Inferencia de información para una población

Distribuciones muestrales y teorema central del límite. Intervalos de confianza. Contrastes de hipótesis para una población

Blanca de la Fuente

PID_00161059

UOC
Universitat Oberta de Catalunya
www.uoc.edu

Índice

Introducción

El objetivo de la inferencia estadística es obtener información acerca de una población, partiendo de la información que contiene la muestra. La selección de la muestra debe garantizar su representatividad, lo que se consigue eligiéndola al azar mediante diferentes procedimientos de muestreo que se estudian en el módulo 5.

Una vez seleccionada una muestra, se dispone de un conjunto de valores, en cuyo caso los *métodos descriptivos* estudiados en el módulo 1 facilitan el análisis de estos valores muestrales. El problema que ahora se aborda es la extensión de estos resultados al conjunto de la población o, en otras palabras, dar respuesta al siguiente interrogante: Dada cierta información muestral ¿qué podemos afirmar de la población?

La solución de este problema será el objetivo de la *inferencia estadística*.

Hasta ahora se había supuesto que los valores de los parámetros de las distribuciones de probabilidad eran conocidos. Pero esto casi nunca ocurre, de manera que tenemos que usar los datos muestrales para estimarlos. Los **estimadores** proveen valores a esos parámetros.

Cuando las inferencias que se realizan se refieren a características poblacionales concretas, es necesaria una etapa de diseño de estimadores. En este módulo se presentan los conceptos básicos para la estimación de la proporción, de la media y de la varianza de la población respectivamente.

Un enfoque alternativo es indicar un rango de valores, entre los cuales tiene que estar el parámetro con una determinada precisión: esta es la idea de un **intervalo de confianza**.

A continuación se plantea en este módulo el problema del **contraste de hipótesis**, desarrollando métodos que permiten contrastar la validez de una conjetura o de una afirmación utilizando datos muestrales. El proceso comienza cuando un investigador formula una hipótesis sobre la naturaleza de una población. La formulación de esta hipótesis implica claramente la elección entre dos opciones; a continuación, el investigador selecciona una opción basándose en los resultados de un estadístico calculado a partir de una muestra aleatoria de datos.

Objetivos

Los objetivos académicos del presente módulo se describen a continuación:

1. Explorar las distribuciones de la media, de la proporción y de la varianza muestral.

2. Aplicar el Teorema central del límite.

3. Crear intervalos de confianza.

4. Usar la distribución *t* en una prueba de hipótesis.

5. Utilizar la distribución chi-cuadrado (χ^2) en una prueba de hipótesis.

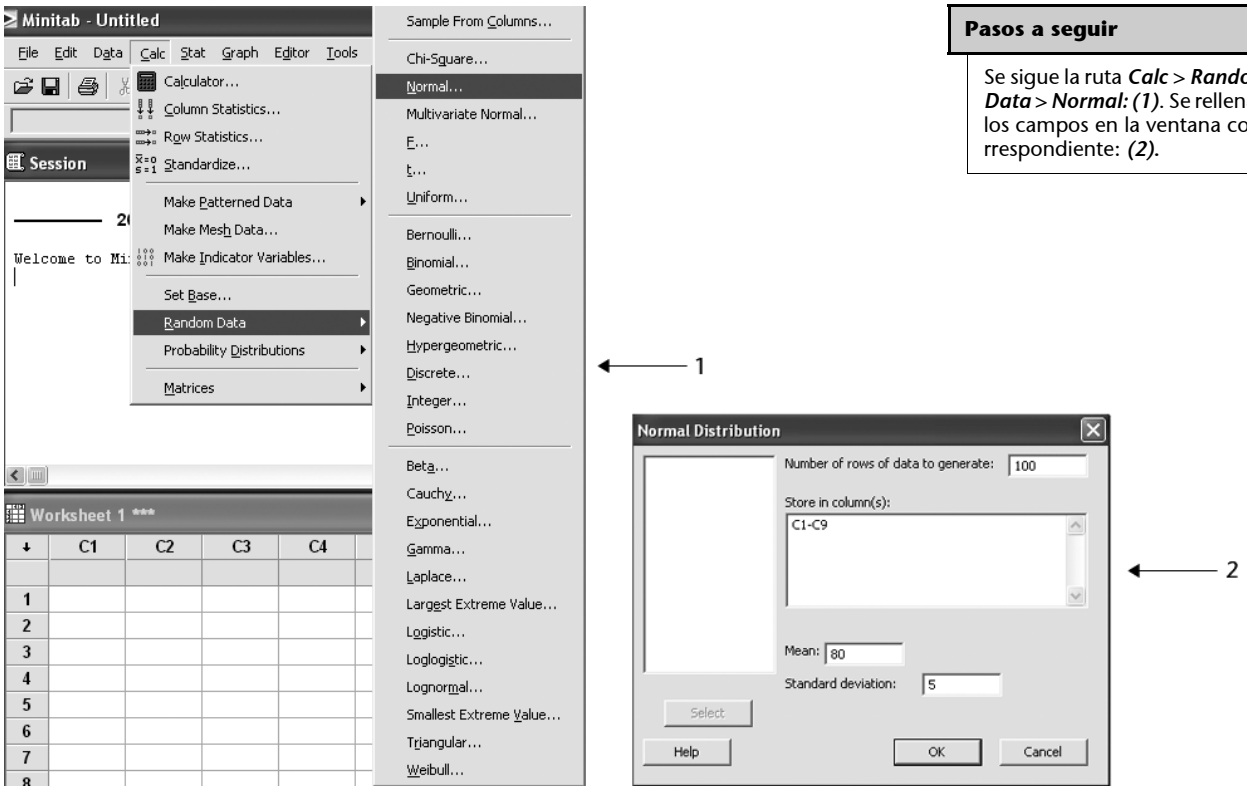
1. Distribuciones muestrales y Teorema central del límite

Una muestra aleatoria permite hacer inferencia sobre ciertas características de la distribución de la población. Esta inferencia estará basada en algún **estadístico**, es decir, alguna función particular de la información muestral. La **distribución muestral** de este estadístico es la distribución de probabilidades de los valores que puede tomar el estadístico a lo largo de todas las posibles muestras con el mismo número de observaciones, que pueden ser extraídas de la población.

Por ejemplo, en la distribución normal, los dos parámetros son la media de la población μ y la desviación estándar poblacional σ. Se puede estimar el valor μ calculando el promedio muestral o media muestral, $\bar{x}$, y el valor de σ mediante el cálculo de la desviación típica muestral, s. En este caso la media muestral, $\bar{x}$ y la desviación típica muestral, s, son los estadísticos. En el caso de la distribución binomial, los parámetros son n y p. Para estimar el parámetro proporción poblacional, p, se utiliza el estadístico proporción muestral, $\hat{p}$.

El estudio de las distribuciones muestrales se puede ilustrar mediante la creación con Minitab de 100 muestras de datos aleatorios normales con media 80 y desviación típica 5, con 9 observaciones de cada muestra (figura 1). A partir de datos aleatorios se crea una columna de datos que contenga el promedio de cada muestra o media muestral.

Figura 1. Pasos a seguir para estudiar una distribución muestral



Pasos a seguir

Se sigue la ruta ***Calc*** > ***Random Data*** > ***Normal: (1)***. Se rellenan los campos en la ventana correspondiente: ***(2)***.

Se ha generado así una matriz de nueve columnas y cien filas (figura 2). Cada componente de esta matriz es una observación aleatoria proveniente de una distribución normal de media 80 y desviación estándar 5.

Figura 2. Resultado de una matriz

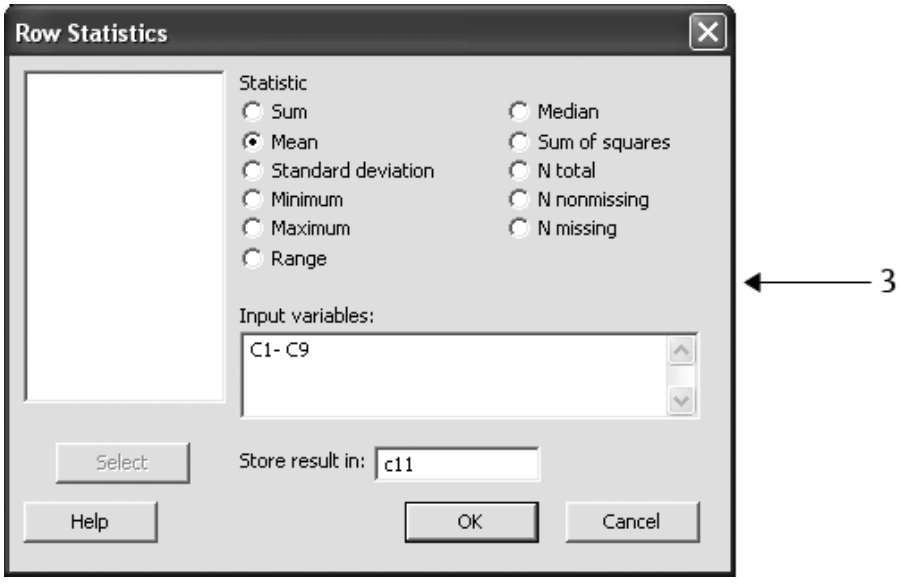
Minitab - Untitled - [Worksheet 1 ***]

File Edit Data Calc Stat Graph Editor Tools Window Help

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	86,2119	79,8202	83,3841	85,2706	84,3977	73,3733	79,2915	79,7839	85,9532	
2	77,9609	77,8749	75,9217	76,5488	80,2022	77,9314	76,3676	75,2757	81,7320	
3	79,4466	81,7029	79,6951	64,8797	77,2981	83,3234	89,4473	74,5262	88,4057	
4	75,2041	79,8637	79,4250	76,8203	77,6463	74,8025	91,9961	81,1118	78,1692	
5	80,3298	80,6675	81,6815	77,6024	71,7586	84,5731	85,6013	78,1734	78,8703	
6	72,6596	75,2371	85,2919	69,2655	83,4426	89,4642	82,6151	74,9094	83,5735	
7	83,4955	79,0883	80,8709	74,2571	80,8703	83,4584	78,2343	76,5349	84,1050	
8	77,7092	76,8970	74,7272	82,4228	84,4440	78,4365	79,6917	83,6371	81,2270	
9	83,4490	81,1309	76,9926	85,5929	84,2468	78,2026	87,5854	83,7920	65,8216	
10	81,0768	85,6596	84,0062	68,5531	70,6466	76,6778	82,9853	70,8620	79,8056	
11	78,5384	77,2805	83,3829	91,9047	76,0708	73,8196	84,3317	75,8071	74,4023	
...										
91	65,7335	73,9428	86,7681	85,7170	83,8731	84,1641	79,1929	79,6862	81,4350	
92	76,0597	83,0876	77,9049	71,7392	83,9294	90,2161	76,4049	74,5141	78,1746	
93	85,8218	83,5000	70,8513	82,4729	85,0008	78,4955	77,3336	79,6959	79,5596	
94	74,7113	68,5013	79,8273	76,5099	70,7168	86,8968	70,3262	81,5301	82,8092	
95	84,2597	70,3432	82,1207	81,9862	80,0716	79,1272	83,2129	80,0331	74,9672	
96	79,5742	78,4683	80,5376	75,5732	83,9279	80,5597	81,9438	80,2280	79,6279	
97	81,5282	81,0194	77,3914	77,9764	73,2830	76,8476	87,2748	86,6303	85,1350	
98	81,8135	82,7231	81,0083	84,1466	79,2423	81,0257	78,1474	84,3158	74,5273	
99	73,7441	79,6640	74,2215	85,9686	75,9545	81,6767	77,6596	77,1299	85,1886	
100	77,8603	80,4039	78,1897	73,7551	72,7348	75,9893	74,9260	69,2976	77,8818	

Se considera que cada una de las filas obtenidas es una muestra, y se calcula la media asociada a cada una de estas cien muestras (figura 3):

Figura 3. Pasos a seguir para calcular las medias



Pasos a seguir

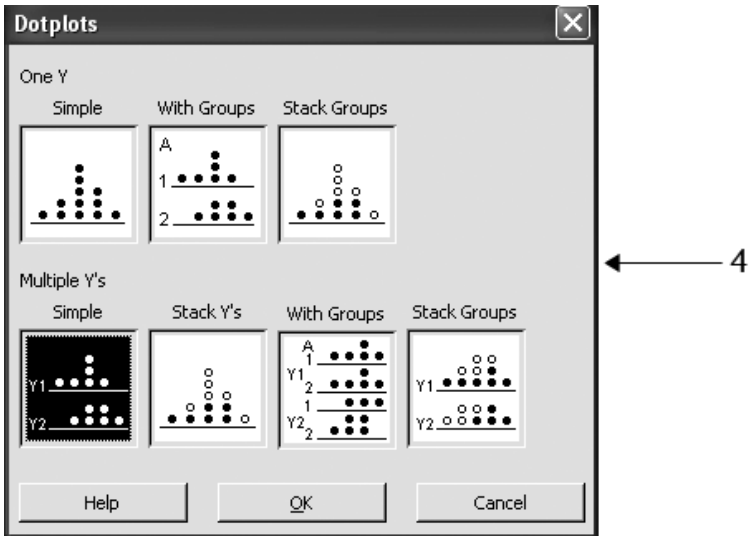
Una vez generados los datos se sigue la ruta ***Calc*** > ***Row Statistics*** y se rellenan los campos en la ventana correspondiente: ***(3)***.

En la columna C11 de la figura 4 hay cien nuevos valores (las medias). En la figura 5 se muestran los *dotplot* asociados a las columnas C1 (que representan cien valores aleatorios obtenidos de una normal 80-5) y C11:

Figura 4. Resultado del análisis

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
											x-barra	
1	86,2119	79,8202	83,3841	85,2706	84,3977	73,3733	79,2915	79,7839	85,9532		81,9429	
2	77,9609	77,8749	75,9217	76,5488	80,2022	77,9314	76,3676	75,2757	81,7320		77,7572	
3	79,4466	81,7029	79,6951	64,8797	77,2981	83,3234	89,4473	74,5262	88,4057		79,8583	
4	75,2041	79,8637	79,4250	76,8203	77,6463	74,8025	91,9961	81,1118	78,1692		79,4488	
5	80,3298	80,6675	81,6815	77,6024	71,7586	84,5731	85,6013	78,1734	78,8703		79,9175	
6	72,6596	75,2371	85,2919	69,2655	83,4426	89,4642	82,6151	74,9094	83,5735		79,6065	
7	83,4955	79,0883	80,8709	74,2571	80,8703	83,4584	78,2343	76,5349	84,1050		80,1016	
8	77,7092	76,8970	74,7272	82,4228	84,4440	78,4365	79,6917	83,6371	81,2270		79,9103	
9	83,4490	81,1309	76,9926	85,5929	84,2468	78,2026	87,5854	83,7920	65,8216		80,7571	
10	81,0768	85,6596	84,0062	68,5531	70,6466	76,6778	82,9853	70,8620	79,8056		77,8081	
11	78,5384	77,2805	83,3829	91,9047	76,0708	73,8196	84,3317	75,8071	74,4023		79,5042	
12	70,0951	78,7096	77,3802	82,9569	72,4905	76,7535	88,9660	85,7643	75,4986		78,7350	

Figura 5. Pasos a seguir para crear el gráfico de puntos de los *dotplot*



Pasos a seguir

Se sigue la ruta ***Graph*** > ***Dotplot*** y se rellenan los campos en la ventana correspondiente: ***(4)***.

La salida de Minitab de la figura 6 muestra que la distribución de la variable aleatoria inicial X (columna C1) era normal y, según el gráfico de puntos, parece que también la distribución de la v.a. X-barra ($\bar{x}$) es normal, de media muy similar y desviación estándar menor (los puntos de la $\bar{x}$ están menos "dispersos" que los de la x).

También podemos hacer un histograma de frecuencias de la distribución de las medias muestrales ($\bar{x}$), como se aprecia en la figura 7.

Figura 6. Gráfico de puntos de valores de los *dotplot*

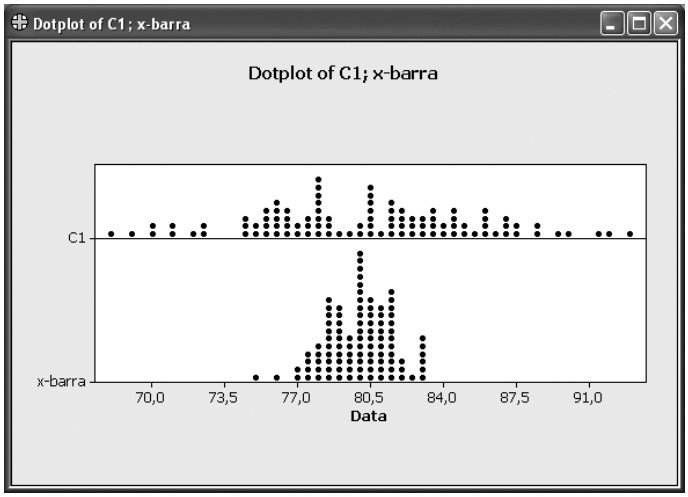
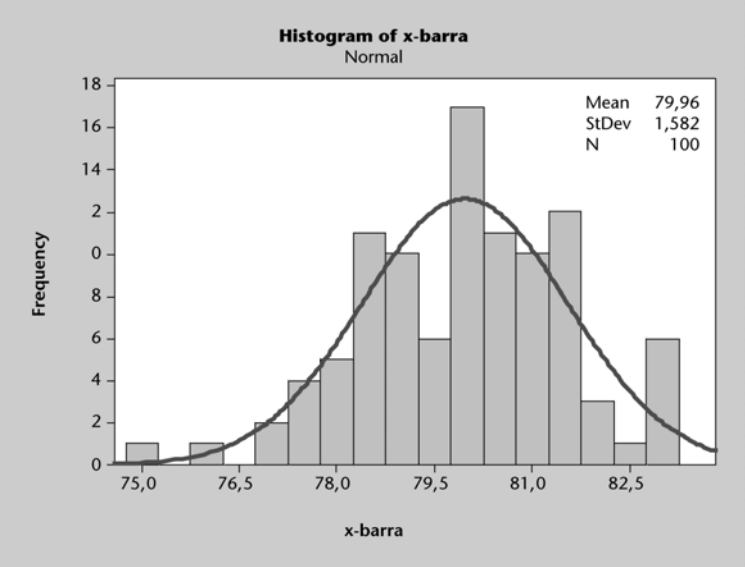


Figura 7. Histograma de frecuencias absolutas de valores de $\bar{x}$ a partir de nueve muestras aleatorias simples, cada una de tamaño cien



Finalmente, en la figura 8 se obtienen los estadísticos que describen la distribución de las medias muestrales.

Figura 8. Resultado del análisis de *X*-barra

Descriptive Statistics: x-barra

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1
Median	Q3						
x-barra	100	0	79,962	0,158	1,582	75,192	78,814
80,146	81,000						

Variable	Maximum
x-barra	83,154

Pasos a seguir

Se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***Display Descriptive Statistics*** y se selecciona la variable ***C11*** (***x-barra***) en la ventana correspondiente.

La media de los cien valores contenidos de la columna C11 (y que es una aproximación a la media de la v.a. *X*-barra) es de 79,962, valor muy similar a la media de *X* (que era de 80). Esto es coherente con lo que la teoría nos indica:

- La media muestral coincide con la media de la población, $\mu_{\bar{X}} = \mu$.

La desviación estándar de los cien valores de la columna C11 (que será una aproximación a la desviación estándar de *X*-barra) es de 1,582. Si tomamos la desviación estándar de *X* (que era de 5) y la dividimos por 3 (raíz de 9, el tamaño de la muestra), obtenemos el valor 1,667.

- Ambos valores son muy parecidos, tal y como la teoría predice:

$$\sigma_{\bar{X}} = \sigma / \sqrt{n}$$

Es interesante señalar que si no se hubiera tomado inicialmente una variable normalmente distribuida, las conclusiones obtenidas serían semejantes siempre que el tamaño muestral *n* fuera lo suficientemente grande tal y como predice el **Teorema central del límite.**

Teorema central del límite

El análisis anterior se aplica sólo a la distribución normal. ¿Qué ocurre si nuestros datos provienen de otra distribución de probabilidad? ¿Podemos decir algo acerca de la distribución muestral de la media en ese caso? Para ello se utiliza el ***Teorema central del límite,*** el cual expresa que si tenemos una muestra tomada de una distribución de probabilidad con media μ y desviación típica de σ, la distribución muestral de $\bar{x}$ es aproximadamente normal con media μ y desviación típica de, $\sigma/\sqrt{n}$ que es el error estándar. Lo notable acerca del teorema central del límite es que la distribución de la media muestral de $\bar{x}$ es más o menos normal, sea cual sea la distribución original de probabilidad. A medida que aumenta el tamaño de la muestra, la aproximación a la distribución normal se acerca cada vez más.

Nota

Consideraremos que n es lo bastante grande cuando, como mínimo, $n > 30$.

Una consecuencia de este teorema es:

Dada cualquier variable aleatoria con esperanza μ y para n suficientemente grande, la distribución de la variable:

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

es una normal estándar $N(0,1)$.

Cálculo del error estándar

Recordemos que si la variable tiene una desviación típica conocida σ, el error estándar se puede calcular como $\sigma/\sqrt{n}$. Cuando σ es desconocida, calculamos el error estándar como $s/\sqrt{n}$, siendo s la desviación típica de la muestra.

Un caso particular es la **aproximación de la binomial a la normal**:

Sea X una variable aleatoria con distribución $B(n, p)$ binomial con n suficientemente grande. Entonces, X es aproximadamente normal con esperanza np y varianza $np(1–p)$.

En este caso, n grande significa que np y $np(1–p)$ son los dos mayores que 5 o bien que $n > 30$.

Por tanto, cuando el tamaño de la muestra, n, es grande, la distribución de la **proporción** es aproximadamente una distribución normal de esperanza p y desviación típica $\sqrt{p(1-p)/n}$. En este caso $\sqrt{p(1-p)/n}$, corresponde al error estándar $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Recordatorio

Si X sigue una distribución **binomial** de parámetros n y p, entonces:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

para los $k \in \{0, \ldots, n\}$

Ejemplo: se hace una encuesta sobre un determinado tema que tiene dos opciones, A y B. La probabilidad de que un individuo concreto opine A es p y n es el número de encuestas hechas. Hemos preguntado a cuatrocientos habi-

tantes y encontramos que el 30% opina A, es decir, que podemos establecer que $p = 0,3$. Entonces, la distribución de la proporción de habitantes que opina A sigue una distribución normal, cuya media es 0,3, que coincide con la proporción del 30% de los habitantes de la población que opinan A, y la desviación estándar es 0,0229, que corresponde a la desviación típica de la población dividida por la raíz cuadrada del tamaño de la muestra.

$$N\left(0,3,\sqrt{\frac{0,3(1-0,3)}{400}}\right) = N(0,3;0,0229)$$

2. Distribución de la media muestral

Se deben considerar dos casos para la distribución de la media muestral.

Caso de desviación típica poblacional conocida

Si la variable que estudiamos sigue una distribución normal con media μ y desviación típica σ conocidas, entonces la media muestral es también normal con la misma media μ y desviación típica $\sigma/\sqrt{n}$, donde n es el tamaño de la muestra.

Nota

Si σ es la desviación típica de la población y n el tamaño de la muestra, se define el **error estándar de la media muestral** como:

$$\sigma/\sqrt{n}$$

Observad

El error estándar es cada vez menor cuanto mayor es el tamaño de la muestra.

Siempre que la distribución de las medias muestrales sea una distribución normal, se puede calcular una **variable aleatoria normal estandarizada**, ***Z***, que tiene una media 0 y una varianza 1:

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

Si la distribución de la población no es normal pero el tamaño muestral n es suficientemente grande, entonces se usará el teorema central del límite y la variable media muestral se aproxima a una normal estándar a medida que el tamaño de la muestra aumenta. En general, dicha aproximación se considera válida para tamaños muestrales superiores a treinta.

En el apartado anterior se vio que la variable aleatoria binomial sigue una distribución normal aproximada cuando aumenta el tamaño de la muestra.

Ejemplo: en la asignatura de *Archivística* de una licenciatura de Documentación se sabe que las calificaciones siguen una distribución normal de media 7,4 y desviación estándar 0,78. Se desea conocer el porcentaje de estudiantes con nota superior a 6,5 e inferior a 8,5. ¿Con qué nota se va a calificar como "excelente" (A), si esta es la calificación del 5% de estudiantes con mejor nota?

Solución:

La variable sigue una distribución $N(7,4; 0,78)$. Primero se calcula el estadístico Z normal estandarizado:

$$P(6,5 \le X \le 8,5) = P\left(\frac{6,5-7,4}{0,78} \le \frac{X-7,4}{0,78} \le \frac{8,5-7,4}{0,78}\right) =$$
$$= P(-1,15 \le Z \le 1,41) =$$
$$= P(Z \le 1,41) - P(Z \le -1,15) = 0,9207 - 0,1251 = 0,7956$$

Los valores de probabilidad se buscan en la tabla $N(0,1)$ o calculándose con cualquier programa estadístico como se muestra en el ejemplo desarrollado en el módulo 1.

A la vista del resultado, se puede decir que el porcentaje de estudiantes con nota superior a 6,5 e inferior a 8,5 es de 79,56%.

Para calcular la nota a partir de la cual se califica como excelente, se calcula el estadístico Z normal estandarizado:

$$P(X \geq A) = P\left(\frac{X-7,4}{0,78} \geq \frac{A-7,4}{0,78}\right) = P(Z \geq z_A) = 0,05$$

En las tablas de la $N(0,1)$ o mediante cualquier programa estadístico se busca un valor z que deje a la derecha un área de 0,05, aproximadamente el valor es: $z_A = 1,645$, de manera que:

$$\frac{A-7,4}{0,78} = 1,645 \qquad \Rightarrow \qquad A = 7,4 + 1,645 \cdot 0,78 = 8,683$$

A partir de una nota de 8,6 se califica como "excelente"(A).

Caso de desviación típica poblacional desconocida

Cuando la desviación poblacional es desconocida y el tamaño de la muestra es pequeño, deberemos hacer una estimación de la desviación típica con la llamada *desviación típica muestral*. Para ello es necesario presentar una nueva distribución de probabilidad. Esta nueva distribución se conoce con el nombre de ***t* de Student** cuyas características se explicaron en el módulo 1.

Para determinar la distribución de la media muestral cuando la desviación poblacional es desconocida, se debe calcular la desviación típica muestral:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Si la variable estudiada sigue una distribución normal con media μ y desviación típica desconocida, entonces el estadístico media muestral sigue una distribución t_{n-1}, es decir, una ***t* de Student con *n*–1 grados de libertad.**

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Los grados de libertad asociados con el valor de t son n–1(tamaño de la muestra menos uno).

Nota

En este caso se define el **error estándar de la media muestral** como:

$$s/\sqrt{n}$$

Ejemplo: el tiempo que han tardado en infectarse de virus cada uno de los ordenadores de una editorial ha sido: 2,5; 7,4; 8,0; 4,5; 7,4 y 9,2 segundos.

Suponemos que el tiempo que tarda un ordenador de esa editorial en infectarse sigue la distribución normal de media 6,5 y se desconoce la varianza poblacional. Se desea calcular la probabilidad de que un ordenador tarde entre 5 y 10 segundos en infectarse.

Solución:

Como se desconoce la varianza de la población, la media muestral seguirá una distribución ***t* de Student con 5 grados de libertad**.

Para calcular el valor del estadístico *t*, se debe calcular la desviación típica muestral. El valor obtenido es $S = 2{,}5$:

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

La probabilidad solicitada será:

$$p(5 \le T \le 10) = p\left(\frac{5-6{,}5}{2{,}5/\sqrt{6}} \le t_5 \le \frac{10-6{,}5}{2{,}5/\sqrt{6}}\right) = p(-1{,}47 \le t_5 \le 3{,}43) =$$
$$p(t_5 \le 3{,}43) - p(t_5 \le -1{,}47) = 0{,}99 - 0{,}1 = 0{,}89$$

Para calcular la probabilidad se utiliza la tabla *t* o un programa estadístico (figura 9).

Figura 9. Resultado de Minitab

```
Cumulative Distribution Function

Student's t distribution with 5 DF

   x  P( X <= x )
3,43     0,990682

Cumulative Distribution Function

Student's t distribution with 5 DF

    x  P( X <= x )
-1,47     0,100758
```

Pasos a seguir

Para calcular las probabilidades de una distribución *t* de Student se sigue la ruta ***Calc*** > ***Probability Distributions*** > ***t*** y se completan los parámetros en la ventana correspondiente. El resultado se muestra en la figura 9.

3. Distribución de la proporción muestral

En el apartado 5 del módulo 1 se dijo que la distribución binomial era la suma de *n* variables aleatorias independientes, cada una de las cuales tiene una probabilidad de éxito *p*. Para caracterizar la distribución se necesita conocer el valor de *p,* que es la proporción de miembros de la población que tienen una característica de interés. La **proporción muestral de éxitos** en una muestra aleatoria extraída de una población en la que la proporción de éxitos *p* será:

Distribución de la proporción muestral

Es una aplicación del **Teorema central del límite**.

$$\hat{p} = \frac{X}{n}$$

Por lo tanto $\hat{p}$ es la media de un conjunto de variables aleatorias independientes. Además puede utilizarse el teorema central del límite para sostener que la distribución de probabilidad de $\hat{p}$ puede considerarse una distribución normal si el tamaño de la muestra es grande.

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Nota

La distribución de $\hat{p}$ tiene una media igual a la proporción poblacional *p*.

La desviación estándar de $\hat{p}$ es el **error estándar de la media muestral** como:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Igual que en el caso de la media muestral, siempre que la distribución de la proporción muestral sea una distribución normal, se puede calcular una **variable aleatoria normal estandarizada**, **Z**, que tiene una media cero y una varianza uno.

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

La proporción muestral tiene muchas aplicaciones, entre las cuales se encuentra el estudio de los resultados de encuestas, la estimación de la cuota porcentual del mercado, el porcentaje de inversiones empresariales que tiene éxito y los resultados electorales entre otros.

Observad

El error estándar es cada vez menor cuanto mayor es el tamaño de la muestra.

Ejemplo: el 22% de los discos se venden por la Red en formato MP3 y el resto se vende en tiendas en formato CD. Se consideran las ventas de los próximos 5.000 discos. Se desea saber ¿qué distribución sigue la proporción muestral de discos vendidos por la Red? ¿Cuál es el número esperado de discos que se venderán por la Red? ¿Cuál es la probabilidad de que se vendan por la Red más de 1.500 discos?

Solución:

En este ejercicio se tiene que $p = 0{,}22$ y $n = 5.000$.

Para determinar la distribución de la proporción muestral, dado que el tamaño de la muestra es grande $n = 5.000$, se aplica el teorema central del límite. La distribución será aproximadamente **normal**, el valor de la media es el de la proporción poblacional (0,22).

Se calculará el error estándar $s_{\hat{p}} = \sqrt{\frac{0,22(1-0,22)}{5.000}} = 0,00586$

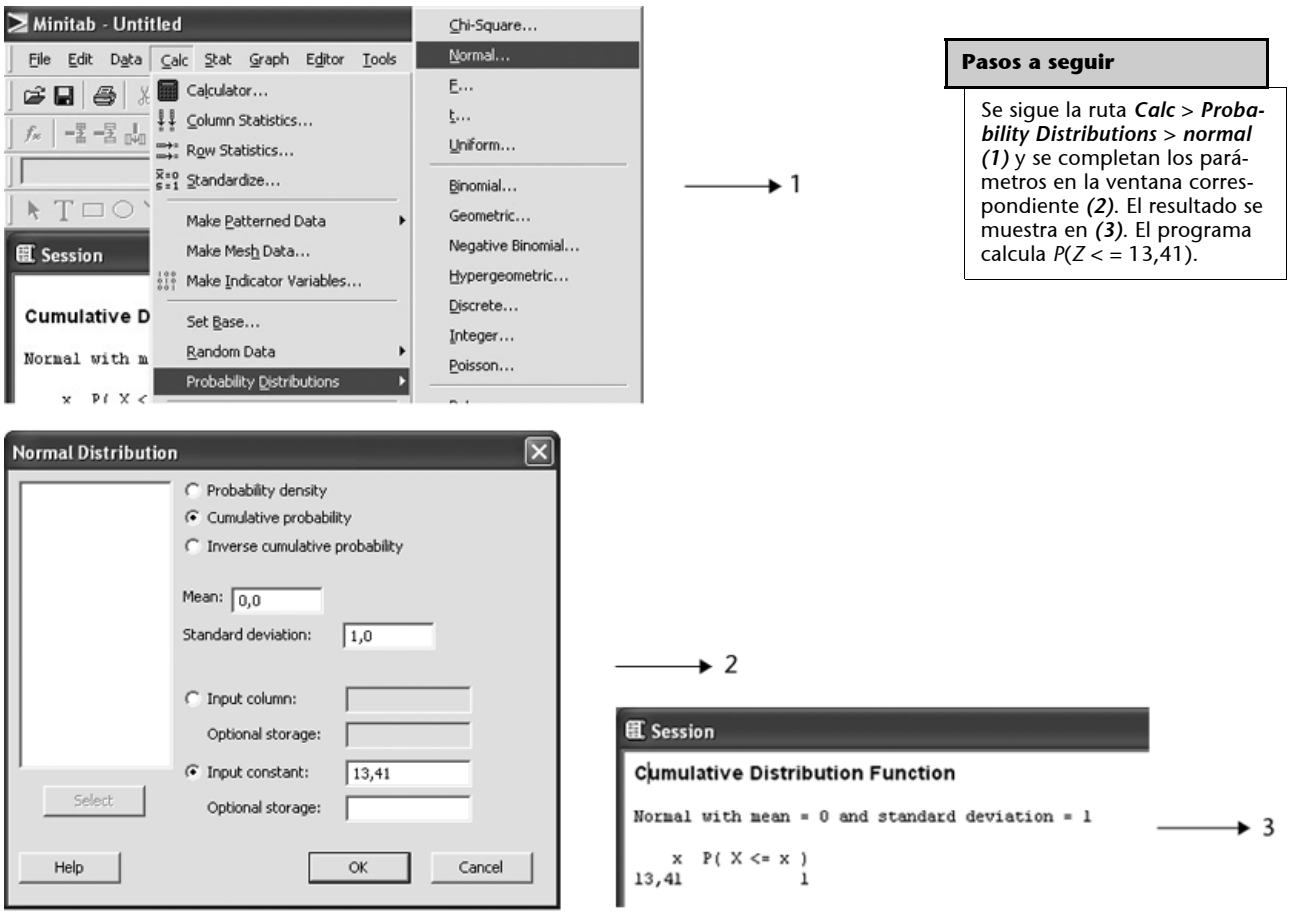
El valor esperado de discos vendidos por la Red será del 22% de los 5.000 que se venden en total, es decir, 1.100 discos en formato MP3.

La probabilidad de que se vendan menos de 1.500 discos por la Red será igual a la probabilidad de que la proporción muestral sea superior o igual al 30%. Para obtener esta probabilidad, primero se calculará el estadístico Z normal estandarizado:

$$P(p > 30\%) = P\left(Z > \frac{0,30 - 0,22}{0,00586}\right) = P(Z > 13,41) = 0$$

La probabilidad de Z se obtiene en la tabla $N(0,1)$. En la práctica, los cálculos probabilísticos anteriores se suelen automatizar con la ayuda de algún software estadístico o de análisis de datos. La figura 10 muestra cómo se pueden calcular probabilidades de una normal con ayuda de Minitab.

Figura 10. Cálculo de probabilidades con Minitab



Pasos a seguir

Se sigue la ruta ***Calc*** > ***Probability Distributions*** > ***normal (1)*** y se completan los parámetros en la ventana correspondiente ***(2)***. El resultado se muestra en ***(3)***. El programa calcula $P(Z < = 13,41)$.

El valor obtenido con Minitab es $P(Z <= 13{,}41)$. Por lo tanto, para obtener la probabilidad deseada calcularemos la probabilidad complementaria $P(Z > 13{,}41) = 1 - P(Z <= 13{,}41) = 1 - 1 = 0$.

4. Distribución de la varianza muestral

Una vez analizadas las distribuciones de las medias muestrales y las proporciones muestrales, se examinarán las distribuciones de las varianzas muestrales. A medida que las empresas y la industria ponen más énfasis en la producción de productos que satisfagan los criterios de calidad, es mayor la necesidad de calcular y reducir la varianza poblacional. Cuando la varianza es alta en un proceso, algunas características de los productos pueden tener una gama más alta de valores, como consecuencia de la cual hay más productos que no tienen un nivel de calidad aceptable. Se pueden obtener productos de calidad si el proceso de producción tiene una varianza baja, de manera que es menor el número de unidades que tienen un nivel de calidad inferior al deseado. Comprendiendo la distribución de las varianzas muestrales podemos hacer inferencias sobre la varianza poblacional.

Si se estudia una muestra aleatoria de tamaño n y varianza muestral s^2 obtenida de una población normal de media μ y varianza σ^2 desconocidas, entonces la varianza muestral se distribuye como una χ^2_{n-1} con n–1 grados de libertad:

$$\chi^2_{n-1} = \frac{(n-1)s_x^2}{\sigma_x^2}$$

Por lo tanto, se pueden hacer inferencias sobre la varianza poblacional σ^2 utilizando s^2 y la distribución chi-cuadrado. Este proceso se muestra en el siguiente ejemplo.

Ejemplo: en una gran ciudad se ha observado que durante el verano las facturas del consumo de electricidad siguen una distribución normal que tiene una desviación típica del 100 euros. Se ha tomado una muestra aleatoria de 25 facturas. Se desea calcular la probabilidad de que la desviación típica muestral sea inferior a 75 euros.

Solución:

En este ejercicio se tiene que $n = 25$ y $\sigma^2 = (100)^2$. Utilizando la distribución chi-cuadrado se puede establecer que:

$$P\left(s^2 < 75^2\right) = P\left(\frac{(25-1)75^2}{(100)^2} < \chi^2_{24\,g.l.}\right) = P\left(13,5 < \chi^2_{24\,g.l.}\right)$$

Los valores de la distribución chi-cuadrado pueden obtenerse en la tabla de dicha distribución con 24 grados de libertad:

$$\chi^2_{24\,g.l.} = 12,401;\ \chi^2_{24\,g.l.} = 13,848$$

El valor de probabilidad estará entre 0,025 y 0,05 (0,0428) exactamente.

5. Intervalos de confianza para una población

En los apartados anteriores hemos considerado la estimación puntual de un parámetro desconocido de la población, es decir, el cálculo de un único número que sea una buena aproximación. En la mayoría de los problemas prácticos, un estimador puntual por sí solo es inadecuado. Por ejemplo, supongamos que un control hecho sobre una muestra aleatoria de manuales procedentes de un gran envío de una editorial nos lleva a estimar que el 10% de todos los manuales son defectuosos. Un gerente que se enfrenta a este dato posiblemente se hará preguntas del tipo: ¿puede estar totalmente seguro de que el verdadero valor del porcentaje de manuales defectuosos está entre el 5% y el 15%? O ¿es muy posible que entre el 9% y el 11% de los manuales sean defectuosos? Esta clase de preguntas requieren información que va más allá de la contenida en una simple estimación puntual; son preguntas que buscan la fiabilidad de dicho estimador. En otras palabras, se trata de la búsqueda de un **estimador por intervalos**, un rango de valores entre los que posiblemente se encuentre la cantidad que se estima.

Debemos medir de alguna manera la confianza que podemos tener en el intervalo. Este porcentaje de muestras que dan lugar a intervalos que contienen el auténtico valor del parámetro es el llamado **nivel de confianza**.

Nivel de confianza

El nivel de confianza también se denota por $(1 - \alpha)$ 100% normalmente consideraremos $(1 - \alpha)$, igual a 90%, 95% o 99%.

Así pues, un intervalo de confianza para cierto parámetro con un nivel de confianza de C% es un intervalo calculado a partir de una muestra de manera que el procedimiento de cálculo garantiza que el C% de las muestras dé lugar a un intervalo que contenga el valor real del parámetro.

La expresión *confianza del 95%* indica confianza en el método utilizado, de manera que el 95% de las veces que apliquemos el método a la misma población obtendremos intervalos que sí contienen el valor del parámetro poblacional.

Intervalo de confianza para la media cuando la población es normal y conocemos la desviación estándar

La variable que queremos estudiar sigue una ley normal de media μ (desconocida) y desviación estándar σ conocida. Disponemos de una muestra aleatoria simple de tamaño n y el valor de la media de la muestra es $\overline{x}$.

Se calculan los intervalos de confianza al nivel de confianza $(1 - \alpha)$% mediante la siguiente expresión:

$$(\text{media de la muestra} - ME, \text{media de la muestra} + ME)$$

donde ME es el **margen de error** que tenemos que calcular, de manera que el $(1 - \alpha)$ % de las muestras produzca un intervalo que contenga el verdadero valor de μ.

El procedimiento que describimos sirve también para variables que no sigan una distribución normal, siempre que la desviación típica sea conocida y que el tamaño de la muestra sea $n > 30$.

Fijamos el nivel de confianza: se acostumbra a considerar $(1 - \alpha)$ igual a 90%, 95% o 99%.

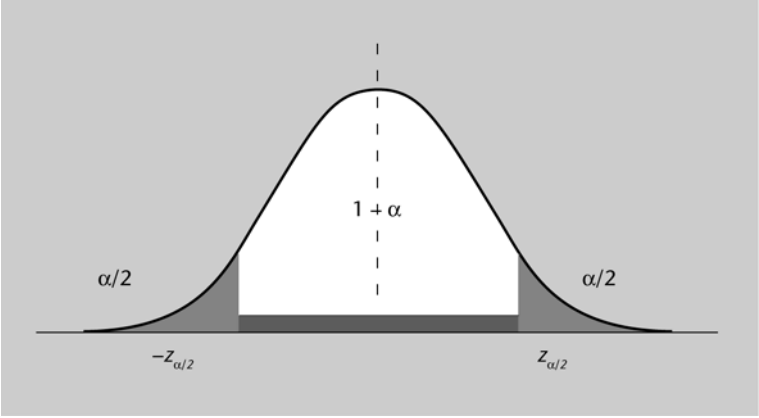
Calculamos el **error estándar** de la media como $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

Obtenemos el **valor crítico**, que es aquel valor $z_{\alpha/2}$ que hace que:

$$P(Z \geq z_{\alpha/2}) = \alpha/2$$

en el que Z es una variable aleatoria normal $N(0,1)$. Se muestra gráficamente en la figura 11.

Figura 11. Gráfico de intervalo de confianza para μ con desviación típica conocida



Para los niveles de confianza usuales, los valores críticos correspondientes son:

- $(1 - \alpha) = 90\% = 0{,}9$, $\alpha = 0{,}1$ y $z_{\alpha/2} = z_{0,05} = 1{,}645$
- $(1 - \alpha) = 95\% = 0{,}95$, $\alpha = 0{,}05$ y $z_{\alpha/2} = z_{0,025} = 1{,}96$
- $(1 - \alpha) = 99\% = 0{,}99$, $\alpha = 0{,}01$ y $z_{\alpha/2} = z_{0,005} = 2{,}575$

Calculamos el denominado ***margen de error*** (también denominado ***precisión de la estimación***) como $z_{\alpha/2}$ para el error estándar, es decir, como:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Nota

Por tanto, el margen de error es la mitad de la longitud del intervalo de confianza.

El intervalo de confianza obtenido con la muestra de partida es:

$$\left(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

o lo que es lo mismo, $\bar{x} \pm ME$.

Es necesario interpretar exactamente los intervalos de confianza. Si se extraen repetida e independientemente muestras aleatorias de n observaciones de la población, entonces el $100(1-\alpha)\%$ de estos intervalos contendrá el verdadero valor de la media poblacional.

El efecto del tamaño de la muestra

En muchas ocasiones, una vez fijado el nivel de confianza nos marcaremos como objetivo dar el valor del parámetro μ con cierta precisión. La única manera de obtener la precisión deseada consiste en modificar de forma adecuada el tamaño de la muestra. Supongamos que deseamos una precisión o margen de error *ME*; puesto que sabemos que:

Tamaño de la muestra

Es fácil ver que si queremos reducir el ancho del intervalo de confianza a la mitad, deberemos tomar una muestra cuatro veces mayor.

$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Se obtiene el tamaño deseado de la muestra para dicha precisión:

$$n \geq \left(z_{\alpha/2} \right)^2 \frac{\sigma^2}{ME^2}$$

Intervalo de confianza para la media cuando la población es normal y desconocemos la desviación estándar

La variable que queremos estudiar sigue una ley normal de media μ (desconocida) y desviación estándar también desconocida. Disponemos de una muestra aleatoria simple de tamaño n y el valor de la media de la muestra es $\bar{x}$. Entonces:

Calculamos los intervalos de confianza al nivel de confianza $(1-\alpha)\%$, mediante la siguiente expresión se fija el **nivel de confianza**, que habitualmente se escribe como $(1-\alpha)\%$.

Calculamos la desviación típica muestral S para obtener el **error estándar** de la media como:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Calculamos el **valor crítico**, que es aquel valor $t_{\alpha/2}$ tal que:

$$P(t_{n-1} \geq t_{n-1,\alpha/2}) = \alpha/2$$

en el que t_{n-1} es una variable aleatoria de Student con $n-1$ grados de libertad.

Como el **margen de error** es:

$$ME = t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

El intervalo de confianza obtenido con la muestra es el siguiente:

$$\bar{x} \pm ME$$

Intervalo de confianza para la proporción

Interesa conocer la proporción de miembros de la población que poseen una característica específica. Si se toma una muestra aleatoria simple de tamaño *n,* la proporción muestral es un buen estimador de la proporción poblacional. En este apartado se desarrollan intervalos de confianza para la proporción.

Cuando el tamaño de la muestra sea bastante grande, en concreto siempre que el tamaño sea superior a cien, se aplicará el teorema centra del límite, y, como se ha visto en apartados anteriores, la distribución de la proporción muestral sigue una distribución normal estándar $N(0,1)$.

Igual que en los intervalos anteriores se calcula el ***margen de error*** como $z_{\alpha/2}$ multiplicado por el error estándar, es decir:

$$ME = z_{\alpha/2}s_{\hat{p}} = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Nota

El parámetro es *p*.

El estadístico es $\hat{p}$.

El intervalo de confianza obtenido con la muestra de partida será:

$$\hat{p} \pm ME$$

El tamaño de la muestra es $n = \left(z_{\alpha/2}\right)^2 \frac{\hat{p}(1-\hat{p})}{ME^2}$

Ejemplo: un servidor de correo ha recibido 2.000 mensajes, de los cuales 250 son “SPAM”. Construid un intervalo de confianza del 96% para la proporción de mensajes “SPAM”, ¿cuántos correos se han de estudiar en el servidor para poder afirmar que el error entre la proporción de mensajes “SPAM” recibidos y la probabilidad de que el servidor reciba un “SPAM” sea menor que 0,03 con una probabilidad del 95%?

Solución:

El intervalo de confianza del 96% para la proporción de la población se obtiene por medio de la ecuación:

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right)$$

Se deduce que $\hat{p} = \frac{250}{2000} = 0{,}125,\ n = 2000,\ z_{\alpha/2} = z_{0{,}02} = 2{,}054.$

Por lo tanto, el intervalo de confianza de la proporción poblacional al 96% es

$$\left(0{,}125 - 2{,}054\sqrt{\frac{0{,}125\cdot 0{,}875}{2000};}\quad 0{,}125 + 2{,}054\sqrt{\frac{0{,}125\cdot 0{,}875}{2000}}\right) = (0{,}1098;\ 0{,}1402).$$

Se podría decir que la proporción de todos los mensajes Spam recibidos de la población estarán entre el 10,98% y el 14,02% (con un margen de error del 1,52% al nivel de confianza del 96%).

Se calculará el mínimo tamaño de la muestra necesario para que el error sea menor que 0,03 con una probabilidad del 95% es:

$$n \geq (z_{\alpha/2})^2 \frac{\hat{p}\cdot(1-\hat{p})}{ME^2} = (z_{0{,}025})^2 \frac{0{,}125\cdot 0{,}875}{0{,}03^2} = 1{,}96^2 \cdot \frac{0{,}109}{0{,}0009} = 466{,}75$$

Por tanto, se deben estudiar 467 mensajes.

Ejemplo con Minitab: en el ejemplo anterior se comparan los intervalos de confianza al 90 y el 99%, manteniendo constante el tamaño de la muestra, para contestar a la siguiente pregunta: Conforme aumenta la amplitud de un intervalo de confianza, ¿aumenta o disminuye el nivel de confianza asociado? En las figuras 12 y 13 utilizamos Minitab para analizar ambos escenarios.

Figura 12. Resultado del Intervalo de confianza del 90% con Minitab

Test and CI for One Proportion

Test of p = 0,125 vs p not = 0,125

Sample	X	N	Sample p	90% CI	Z-Value	P-Value
1	250	2000	0,125000	(0,112836; 0,137164)	0,00	1,000

Using the normal approximation.

Figura 13. Resultado del Intervalo de confianza del 99% con Minitab

Test and CI for One Proportion

Test of p = 0,125 vs p not = 0,125

Sample	X	N	Sample p	99% CI	Z-Value	P-Value
1	250	2000	0,125000	(0,105951; 0,144049)	0,00	1,000

Using the normal approximation.

Notar que al aumentar el nivel de confianza, deberemos ampliar la amplitud del intervalo a fin de "abarcar" un rango mayor para el parámetro poblacional estimado.

Intervalo de confianza para la varianza

¿Cómo se puede construir un intervalo de confianza para la varianza poblacional?

Primero se fijará el nivel de confianza $1 - \alpha$. Se calculará el estadístico.

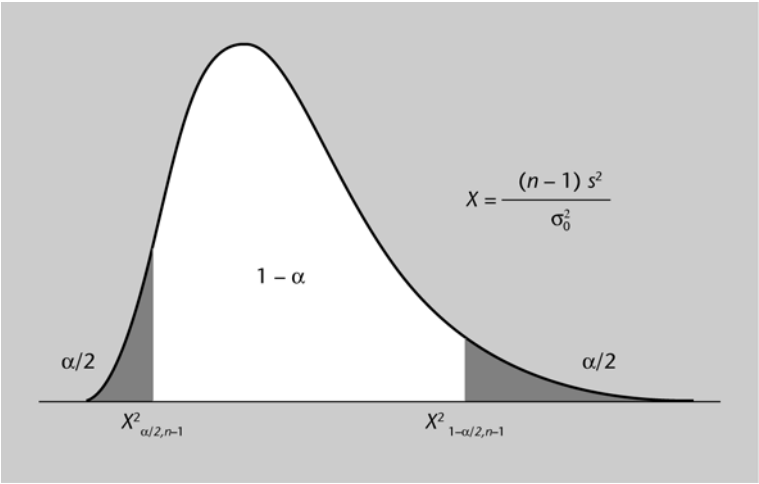
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

es una observación de una variable aleatoria χ^2 con $n - 1$ grados de libertad.

Donde s^2 es la varianza muestral de una muestra aleatoria de tamaño n tomada de una población normal de varianza σ^2.

La figura 14 muestra los valores de la distribución χ^2_{n-1} que cortan una probabilidad de $\alpha/2$ en las dos colas, es decir, los puntos críticos $\chi^2_{n-1,\alpha/2}$ y $\chi^2_{n-1,\,1-\alpha/2}$.

Figura 14. Gráfico de intervalo de confianza de la varianza



Ejemplo de intervalo de confianza para la varianza

Una empresa de autobuses urbanos espera que las horas de llegada en diversas paradas tengan poca variabilidad. La varianza de la muestra de 10 tiempos de llegada de autobús fue $s^2 = 4{,}8$ minutos^2. Suponiendo que la población de tiempos de llegada tiene una distribución normal, se desea determinar un intervalo de confianza del 95% para la varianza poblacional de los tiempos de llegada.

El estadístico de prueba: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ tiene una distribución chi-cuadrado con $n - 1 = 9$ grados de libertad. Determinamos los valores $\chi^2_{9,0,975} = 16{,}0471$ y $\chi^2_{9,0,025} = 45{,}7222$.

El intervalo de confianza para la varianza de la población será:

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}};\frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}\right]=\left[\frac{9\cdot 4,8}{45,7222};\frac{9\cdot 4,8}{16,0471}\right]=[0,94;2,69]\text{minutos}$$

La raíz cuadrada de esos valores será el intervalo de confianza de 95% para la desviación estándar: $0,97 \leq \sigma \leq 1,64$.

6. Contrastes de hipótesis para una población

En este apartado se desarrollan métodos para contrastar hipótesis que permiten comparar la validez de una conjetura o afirmación utilizando datos muestrales. El proceso comienza cuando un investigador formula una hipótesis sobre la naturaleza de una población. La formulación de esta hipótesis implica la elección entre dos opciones; a continuación, el investigador selecciona una opción basándose en los resultados de un estadístico calculado a partir de una muestra aleatoria de datos.

He aquí algunos ejemplos de problemas representativos:

1) Un investigador quiere saber si una propuesta de reforma fiscal es acogida de igual forma por hombres y mujeres. Para analizar si es así, recoge las opiniones de una muestra aleatoria de hombres y mujeres.

2) Una compañía recibe un cargamento de piezas. Sólo puede aceptar el envío si no hay más de un 5% de piezas defectuosas. La decisión de si aceptar la remesa puede basarse en el examen de una muestra aleatoria de piezas.

3) Una profesora está interesada en valorar la utilidad de hacer controles regularmente en un curso de estadística. El curso consta de dos partes y la profesora realiza estos controles sólo en una de ellas. Cuando acaba el curso, compara los conocimientos de los estudiantes en las dos partes del curso mediante un examen final y analiza la hipótesis de que los controles aumentan el nivel medio de conocimientos.

Los ejemplos propuestos tienen algo en común. La hipótesis se formula sobre la población y las conclusiones sobre la validez de esta hipótesis se basan en la información muestral. El test o contraste será la herramienta que nos permitirá extraer conclusiones a partir de la diferencia entre las observaciones y los resultados que se deberían obtener si la hipótesis de partida fuese cierta.

Planteamiento del contraste de hipótesis

En la prueba de hipótesis se comienza proponiendo una hipótesis de partida acerca de un parámetro poblacional. Esta hipótesis se llama **hipótesis nula** y se representa como $\boldsymbol{H_0}$. A continuación se define otra hipótesis, la **hipótesis alternativa**, que es la opuesta de lo que se afirma en la hipótesis nula. La hipótesis alternativa se representa como $\boldsymbol{H_1}$. El procedimiento para probar una hipótesis comprende el uso de datos de una muestra para probar las dos aseveraciones representadas por H_0 y H_1.

Hipótesis

Con la misma hipótesis nula podemos estudiar varias hipótesis alternativas.

Las hipótesis expresan una afirmación sobre el valor del parámetro. Podemos tener una hipótesis nula del tipo $\boldsymbol{H_0}$**:** $\theta = \theta_0$.

La hipótesis alternativa puede ser unilateral, como $\boldsymbol{H_1}$**:** $\theta > \theta_0$ o $\boldsymbol{H_1}$**:** $\theta < \theta_0$, o bilateral, como $\boldsymbol{H_1}$**:** $\theta \neq \theta_0$**.**

Una vez planteadas las hipótesis nula y alternativa, debemos tomar una decisión a partir de las observaciones. Por otro lado, existen dos decisiones posibles:

1) Aceptar la hipótesis nula.

2) Rechazar la hipótesis nula.

Errores en el contraste

Con el fin de llegar a una de estas dos conclusiones, se adopta una **regla de decisión** basada en la evidencia muestral. Por consiguiente, *no se puede saber con seguridad* si la hipótesis nula es cierta o falsa. Por tanto, cualquier regla de decisión adoptada tiene cierta probabilidad de llegar a una conclusión falsa. Como se indica en la tabla 1, pueden cometerse dos tipos de errores. Un error que se puede cometer, llamado **error de tipo I**, es rechazar una hipótesis nula cierta. Si la regla de decisión es tal que la probabilidad de rechazar la hipótesis nula cuando es cierta es α, entonces α se llama **nivel de significación** del contraste. La probabilidad de aceptar la hipótesis nula cuando es cierta es $(1 - \alpha)$. El otro error posible, llamado **error de tipo II**, ocurre cuando se acepta una hipótesis nula falsa. La probabilidad de cometer este tipo de error, cuando la hipótesis nula es falsa, se denota por β. Entonces, la probabilidad de rechazar una hipótesis nula falsa es $(1 - \beta)$, y se denomina **potencia del contraste.**

Regla de decisión

Error de tipo I: rechazar una hipótesis nula cierta.

Error de tipo II: aceptar una hipótesis nula falsa.

Nivel de significación: la probabilidad de rechazar una hipótesis nula que es cierta (esta probabilidad a veces se expresa en %, con lo que nos referimos a un contraste de significación α como un contraste al nivel 100 α%).

Potencia: la probabilidad de rechazar una hipótesis nula que es falsa.

Tabla 1. Errores y decisiones correctas en contrastes de hipótesis

		Condición de la población	
		H_0 **verdadera**	H_0 **falsa**
Decisión	**Aceptar** H_0	Decisión correcta	Error de tipo II
	Rechazar H_0	Error de tipo I	Decisión correcta

Atención

Un nivel α **= 0,05** significa que aunque la hipótesis nula sea cierta, los datos de cinco de cada cien muestras nos la harán rechazar. Es decir, aceptamos que podemos rechazar la hipótesis nula equivocadamente cinco de cada cien veces.

Para plantear y resolver un contraste de hipótesis, es necesario:

1) Fijar las hipótesis nula y alternativa.

2) Fijar un nivel de significación.

3) Determinar el estadístico de contraste y su ley.

4) A partir de aquí, tenemos dos métodos posibles:

4a) Calcular el *p*-valor asociado a nuestro estadístico de contraste calculado. Comparar el *p*-valor con el nivel de significación y tomar una decisión.

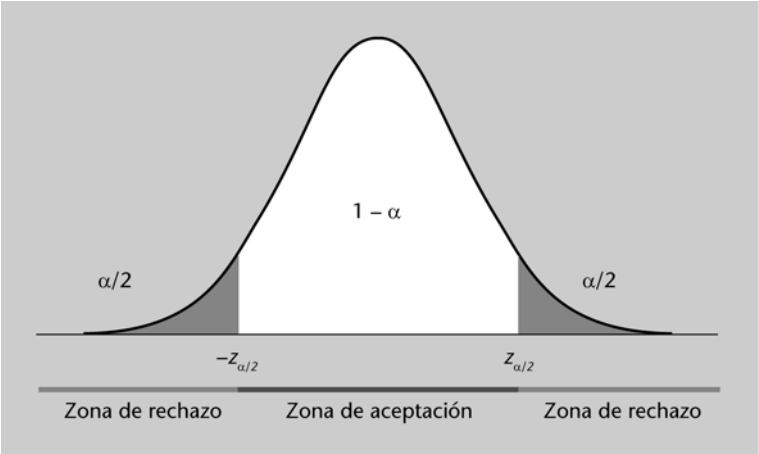
4b) Calcular el valor crítico. Comparar el valor crítico con el estadístico de contraste y tomar una decisión.

Zona de aceptación y zona de rechazo de la hipótesis nula

Ejemplo 1. "Contraste bilateral"

La parte del gráfico (figura 15) sombreada en rojo corresponde a la zona en la que rechazamos la hipótesis nula. La zona sin sombrear corresponde a la región de aceptación de la hipótesis nula.

Figura 15. Gráfico que muestra la zona de aceptación y de rechazo de la hipótesis nula en un contraste bilateral



Recordad

Si tenemos una muestra de tamaño n de una distribución $N(\mu, \sigma^2)$, entonces

$$Z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

sigue una distribución normal estándar.

Para determinar el valor $z_{\alpha/2}$, sólo hay que imponer que el error de tipo I (probabilidad de rechazar H_0 cuando es cierta) sea menor o igual que el nivel de significación α. Por ejemplo, para $\alpha = 0{,}05$ encontramos (por ejemplo, en las tablas de la normal) que $z_{\alpha/2} = 1{,}96$.

Para decidir si rechazamos la hipótesis nula o no, usaremos el llamado **estadístico de contraste.** Un estadístico de contraste es una función de la muestra cuya distribución conocemos bajo la hipótesis nula.

Validez del método

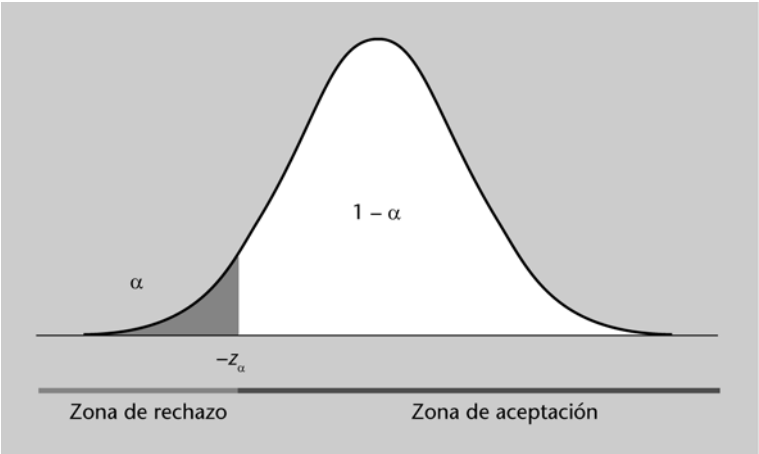
El método es el mismo para cualquier distribución simétrica, así que también sirve si el estadístico de contraste sigue una distribución t de Student.

- Aceptaremos H_0 si $|z| \leq z_{\alpha/2}$
- Rechazaremos H_0 si $|z| \geq z_{\alpha/2}$

Ejemplo 2. "Contraste unilateral inferior"

La parte del gráfico (figura 16) sombreada corresponde a la zona de rechazo de la hipótesis nula. La zona sin sombrear corresponde a la región de aceptación de la hipótesis nula.

Figura 16. Gráfico que muestra la zona de rechazo de la hipótesis nula en un contraste unilateral inferior



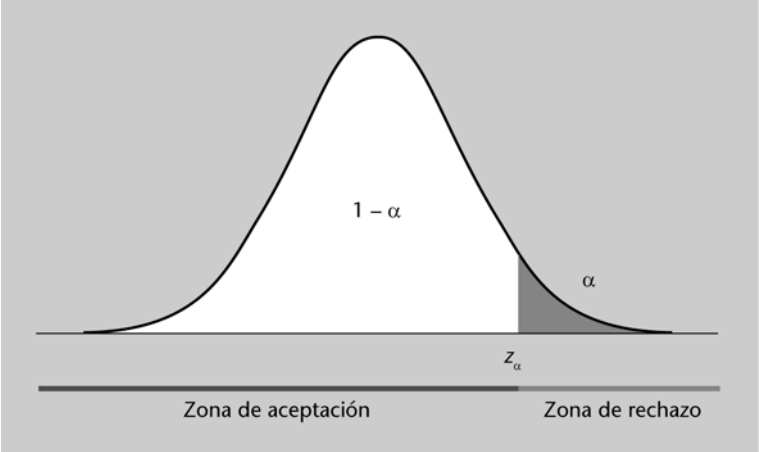
Para α= 0,05 encontramos que $-z_\alpha = -1{,}65$. En este contraste unilateral se dice que la probabilidad de la cola de la izquierda debe ser α.

- Aceptaremos H_0 si $\boldsymbol{Z} \geq -z_\alpha$
- Rechazaremos H_0 si $\boldsymbol{Z} < -z_\alpha$

Ejemplo 3. "Contraste unilateral superior"

La parte del gráfico (figura 17) sombreada en rojo corresponde a la zona en la que rechazamos la hipótesis nula. La zona sin sombrear corresponde a la región de aceptación de la hipótesis nula.

Figura 17. Gráfico que muestra la aceptación o no de la hipótesis nula en un contraste unilateral superior



Para $\alpha = 0{,}05$ encontramos que $z_\alpha = 1{,}65$. En este contraste unilateral se dice que la probabilidad de la cola de la derecha debe ser α.

- Aceptaremos H_0 si $\mathbf{Z} \leq z_\alpha$
- Rechazaremos H_0 si $\mathbf{Z} > z_\alpha$

El *p*-valor

Existe otro método para examinar el contraste de la hipótesis nula. Obsérvese que si se utiliza un nivel de significación bajo se reduce la probabilidad de rechazar una hipótesis nula verdadera. Eso modificaría la regla de decisión para que fuera menos probable que se rechazara la hipótesis nula, independientemente de que fuera verdadera o no. Evidentemente, cuanto menor es el nivel de significación al que se rechaza una hipótesis nula mayores son las dudas sobre su veracidad. En lugar de contrastar hipótesis a los niveles preasignados de significación, los investigadores a menudo hallan el nivel menor de significación al que se puede rechazar una hipótesis nula.

El *p*-valor es el menor nivel de significación al que puede rechazarse una hipótesis nula.

El criterio del *p*-valor es: rechazar H_0 si el *p*-valor < α.

Interpretación del *p*-valor

Se considera una muestra aleatoria de *n* observaciones procedentes de una población que sigue una distribución normal de media μ y desviación estándar σ y la media muestral calculada $\bar{x}$. Se ha contrastado la hipótesis nula $H_0 : \mu = \mu_0$ frente a la alternativa $H_1 : \mu > \mu_0$

El *p*-valor del contraste es:

$$p\text{-valor} = P\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_p \,|\, H_0 : \mu = \mu_0 \right)$$

donde z_p es el valor normal estándar correspondiente al menor valor de significación al que puede rechazarse la hipótesis nula. La mayoría de los programas informáticos estadísticos calculan el *p*-valor, este suministra más información sobre el contraste basándose en la media muestral observada, por lo que se utiliza frecuentemente en muchas aplicaciones estadísticas.

Ejemplo de aplicación del *p*-valor: un grupo editorial emite un periódico especializado en información económica. El director del periódico desea saber si el número medio de ejemplares diarios producidos y no vendidos es menor de 400. Para dar respuesta a esta pregunta, se toma una muestra formada por los resultados correspondientes a 172 días elegidos de forma aleatoria. La media de dicha muestra es de 407 ejemplares no vendidos, con una desviación estándar de 38.

Utilizando un nivel de significación de 0,05, realizad un contraste de hipótesis para responder razonadamente a la pregunta del director del periódico.

Solución:

1) Si se hace el contraste H_0: media poblacional = 400 contra H_1: media poblacional $\neq$ 400.

Primero se calcula el estadístico de contraste para decidir si rechazamos la hipótesis nula o no.

La desviación estándar de la muestra es: $\frac{S}{\sqrt{n}} = \frac{38}{\sqrt{172}} = 2,89.$

El estadístico será $z = \frac{407 - 400}{2,89} = 2,42$, este valor es una observación de una distribución $N(0,1)$.

En este caso, por ser un contraste bilateral se divide el nivel de significación α por igual entre las dos colas de la distribución normal. Por lo tanto, la probabilidad de que Z sea superior $z_{\alpha/2}$ o inferior a $-z_{\alpha/2}$ es α. En este caso, el

p-valor es la suma de las probabilidades de la cola superior y la cola inferior. El *p*-valor correspondiente al contraste de dos colas es:

$$p - valor = 2P\left(\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}\right);$$

$$P\left(Z > |2,42|\right) = P\left(Z > 2,42\right) + P\left(Z < -2,42\right) = 2 \cdot 0,00776 = 0,01552$$

Como 0,01552 es menor que el nivel de significación propuesto ($\alpha = 0,05$), se rechaza la hipótesis nula. No se puede afirmar que el número medio de ejemplares diarios producidos y no vendidos sea de 400. Se acepta que es distinto de 400.

2) Si se hace el contraste H_0: media poblacional = 400 contra H_1: media poblacional > 400, entonces el *p*-valor es la probabilidad "es la cola de la derecha":

$$p - valor = P\left(Z\right) > z_{\alpha}$$

$P\left(Z > 2,42\right) = 0,00776 < \alpha \Rightarrow$ Se rechaza la hipótesis nula.

Se acepta la hipótesis alternativa, por lo tanto, se acepta que el número medio de ejemplares diarios producidos y no vendidos es mayor de 400.

3) Si se realiza el contraste H_0: media poblacional = 400 contra H_1: media poblacional < 400, entonces el p-valor es la probabilidad "es la cola de la izquierda":

$$p - valor = P\left(Z\right) < z_{\alpha}$$

$P\left(Z < 2,42\right) = 1 - 0,00776 = 0,99224 > \alpha \Rightarrow$ No se puede rechazar la hipótesis nula.

Se rechazará la hipótesis alternativa, luego el número medio de ejemplares diarios producidos y no vendidos no es menor de 400.

Por tanto, a la vista de los resultados de los tres contrastes, la contestación a la pregunta del director sería:

"El número medio de ejemplares diarios producidos y no vendidos es mayor de 400".

Para calcular el *p*-valor se suele utilizar un software estadístico, como se verá en ejemplos resueltos con Minitab.

Otro procedimiento: para resolver contrastes bilaterales utilizando intervalos de confianza.

Ejemplo: supongamos que se plantea el siguiente contraste bilateral:

$$H_0: \mu = 280,\ H_1: \mu \neq 280$$

Para probar esta hipótesis con un nivel de significación $\alpha = 0{,}05$, el tamaño de la muestra es 36 y se determinó que la media muestral $\bar{x} = 278{,}5$ y la desviación estándar de las muestras $s = 12$. Sustituyendo estos resultados con $z_{0{,}025} = 1{,}96$, vemos que el intervalo de confianza del 95% para la media de la población es:

$$\bar{x} \pm 1{,}96\frac{s}{\sqrt{n}} \;;\; 278{,}5x \pm 1{,}96\frac{12}{\sqrt{36}} \;;\; 278{,}5 \pm 3{,}92$$

El intervalo será: (274,58; 282,42).

El resultado permite llegar a la conclusión de que, con un 95% de confianza, la media para la población está entre 274,58 y 282,42. Como el valor supuesto de la media de la población $\mu_0 = 280$ está en el intervalo de confianza, la conclusión del contraste es que no se puede rechazar la hipótesis nula, por tanto, aceptamos la hipótesis de que: H_0: $\mu = 280$.

Ejemplo de inferencia para una población (utilizando Minitab)

Una característica importante en el diseño de una página web es el tiempo que el usuario tardará en abrir la página, que se considera una variable normal. Con el objetivo de estimar el tiempo medio, se seleccionan al azar 101 páginas, entre las que ha diseñado una empresa el último año, obteniéndose los datos siguientes (en centésimas de segundo):

Tabla 2. Tiempo de descarga de páginas web

Tiempo de descarga	55	60	62	64	65	69
Número de páginas	11	21	26	19	15	9

Observación: se crea un fichero de datos en la hoja de Minitab, introduciendo los datos de forma unitaria.

a) Se comprueba que la colección de datos sigue una distribución aproximadamente normal.

b) Puede considerarse que el tiempo medio de apertura de las páginas de esta empresa es de 62 centésimas de segundo, con un nivel de confianza del 90%. ¿Qué resultado se obtiene? Razónese la respuesta del contraste a través del *p*-valor.

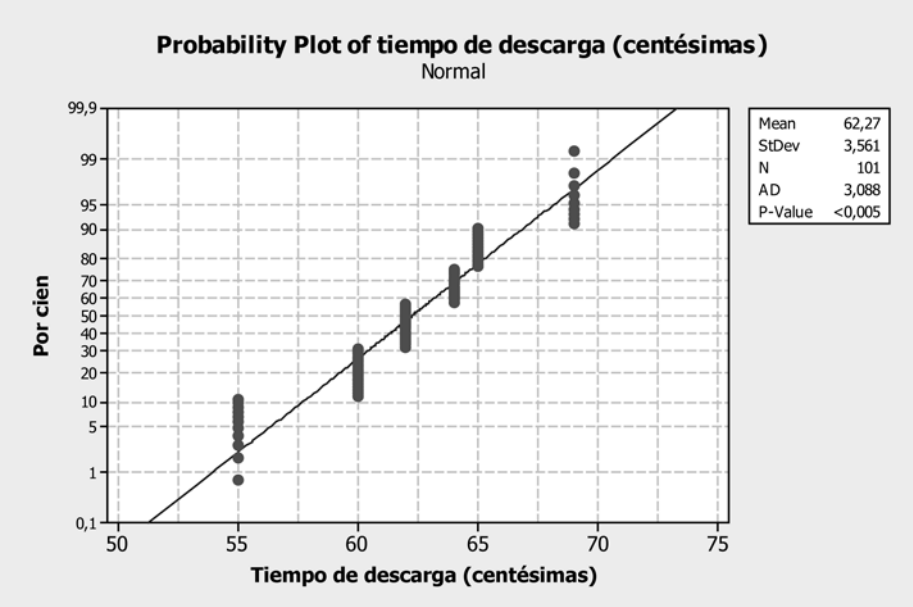
c) Calcúlese un intervalo de confianza a nivel del 90% para el tiempo medio y coméntese si el resultado obtenido es coherente con el resultado esperado.

d) Finalmente, se realizará el mismo contraste que en el apartado b), pero suponiendo esta vez que no se conoce la desviación estándar.

Solución:

a) Para comprobar la normalidad de los datos, se selecciona ***Stat*** > ***Basic Statistics*** > ***Normality Test***. Así se obtiene el gráfico de la figura 18.

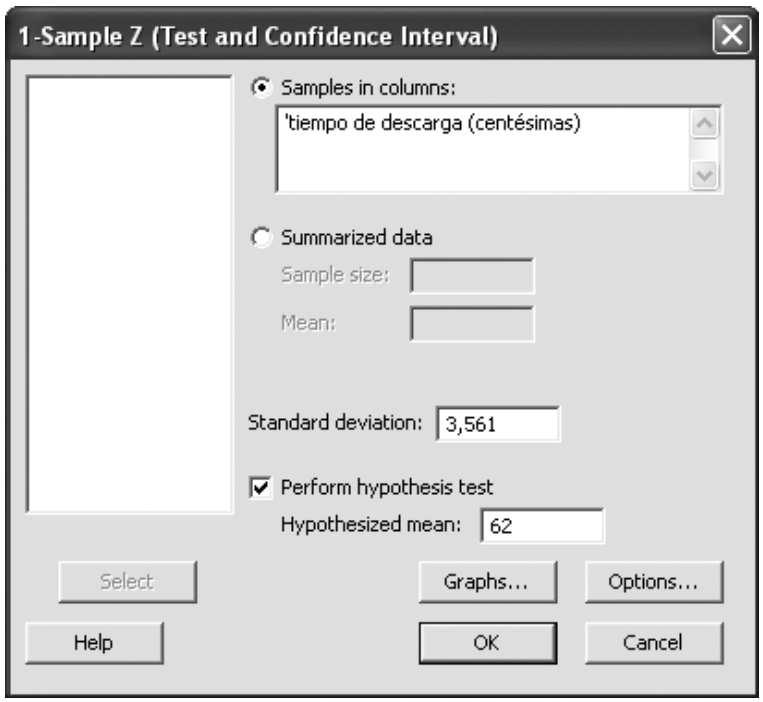
Figura 18. Gráfico de normalidad



Observando el *p*-valor se puede concluir que los datos siguen una distribución normal. Pudiendo asegurar que X sigue una distribución normal, la media muestral también sigue una distribución normal.

b) El contraste de hipótesis será H_0: $\mu = 62$ vs. H_1: $\mu \neq 62$. Es un contraste bilateral a un nivel de confianza de 0,90 (figura 19).

Figura 19. Pasos a seguir para realizar el contraste de hipótesis



Los resultados de Minitab son los que muestra la figura 20.

Figura 20. Resultados del contraste de hipótesis e intervalo de confianza del 90% (desviación típica población conocida)

One-Sample Z: tiempo de descarga (centésimas)

```
Test of mu = 62 vs not = 62
The assumed standard deviation = 3,561

Variable  N    Mean  StDev  SE Mean       90% CI          Z
tiempo  101  62,267  3,561    0,354  (61,685; 62,850)  0,75

Variable                      P
tiempo de descarga (cent.)  0,451
```

Se observa que el *p*-valor es 0,451, por lo tanto, como *p*-valor > $\alpha = 0{,}10$, no se puede rechazar la hipótesis nula, luego se acepta que el tiempo medio es de 62 centésimas por segundo.

c) El intervalo de confianza para el tiempo medio es (**61,685**; **62,850**), es coherente con los resultados esperados, ya que contiene al valor medio de 62 centésimas de segundo.

d) Análogamente se realiza el contraste de hipótesis para la media de la población con desviación típica desconocida, se selecciona ***Stat*** > ***Basic Statistic*** > ***1-Sample t***, obteniéndose los resultados de la figura 21.

Figura 21. Resultados del contraste de hipótesis e intervalo de confianza del 90% (desviación típica población desconocida)

One-Sample T: tiempo de descarga (centésimas)

```
Test of mu = 62 vs not = 62

Variable  N    Mean  StDev  SE Mean       90% CI          T
tiempo   101  62,267  3,561    0,354  (61,679; 62,856)  0,75

Variable                      P
tiempo de descarga (cent.)  0,452
```

El *p*-valor es 0,452 > 0,10, nos indica que se puede aceptar la hipótesis de que el tiempo medio es de 62 centésimas por segundo.

Continuando con el mismo ejemplo, se va a considerar que una página no es satisfactoria cuando tarde en ser descargada más de 68 centésimas. Los programadores afirman que el porcentaje de páginas para las que el tiempo de descarga no es satisfactorio no supera el 10%.

e) Se calculará un intervalo de confianza para la proporción de páginas no satisfactorias, a un nivel de confianza del 95%.

f) ¿Hay evidencias, al nivel 0,05, para rechazar la afirmación de los programadores? Se plantearán las hipótesis que se deben contrastar y se efectuará el contraste.

e) Para calcular el intervalo de confianza de la proporción de páginas no satisfactorias, a un nivel de confianza del 95%, se selecciona ***Stat*** > ***Basic Statistics*** >***1 Proportion*** (figura 22).

Observando la figura 23 de datos, se ve que únicamente hay 9 páginas que superan las 68 centésimas de segundo, o lo que es lo mismo, 9 páginas de las 101 se considera el tiempo de descarga no satisfactorio.

Figura 22. Pasos a seguir para obtener un intervalo de confianza del 95% para la proporción

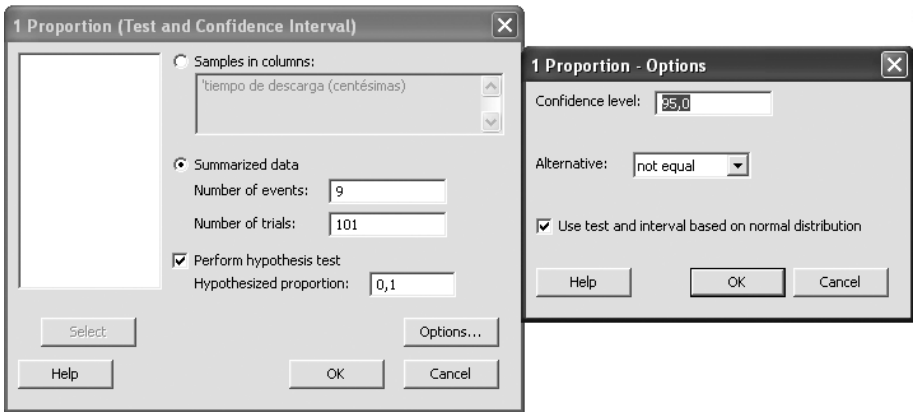


Figura 23. Resultados del intervalo de confianza del 95% para la proporción de páginas no satisfactorias

Test and CI for One Proportion

Test of p = 0,1 vs p not = 0,1

Sample	X	N	Sample p	95% CI	Z-Value	P-Value
1	9	101	0,089109	(0,033546; 0,144671)	-0,36	0,715

Using the normal approximation.

El intervalo de confianza obtenido con un nivel de confianza del 95% es **(0,033546; 0,144671)**.

f) Debemos plantear un contraste unilateral para la proporción de páginas no satisfactorias:

$\left.\begin{array}{l} H_0 : p = 0,1, \\ H_1 : p > 0,1, \end{array}\right\}$, donde p representa la proporción de páginas para las que el tiempo de descarga no es satisfactorio (figura 24).

Figura 24. Pasos a seguir para realizar el contraste de hipótesis

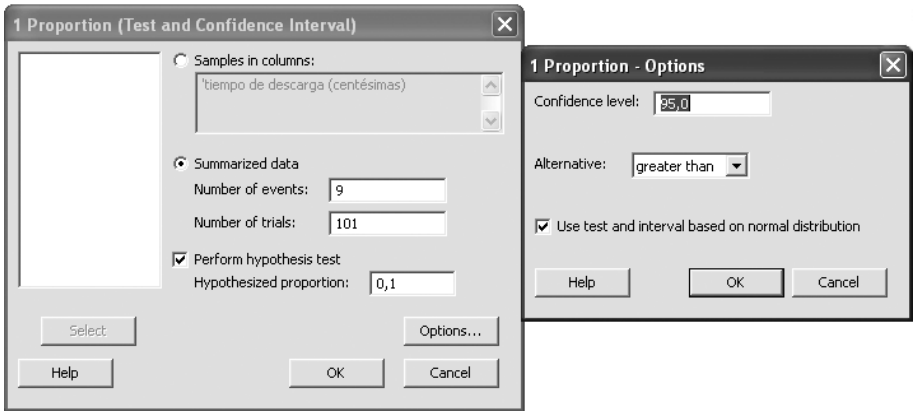


Figura 25. Resultados del contraste de hipótesis para la proporción de páginas

Test and CI for One Proportion

Test of p = 0,1 vs p > 0,1

Sample	X	N	Sample p	95% Lower Bound	Z-Value	P-Value
1	9	101	0,089109	0,042479	-0,36	0,642

Using the normal approximation.

Según se muestra en la figura 25, el *p*-valor del contraste vale lo siguiente: *p*-valor = 0,642. Como es mayor que 0,05, se acepta la hipótesis nula, luego se acepta la afirmación de los programadores de que el porcentaje de páginas no supera el 10%.

Resumen

En este módulo se presentan las distribuciones muestrales. Se analiza cómo seleccionar una muestra aleatoria simple, cómo se pueden emplear los datos obtenidos con ella para desarrollar estimaciones puntuales de los parámetros de población. La distribución de probabilidad de estas variables aleatorias se llama *distribución muestral*. En particular, se describen las distribuciones de la media de la muestra $\bar{x}$, de la proporción muestral $\hat{p}$ y de la varianza muestral s^2. Después de desarrollar las fórmulas de la desviación típica o error estándar para esos estimadores, se indica que el teorema central del límite es la base para usar una distribución normal de probabilidades y aproximar esas distribuciones muestrales en el caso de muestra grande.

Además, también se desarrollan estimaciones de intervalos de confianza de parámetros de una población. En este módulo se han utilizado la distribución Z normal estándar, la t de Student y la chi-cuadrado χ^2 para construir intervalos de confianza. Se determina el tamaño de muestra necesario para que los estimadores de intervalo de μ y de p tengan un nivel especificado de precisión.

Finalmente, en este módulo se ha presentado la metodología para realizar contrastes clásicos de hipótesis, comenzando con los argumentos para tomar decisiones en condiciones de incertidumbre. Las decisiones se toman rechazando una hipótesis nula si hay pruebas contundentes a favor de la hipótesis alternativa. Pueden cometerse dos tipos de error: un error de tipo I, que se comete cuando se rechaza la hipótesis nula, cuando es verdadera, y un error de tipo II, que se comete cuando no se rechaza la hipótesis nula, cuando no es verdadera, presentando diversos métodos y reglas de decisión específicos para realizar contrastes. La regla de rechazo para todos los procedimientos implica comparar el valor del estadístico con un valor crítico y también utilizando el *p*-valor para pruebas de hipótesis, la regla es rechazar la hipótesis nula siempre que el *p*-valor sea menor que α.

Ejercicios de autoevaluación

1) Una biblioteca presta un promedio de μ =320 libros por día, con desviación estándar σ = 75 libros. Se tiene una muestra de 30 días de funcionamiento, y $\bar{x}$ es la cantidad de la media de la muestra de libros prestados en un día.

a) Presente la distribución muestral de $\bar{x}$.
b) ¿Cuál es la distribución estándar de $\bar{x}$?
c) ¿Cuál es la probabilidad de que la media de una muestra de 30 días sea entre 300 y 400 libros?
d) ¿Cuál es la probabilidad de que la media de una muestra sea de 325 o más prestamos diariamente?

2) Un investigador informa los resultados de una encuesta diciendo que el error estándar de la media es de 20.La desviación estándar de la población es de 500.

a) ¿De qué tamaño fue la muestra que se usó en esta encuesta?
b) ¿Cuál es la probabilidad de que el error estimado quede a ± 25 o menos de la media de la población?

3) Cada curso escolar, una prestigiosa universidad oferta becas a sus estudiantes para ampliar estudios en el extranjero. De la experiencia recogida en anteriores convocatorias, se observa que las calificaciones medias de los expedientes aspirantes a obtener una beca se distribuyen según una normal de media 6,9 puntos y desviación estándar 0,7 puntos. Para entender la aplicación del teorema central del límite, generar con Minitab 50 muestras aleatorias de 100 observaciones cada una, que corresponden a la población normal anterior $N(6{,}9,\ 0{,}7)$.

a) Calcular en una nueva columna la media de las 50 muestras anteriores.
b) Comentar los resultados haciendo referencia al teorema central del límite.
c) Realiza el *dotplot* asociado a una de las muestras.
d) Compara estos resultados con la media de la población, y el valor de la desviación estándar de la media muestral con la desviación estándar de la población y explica la relación entre ambos valores.

4) Un estudio previo nos dice que el servicio de préstamo diario de libros de las bibliotecas de una ciudad sigue una distribución normal con una media de 300 ejemplares prestados, con una desviación estándar de 10. Una inspección quiere verificar si estos datos son correctos. Para hacerlo, coge una muestra de los préstamos diarios de 10 bibliotecas y obtiene una media de 285 ejemplares prestados.

a) ¿Cuál es la probabilidad de que si la media es verdaderamente de 300 ejemplares prestados se obtenga una media de préstamos igual o inferior a los 285 ejemplares en las 10 bibliotecas que componen la muestra?
b) Determinar un intervalo de confianza del 90% para la media de préstamos teniendo en cuenta los datos de la muestra.
c) ¿Qué decisión lógica debería tomar el inspector?

5) En la página web de una editorial aparecen dos números de teléfono. Hemos comprobado, después de analizar 400 llamadas del teléfono, que el intervalo entre llamadas tiene una varianza de 2.

Suponiendo normalidad, indicad si podemos considerar, a un nivel de confianza del 90%, que la varianza del intervalo entre llamadas del primer número es inferior a 1,7.

6) El responsable de comunicaciones de un centro de documentación afirma que la media del tiempo de transferencia de un fichero de tamaño 2Mb es superior a 30 segundos. Para comprobar esta afirmación se tomó una muestra de tiempos de transferencia de 12 ficheros de 2Mb, obteniendo que la media y la desviación estándar muestrales valen $\bar{x} = 30{,}2,\ s = 1{,}833$ (en segundos).

a) Suponiendo que el tiempo de transferencia se distribuye normalmente a partir de los datos muestrales obtenidos, ¿tenemos suficientes evidencias para aceptar la afirmación del responsable? (Tomad $\alpha = 0{,}05$). Encontrad el *p*-valor del contraste.

Si además de disponer de estas observaciones nos hubiesen dado como información adicional (obtenida de experiencias previas) que la varianza del tiempo de transferencia es de $\sigma^2 = 9{,}2$ segundos^2, ¿hubiéramos llegado a la misma conclusión que en el apartado anterior? Encontrad el *p*-valor del contraste (Tomad $\alpha = 0{,}05$).

Solucionario

1)

a) Normal con $\mu = 320$ y desviación típica 13,69
b) 13,69
c) 0,8558
d) 0,3557

2)

a) 625
b) 0,7888

3)

De esta manera obtenemos las 50 muestras con 100 observaciones cada una.

Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
1	6,02255	7,51039	7,56941	6,73114	6,89504	7,27531	6,49352	7,07449	7,72782	8,01131	6,51824	7,31135	5,70901	7,18919	6,23128	7,986(
2	5,82796	7,23382	8,63014	7,67923	6,82915	7,50193	6,80251	6,83211	6,08951	7,08834	7,44647	7,36977	6,94125	6,82865	8,08839	6,073(
3	5,89460	6,19736	7,48944	7,16782	7,30042	6,23994	5,55889	5,92865	6,65836	7,20623	6,29604	6,50873	7,42008	6,63113	7,64693	7,277
4	7,91373	6,63874	6,88954	7,74717	7,75320	6,93165	6,16663	5,94975	7,84482	7,35052	6,62743	7,61428	6,31124	6,59273	6,66213	5,852(
5	6,96531	6,54096	7,83251	6,71421	5,76998	6,48673	6,47134	6,87821	7,57085	6,96200	7,26595	7,65586	7,88420	8,22074	7,28688	5,476(
6	6,38379	8,14155	5,26726	7,28067	6,89716	6,26790	7,12585	5,82249	6,81010	7,49986	6,99193	5,76598	7,32813	5,52918	7,05477	5,853
7	6,98679	7,50650	7,06138	7,30100	6,61602	7,20820	8,00421	6,49912	8,02260	7,28351	5,36631	6,93591	7,84171	6,61069	6,50505	6,263
8	6,28190	8,55299	8,20722	7,29348	6,51880	7,74509	6,95879	7,46706	7,55387	7,82200	7,22971	7,12365	7,03669	6,01653	8,28156	6,957(
9	5,85308	6,42670	7,24762	7,50398	6,46682	7,32013	6,42494	5,69820	6,13376	6,79857	7,15053	6,40466	6,38444	6,24852	6,05964	6,250

Project Man...

Current Worksheet: Worksheet 1 18:19

a) En la columna C101 se muestran las medias maestrales.

Worksheet 1 ***

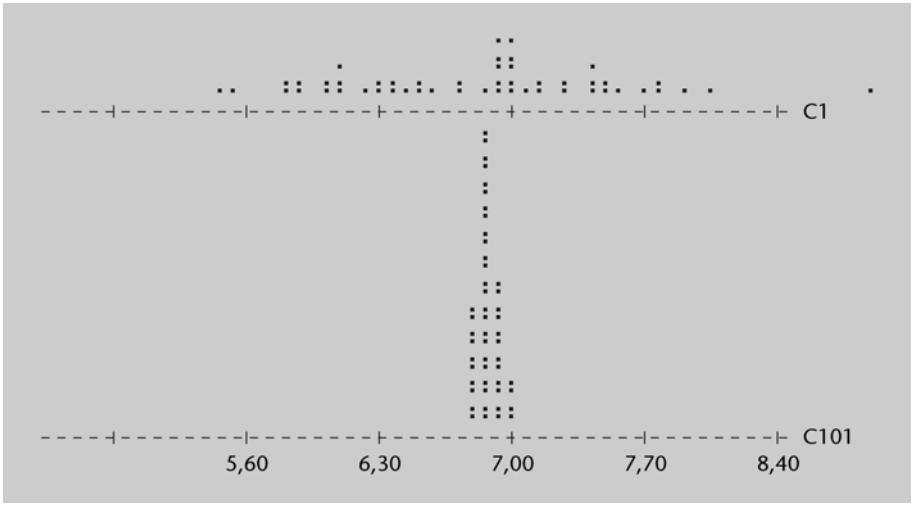
↓	C98	C99	C100	C101
1	5,49495	7,33352	6,49861	6,84032
2	7,27693	5,57569	5,13667	6,90402
3	6,86654	6,96175	7,56928	6,87066
4	7,85956	6,10293	7,09721	6,87309
5	7,68902	4,92515	7,13723	6,89323
6	6,51449	5,78576	7,18556	6,81035
7	7,46093	6,67470	5,89898	6,87570
8	5,41243	7,05719	7,60637	6,98587
9	8,66592	7,03988	7,55059	6,85254

Project Man...

Current Worksheet: Worksheet 1

b)

***dotplot*: C1; C101**



c) Tras haber generado 50 muestras de datos provenientes de una distribución normal de media 6,9 y desviación estándar 0,7, observamos que el primer *dotplot* parece corresponder a una distribución normal.

Asimismo, el segundo *dotplot* corresponde a la distribución de las medias de las muestras y también corresponde a una distribución normal.

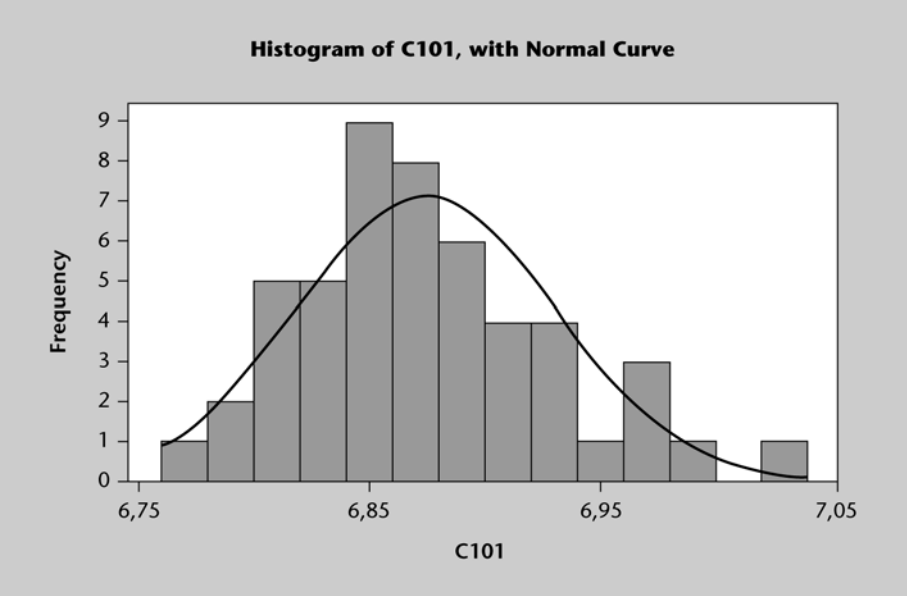
Esto indica que las medias de estas muestras siguen una distribución normal. Esta propiedad es la que enuncia el TCL, sea cual sea la distribución de los datos, la media muestral (con un tamaño de muestra *n* suficientemente grande) de una colección de datos sigue una distribución normal.

d) Estudiaremos la distribución de estas medias muestrales:

Descriptive Statistics: C101

Variable	N	Mean	Median	TrMean	StDev	SE Mean
C101	50	6,9035	6,8914	6,9016	0,0660	0,0093

Variable	Minimum	Maximum	Q1	Q3
C101	6,7837	7,0412	6,8507	6,9603



El histograma de frecuencias se aproxima a la curva normal, es simétrica.

La media muestral coincide con la media de la población, $\mu = \overline{x} = 6,9$.

La desviación estándar de la media muestral será aproximadamente el error estándar.

Si la variable tiene desviación estándar conocida *s* (en la población), el error estándar se puede calcular como:

$$\frac{\sigma}{\sqrt{n}}$$

Como consecuencia, podemos decir que la media muestral sigue una distribución normal $N(\mu, \frac{\sigma}{\sqrt{n}})$, que se puede aproximar a una $N(0,1)$, realizando un cambio de variable (tipificación): $Z = \frac{X - \mu}{\sigma / \sqrt{n}}$.

4)

a) Si estandarizamos la puntuación de 285, resulta un valor *z* de –4,74, lo que supone (mirando las tablas de la normal) aproximadamente que el 0% es la probabilidad de obtener dicha puntuación.

$$p(X < 285) = p(Z < \frac{285 - 300}{10 / \sqrt{10}}) = p(Z < -4,74) \approx 0$$

b) El intervalo de confianza es:

$$z_{\alpha/2,} = 1{,}64$$

$$I = 285 \pm 1{,}64 \cdot \frac{10}{\sqrt{10}} = 285 \pm 5{,}17 \qquad [279{,}83;\ 290{,}19]$$

c) 300 está fuera del intervalo y, por lo tanto, con un nivel de confianza del 90%, podremos afirmar que la media no llega a 300 ejemplares, sino que está por debajo.

5) La hipótesis nula es $\sigma^2 = 1{,}7$ y la alternativa es $\sigma^2 < 1{,}7$.

El estadístico de contraste es: $\chi^2 = \frac{(400-1)s^2}{1{,}7}$, donde s^2 es la varianza muestral. Entonces $\chi^2 = 469{,}412$ y su distribución es la de χ^2 la con $400-1 = 399$ grados de libertad.

En este caso, el *p*-valor vale $P(\chi^2 < 469{,}412) = 0{,}991406$ y por lo tanto, no rechazamos la hipótesis nula: no podemos afirmar que sea inferior a 1,7. El valor crítico es 363,253.

6)

a) Hemos de hacer el contraste de una media con varianza desconocida. Las hipótesis nula y alternativa son: $\begin{matrix} H_0 : \mu = 30 \\ H_1 : \mu > 30 \end{matrix}$, donde μ representa la media del tiempo de transferencia de un fichero de 2Mb. El estadístico de contraste es $t = 0{,}378$. El valor crítico valdrá: $t_{0.05,11} = 1{,}80$. Como que $t < t_{0.05,11}$, aceptamos la hipótesis nula y concluimos que la afirmación del responsable es cierta. Si quisiéramos hallar el *p*-valor, éste sería: $p = p(t_{11} > 0{,}378) \approx 0{,}36$. Como es un *p*-valor alto, mayor que 0,05, aceptamos la hipótesis nula tal y como hemos hecho antes.

b) Hemos de hacer el contraste de una media con varianza conocida.

La hipótesis nula y la alternativa son: $\begin{matrix} H_0 : \mu = 30 \\ H_1 : \mu > 30 \end{matrix}$, donde μ representa la media del tiempo de transferencia de un fichero de 2Mb.

El estadístico de contraste es: $z = \frac{\bar{x} - 30}{\sigma/\sqrt{12}}$, donde $\bar{x}$ es la media muestral y σ es la desviación estándar poblacional. La distribución de *z* es la de una normal $N(0,1)$. La media y la desviación estándar poblacionales valen respectivamente: $\bar{x} = 30{,}2$, $\sigma = \sqrt{9{,}2} \approx 3{,}03$. El valor del estadístico de contraste es: $z \approx 0{,}228$.

El valor crítico valdrá: $z_{0{,}05} \approx 1{,}645$. Como $z < z_{0{,}05}$, volvemos a aceptar la hipótesis nula y concluimos que la afirmación del responsable no es cierta. Si quisiéramos hallar el *p*-valor, éste sería: $p = p(z > 0{,}228) \approx 0{,}41$. Como es un *p*-valor alto, mayor que 0,05, aceptamos la hipótesis nula como hemos hecho anteriormente. Por tanto, hemos llegado a la misma conclusión que en el apartado anterior.