

Guía 6

Tema: Pruebas de Hipótesis

Ingeniería Industrial. Facultad de Ciencias Agrarias

Ejercicio 1

Ante una denuncia de los consumidores a la Dirección de Lealtad Comercial, una empresa quiere saber si puede garantizar que las cajas de un detergente en polvo que vende, contiene 500 gramos del producto. Por experiencia, la firma sabe que la cantidad de detergente que contienen las cajas tiene distribución normal, por lo que se toma una muestra aleatoria de 25 cajas y se calcula la media de detergente contenido en ellas, lo que da un valor de 480 gramos, con un desvío estándar de 75 gramos. Con esos datos, pruebe si el contenido no supera los 500 gramos con un nivel de significación del 5 %.

Ejercicio 2

Supongamos que una tienda de comestibles vende cajas de cereal Captain Crisp con un peso de 16 onzas. El departamento de control de calidad, requiere asegurarse que la media de las cajas de cereal pesan al menos 16 onzas. Se tomó una muestra aleatoria de tamaño 9 (cajas) y se pesó. Trabaje con un nivel de significancia (alfa) de 0.05

Pesos		
15.5	16.2	16.1
15.8	15.6	16.0
15.8	15.9	16.2

La compañía asegura que la media de las cajas de cereal es **al menos** de 16 onzas.

- ¿Cuántas muestras involucra el problema?
- Plantee las hipótesis y realice el test. Trabaje con valor de significancia del $\alpha=0.05$
- Concluir sobre los resultados obtenidos.

En R

En primer lugar, ingreso los datos

```
caja_cereal<-c(15.5, 16.2, 16.1, 15.8, 15.6, 16.0, 15.8, 15.9, 16.2)
```

En R, podemos hacerlo paso a paso de la siguiente manera:

- Realizo cálculos de media y desviación típica

```
x_bar<-mean(caja_cereal)
s<-sd(caja_cereal)
mu_0<-16
n=9
```

2. Cálculo de t de prueba

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

```
t<-(x_bar-mu_0)/(s/sqrt(n))
t
```

```
## [1] -1.2
```

3. p-valor

```
pt(t, df = n - 1)
```

```
## [1] 0.1322336
```

La otra forma es hacerlo con la funcion t.test, ingresando los valores apropiados:

```
t.test(x = caja_cereal , mu = 16, alternative = c("less"), conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data:  caja_cereal
## t = -1.2, df = 8, p-value = 0.1322
## alternative hypothesis: true mean is less than 16
## 95 percent confidence interval:
##      -Inf 16.05496
## sample estimates:
## mean of x
##      15.9
```

Ejercicio 3

Prueba t de muestras independientes

Se tienen 2 sets de datos que corresponden a pesos de hombres y mujeres de dos muestras independientes. Queremos saber si existen diferencias significativas entre los dos grupos de estudio.

Muestra	Hombres	Mujeres
1	70	64
2	82	72
3	78	60
4	74	76
5	94	72
6	82	80
7	-	84
8	-	68

- ¿Cuántas muestras involucra el problema?
- Plantee las hipótesis del experimento y el estadístico de prueba.
- Realice la prueba de hipótesis con un intervalo de 95% de confianza.
- ¿Existen diferencias significativas entre ambos grupos?

En R

- Ingreso de datos y cálculo de n y m

```
varones = c(70, 82, 78, 74, 94, 82)
n = length(varones)
mujeres = c(64, 72, 60, 76, 72, 80, 84, 68)
m = length(mujeres)
```

- Cálculo de media y desviación típica

```
x_bar<-mean(varones)
s_x<-sd(varones)
y_bar<-mean(mujeres)
s_y<-sd(mujeres)
```

- Cálculo de la desviación estándar agrupada

$$s_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}$$

```
s_p = sqrt(((n - 1) * s_x ^ 2 + (m - 1) * s_y ^ 2) / (n + m - 2))
```

- Cálculo de t

$$t = \frac{(\bar{x} - \bar{y}) - \mu_0}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

```
t = ((x_bar - y_bar) - 0) / (s_p * sqrt(1 / n + 1 / m))
t
```

```
## [1] 1.823369
```

5. Cálculo de p-valor

```
1 - pt(t, df = n + m - 2)
```

```
## [1] 0.04661961
```

O podemos hacerlo mediante la función `t.test()`

```
t.test(varones, mujeres, mu=0, alternative = c("greater"), var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data:  varones and mujeres
## t = 1.8234, df = 12, p-value = 0.04662
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.1802451      Inf
## sample estimates:
## mean of x mean of y
##      80      72
```

Ejercicio 4

Prueba t de muestras dependientes o prueba t-apareada

Una escuela de atletismo ha tomado un nuevo instructor, y quiere testear la efectividad del nuevo entrenamiento propuesto comparando las medias de 10 corredores en los 100 metros. Se presentan los tiempos antes y después del entrenamiento de cada atleta.

Tiempos										
Antes	12.9	13.5	12.8	15.6	17.2	19.2	12.6	15.3	14.1	11.3
Después	12.0	12.2	11.2	13.0	15.0	15.8	12.2	13.4	12.9	11.0

- ¿Cuántas muestras involucra el problema?
- ¿Cómo plantearía las hipótesis y cuál es el estadístico de prueba?
- Realizar una prueba de hipótesis con un 95% de confianza.

En R

```
#Ingreso los datos
a<-c(12.9, 13.5, 12.8, 15.6, 17.2, 19.2, 12.6, 15.3, 14.4, 11.3)
```

```
b<-c(12.0, 12.2, 11.2, 13.0, 15.0, 15.8, 12.2, 13.4, 12.9, 11.0)
```

```
#Cálculo la diferencia y la media de esas diferencias con la función mean()  
diferencia<-a-b  
diferencia
```

```
## [1] 0.9 1.3 1.6 2.6 2.2 3.4 0.4 1.9 1.5 0.3
```

```
mean(diferencia)
```

```
## [1] 1.61
```

```
#Cálculo de t  
t<-mean(diferencia)/(sd(diferencia)/sqrt(10))  
t
```

```
## [1] 5.267141
```

```
#Cálculo del p-valor  
1- pt(t, df = 10 - 1)
```

```
## [1] 0.000257911
```

```
t.test(a, b, mu=0, alternative="greater", paired=TRUE, var.equal=T)
```

```
##  
## Paired t-test  
##  
## data: a and b  
## t = 5.2671, df = 9, p-value = 0.0002579  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 1.049675 Inf  
## sample estimates:  
## mean of the differences  
## 1.61
```

Ejercicio 5

Una empresa necesita comprar engranajes plásticos para un tipo impresora láser; en la ciudad donde existen dos proveedores que fabrican este tipo de engranajes. Para decidir a cuáles de los proveedores comprarle, la empresa analiza una característica importante de los mismos que es la resistencia al impacto de esos engranajes (medida en pies/libra). Para ello, tomó una muestra aleatoria de engranajes de cada proveedor.

Proveedor	n	Media	Desviación estándar
1	10	290	12
2	16	321	22

Se desea estudiar si existe evidencia para apoyar la afirmación de que el proveedor 2 ofrece engranajes con mayor resistencia al impacto medio que el proveedor 1.

- En primer lugar, con un nivel de significación (alfa) de 0.05, evalúe si las varianzas de las poblaciones son iguales.
- Defina las hipótesis en este problema y escriba la expresión del estadístico de prueba.
- Verifique las hipótesis planteadas en b) utilizando $\alpha = 0.05$. ¿Qué conclusión obtiene?

Ejercicio 6

Una empresa de software está investigando la posibilidad de cambiar el lenguaje de programación usado comúnmente, de forma de sustituir el lenguaje C (Leng C) por Visual Basic (VBA), a fin de mejorar la rapidez de programación y aumentar así la productividad de la empresa. Para ello se pide a 12 programadores, todos familiarizados con ambos lenguajes, que programen un cierto algoritmo en ambos lenguajes. Se registró el tiempo que tardan en completar la tarea en cada caso, produciendo los siguientes datos medidos en minutos:

Programador												
Lenguaje	1	2	3	4	5	6	7	8	9	10	11	12
VBA	17	16	21	14	18	24	16	14	21	23	13	18
Leng C	18	14	19	11	23	21	10	13	19	24	15	20

- Enuncie formalmente las hipótesis que le interesa testear a la empresa de software y calcule el estadístico de prueba.
- A partir de los dos puntos anteriores, ¿Qué lenguaje erigiría si usted fuese el dueño de la empresa?