las siguientes:

```
> (p.aprobadoM.muestra.grande=ng.aprobadoM.muestra.grande/n.notaM.muestra.grande)
[1] 0.07142857
> (p.aprobadoT.muestra.grande=ng.aprobadoT.muestra.grande/n.notaT.muestra.grande)
[1] 0.25
```

Como la proporción muestral del turno de mañana es bastante más pequeña que la del turno de tarde, podemos plantear directamente la hipótesis alternativa de que la proporción poblacional también es menor:

$$H_0: p_{ApMañana} = p_{ApTarde} \text{ vs } H_A: p_{ApMañana} < p_{ApTarde}$$

```
> prop.test(c(ng.aprobadoM.muestra.grande,ng.aprobadoT.muestra.grande),
c(n.notaM.muestra.grande,n.notaT.muestra.grande), alternative="less",
conf.level=0.95, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.aprobadoM.muestra.grande, ng.aprobadoT.muestra.grande) out of
c(n.notaM.muestra.grande, n.notaT.muestra.grande)
X-squared = 4.5764, df = 1, p-value = 0.01621
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.03670742
sample estimates:
  prop 1    prop 2
0.07142857 0.25000000
```

Como $p\text{-valor} < 0.05$, se puede aceptar la hipótesis alternativa.

5.2.1.2 Muestras pequeñas

Las proporciones para las muestras pequeñas son las siguientes:

```
>
(p.aprobadoM.muestra.pequeña=ng.aprobadoM.muestra.pequeña/n.notaM.muestra.pequeña)
[1] 0.1
>
(p.aprobadoT.muestra.pequeña=ng.aprobadoT.muestra.pequeña/n.notaT.muestra.pequeña)
[1] 0.25
```

Como la proporción muestral del turno de mañana es menor que la del turno de tarde, planteamos la hipótesis alternativa de que la proporción poblacional también es menor:

$$H_0: p_{ApMañana} = p_{ApTarde} \text{ vs } H_A: p_{ApMañana} < p_{ApTarde}$$

```
> prop.test(c(ng.aprobadoM.muestra.pequeña,ng.aprobadoT.muestra.pequeña),
c(n.notaM.muestra.pequeña,n.notaT.muestra.pequeña), alternative="less",
conf.level=0.95, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.aprobadoM.muestra.pequeña, ng.aprobadoT.muestra.pequeña) out of
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña)
X-squared = 1.5584, df = 1, p-value = 0.1059
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000  0.04375091
sample estimates:
prop 1 prop 2
 0.10  0.25
```

Como $p\text{-valor} > 0.05$, no se puede aceptar la hipótesis alternativa en el caso de la muestra pequeña.

Podemos confirmarlo, planteando el contraste con la hipótesis alternativa de que son diferentes.

$$H_0: p_{ApMañana} = p_{ApTarde} \text{ vs } H_A: p_{ApMañana} \neq p_{ApTarde}$$

```
> prop.test(c(ng.aprobadoM.muestra.pequeña, ng.aprobadoT.muestra.pequeña),
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña), alternative="two.sided",
conf.level=0.95, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.aprobadoM.muestra.pequeña, ng.aprobadoT.muestra.pequeña) out of
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña)
X-squared = 1.5584, df = 1, p-value = 0.2119
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.38086845  0.08086845
sample estimates:
prop 1 prop 2
 0.10  0.25
```

Como $p\text{-valor} > 0.05$, se confirma que no se puede afirmar que haya diferencia entre las proporciones si se usan muestras pequeñas.

5.2.2 Diferencia de proporciones de alumnos con calificación de Notable

5.2.2.1 Muestras grandes

Las proporciones para las muestras grandes son las siguientes:

```
> (p.notableM.muestra.grande=ng.notableM.muestra.grande/n.notaM.muestra.grande)
[1] 0.6904762
> (p.notableT.muestra.grande=ng.notableT.muestra.grande/n.notaT.muestra.grande)
[1] 0.625
```

Como la proporción muestral del turno de mañana es muy parecida a la del turno de tarde, podemos empezar probando si hay alguna diferencia significativa.

$$H_0: p_{\text{NotMañana}} = p_{\text{NotTarde}} \text{ vs } H_A: p_{\text{NotMañana}} \neq p_{\text{NotTarde}}$$

```
> (prop.test(c(ng.notableM.muestra.grande,ng.notableT.muestra.grande),
c(n.notaM.muestra.grande,n.notaT.muestra.grande), alternative="two.sided",
conf.level=0.95, correct=FALSE))
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.notableM.muestra.grande, ng.notableT.muestra.grande) out of
c(n.notaM.muestra.grande, n.notaT.muestra.grande)
X-squared = 0.34807, df = 1, p-value = 0.5552
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1528884  0.2838408
sample estimates:
  prop 1    prop 2
0.6904762 0.6250000
```

Como $p\text{-valor} > 0.05$, no se puede aceptar la hipótesis alternativa, así que no hay diferencia significativa entre las proporciones de Notables en los dos turnos.

5.2.2.2 Muestras pequeñas

Las proporciones para las muestras pequeñas son las siguientes:

```
>
(p.notableM.muestra.pequeña=ng.notableM.muestra.pequeña/n.notaM.muestra.pequeña)
[1] 0.55
>
(p.notableT.muestra.pequeña=ng.notableT.muestra.pequeña/n.notaT.muestra.pequeña)
[1] 0.6
```

Como en el caso anterior, empezamos probando si son diferentes.

$$H_0: p_{\text{NotMañana}} = p_{\text{NotTarde}} \text{ vs } H_A: p_{\text{NotMañana}} \neq p_{\text{NotTarde}}$$

```
> (prop.test(c(ng.notableM.muestra.pequeña,ng.notableT.muestra.pequeña),
c(n.notaM.muestra.pequeña,n.notaT.muestra.pequeña), alternative="two.sided",
conf.level=0.95, correct=FALSE))
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.notableM.muestra.pequeña, ng.notableT.muestra.pequeña) out of
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña)
X-squared = 0.1023, df = 1, p-value = 0.7491
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.3559993  0.2559993
sample estimates:
 prop 1 prop 2
```

0.55 0.60

Como $p\text{-valor} > 0.05$, tampoco con muestras pequeñas hay evidencias de que haya diferencia significativa entre la proporción de notables en ambos turnos.

5.2.3 Diferencia de proporciones de alumnos con calificación de Sobresaliente

5.2.3.1 Muestras grandes

Las proporciones para las muestras grandes son las siguientes:

```
> (p.sobM.muestra.grande=ng.sobM.muestra.grande/n.notaM.muestra.grande)
[1] 0.2380952
> (p.sobT.muestra.grande=ng.sobT.muestra.grande/n.notaT.muestra.grande)
[1] 0.125
```

Podemos empezar probando si hay alguna diferencia significativa.

$$H_0: p_{SobMañana} = p_{SobTarde} \text{ vs } H_A: p_{SobMañana} \neq p_{SobTarde}$$

```
> prop.test(c(ng.sobM.muestra.grande, ng.sobT.muestra.grande),
c(n.notaM.muestra.grande, n.notaT.muestra.grande), alternative="two.sided",
conf.level=0.95, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.sobM.muestra.grande, ng.sobT.muestra.grande) out of
c(n.notaM.muestra.grande, n.notaT.muestra.grande)
X-squared = 1.5144, df = 1, p-value = 0.2185
alternative hypothesis: two.sided
95 percent confidence interval:
-0.05930537 0.28549584
sample estimates:
prop 1 prop 2
0.2380952 0.1250000
```

Como $p\text{-valor} > 0.05$, no se puede aceptar la hipótesis alternativa, así que no hay diferencia significativa entre las proporciones de Sobresalientes en los dos turnos.

5.2.1.2 Muestras pequeñas

Las proporciones para las muestras pequeñas son las siguientes:

```
> (p.sobM.muestra.pequeña=ng.sobM.muestra.pequeña/n.notaM.muestra.pequeña)
[1] 0.35
> (p.sobT.muestra.pequeña=ng.sobT.muestra.pequeña/n.notaT.muestra.pequeña)
[1] 0.15
```

Como en el caso anterior, podemos empezar probando si hay alguna diferencia.

$$H_0: p_{SobMañana} = p_{SobTarde} \text{ vs } H_A: p_{SobMañana} \neq p_{SobTarde}$$

```
> prop.test(c(ng.sobM.muestra.pequeña, ng.sobT.muestra.pequeña),
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña), alternative="two.sided",
conf.level=0.95, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: c(ng.sobM.muestra.pequeña, ng.sobT.muestra.pequeña) out of
c(n.notaM.muestra.pequeña, n.notaT.muestra.pequeña)
X-squared = 2.1333, df = 1, p-value = 0.1441
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.06112429  0.46112429
sample estimates:
prop 1 prop 2
 0.35  0.15
```

Como $p\text{-valor} > 0.05$, no se puede aceptar la hipótesis alternativa, así que no hay diferencia significativa entre las proporciones de Sobresalientes en los dos turnos.

6. Escribir el informe

Se debe crear un documento con los apartados habituales en este tipo de informes: Título, Introducción, Metodología, Resultados, Análisis de resultados, Conclusiones y Referencias.

6.1 Título

Debe elegirse un título que refleje claramente sobre qué trata el informe.

Ejemplo:

Análisis de las notas de acceso de los estudiantes del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá del curso 2021-22.

6.2 Introducción

En informe comienza situando el trabajo en su contexto, y a continuación se plantea el objetivo del mismo y su alcance.

Ejemplo:

Contexto

En la Universidad de Alcalá se imparten diferentes estudios de grado relacionados con las Tecnologías de la Información y las Comunicaciones, entre los que se encuentra el Grado en Ingeniería en Sistemas de Información. Dado el elevado número de alumnos existente, los estudios se organizan en dos turnos: Mañana y Tarde, y en cada turno los alumnos asisten a un grupo de teoría de cada asignatura, pero se reparten en dos grupos diferentes para realizar las prácticas.

Objetivo

El objetivo de la investigación llevada a cabo ha sido “realizar un análisis estadístico inferencial de las notas de acceso a la Universidad de Alcalá de los alumnos del Grado en Ingeniería en Sistemas de Información del curso 2021-22, comprobando si hay diferencias entre los alumnos del turno de mañana y de tarde”.

Alcance

La investigación se limita a los alumnos del curso 2021-22 matriculados en la asignatura Estadística de primero del Grado en Ingeniería en Sistemas de Información.

6.3 Metodología

Hay que describir las poblaciones, la propiedad de las poblaciones objeto de estudio, las muestras seleccionadas de individuos de las poblaciones y cómo se han conseguido, las hipótesis que se pretenden comprobar, así como las herramientas utilizadas.

Ejemplo:

Poblaciones y propiedad estudiada

Las poblaciones sobre las que se realizará la inferencia son dos y se desconoce su tamaño:

- *Alumnos matriculados en el turno de mañana*
- *Alumnos matriculados en el turno de tarde*

La propiedad de los individuos de las poblaciones que se ha utilizado en el estudio es “Nota de acceso”, numérica continua, con valores positivos que pueden ser mayores o iguales a 5. Se desconoce su varianza poblacional.

Muestras y origen de los datos

Se han seleccionado dos muestras de cada población para realizar el estudio.

- *Muestras grandes: Se ha seleccionado una muestra de 42 alumnos del turno de mañana y una muestra de 32 alumnos del turno de tarde, obteniendo sus notas de acceso mediante una encuesta. Los datos están disponibles en el archivo “encuesta.csv”, en el que los alumnos del turno de mañana son los de los grupos A1 y A2, y los alumnos del turno de tarde son los alumnos de los grupos B1 y B2. Estos datos han sido obtenidos mediante una encuesta.*
- *Muestras pequeñas: Para comparar resultados utilizando también muestras pequeñas, se ha seleccionado aleatoriamente una muestra de 20 alumnos del turno de mañana y una muestra de 20 alumnos del turno de tarde, a partir de las respectivas muestras grandes. Las notas de los alumnos de la muestra pequeña del turno de mañana están en el archivo “20notasmañana.csv”, y las de los alumnos de la muestra pequeña del turno de tarde están en el archivo “20notastarde.csv”.*

Hipótesis

La investigación pretende comprobar si se cumplen las siguientes hipótesis.

1. *Hipótesis: Existe una diferencia significativa entre la media de las notas de acceso de los alumnos del turno de mañana y la media de las notas de acceso de los alumnos del turno de tarde.*
2. *Hipótesis: Existe una diferencia significativa entre la proporción de alumnos que tienen una calificación de “Aprobado” en el turno de mañana y la proporción de alumnos que tienen una calificación de “Aprobado” en el turno de tarde.*

3. *Hipótesis: Existe una diferencia significativa entre la proporción de alumnos que tienen una calificación de “Notable” en el turno de mañana y la proporción de alumnos que tienen una calificación de “Notable” en el turno de tarde.*
4. *Hipótesis: Existe una diferencia significativa entre la proporción de alumnos que tienen una calificación de “Sobresaliente” en el turno de mañana y la proporción de alumnos que tienen una calificación de “Sobresaliente” en el turno de tarde.*

Herramientas

Se han procesado los datos utilizando la aplicación RStudio para Windows, versión 2023.06.1 y el paquete “BSDA” para utilizar la función “z.test” para calcular intervalos de confianza y contrastes de hipótesis sobre muestras grandes.

6.4 Resultados

En este apartado se deben presentar los resultados obtenidos, que son los intervalos de confianza y los p-valores de los contrastes de hipótesis planteados.

Ejemplo:

Intervalos de confianza

En la tabla 1 se muestran los resultados del cálculo de los intervalos de confianza utilizando las muestras grandes, con una confianza del 95%, es decir, con una significación de 0.05. Se han calculado intervalos para la media poblacional de las notas de los alumnos de cada turno, y para las proporciones de diferentes grupos de notas:

- *Aprobado: notas en el rango [5, 7)*
- *Notable: notas en el rango [7,9)*
- *Sobresaliente: notas en el rango [9,11]*

Tabla 1. Intervalos de Confianza usando muestras grandes

Medidas	Turno Mañana	Turno Tarde
Tamaño muestra	42	32
Media (muestra)	8.35	7.59
IC (95%) media	[8.02, 8.67]	[7.22, 7.97]
Prop. Aprobados (muestra)	7.14%	25%
IC (95%) Prop, Aprobados	[2.46%, 19.01%]	[13.25%, 42.11%]
Prop. Notables (muestra)	69.05%	62.5%
IC (95%) Prop, Notables	[53.97%, 80.93%]	[45.25%, 77.07%]
Prop. Sobresalientes (muestra)	23.81%	12.5%
IC (95%) Prop, Sobresalientes	[13.48%, 38.53%]	[4.97%, 28.07%]

En la tabla 2 se muestran los mismos cálculos, pero al utilizar las muestras pequeñas.

Tabla 2. Intervalos de Confianza usando muestras pequeñas

Medidas	Turno Mañana	Turno Tarde
Tamaño muestra	20	20
Media (muestra)	8.45	7.68
IC (95%) media	[7.93, 8.98]	[7.16, 8.19]
Prop. Aprobados (muestra)	10%	25%
IC (95%) Prop, Aprobados	[2.79%, 30.10%]	[11.19%, 46.87%]
Prop. Notables (muestra)	55%	60%
IC (95%) Prop, Notables	[34.21%, 74.18%]	[38.66%, 78.12%]
Prop. Sobresalientes (muestra)	35%	15%
IC (95%) Prop, Sobresalientes	[18.12%, 56.71%]	[5.24%, 36.04%]

En las figuras 1 a 4 se muestran los intervalos de confianza para la media, la proporción de aprobados, la proporción de notables y la proporción de sobresalientes.

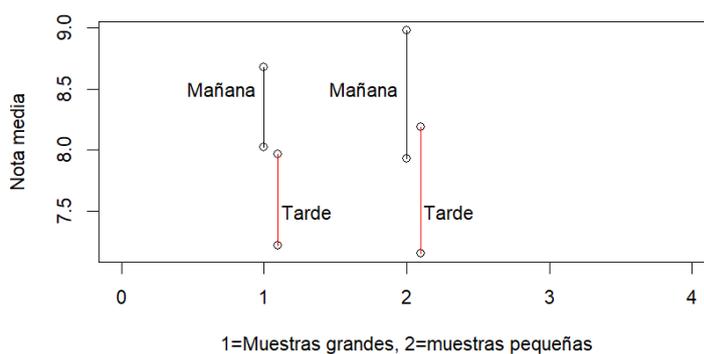


Figura 1. Intervalos de confianza para la media

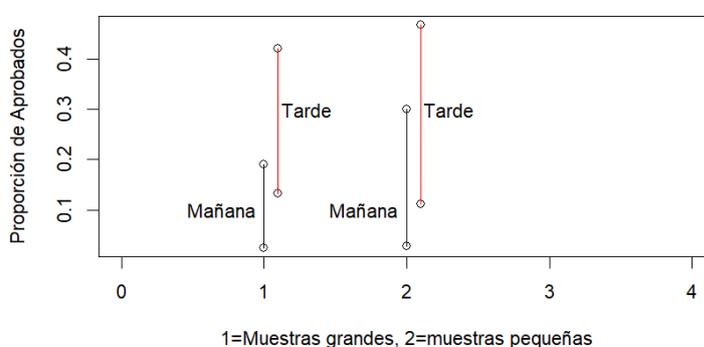


Figura 2. Intervalos de confianza para la proporción de aprobados

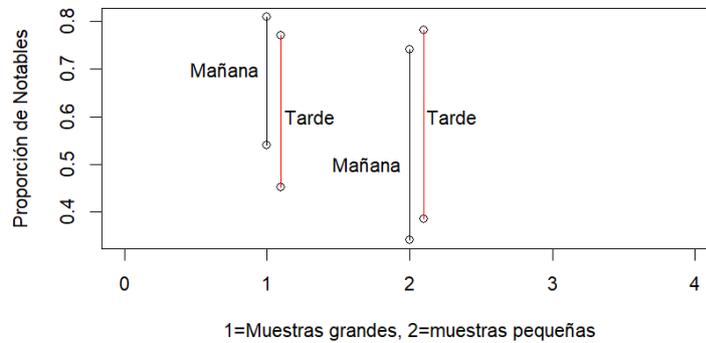


Figura 3. Intervalos de confianza para la proporción de notables

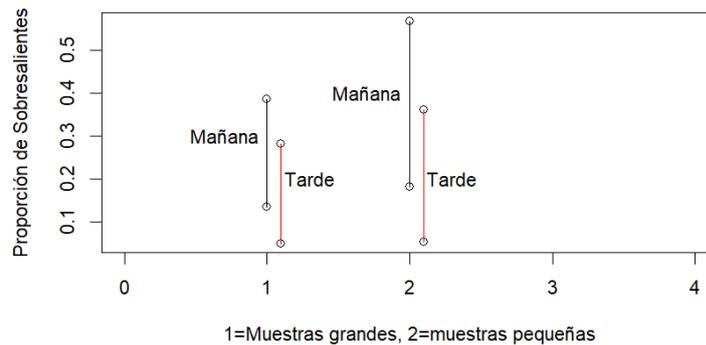


Figura 4. Intervalos de confianza para la proporción de sobresalientes

Contrastes de hipótesis

En la tabla 3 se muestran los resultados que se han obtenido, utilizando las muestras grandes, en los contrastes de hipótesis que se han planteado sobre la diferencia entre las medias poblacionales de las notas de acceso de ambos turnos, y entre las proporciones de alumnos con una calificación de Aprobado, Notable y Sobresaliente.

Tabla 3. Contraste de hipótesis usando muestras grandes

Hipótesis nula (H_0)	Hipótesis alternativa (H_A)	p-valor
Medias iguales	Media del turno de mañana mayor que media del turno de tarde	0.0028
Proporciones de Aprobados iguales	Proporción de aprobados por la mañana menor que por la tarde	0.0162
Proporciones de Notables iguales	Proporciones de Notables diferentes	0.5552

Hipótesis nula (H0)	Hipótesis alternativa (HA)	p-valor
Proporciones de Sobresalientes iguales	Proporciones de Notables diferentes	0.2185

En la tabla 4 se muestran los resultados de los contrastes que se han obtenido al utilizar las muestras pequeñas.

Tabla 4. Contraste de hipótesis usando muestras pequeñas

H0	HA	p-valor
Medias iguales	Media del turno de mañana mayor que media del turno de tarde	0.0163
Proporciones de Aprobados iguales	Proporciones de Aprobados diferentes	0.2119
Proporciones de Notables iguales	Proporciones de Notables diferentes	0.7491
Proporciones de Sobresalientes iguales	Proporciones de Notables diferentes	0.1441

6.5 Análisis de resultados

En este apartado hay que realizar la interpretación de los resultados obtenidos, resaltando las diferencias entre los grupos en los que se haya dividido la muestra. Debe hacerse referencia a todas las tablas y figuras incluidas en el apartado de resultados.

Ejemplo

De los resultados obtenidos (tablas 1 y 2) y de la observación de los intervalos de confianza de la figura 1, se aprecia diferencia entre la nota media de los turnos de mañana y tarde, siendo mayor la media de los alumnos de la mañana. En el caso de las muestras grandes ($n=42$, $m=32$) se observa que la media muestral de la mañana es 8.35 frente a 7.59 por la tarde y no hay solapamiento entre los intervalos de confianza para la media de ambos turnos. Si embargo, en el caso de las muestras pequeñas ($n=20$), en la figura 1 se observa que hay solapamiento entre los intervalos, por lo que no queda claro si la diferencia es significativa, y es necesario comprobarlo con un contraste de hipótesis.

Como puede en la tabla 3, en el caso de las muestras grandes, se ha obtenido un $p\text{-valor}=0.0028$, inferior al nivel de significación 0.05, al plantear la hipótesis alternativa de que la media poblacional del turno de mañana es superior a la del turno de tarde. Y también se confirma con las muestras pequeñas, ya que en la tabla 4 puede verse que el $p\text{-valor}=0.0163$, también menor al nivel de significación 0.05, por lo que tanto con las muestras grandes como con las pequeñas se confirma con un 95% de confianza, que existe una diferencia significativa entre las notas de los turnos de mañana y tarde, y que la nota media del turno de mañana es superior a la nota media del turno de tarde.

En relación son los porcentajes de calificaciones de Aprobado (notas inferiores a 7), Notable (mayores o iguales a 7 y menores que 9) y Sobresaliente (superiores o iguales a 9); en las tablas 1

y 2 se comprueba que, tanto con muestras grandes como pequeñas, los porcentajes muestrales de alumnos con calificación de Notable son muy parecidos: aproximadamente un 69% por la mañana frente a un 62% por la tarde en las muestras grandes, y un 55% frente a un 60% en las muestras pequeñas. Esto también se puede observar gráficamente en la figura 3, donde se aprecia que hay un gran solapamiento entre los intervalos de confianza. Se ha planteado un contraste de hipótesis, con una hipótesis alternativa de que existen diferencias, que se puede rechazar porque el p-valor obtenido, tanto con muestras grandes (0.5552) como con muestras pequeñas (0.7491), es superior al nivel de significación 0.05; lo cual confirma que no existe diferencia significativa entre la proporción de alumnos con calificación de Notable en el turno de mañana y el de tarde.

En el caso las calificaciones de Aprobado, en la figura 2, se observa un cierto solapamiento entre los intervalos de confianza, pero para comprobar si se puede afirmar las proporciones son similares, se ha planteado un contraste de hipótesis, pero no se ha obtenido el mismo resultado utilizando muestras grandes y muestras pequeñas. Con las muestras grandes, el p-valor=0.162 cuando la hipótesis alternativa es que la proporción de aprobados es por la mañana es inferior a la de la tarde, por tanto, como es inferior al nivel de significación 0.05. Con las muestras pequeñas, sin embargo, el p-valor 0.1059 es superior a 0.05, por lo que con estas muestras no hay evidencias de que haya diferencias significativas.

En relación con las calificaciones de Sobresaliente, no hay disparidad en los resultados con muestras grandes o pequeñas; en ambos casos, gráficamente en la figura 4 se observa un gran solapamiento de los intervalos de confianza, y en los contrastes de hipótesis se obtiene en ambos casos un p-valor superior al nivel de significación 0.05, tanto con muestras grandes (p-valor=0.2185) como con muestras pequeñas (p-valor=0.1441). Por lo que se puede afirmar con un nivel de confianza del 95% que no hay diferencia significativa entre la proporción de sobresalientes en ambos turnos.

6.6 Conclusiones

En este apartado debe realizarse un resumen de la investigación explicando los hallazgos más relevantes.

Ejemplo

A partir del análisis de los resultados obtenidos, puede concluirse que existe una diferencia significativa entre las notas medias de los alumnos de los turnos de mañana y tarde, siendo superior la nota de los alumnos del turno de mañana. Por otra parte, no se aprecian diferencias significativas entre ambos turnos en cuanto a la proporción de alumnos con calificaciones de Notables y la proporción de calificaciones de Sobresaliente. Sin embargo, sí se comprueba que hay evidencias, usando muestras grandes, que la proporción de alumnos con calificación de Aprobado es menor en el turno de mañana, lo cual es razonable teniendo en cuenta que en ese turno las notas son más elevadas, ya que hay evidencias de que a media es superior a la de la tarde.

Por tanto, sólo se han podido demostrar dos de las cuatro hipótesis planteadas sobre la existencia de diferencias significativas entre ambas poblaciones, son las hipótesis que afirmaban que había diferencias entre las medias y entre las proporciones de Aprobados.

6.7 Referencias

Si en el informe se cita algún documento o página web, en este apartado debe incluirse la referencia completa, utilizando un formado normalizado, como por ejemplo el formato [APA](#).