BLOQUE I

Estimación de parámetros y contraste de hipótesis en los diseños de una muestra

Soluciones a los ejercicios de los capítulos 1 y 2

1

SITUACIÓN 1. Un sociólogo conoce, por investigaciones anteriores, que el sueldo medio de la población de hombres para un determinado trabajo *T*, se distribuye normalmente con media igual a 1500 euros mensuales y una desviación típica de 180 euros. Extrae una muestra aleatoria de 81 mujeres que desempeñan el mismo trabajo *T*, observando que el sueldo medio en dicha muestra es igual a 1400 euros mensuales con una desviación típica de 210 euros. Establece un nivel de confianza del 99% y quiere comprobar:

1.1. Si la variabilidad del sueldo de las mujeres es mayor que el que presentan los hombres para el mismo trabajo.

1.2. Si el sueldo medio de las mujeres es inferior al de los hombres y la potencia estadística de este contraste.

SOLUCIÓN:

1.1. Para el primer contraste de hipótesis, seguiremos los siguientes pasos:

CONDICIONES Y SUPUESTOS: El diseño de este trabajo utiliza una muestra aleatoria de 81 mujeres en el que la variable de análisis, el sueldo, es de naturaleza cuantitativa que se distribuye normalmente.

HIPÓTESIS: El investigador quiere comprobar si la variabilidad del sueldo de las mujeres es mayor que el que presentan los hombres para el mismo trabajo para lo que plantea un contraste unilateral derecho con las siguientes hipótesis:

$$H_0 : \sigma^2 \leq 180 \qquad\qquad H_1 : \sigma^2 > 180$$

ESTADÍSTICO DE CONTRASTE: El estadístico chi-cuadrado toma el siguiente valor:

$$\chi^2 = \frac{n \cdot S_n^2}{\sigma_0^2} = \frac{81 \cdot 210^2}{180^2} = 110,25$$

REGLA DE DECISIÓN: En una distribución chi-cuadrado hay que buscar el valor crítico correspondiente a un contraste unilateral derecho con nivel de confianza del 99% y con 80 grados de libertad. El valor que deja por debajo una probabilidad de 0,99 es 112,328.

g.l.	Probabilidad									
	0,005	**0,010**	**0,025**	**0,050**	**0,100**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	14,8603
50	27,9907	29,7067	32,3574	34,7643	37,6886	63,1671	67,5048	71,4202	76,1539	79,4900
60	35,5345	37,4849	40,4817	43,1880	46,4589	74,3970	79,0819	83,2977	88,3794	91,9517
70	43,2752	45,4417	48,7576	51,7393	55,3289	85,5270	90,5312	95,0232	100,4252	104,2149
80	51,1719	53,5401	57,1532	60,3915	64,2778	96,5782	101,8795	106,6286	112,3288	116,3211
90	59,1963	61,7541	65,6466	69,1260	73,2911	107,5650	113,1453	118,1359	124,1163	128,2989
100	67,3276	70,0649	74,2219	77,9295	82,3581	118,4980	124,3421	129,5612	135,8067	140,1695

Tabla 1: *Representación parcial de la tabla de la distribución chi-cuadrado.*

CONCLUSIÓN: Como el valor del estadístico de contraste obtenido no supera el valor crítico (110,25 < 112,3288), con un nivel de confianza del 99% no podemos rechazar la hipótesis nula.

A la misma conclusión llegaríamos calculando el nivel crítico p correspondiente al estadístico de contraste obtenido de 110,25, que en una distribución chi-cuadrado con 80 gl es un valor comprendido entre 106,63 y 112,33. Estos valores dejan por encima una probabilidad de 0,025 y 0,01, respectivamente, por lo que el nivel crítico es mayor de 0,01 ($p > 0,01$).

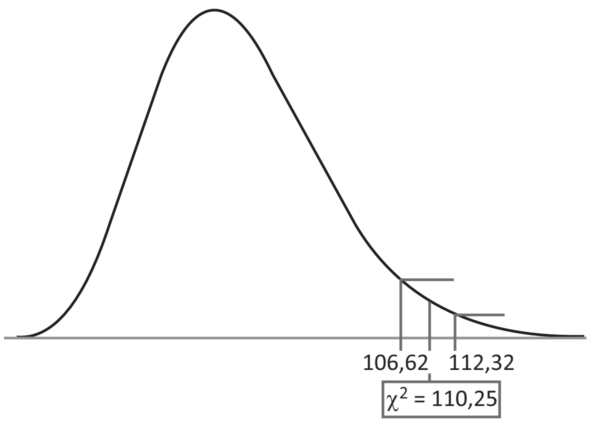


Figura 1: *Representación gráfica del nivel crítico del estadístico chi-cuadrado =110,5.*

INTERPRETACIÓN: La variabilidad del sueldo de las mujeres no es significativamente mayor que el de los hombres (χ^2 = 110,25; p > 0,01) con un nivel de confianza del 99%.

1.2. Para el segundo contraste de hipótesis, sobre el sueldo medio, seguiremos los mismos pasos:

CONDICIONES Y SUPUESTOS. El diseño de este trabajo utiliza una muestra aleatoria de 81 mujeres en el que la variable de análisis, el sueldo en la muestra de mujeres, es de naturaleza cuantitativa que se distribuye normalmente en la población con varianza conocida e igual a la de los hombres, como hemos contrastado en el paso anterior.

HIPÓTESIS: El investigador quiere comprobar si el sueldo medio que perciben las mujeres es inferior a 1.500 euros mensuales que, por investigaciones anteriores, corresponde al que reciben los hombres. Formula, por tanto, un contraste unilateral izquierdo.

$$H_0 : \mu \geq 1.500 \qquad\qquad H_1 : \mu < 1.500$$

ESTADÍSTICO DE CONTRASTE: La distribución muestral de la media de muestras extraídas de una población normal con varianza conocida es normal, por lo que el estadístico de contraste es la *Z*:

$$Z = \frac{\overline{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1400 - 1500}{\frac{180}{\sqrt{81}}} = -5$$

REGLA DE DECISIÓN: En la tabla de la distribución normal tipificada de puntuaciones *Z* buscamos el valor crítico para un nivel de confianza del 99% para un contraste unilateral izquierdo, es decir, el valor de *Z* que deja por debajo una probabilidad de 0,01, y es –2,33.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,50	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,40	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,30	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,20	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,10	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,00	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,90	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,80	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,70	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,60	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,50	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,40	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,30	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,20	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,10	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,00	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,90	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,80	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,70	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,60	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,50	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,40	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,30	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823

Tabla 2: *Reproducción parcial de la tabla de la distribución normal tipificada.*

CONCLUSIÓN: Como el estadístico de contraste (*Z* = –5) supera la máxima diferencia que cabe esperar por simple azar (el valor crítico –2,33 para un nivel de confianza del 99%), rechazamos la hipótesis nula. A la misma conclusión llegamos comparando el nivel crítico p con el nivel de significación «alfa». Obsérvese que en la distribución normal tipificada el valor más extremo que podemos ver es –3,59 que deja por debajo una probabilidad de 0,0002. Como el estadístico de contraste *Z* = –5 es aún más extremo, el nivel crítico p es menor de 0,0002 que a su vez es menor que el nivel de significación «alfa» de 0,01.

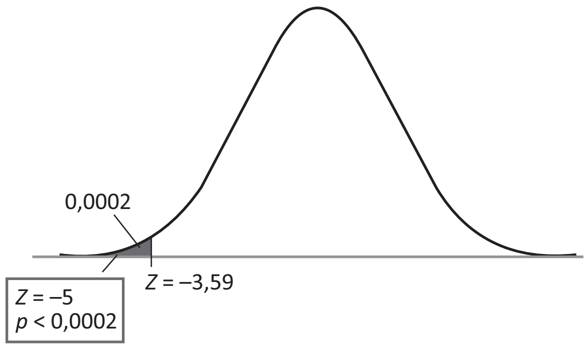


Figura 2: *Nivel crítico p del estadístico Z = –5*

INTERPRETACIÓN: Con un nivel de confianza del 99% se puede afirmar que el sueldo medio de las mujeres es menor de 1.500 euros mensuales que es el que tienen los hombres ($Z = -5$; $p < 0{,}01$).

POTENCIA DEL CONTRASTE: Se calcula para la H_0: $\mu = 1.500$ frente a la hipótesis alternativa H_1: $\mu = 1.400$.

Fijamos un nivel de confianza del 99% por lo que, suponiendo que la H_0 es verdadera, cualquier muestra con una media cuya puntuación típica sea menor o igual que $Z = -2{,}33$ nos llevaría a rechazar la H_0. El valor de esta media es:

$$Z = \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow -2{,}33 = \frac{\bar{Y} - 1500}{20}$$

$$\bar{Y} = -2{,}33 \cdot 20 + 1500 = 1453{,}4 \textit{ euros}$$

En la distribución correspondiente a la H_1, a la puntuación 1453 le corresponde una puntuación típica de:

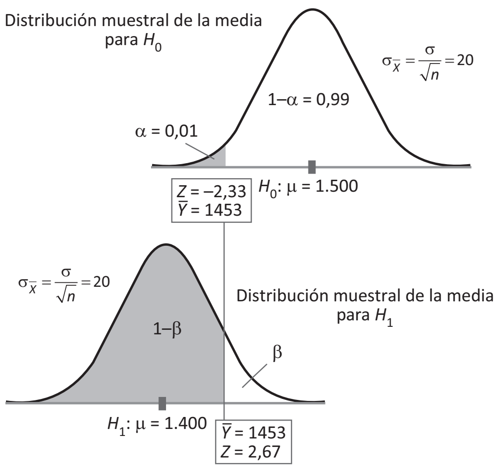


Figura 3: *Representación gráfica de la potencia del contraste.*

$$Z = \frac{\bar{Y} - \mu_1}{\frac{\sigma}{\sqrt{n}}} = \frac{1453{,}4 - 1400}{\frac{180}{\sqrt{81}}} = \frac{53{,}4}{20} = 2{,}67$$

La potencia del contraste es la probabilidad de rechazar una H_0 que es falsa. Y esto ocurrirá cuando el estadístico de contraste se sitúe en la zona representada por 1–«beta» que corresponde a la probabilidad de obtener puntuaciones menores o iguales a Z = 2,67. Buscamos esta probabilidad en la tabla de la distribución normal y es 0,9962. Su complementario «beta» es la probabilidad de cometer un error de tipo II o probabilidad de rechazar una H_0 que es verdadera y vale 1 – 0,9962 = 0,0038.

SITUACIÓN 2. Un investigador social tiene la hipótesis de que la edad media de inicio en el consumo de alcohol de los jóvenes de una determinada Comunidad es más tardía que la media de la población general, establecida en 13 años. Para contrastar su hipótesis selecciona una muestra de 25 jóvenes de su Comunidad, encontrando que la edad media es de 14 años con una desviación típica insesgada de 2,8. Sabemos que la variable, edad de inicio en el consumo de alcohol, se distribuye normalmente en la población.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS: Se utiliza una muestra extraída de una población con distribución normal y varianza desconocida. Se analiza la variable «edad de inicio del consumo de alcohol» en los jóvenes que es una variable medida con escala de razón. Aunque la muestra es pequeña como la variable tiene distribución normal en la población podremos utilizar un contraste paramétrico.

HIPÓTESIS: El investigador quiere comprobar que la edad media de inicio en su Comunidad es más tardía (por tanto, mayor) que la media de la población general establecida en 13 años. Para ello, formulará un contraste unilateral derecho en el que la hipótesis nula que establece que no hay diferencia significativas entre la edad media en su Comunidad y la media de la población general

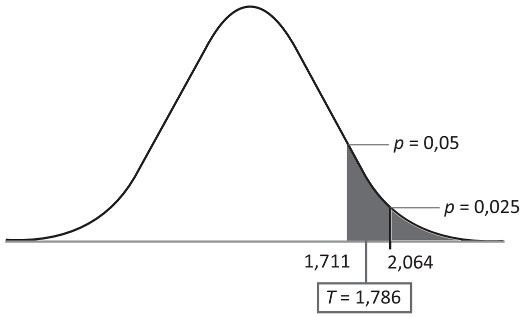
$$H_0 : \mu_{edad} \leq 13 \text{ } años \qquad H_1 : \mu_{edad} > 13 \text{ } años$$

ESTADÍSTICO DE CONTRASTE: Siendo los datos provenientes de una población con distribución normal y varianza desconocida, la distribución muestral de la media es la distribución t de Student con n–1 = 24 grados de libertad.

Conocemos la desviación típica insesgada que utilizaremos como estimador de la desviación típica poblacional en la siguiente expresión:

$$T=\frac{\bar{Y}-\mu_0}{\hat{\sigma}/\sqrt{n}} \rightarrow \begin{cases} T=\dfrac{\bar{Y}-\mu_0}{S_{n-1}/\sqrt{n}}=\dfrac{14-13}{2{,}8/\sqrt{25}}=1{,}786 \\[2ex] T=\dfrac{\bar{Y}-\mu_0}{S_n/\sqrt{n-1}} \end{cases}$$

REGLA DE DECISIÓN: En la distribución t de Student con 24 gl y para un contraste unilateral derecho, buscamos el nivel crítico *p* asociado al estadístico T = 1,786. Los valores más próximos a este valor del estadístico de contraste son 1,711 y 2,064 a los que corresponden unos niveles críticos *p* de 0,05 y 0,025 respectivamente. Al ser el estadístico de contraste 1,786 un valor comprendido entre 1,711 y 2,065 su nivel crítico p es un valor comprendido entre 0,025 y 0,05 ($p < 0{,}05$ y $p > 0{,}025$). Por otra parte, los valores críticos de la distribución muestral que delimitan la zona de rechazo y de mantenimiento de la H_0 son 1,711 y 2,492 para niveles de confianza de 0,95 y 0,99 respectivamente, que son los utilizados habitualmente en el campo de la psicología.



g.l.	Probabilidad											
	0,550	**0,600**	**0,650**	**0,700**	**0,750**	**0,800**	**0,850**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	$\boxed{1{,}711}$	$\boxed{2{,}064}$	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763

Tabla 3: *Representación gráfica del nivel crítico del estadístico t = 1,786 y su localización en la tabla de la distribución t.*

CONCLUSIÓN: Se rechazaría la Ho con un nivel de confianza del 95% pero no se podría rechazar con un nivel de confianza del 99%. Con otras palabras, el resultado obtenido es significativo con α = 0,05 pero no con α = 0,01

INTERPRETACIÓN: A partir de los resultados de esta investigación, la edad media de inicio de alcohol de los jóvenes de esta Comunidad es más tardía que la media general con un nivel de confianza del 95%.

SITUACIÓN 3. La empresa SND's de sondeos electorales ha pronosticado que el nivel de apoyo que recibirá el partido X en las próximas elecciones será del 40%. Desde el propio partido X se promueve un nuevo sondeo con el fin de contrastar la veracidad de esta afirmación. Para ello, selecciona una muestra aleatoria de 400 personas, con derecho a voto, de los cuales 128 manifiestan su intención de votar al partido X.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS: Se utiliza una muestra de 400 personas y la variable dependiente es el «apoyo que recibirá el partido X en las próximas elecciones». Es una variable dicotómica con distribución binomial. La muestra es grande por lo que esta distribución se aproximará a la normal.

HIPÓTESIS: En la muestra utilizada en el sondeo encargado por el partido X, la proporción de personas que les votarían es del 32% (128/400 = 0,32). A partir de esta evidencia mostrada en los datos, se quiere contrastar si «el nivel de apoyos será del 40%». Como no se marca si la diferencia será mayor o menor que ese valor se plantea un contraste bilateral, con las siguientes hipótesis:

$$H_0 : \pi = 0,40 \qquad\qquad H_1 : \pi \neq 0,40$$

ESTADÍSTICO DE CONTRASTE: Por aproximación de la distribución binomial a la normal, utilizaremos el estadístico Z

$$Z = \frac{P - \pi_0}{\sigma_p}$$

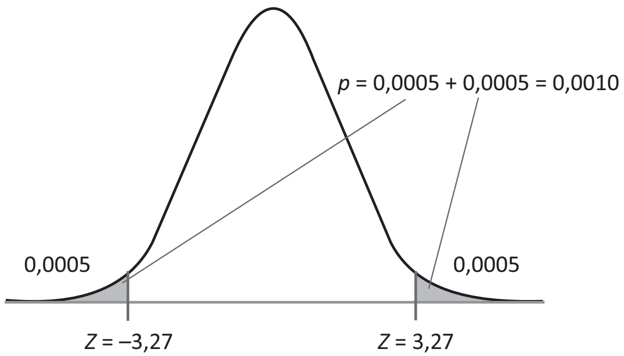
Siendo el error típico de la proporción:

$$\sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0,40(1-0,40)}{400}} = 0,0245$$

Y el estadístico:

$$Z = \frac{P - \pi_0}{\sigma_p} = \frac{0,32 - 0,40}{0,0245} = -3,266$$

REGLA DE DECISIÓN: En la distribución normal y para un contraste bilateral, buscamos el nivel crítico *p* asociado al estadístico $Z = -3,266$. En la tabla el valor más próximo es $Z = -3,27$, que deja por debajo una probabilidad de 0,0005 en uno de los extremos de la distribución. Como se trata de un contraste de bilateral (de dos lados) tendremos que sumar también la probabilidad de $P(Z > 3,279) = 0,0005$ y el nivel crítico *p* es la suma de las probabilidades de los dos extremos de la distribución: $p = 0,0010$.



z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,50	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,40	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,30	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,20	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,10	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,00	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,90	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,80	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,70	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026

Tabla 4: *Reproducción parcial de la tabla de la distribución normal N(0;1) para encontrar el nivel crítico p, P(Z < = –3,27) representado en la figura superior.*

CONCLUSIÓN: Aunque el enunciado de este ejercicio no establece el nivel de confianza, recurriremos a los utilizados habitualmente, que son el 95% o el 99%. Al ser el nivel crítico $p = 0,0010\,(0,0005 + 0,0005)$ menor que el nivel de significación 0,01 y menor que 0,05, rechazamos la H_0 con un nivel de confianza tanto del 95% como del 99%. A la misma conclusión llegamos comparando el valor del estadístico de contraste con los valores críticos que son ±1,96 y ±2,58 correspondientes a los niveles de confianza del 95% o del 99%, respectivamente, para un contraste bilateral.

INTERPRETACIÓN: El nivel de apoyo que recibirá el partido X en las próximas elecciones será significativamente distinto del 40% ($Z = -3,266$; $p < 0,001$).

INTERVALO DE CONFIANZA: A esta misma conclusión se llega utilizando el intervalo de confianza para estimar, a partir de la proporción observada en la muestra, la proporción poblacional de votantes que tendrá el partido.

Con un nivel de confianza del 95%, el error máximo de estimación es:

$$E_{max} = Z_{1-\alpha/2}\sqrt{\frac{p\cdot(1-p)}{n}} = 1,96\cdot\sqrt{\frac{0,32\cdot(1-0,32)}{400}} = 0,0457$$

Y el intervalo de confianza es:

$$L_{inf} = p - E_{max} = 0,32 - 0,0457 = 0,274$$
$$L_{sup} = p + E_{max} = 0,32 + 0,0457 = 0,366$$

De acuerdo con esta estimación, la proporción poblacional de votos para el partido es un valor comprendido entre el 27,4% y el 36,6%. Como este intervalo no contiene el valor planteado en la H_0 de 0,40 no podemos asumir este valor y se rechaza H_0. Compruebe ahora que con un nivel de confianza del 99% también se rechazaría H_0, ya que el intervalo de confianza estima como proporción apoyos al partido X un valor comprendida entre 0,26 y 0,38 que no incluye el valor 0,40 planteado en H_0.

TAMAÑO DE LA MUESTRA: En el punto anterior se ha visto que el error máximo de estimación, con un nivel de confianza del 95%, era de 0,0457. Si quisiéramos fijar este error máximo de estimación en dos puntos porcentuales (0,02) ¿cuál debería ser el tamaño de la muestra?

Para determinar el tamaño de la muestra con esas condiciones, debemos despejar el valor de n de la expresión:

$$E_{max} = Z_{1-\alpha/2}\sqrt{\frac{p\cdot(1-p)}{n}} \rightarrow 0,02 = 1,96\cdot\sqrt{\frac{0,32\cdot(1-0,32)}{n}}$$

Y la expresión que resulta es:

$$n = p(1-p)\frac{z^2_{1-\alpha/2}}{E^2_{max}} = 0,32 \cdot (1-0,32)\frac{1.96^2}{0,02^2} = 2089,83 \approx 2090 \ personas$$

SITUACIÓN 4. *«La formación alcanzada por la población adulta española ha mejorado, de forma continua, en los últimos 10 años. Desde 1998 el porcentaje de españoles de 25 a 64 años que poseen estudios superiores a los obligatorios ha pasado de 33% al 51% en 2008. En la misma proporción ha disminuido, por tanto, el porcentaje de españoles que sólo poseen estudios obligatorios, que ha pasado del 67% en 1998 al 49% en 2008»* (Panorama de la Educación. Informe OCDE 2010). Ante esta situación, imagine que usted quiere confirmar que la proporción de adultos de su Comunidad que solo tienen los estudios obligatorios es significativamente inferior que el indicador de España en 2008. Establece un nivel de confianza del 95% y utiliza una muestra aleatoria de 900 adultos con edades comprendidas entre 25 y 64 años, encontrando que 378 de ellos tienen solo los estudios obligatorios.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS: Utilizamos una muestra de sujetos en los que la variable de estudio «estar en posesión de estudios superiores» es de naturaleza dicotómica (se tienen o no se tienen estos estudios) sobre la que se determina la proporción de personas que poseen dicha cualidad.

Los datos de los que partimos son:

	Estudios superiores a los obligatorios	*Estudios obligatorios*	*Muestra*
1998	33%	67%	n = 900
2008	51%	49%	378 *con estudios obligatorios*

Proporción de personas en la muestra con estudios obligatorios: $p = \frac{378}{900} = 0,42$

HIPÓTESIS: El investigador quiere comprobar que la proporción de adultos de su Comunidad que solo tienen los estudios obligatorios es significativamente inferior que el indicador de España en 2008 establecido en el 49%. Las hipótesis nula y alternativa son:

$$H_0 : \pi \geq 0,49; \qquad H_1 : \pi < 0,49;$$

ESTADÍSTICO DE CONTRASTE: La variable de estudio «estar en posesión de estudios obligatorias» es de naturaleza dicotómica (se tienen o no se tienen estos estudios) con distribución binomial. Al tratarse de una muestra grande (n = 900) la distribución binomial se aproxima a la normal, de forma que el estadístico de contraste es una puntuación típica, Z:

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1-\pi_0)}{n}}} = \frac{0,42 - 0,49}{\sqrt{\frac{0,49 \cdot 0,51}{900}}} = -4,20$$

REGLA DE DECISIÓN: En la distribución normal el valor crítico, que representa la máxima discrepancia que puede observarse por simple azar entre el valor obtenido en la muestra (p = 0,42) y el formulado en H_0 como parámetro poblacional (H_0 = 0,49) correspondiente a un nivel de confianza del 95% para un contraste unilateral izquierdo es Z = –1,64. Como el estadístico de contraste obtenido Z = –4,20 representa una discrepancia mayor, rechazamos H_0.

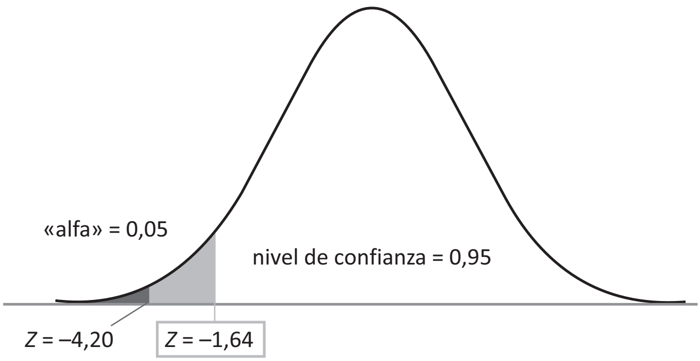


Figura 4: *Representación del valor crítico para un NC del 95% en un contraste unilateral izquierdo.*

De otra forma: buscamos el nivel crítico p asociado al estadístico Z = –4,20. En la tabla el valor más próximo es Z = –3,59, que deja por debajo una probabilidad de 0,0002. Como el estadístico de contraste Z = –4,20 es menor que este valor crítico deducimos que la probabilidad de encontrar valores menores de Z = –4,20 es menor de 0,0002.

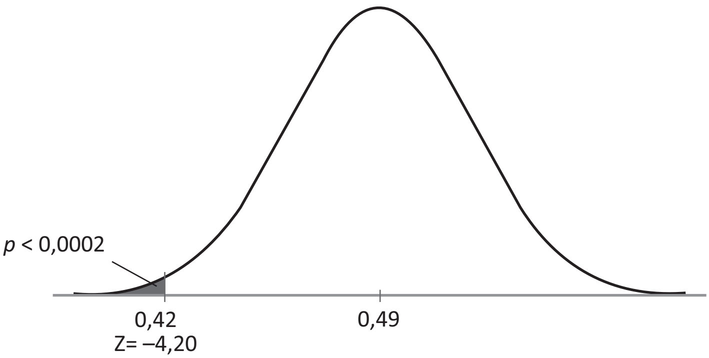


Figura 5: *Nivel crítico p del estadístico* ***Z = –4,20***

CONCLUSIÓN: Con un nivel de confianza, previamente establecido en el 95%, rechazamos H_0 ya que el estadístico de contraste (Z = –4,20) representa una discrepancia «mayor» que el valor crítico correspondiente a ese nivel de confianza del 95%. Por otra parte, el nivel crítico *p* (*p* < 0,0002) correspondiente al estadístico de contraste (Z = 4,20) nos permite concluir que la diferencia es significativa tanto con un nivel de significación de 0,05 como del 0,01.

INTERPRETACIÓN: Se confirma la hipótesis del investigador de que la proporción de adultos de su Comunidad que solo tienen los estudios obligatorios es significativamente inferior que el indicador de España en 2008 con un nivel de significación menor de 0,01.

INTERVALO DE CONFIANZA: Con un nivel de confianza del 95% el intervalo de confianza nos permite inferir, a partir de los datos obtenidos en la muestra, la proporción de adultos con estudios superiores. Este intervalo es:

$$p \pm Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \rightarrow 0,42 \pm 1,96\sqrt{\frac{0,42 \cdot 0,58}{900}} \rightarrow \begin{cases} 0,388 \\ 0,452 \end{cases}$$

Por tanto, la proporción de adultos de esa Comunidad con estudios superiores es un valor comprendido entre 0,388 (38,8%) y 0,452 (45,2%) que no cubre el valor formulado en H_0 de 0,49.

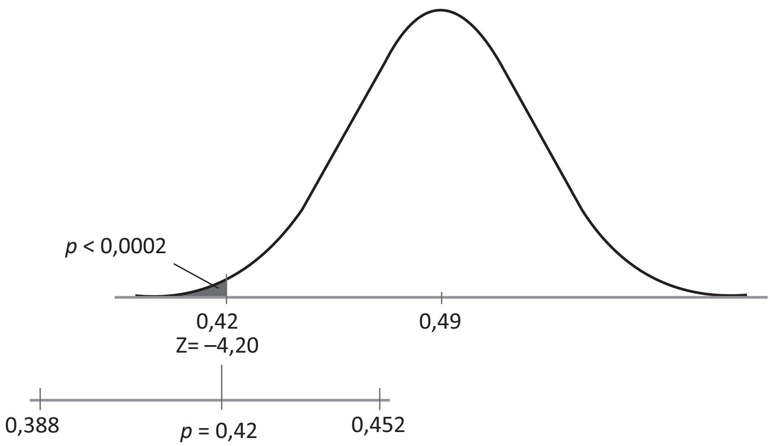


Figura 6: *Representación gráfica del intervalo de confianza de una proporción.*

SITUACIÓN 5. «*En Educación Primaria, la media de alumnos por clase en los centros públicos de España (19,7) es más baja que en la OCDE (21,6) y que en la Unión Europea (20,3). En los centros privados ocurre lo contrario, pues la media en España es de 24,4 frente a 20,8 de media de la OCDE y 19,1 de la UE*». (Panorama de la Educación. Informe OCDE 2010). Imagine que:

5.1. Usted sospecha que la media de los centros públicos de su Comunidad es significativamente mayor que la media de España.

5.2. Y también que la media de los centros públicos de su Comunidad es significativamente menor que la media de la OCDE.

Para contrastar estas hipótesis, selecciona una muestra aleatoria de 100 aulas de educación primaria en centros públicos, encontrando que la media de alumnos por clase es de 20,9 con una desviación típica poblacional de 5,8. Se asume que la variable número de alumnos por clase se distribuye normalmente y se fija el nivel de confianza del 95%.

SOLUCIÓN:

De forma resumida, los datos facilitados en el enunciado son:

	Media de alumnos por clase		
	España	OCDE	UE
Públicos	19,7	21,6	20,3
Privados	24,4	20,8	19,1

Y para la muestra: n = 100, $\overline{Y} = 20{,}9$, σ = 5,8. Para contrastar ambas hipótesis seguimos los siguientes pasos.

CONDICIONES Y SUPUESTOS: Se utiliza una muestra en la que la variable de estudio «número de estudiantes por clase» es de naturaleza continua que se mide con escala de razón, cuya distribución poblacional asumimos que es normal con varianza conocida, según nos indica el enunciado.

HIPÓTESIS: Se plantean dos contrastes de hipótesis sobre dos medias poblacionales. Los dos son unilaterales ya que el investigador está marcando el sentido (mayor o menor) de la diferencia:

5.1. El primero es un contraste unilateral derecho ya que el investigador quiere comprobar si la media de los centros públicos de su Comunidad es significativamente mayor que la media de España establecida en 19,7 alumnos por aula.

$$H_0 : \mu \leq 19{,}7 \qquad H_1 : \mu > 19{,}7$$

5.2. El segundo contraste es unilateral izquierdo, ya que el investigador está marcando el sentido de la diferencia al querer comprobar si la media de los centros públicos de su Comunidad es significativamente menor que de la media de la OCDE, establecida en 21,6 alumnos por aula.

$$H_0 : \mu \geq 21{,}6 \qquad H_1 : \mu < 21{,}6$$

ESTADÍSTICO DE CONTRASTE: Tratándose de una variable que se distribuye normalmente en la población con varianza poblacional conocida, la distribución muestral de la media es normal y el estadístico de contraste que utilizaremos es la Z.

5.1. Para la primera hipótesis el estadístico de contraste es:

$$Z = \frac{\overline{Y} - \mu_0}{\sigma / \sqrt{n}} = \frac{20,9 - 19,7}{5,8 / \sqrt{100}} = 2,069 \approx 2,07$$

5.2. Y para la segunda hipótesis el estadístico de contraste es:

$$Z = \frac{\overline{Y} - \mu_0}{\sigma / \sqrt{n}} = \frac{20,9 - 21,6}{5,8 / 10} = -1,21$$

REGLA DE DECISIÓN: Con un nivel de confianza del 95%, la máxima diferencia que puede encontrarse por simple azar (el valor crítico) es +1,64 para un contraste unilateral derecho y –1,64 para un contraste unilateral izquierdo:

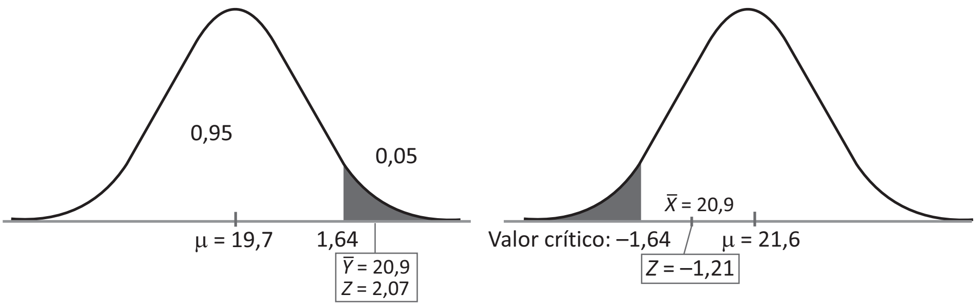


Figura 7: *Representación gráfica de los valores críticos para la regla de decisión dos contrastes unilaterales.*

CONCLUSIÓN: Para el primer contraste, se rechazaría la H_0 ya que el estadístico de contraste obtenido (Z = 2,07) supera la máxima diferencia que cabe esperar por simple azar (el valor crítico: 1,64). Por el contrario, para el segundo contrastes no hay evidencia suficiente para contrastar la H_0 ya que el estadístico de contraste obtenido (Z = –1,21) no supera la máxima diferencia que cabe esperar por simple azar (–1,64) con un nivel de confianza del 95% en ambas situaciones.

De otra forma, los niveles críticos p de ambos estadísticos son: p = 0,0192 y 0,1131 respectivamente, tal y como se representan en las siguientes figuras.

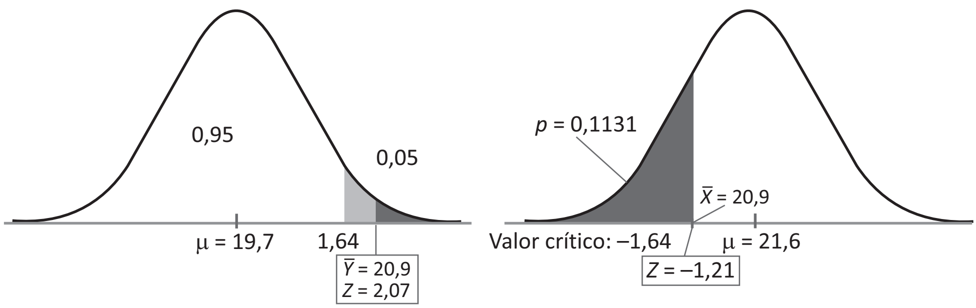


Figura 8: *Representación gráfica de los niveles críticos p para la toma de decisión en dos contrastes unilaterales.*

Por tanto, para el primer contraste y partiendo de que la hipótesis nula (que establece que la media de alumnos por clase de los alumnos de los colegios públicos de esa comunidad es igual o menor de 19,7) es verdadera, la probabilidad de obtener una muestra de 100 aulas de colegios públicos con una media de alumnos por clase de 20,9 es de 0,0192 (el nivel crítico *p*). Esta es una probabilidad pequeña en comparación al nivel de significación de 0,05 pero grande en comparación al nivel de significación de 0,01 por lo que rechazaríamos la H_0 con un nivel de confianza del 95% pero no podríamos tomar la misma decisión con un nivel de confianza del 99%.

Para el segundo contraste, la probabilidad de que siendo cierta la Ho se obtenga una muestra de 100 aulas con una media de estudiantes de 20,9 es de 0,1131. Esta probabilidad es grande en comparación con los niveles de significación del 0,01 y 0,05 por lo que no hay evidencia suficiente para rechazar la H_0 con un nivel de confianza del 95% ni tampoco al 99%.

INTERPRETACIÓN: La media de alumnos por aula en los colegios públicos de la Comunidad del investigador es significativamente mayor que la media general de España con un nivel de confianza del 95% pero no del 99%, ($Z = –2{,}07$; $p = 0{,}0192$), pero no difiere significativamente de la media de los países de la OCDE con un nivel de confianza del 95% y del 99% ($Z = –1{,}21$; $p = 0{,}1131$).

Por otra parte y de forma adicional el investigador podría calcular, a partir de los datos de su estudio, el intervalo de confianza de la media de alumnos por clase en los colegios públicos de su Comunidad. Para ello, establece un nivel de confianza del 95%. En estas circunstancias el error máximo de estimación es:

$$E_{\max} = Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} = 1{,}96 \frac{5{,}8}{\sqrt{100}} = 1{,}1368$$

Y el intervalo de confianza se obtiene sumando y restando a la media de la muestra, este error máximo de estimación:

$$IC = \bar{Y} \pm E_{\max} = 20{,}9 \pm 1{,}1368 \rightarrow (19{,}76; 22{,}04)$$

El intervalo de confianza de alumnos por clase de los colegios públicos de su Comunidad es un valor comprendido entre 19,76 y 22,04 alumnos, que incluye la media de la OCDE (21,6) y la media de la UE (20,3) de los que NO difiere significativamente, con un nivel de confianza del 95%, pero no incluye la media global de España (19,7) de la que SÍ difiere significativamente.

Si, por otra parte, y una vez realizado este análisis el investigador decide establecer el error máximo de estimación en 1 punto y un nivel de confianza del 95%, ¿Qué tamaño de muestra debería haber utilizado?

El error máximo es igual a:

$$E = Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Si hacemos E = 1, entonces:

$$1 = 1{,}96 \frac{5{,}8}{\sqrt{n}}$$

Despejando «n»:

$$n = \sigma^2 \frac{\left(Z_{1-\alpha/2}\right)^2}{E^2} = 5{,}8^2 \frac{1{,}96^2}{1^2} = 129{,}23 \approx 129$$

Debería utilizar una muestra de 129 aulas de colegios públicos de su Comunidad.

SITUACIÓN 6. *Según el último estudio del Observatorio Español sobre Drogas (2009), realizado en estudiantes de Secundaria de 14 a 18 años, el 5,1% ha consumido cocaína alguna vez en la vida y el 2,7% éxtasis. Además, el inicio en el consumo de cocaína y éxtasis tiene lugar cada vez a edades más tempranas. Así, mientras que en el año 2004 la edad media de inicio para la cocaína era de 15,9 años en los hombres y 15,7 en las mujeres, en el año 2008 disminuyó a 15,3 años y 15,2, respec-*

tivamente (Encuesta estatal sobre uso de drogas en enseñanzas secundaria, 2009). Imagine que usted forma parte de un equipo de atención primaria que cubre a un determinado sector de su municipio y desea estimar, con un nivel de confianza del 95%:

6.1. La proporción de jóvenes de enseñanzas secundarias de su municipio que han consumido cocaína al menos una vez

6.2. La edad media del consumo de cocaína de las mujeres de su municipio.

6.3. Además sospecha que la edad media del inicio en el consumo de cocaína en los hombres de su municipio ha disminuido respecto a los datos generales del 2004.

6.4. Finalmente desea contrastar la hipótesis de que la varianza poblacional en la edad de inicio del consumo de cocaína en los hombres de su municipio es significativamente mayor que 1,1 establecido para las mujeres.

Para ello, dispone de los datos de una muestra de 37 jóvenes varones y 41 mujeres atendidos el pasado año en los que el 8% habían consumido cocaína, al menos un vez, siendo la edad media de inicio en el consumo de cocaína de 15,4 años en los hombres y de 15,2 años en las mujeres con una desviación típica de 1,3 para los hombres y 1,1 en las mujeres.

SOLUCIÓN:

De forma resumida, los datos proporcionados en esta situación vienen recogidos en las siguientes tablas:

Datos de encuestas estatales:

2009	5,1% Cocaína	2,7% Éxtasis
	Edad media consumo cocaína	
	Hombres	**Mujeres**
2004	15,9	15,7
2008	15,3	15,2

Datos de la muestra de su municipio:

	n	$\bar{Y}$	S_n	Consumo
Hombres	37	15,4	1,3	8% Cocaína
Mujeres	41	15,2	1,1	

6.1. A partir del 8% de jóvenes (hombres y mujeres) que han consumido cocaína, el intervalo de confianza de la proporción poblacional de jóvenes consumidores con un nivel de confianza del 95%, es:

$$IC = p \pm E_{\max}$$

Siendo el error máximo de estimación de la proporción poblacional:

$$E_{\max} = Z_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 1{,}96 \cdot \sqrt{\frac{0{,}08 \cdot (1-0{,}08)}{37+41}} = 1{,}96 \cdot 0{,}0307 = 0{,}06$$

Y, finalmente, el intervalo de confianza para la media poblacional de la proporción de jóvenes que han consumido cocaína, es:

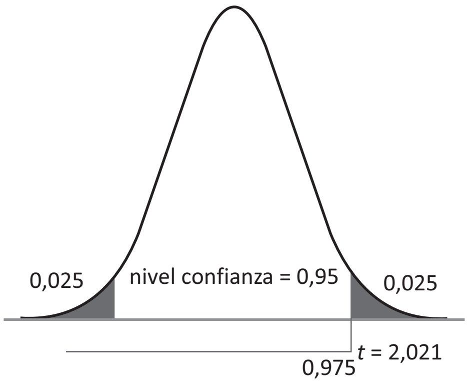
$$IC = p \pm E_{\max} = 0{,}08 \pm 0{,}06 = (0{,}02; 0{,}14)$$

Es decir, entre un 14% y un 2% de los jóvenes de su municipio han consumido cocaína al menos una vez.

6.2. A partir de la información obtenida en la muestra respecto a la edad media de consumo de cocaína en las mujeres que ha sido de 15,2 años, obtenemos el intervalo de confianza para la media poblacional. Para ello, consideramos que la edad es una variable continua medida con escala de razón pero de la que desconocemos la forma de su distribución en la población y su varianza. Bajo estas condiciones y al tratarse de una muestra grande, la distribución muestral de la edad media es la t de Student con n-1 grados de libertad.

$$E_{\max} = t_{n-1;1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} = t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n-1}} = 2{,}021 \frac{1{,}1}{\sqrt{41-1}} = 0{,}35$$

El valor de t = 2,021 se busca en las tablas con $n - 1 = 41 - 1 = 40$ grados de libertad:



g.l.	Probabilidad											
	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	$\boxed{0,975}$	0,990	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
$\boxed{40}$	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	$\boxed{2,021}$	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660

Tabla 5: *Reproducción parcial de la tabla de la distribución t y su representación gráfica con los valores críticos para el cálculo del intervalo de confianza.*

Y el intervalo de confianza, es:

$$\bar{Y} \pm E_{\text{max}} \rightarrow 15,2 \pm 0,35 \rightarrow \begin{cases} 14,84 \\ \\ 15,55 \end{cases}$$

6.3. Para comprobar si la edad media del inicio en el consumo de cocaína en los hombres de su municipio ha disminuido respecto a los datos generales del 2004, formularemos un contraste de hipótesis, siguiendo los siguientes pasos:

CONDICIONES Y SUPUESTOS: Partimos de la información proporcionada por una muestra de 37 hombres con una edad media de 15,4 años. Desconocemos la forma de la distribución poblacional de la edad de inicio en el consumo y su varianza. Por tanto, la distribución muestral de la media es una distribución *t* de Student con n – 1 = 36 grados de libertad.

HIPÓTESIS: El investigador sospecha que la edad media del inicio en el consumo de cocaína en los hombres de su municipio ha disminuido respecto a los datos generales del 2004 que era de 15,9 para los hombres, por lo que plantea un contraste unilateral izquierdo

$$H_o : \mu \geq 15,9 \qquad H_1 : \mu < 15,9$$

ESTADÍSTICO DE CONTRASTE: Al tratarse de un diseño de una muestra extraída de una población con varianza desconocida, el estadístico de contraste es la *t* de Student con *n* – 1 gl.

$$T = \frac{\overline{Y} - \mu_0}{\frac{S_n}{\sqrt{n-1}}} = \frac{15,4 - 15,9}{\frac{1,3}{\sqrt{37-1}}} = -2,31$$

REGLA DE DECISIÓN: Como en la tabla de la distribución *t* no figura la distribución *t* con 36 gl el valor crítico correspondiente a esta distribución deberemos buscarlo por aproximación con el valor de una distribución con 40gl. Para ello, el valor crítico *t* que deja por debajo una probabilidad de 0,05 es el mismo, pero de signo contrario, al que deja por encima una probabilidad de 0,05 (y por debajo 0,95) que corresponde a un nivel de confianza del 95% para un contraste unilateral derecho. Este valor es –1,684 que es el que encontramos en las tablas con 40 gl como valor más aproximado a los 36 gl de este ejemplo.

g.l.	Probabilidad											
	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	0,975	0,990	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660

Tabla 6: *Reproducción parcial de la tabla de la distribución t.*

CONCLUSIÓN: El valor crítico representa la máxima diferencia que puede darse por simple azar. Como el estadístico de contraste ($T = –2,31$; $p < 0,025$) supera esta máxima diferencia (–1,684), rechazamos la hipótesis nula con un nivel de confianza del 95%.

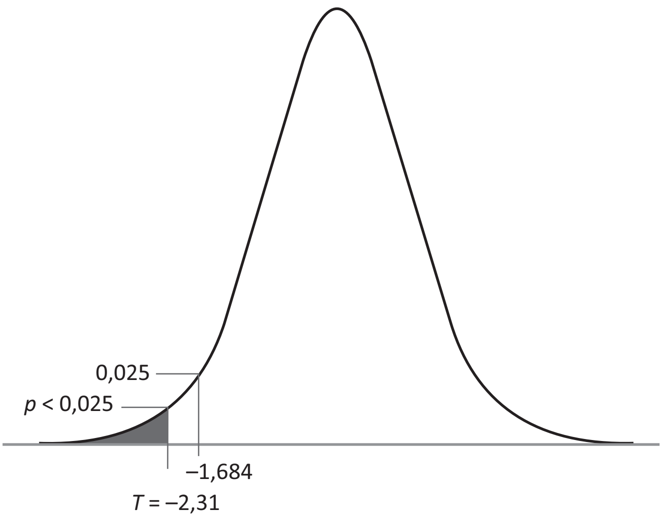


Figura 9: *Nivel crítico del estadístico* ***T = –2,31*** *en la distribución t con 30 gl.*

INTERPRETACIÓN: La edad media de inicio en el consumo de cocaína en los hombres es significativamente menor de 15,9 años ($T = –2,31$; $p < 0,025$).

6.4. Para analizar si la varianza poblacional en la edad de inicio del consumo de cocaína en los hombres de su municipio es significativamente mayor que 1,1 establecido para las mujeres seguiremos los mismos pasos:

CONDICIONES Y SUPUESTOS: Partimos de la información proporcionada por una muestra de 37 hombres con una desviación típica de 1,3 (Varianza de $1,3^2 = 1,69$). La distribución muestral de la varianza es una distribución chi-cuadrado con $n – 1$ gl.

HIPÓTESIS: El investigador quiere comprobar que la varianza poblacional de la edad de inicio en el consumo de cocaína de los hombres es mayor de $1,1^2 = 1,21$ para lo que formula las siguientes hipótesis:

$$H_o : \sigma^2 \leq 1,21 \qquad H_1 : \sigma^2 > 1,21$$

ESTADÍSTICO DE CONTRASTE: Utilizamos el estadístico chi-cuadrado, que vale:

$$\chi^2 = \frac{n \cdot S^2}{\sigma_o^2} = \frac{37 \cdot 1{,}3^2}{1{,}1^2} = 51{,}67$$

REGLA DE DECISIÓN: Con un nivel de confianza del 95%, la máxima diferencia que cabe esperar por simple azar entre el valor de la varianza obtenido en la muestra respecto al valor establecido en la hipótesis nula, viene determinada por el valor crítico de la distribución chi-cuadrado con 36 gl. El valor más aproximado lo buscamos en las tablas con 40 gl y es 55,7585. Como el estadístico de contraste (51,67) no supera el valor crítico 55,7585, no podemos rechazar la hipótesis nula con ese nivel de confianza.

g.l.	Probabilidad									
	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	14,8603
29	13,1211	14,2565	16,0471	17,7084	19,7677	39,0875	42,5570	45,7223	49,5879	52,3356
30	13,7867	14,9535	16,7908	18,4927	20,5992	40,2560	43,7730	46,9792	50,8922	53,6720
40	20,7065	22,1643	24,4330	26,5093	29,0505	51,8051	$\boxed{55,7585}$	59,3417	63,6907	66,7660
50	27,9907	29,7067	32,3574	34,7643	37,6886	63,1671	67,5048	71,4202	76,1539	79,4900
60	35,5345	37,4849	40,4817	43,1880	46,4589	74,3970	79,0819	83,2977	88,3794	91,9517

Tabla 7: *Tabla de la distribución chi-cuadrado.*

INTERPRETACIÓN: La varianza poblacional de la edad de inicio del consumo de cocaína en los hombres es igual que el de las mujeres (χ^2 = 51,67; p > 0,05).

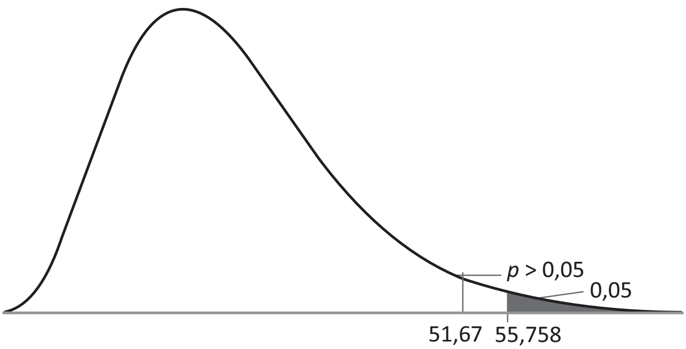


Figura 10: *Representación gráfica de la distribución chi-cuadrado, el estadístico de contraste y el nivel crítico.*

SITUACIÓN 7. *El Centro de Investigaciones Sociológicas (CIS) realiza constantes estudios sobre la ideología e intención de voto de los españoles. Uno de los ítems que incluye en sus cuestionarios es el de la «identificación ideológica», para lo que se le pide al entrevistado que se ubique ideológicamente en una escala de 1 a 10 (donde 1 significa extrema izquierda y 10 extrema derecha). En su última encuesta a la población española la media aritmética obtenida para la variable identificación ideológica fue de un 4,86.* Un psicólogo social interesado en estudiar esta temática, sospecha que en su localidad la intención de voto es significativamente «más conservadora», para lo que diseña una encuesta dirigida a confirmar su hipótesis. Para ello, selecciona una muestra de 31 personas a las que aplica el ítem de identificación ideológica, obteniendo una media de 5,40 y una cuasi-desviación típica de 1,2. Quiere además saber cuál sería el tamaño de la muestra que debería utilizar para estimar la media poblacional fijando un error máximo de estimación de 0,2 puntos.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS: El estudio corresponde a un diseño de una muestra en la que la variable de estudio «Identificación ideológica» es una variable cuantitativa medida con escala de intervalo en la que se obtiene la media y la varianza de la muestra de 31 personas que han respondido a la encuesta. Se desconoce la distribución poblacional de esta variable y su varianza por lo que la distribución muestral de la media es la *t* de Student.

HIPÓTESIS: El investigador sospecha que en su localidad la intención de voto es significativamente «más conservadora», por lo que plantea un contraste unilateral en el que la media en su locaclidad debe ser mayor de 4,86 que es la media de la población española:

$$H_o : \mu \leq 4{,}86; \qquad H_1 : \mu > 4{,}86$$

ESTADÍSTICO DE CONTRASTE: Al desconocer la forma de la distribución poblacional y su varianza el estadístico de contraste es:

$$T = \frac{\bar{Y} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{5{,}40 - 4{,}86}{\frac{1{,}2}{\sqrt{31}}} = 2{,}51$$

REGLA DE DECISIÓN: Para los niveles de confianza utilizados habitualmente (95% y 99%), localizamos en la tabla de la distribución *t* de Student con 30 grados de libertad estos valores críticos y son 2,457 para un nivel de confianza del 99% en un contraste unilateral derecho y 1,697 para un nivel de confianza del 95% de un contraste unilateral derecho.

g.l.	Probabilidad											
	0,550	**0,600**	**0,650**	**0,700**	**0,750**	**0,800**	**0,850**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	$\boxed{1,697}$	2,042	$\boxed{2,457}$	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660

Tabla 8: *Reproducción parcial de la tabla de la distribución t.*

CONCLUSIÓN: El estadístico T = 2,51 supera ambos valores los que nos conduce a rechazar la H_0 con un nivel de confianza del 99%.

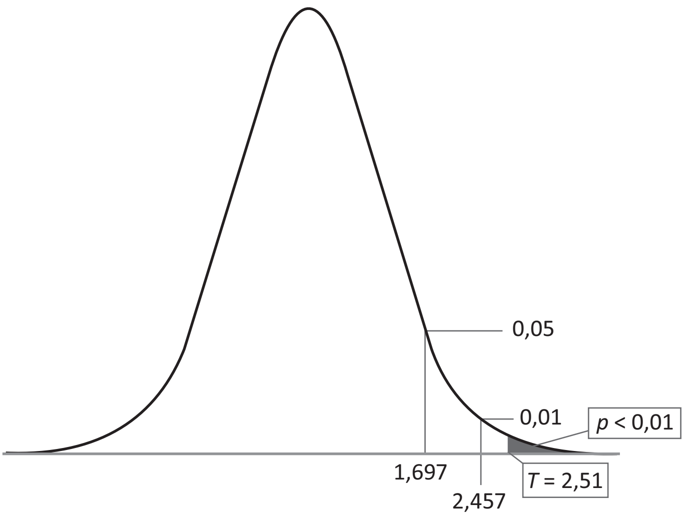


Figura 11: *Distribución t de Student.*

INTERPRETACIÓN: Se confirma la sospecha del investigador de que la identificación ideológica en su localidad la es «más conservadora» que en la población general (T = 2,51; 0,005 < p < 0,01).

Si el investigador ha fijado un error máximo de 0,2 puntos para estimar la media poblacional, con un NC del 95%, el tamaño de la muestra utilizada sería:

$$E_{\max} = t^2_{n-1;1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} = 0,2$$

Y despejando el valor de *n*:

$$n = S^2_{n-1} \frac{t^2_{n-1;1-\alpha/2}}{E^2_{\max}} = 1,2^2 \cdot \frac{2,042^2}{0,2^2} = 150,11 \approx 150$$

SITUACIÓN 8. *El «Síndrome Jubilación» es la situación que experimentan ciertas personas ante esta nueva etapa vital con manifestaciones somáticas, psíquicas y sociales negativas que afectan la calidad de vida del jubilado. Un estudio publicado por el GIE (Grupo de Investigación del Envejecimiento) mediante una encuesta realizada en el 2006 utilizando una muestra de jubilados con edad media de 77,6 años y desviación típica de 8,79 años encuentra que las manifestaciones psíquicas más frecuentes eran la ansiedad (82%), el pesimismo (13,3%) y la depresión (4,7%) y que para el 32% de los expertos consultados el apoyo psicológico constituye la estrategia de intervención más adecuada para superar estos estados.* Imagine que usted quiere estudiar la situación de los jubilados de su localidad respecto a este «síndrome», para lo que utiliza una muestra aleatoria de 362 jubilados, con una edad media de 71,2 años y una desviación típica de 12,5 y de los cuales, el 59,8% presenta signos de ansiedad, el 35% pesimismo y el 5,2% depresión. Adicionalmente le indican que, con los datos del estudio anterior, el intervalo de confianza de la proporción de jubilados con manifestaciones de depresión, es un valor comprendido entre 0,0387 (3,87%) y 0,0553 (5,53%) con un nivel de confianza del 95%. Con esta información se propone realizar los siguientes análisis:

8.1. Conocer el tamaño aproximado de la muestra que se ha utilizado en el estudio anterior.

8.2. Con los datos de su estudio, calcular el intervalo de confianza de la edad media de los jubilados de su localidad con un nivel de confianza del 95%.

8.3. Calcular el intervalo de confianza para la proporción de jubilados con manifestaciones psíquicas de ansiedad, con un nivel de confianza del 99%.

8.4. Contrastar la hipótesis de que la proporción de jubilados de su localidad con manifestaciones psíquicas de pesimismo es significativamente mayor que el valor 0,133 facilitado por el GIE en el 2006.

SOLUCIÓN:

8.1. Para responder a la primera pregunta conocemos los límites superior e inferior del intervalo de confianza. La diferencia entre ambos límites corresponde a la amplitud del intervalo y el error máximo de estimación es la mitad de esta diferencia:

$$\left.\begin{aligned} l_i &= p - E_{\max} \rightarrow 0{,}0387 = 0{,}047 - E_{\max} \\ l_s &= p + E_{\max} \rightarrow 0{,}0553 = 0{,}047 + E_{\max} \end{aligned}\right\} \rightarrow$$

$$E = \frac{l_s - l_i}{2} = \frac{0{,}0553 - 0{,}0387}{2} = 0{,}0083$$

El error máximo de estimación para obtener el intervalo de confianza de una proporción, es:

$$E_{\max} = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Y despejando el valor de n, obtenemos el tamaño de la muestra con un nivel de confianza del 95% (Z = 1,96) siendo p = 0,047 la proporción de jubilados con síndrome de depresión en el estudio anterior que se ha utilizado para obtener el intervalo:

$$n = p(1-p)\frac{z^2_{\alpha/2}}{E^2_{\max}} = 0{,}047(1-0{,}047)\frac{1{,}96^2}{0{,}0083^2} = 2497{,}73 \approx 2500 \text{ jubilados}$$

8.2. La edad media de los jubilados es una variable cuantitativa medida con escala de razón y de la que desconocemos su distribución poblacional y su varianza. En estas condiciones la distribución muestral de la media es la distribución t de Student con n – 1 grados de libertad. Con un nivel de confianza del 95%, el intervalo de confianza de la edad media de los jubilados de su localidad se obtiene sumando y restando a la media de la muestra el error máximo de estimación. Al tratarse de una muestra grande (n = 362) la distribución t se aproxima a la normal por lo que el valor de t se busca en la tabla de la distribución normal tipificada de puntuaciones Z :

$$E_{\max} = t_{1-\alpha/2} \cdot \frac{S_n}{\sqrt{n-1}} = 1{,}96 \cdot \frac{12{,}5}{\sqrt{362-1}} = 1{,}289$$

$$IC = \bar{Y} \pm E_{\max} = 71{,}2 \pm 1{,}289 \rightarrow \begin{cases} 69{,}91 \\ \\ 72{,}49 \end{cases}$$

8.3. La proporción de jubilados con manifestaciones psíquicas de ansiedad es una variable con distribución binomial. Al tratarse de una muestra grande la distribución binomial se aproxima a la normal, por lo que utilizaremos esta distribución para determinar el intervalo de confianza que se obtendrá sumando y restando a la proporción obtenida en la muestra (0,598) el error máximo de estimación. Con un nivel de confianza del 99% el valor de *Z* es 2,58:

$$E_{\max} = Z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} = 2{,}58 \sqrt{\frac{0{,}598(1-0{,}598)}{362}} = 0{,}0665$$

$$IC = p \pm E_{\max} = 0{,}598 \pm 0{,}0665 \rightarrow \begin{cases} 0{,}5315 \\ 0{,}6645 \end{cases}$$

8.4. Para contrastar la hipótesis de que la proporción de jubilados de su localidad con manifestaciones psíquicas de pesimismo es significativamente mayor que el valor 0,133 facilitado por el GIE en el 2006 seguimos los siguientes pasos:

CONDICIONES Y SUPUESTOS: El estudio corresponde a un diseño de una muestra de 362 jubilados en la que la variable de estudio «manifestaciones psíquicas de pesimismo» de los jubilados de su localidad es una variable dicotómica con distribución binomial. Como la muestra es grande la distribución binomial se aproxima a la normal.

HIPÓTESIS: El investigador quiere contrastar que la proporción de jubilados con manifestaciones psíquicas de pesimismo es mayor de 0,133 que es el valor encontrado por el GIE para la población general. Plantea, por tanto un contraste unilateral.

$$H_0 : \pi \leq 0{,}133; \qquad H_1 : \pi > 0{,}133$$

ESTADÍSTICO DE CONTRASTE: Aunque la distribución muestral de la proporción es una distribución binomial, al tratarse de una muestra grande, la distribución binomial se aproxima a la normal, por lo que utilizaremos el estadístico Z, siendo la proporción de jubilados con manifestaciones psíquicas de pesimismo encontrado en la muestra de estudio ha sido p = 0,35.

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0,35 - 0,133}{\sqrt{\frac{0,133(1-0,133)}{362}}} = 12,158$$

REGLA DE DECISIÓN: El estadístico de contraste es un valor muy extremo (Z = 12,158). El valor máximo que podemos consultar en las tablas de curva normal es z = 3,59, que deja por encima de sí una proporción: p = 0,0002.

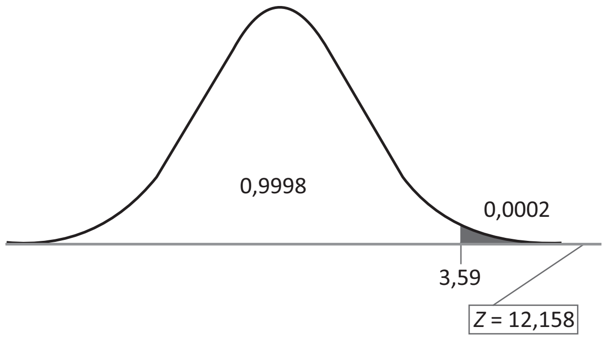


Figura 12: *Representación gráfica del nivel crítico p en la distribución t de Student.*

Por tanto, el valor crítico asociado a este estadístico de contraste es un valor menor de 0,0002

CONCLUSION: Se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

INTERPRETACIÓN: Se confirma la hipótesis del investigador de que la proporción de jubilados de su localidad con manifestaciones psíquicas de pesimismo es significativamente mayor de 0,133 formulada para la población general en el estudio anterior (Z = 12,158; p < 0,0002).

SITUACIÓN 9. El barómetro del CIS de marzo de 2012 realizado en 240 municipios de 48 provincias señalaba que el 23,4% estaba en situación de paro y de éstos, a la pregunta *¿Y cree Ud. que es muy probable, bastante, poco o nada probable que durante los próximos doce meses encuentre Ud. trabajo?*, el 22,6% manifestaba que «bastante probable», frente al 43,1% que creía que «poco probable» y el 19,2%

que «nada probable» y el resto «NS/NC». Imagine que usted quiere estudiar si estos resultados se reproducen actualmente en su localidad, para lo que realiza una encuesta sobre una muestra de 100 personas en situación de paro con una edad media de 39 años y desviación típica de 8,6 años de los cuales 25 le responden que «bastante probable», 35 responden que «poco probable» y 20 que «nada probable» mientras que el resto «no saben o no contestan». Con esa información se propone realizar los siguientes análisis:

9.1. Calcular el intervalo de confianza de la varianza de la edad de las personas en situación de paro de su localidad con un nivel de confianza del 95%.

9.2. Con los datos obtenidos en su muestra y fijando un nivel de confianza del 95%, calcular el intervalo de confianza para la proporción de personas en situación de paro que considera «poco o nada probable» encontrar trabajo en los próximos doce meses.

9.3. Contrastar si la proporción de personas en paro que consideran «poco probable» encontrar trabajo en los próximos doce meses en su localidad es significativamente menor que el valor 0,431 (43,1%) proporcionados en el estudio del CIS.

9.4. Contrastar si la desviación típica de la edad media de las personas en situación de paro es significativamente mayor de 7,6 años.

SOLUCIÓN:

Los datos de la encuesta del CIS, aparecen resumidos en la siguiente tabla:

Bastante	Poco	Nada	NS/NC
0,226	0,431	0,192	1 – (0,226 + 0,431 + 0,192) = 0,151

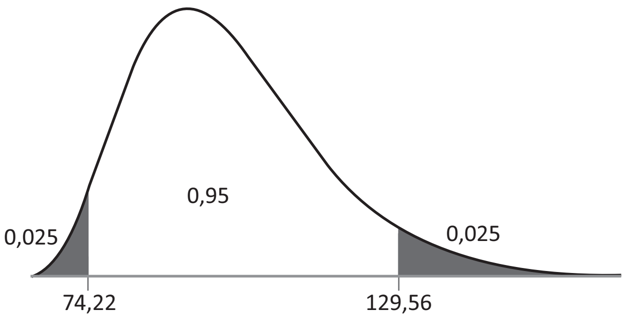
En la muestra con n = 100, media 39 y desviación típica 8,6, los resultados que obtiene son:

Bastante	Poco	Nada	NS/NC
0,25	0,35	0,20	0,20

9.1. La distribución muestral de la varianza es una distribución chi-cuadrado con $n - 1$ grados de libertad. Los límites del intervalo de confianza de la varianza poblacional son:

$$\frac{nS_n^2}{\chi^2_{1-\alpha/2,\,n-1}} < \sigma^2 < \frac{nS_n^2}{\chi^2_{\alpha/2,\,n-1}}$$

Con un nivel de confianza del 95% buscamos en la tabla de chi-cuadrado con 100 – 1 = 99 grados de libertad los valores que delimitan una probabilidad central de 0,95. Al no figurar en la tabla los valores correspondientes a una distribución con 99 gl, utilizamos, por aproximación los que figuran en la fila de 100 gl.



g.l.	Probabilidad									
	0,005	0,010	$\boxed{0,025}$	0,050	0,100	0,900	0,950	$\boxed{0,975}$	0,990	0,995
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	14,8603
5	0,4117	0,5543	0,8312	1,1455	1,6103	9,2364	11,0705	12,8325	15,0863	16,7496
6	0,6757	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,4494	16,8119	18,5476
70	43,2752	45,4417	48,7576	51,7393	55,3289	85,5270	90,5312	95,0232	100,4252	104,2149
80	51,1719	53,5401	57,1532	60,3915	64,2778	96,5782	101,8795	106,6286	112,3288	116,3211
90	59,1963	61,7541	65,6466	69,1260	73,2911	107,5650	113,1453	118,1359	124,1163	128,2989
100	67,3276	70,0649	$\boxed{74,2219}$	77,9295	82,3581	118,4980	124,3421	$\boxed{129,5612}$	135,8067	140,1695

Figura 13: *Representación de la distribución chi-cuadrado con 100 gl y los valores críticos para un nivel de confianza del 95% y reproducción parcial de la tabla con la localización de estos valores críticos.*

Y los límites que se obtienen son:

$$l_{\text{inf}} = \frac{100 \cdot 8{,}6^2}{129{,}5612} = 57{,}085 \qquad l_{\text{sup}} = \frac{100 \cdot 8{,}6^2}{74{,}2219} = 99{,}65$$

Y la varianza poblacional es un valor comprendido entre: 57,085 < σ < 99,65

9.2. La distribución muestral de la proporción es una distribución binomial que se aproxima a la normal cuando la muestra es grande. Para calcular el intervalo de confianza de la proporción de personas en situación de paro que considera «poco o nada probable» encontrar trabajo en los próximos doce meses a partir de la información obtenida en una muestra de 100 personas, utilizaremos la aproximación a la normal.

En la muestra de 100 personas utilizada por el investigador, la proporción de personas que consideran «poco o nada probable» encontrar trabajo es: p = 0,35 + 0,20 = 0,55. El intervalo de confianza, con un nivel de confianza del 95%, es:

$$IC = p \pm E_{\max}$$

$$E_{\max} = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 1{,}96\sqrt{\frac{0{,}55(1-0{,}55)}{100}} = 0{,}0975$$

$$IC = p \pm E_{\max} = 0{,}55 \pm 0{,}0975 = \begin{cases} 0{,}4525 \\ \\ 0{,}6475 \end{cases}$$

9.3. Para contrastar si la proporción de personas en paro que consideran «poco probable» encontrar trabajo en los próximos doce meses en su localidad es significativamente menor que el valor 0,431 (43,1%) proporcionados en el estudio del CIS, seguiremos los siguientes pasos:

CONDICIONES Y SUPUESTOS. El diseño de investigación de este trabajo utiliza una muestra en el que la variable de análisis es de naturaleza dicotómica con distribución binomial. Como la muestra es grande (n = 100 personas) la distribución muestral de la proporción de persona que consideran «poco probable» encontrar trabajo se aproxima a la normal.

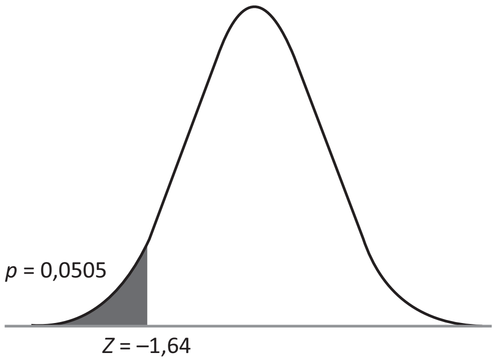
HIPÓTESIS: El objetivo del investigador es demostrar que la proporción de personas en paro que consideran «poco probable» encontrar trabajo en los próximos doce meses en su localidad es significativamente menor que el valor 43,1%. Por lo que plantea un contraste unilateral con las siguientes hipótesis.

$$H_0 : \pi \geq 0{,}431; \qquad H_1 : \pi < 0{,}431$$

ESTADÍSTICO DE CONTRASTE: Por aproximación de la distribución binomial a la normal, aplicamos el estadístico Z:

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}} = \frac{0,35 - 0,431}{\sqrt{\frac{0,431(1 - 0,431)}{100}}} = -1,636 \approx -1,64$$

REGLA DE DECISIÓN: Como *a priori* no se establece un nivel de confianza, recurrimos a determinar el nivel crítico p asociado al estadístico de contraste. En la distribución normal, el nivel crítico p asociado al estadístico de contraste $Z = -1,64$ es 0,0505.



z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,50	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,40	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,30	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,20	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-1,80	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,70	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,60	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,50	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,40	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,30	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823

Figura 14: *Representación del nivel crítico p para el estadístico Z = –1,64 y reproducción parcial de la tabal de la distribución N(0;1) con la localización del nivel crítico para un nivel de significación de 0,05*

CONCLUSIÓN. Para este contraste no se ha establecido a priori un nivel de confianza, pero como el valor crítico $p = 0,0505$ es mayor que los niveles de significación de 0,05 y del 0,01 utilizado habitualmente en este tipo de investigaciones, no tenemos evidencia suficiente para rechazar la hipótesis nula. Por otra parte, en un contraste unilateral izquierdo el valor crítico que deja por debajo una probabilidad de 0,05 es –1,64 que representa la máxima diferencia que cabe esperar por simple azar entre el valor encontrado en la muestra y el valor formulado en la H_0. Como el estadístico de contraste, $Z = –1{,}636$ no supera este valor, no podemos rechazar la H_0.

INTERPRETACIÓN: No hay evidencia para demostrar la hipótesis del investigador (la hipótesis alternativa) de que la proporción de personas en paro que consideran «poco probable» encontrar trabajo en los próximos doce meses en su localidad es significativamente menor que el valor 0,431 (43,1%) ($Z = –1{,}64$; $p > 0{,}05$).

9.4. Para contrastar si la desviación típica de la edad media de las personas en situación de paro es significativamente mayor de 7,6 años, seguimos los mismos pasos que en los anteriores contrastes.

CONDICIONES Y SUPUESTOS: El diseño de investigación de este trabajo utiliza una muestra de 100 personas en el que la variable objeto de estudio es la edad de las personas en situación paro que es una variable cuantitativa que se mide con escala de razón.

HIPÓTESIS: El investigador quiere contrastar si la desviación típica de la edad media de las personas en situación de paro es significativamente mayor de 7,6 años, o lo que es lo mismo, si la varianza es significativamente mayor de $57{,}76 = 7{,}6^2$, para lo que formula las siguientes hipótesis:

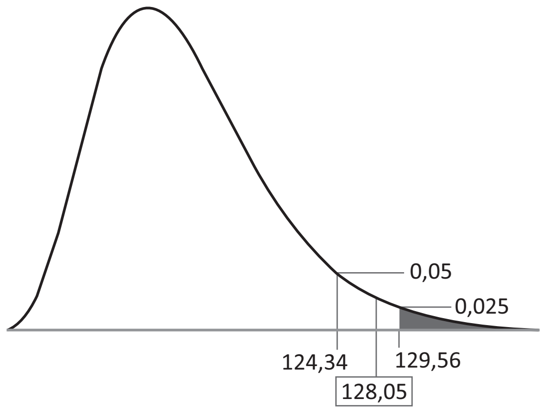
HIPÓTESIS: Se formula un contraste unilateral ya que el investigador está marcando el sentido de la diferencia.

$$H_0 : \sigma^2 \leq 57{,}76; \qquad H_1 : \sigma^2 > 57{,}76$$

ESTADÍSTICO DE CONTRASTE:

$$X^2 = \frac{nS_n^2}{\sigma_o^2} = \frac{100 \cdot 8{,}6^2}{57{,}76} = \frac{7396}{57{,}76} = 128{,}05$$

REGLA DE DECISIÓN: Consultamos en la distribución chi-cuadrado con $n – 1$ grados de libertad (100 – 1 = 99 gl) el valor más próximo al estadístico de contraste obtenido. En la tabla y para 100 gl, los valores más próximos son 124, 34 y 129,56 que dejan por encima probabilidades de 0,05 y 0,025 respectivamente y entre los cuales se encuentra el estadístico de contraste obtenido 128,05. Por tanto el nivel crítico p es un valor menor de 0,05 y mayor de 0,025.



g.l.	Probabilidad									
	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	14,8603
5	0,4117	0,5543	0,8312	1,1455	1,6103	9,2364	11,0705	12,8325	15,0863	16,7496
6	0,6757	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,4494	16,8119	18,5476
70	43,2752	45,4417	48,7576	51,7393	55,3289	85,5270	90,5312	95,0232	100,4252	104,2149
80	51,1719	53,5401	57,1532	60,3915	64,2778	96,5782	101,8795	106,6286	112,3288	116,3211
90	59,1963	61,7541	65,6466	69,1260	73,2911	107,5650	113,1453	118,1359	124,1163	128,2989
100	67,3276	70,0649	74,2219	77,9295	82,3581	118,4980	124,3421	129,5612	135,8067	140,1695

Figura 15: *Representación gráfica del nivel crítico p en una distribución t con 100 gl reproducción parcial de la tabla con la localización de los valores más próximos al estadístico de contraste.*

CONCLUSIÓN: Para este contraste no se ha establecido un nivel de confianza previo, pero al ser el nivel crítico p menor de 0,05 podremos rechazar la hipótesis nula con un nivel de confianza del 95%, pero al ser mayor de 0,025 no podríamos tomar la misma decisión con un nivel de confianza del 99% ni tampoco del 98%.

INTERPRETACIÓN: Se comprueba la hipótesis del investigador de que la desviación típica de la edad de las personas en situación de paro es mayor de 7,6 años (o su varianza mayor de 57,76) con un nivel de confianza del 95% (X^2 = 128,05; p < 0,05).

SITUACIÓN 10. Un psicólogo que investiga la percepción del tiempo, considera que dicha habilidad se encuentra deteriorada en los fumadores durante la retirada de nicotina. Para comprobarlo selecciona una muestra aleatoria de 41 fumadores a los que somete a una abstinencia de tabaco con una duración de 24 horas, pidiéndoles que estimen el tiempo en segundos que había transcurrido en un periodo que, objetivamente, fue de 45 segundos. La media aritmética de la muestra fue igual a 51 segundos con una cuasivarianza igual a 164. El psicólogo desea determinar si la abstinencia tiene un impacto negativo sobre la percepción temporal, provocando que el tiempo se sobre-estime y establece un nivel de significación de 0,05 como regla de decisión para su hipótesis.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS. El estudio corresponde a un diseño de una muestra en el que la variable de estudio «la estimación del tiempo» es una variable continua, medida con escala de razón, de la que se desconoce la forma de su distribución poblacional y la varianza.

HIPÓTESIS: Dado que el psicólogo desea comprobar si el tiempo se sobre-estima, hemos de plantear un contraste unilateral derecho, cuyas hipótesis son:

$$H_0 : \mu \leq 45$$
$$H_1 : \mu > 45$$

ESTADÍSTICO DE CONTRASTE: Dado que se desconoce la varianza poblacional, el estadístico de contraste es:

$$T = \frac{\overline{Y} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{51 - 45}{2} = 3$$

$$\sigma = \frac{S_{n-1}}{\sqrt{n}} = \frac{\sqrt{164}}{\sqrt{41}} = 2$$

REGLA DE DECISIÓN: Buscando en las tablas *T* de Student con 40 grados de libertad y un nivel de confianza del 95%, el valor critico es 1,684 que deja por debajo una probabilidad de 0,95 correspondiente a un contraste unilateral derecho.

g.l.	Probabilidad											
	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	0,975	0,990	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	$\boxed{1,684}$	2,021	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678

Tabla 9: *Reproducción parcial de la tabla de la distribución t de Student.*

Por otra parte, en la distribución *t* de Studente con 40 gl, el valor más alto que podemos ver en la tabla es 2,704 que deja por encima una probabilidad de 0,005.

CONCLUSIÓN: El estadístico de contraste (T = 3) supera el valor crítico con un nivel de confianza del 95% (3 > 1,684), pero también el nivel crítico p < 0,005 es menor que el nivel de significación α = 0,05. Por tanto, rechazamos H_0.

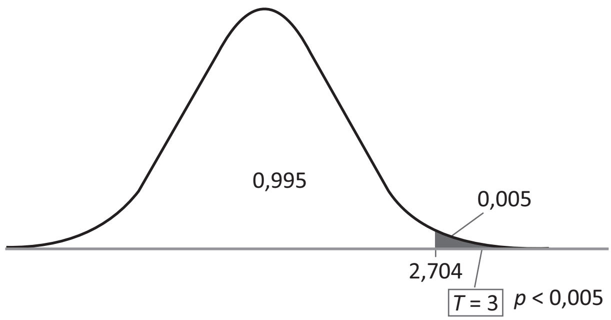


Figura 16: *Representación gráfica del nivel crítico p para el estadístico T = 3.*

INTERPRETACIÓN: Se confirma la hipótesis del investigador de que la percepción del tiempo se encuentra deteriorada en los fumadores durante la retirada de nicotina (T = 3; p < 0,005).

SITUACIÓN 11. Un sociólogo sabe, por investigaciones anteriores, que el sueldo medio de la población de hombres en el trabajo *T* es igual a 1500 euros mensuales. Extrae una muestra aleatoria de 160 mujeres que desempeñan el trabajo *T*, observando que el sueldo medio en dicha muestra es igual a 1400 euros mensuales con una cuasivarianza igual a 64000. Quiere comprobar si el sueldo medio de las mujeres es inferior al de los hombres. Nivel de confianza 95%.

SOLUCIÓN:

CONDICIONES Y SUPUESTOS: En este trabajo el investigador utiliza una muestra de 160 mujeres en el que la variable de estudio, el salario mensual, es una variable cuantitativa que se mide con escala de razón de la que se desconoce la forma de su distribución poblacional y su varianza. (Observe el estudiante que siendo esta situación muy similar a la presentada en la situación 1, las condiciones y supuestos son diferentes, lo que afecta a la elección del estadístico de contraste que hay que aplicar)

HIPÓTESIS: Como el investigador quiere comprobar si el sueldo medio de las mujeres es inferior a 1500 euros, plantea un contraste unilateral izquierdo

$$H_0 : \mu \geq 1.500 \qquad H_1 : \mu < 1.500$$

ESTADÍSTICO DE CONTRASTE: Bajo las condiciones y supuestos expuestos, la distribución muestral de la media es una distribución t de Student con n – 1 grados de libertad y el estadístico de contraste es:

$$T = \frac{\overline{Y} - \mu_0}{\sqrt{\frac{S_{n-1}^2}{n}}} = \frac{1400 - 1500}{\sqrt{\frac{64000}{160}}} = -5$$

La distribución t de Student se aproxima a la normal a medida que el tamaño de la muestra aumenta. En la tabla de la distribución t disponemos de información hasta 100 gl. Para valores mayores, las puntuaciones t y Z son prácticamente idénticas por lo que buscaríamos tanto el valor crítico como el nivel crítico p en la distribución normal tipificada de puntuaciones Z. (Ver el ejemplo de la situación 1).

CONCLUSIÓN: Al estadístico, $Z = -5$, le corresponde un nivel crítico $p < 0{,}0002$ por lo que se rechazaría la H_0. Por otra parte, el estadístico $Z = -5$, supera los valores críticos correspondientes tanto a un nivel de confianza del 95% como del 99% que nos lleva a la misma conclusión.

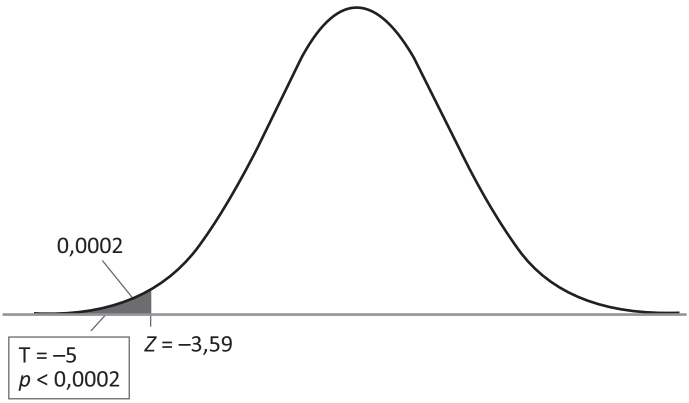


Figura 17: *: Representación del nivel crítico p en una distribución t con 159 gl para el estadístico t = –5.*

INTERPRETACIÓN: Con un nivel de confianza mayor del 99,98% se confirma que el sueldo medio de las mujeres en el trabajo *T* es significativamente menor de 1.500 euros, (T = –5; p < 0,0002) que es el que corresponde a la población de hombres para el mismo trabajo.

SITUACIÓN 12. Una preocupación creciente de los empresarios es el tiempo que los empleados dedican a actividades como navegar por Internet o enviar e-mails a los amigos durante las horas de trabajo. Para disponer de esta información, selecciona una muestra aleatoria de 26 empleados, observando que el promedio de tiempo dedicado a estas actividades durante una jornada laboral semanal fue de 126 minutos y una desviación típica de 15 minutos. Asumiendo que esta variable se distribuye normalmente en la población y con un nivel de confianza 95% desea conocer el intervalo de confianza del tiempo medio que dedican los empleados a estas actividades. Además desea comprobar si la cantidad promedio de tiempo perdido por sus empleados durante una jornada semanal de 40 horas es mayor de 120 minutos que es el valor obtenido el año anterior.

SOLUCIÓN:

El intervalo se obtiene sumando y restando a la media de la muestra el error máximo de estimación:

$$IC = \overline{Y} \pm E_{\max}$$

El error máximo de estimación se obtiene bien a partir de la cuasi-desviación típica de la muestra, $S_n - 1$, o de la desviación típica, S_n:

$$E_{\max} = t_{n-1;1-\alpha/2}\sigma_{\bar{Y}} = t_{n-1;1-\alpha/2}\frac{S_{n-1}}{\sqrt{n}} = t_{n-1;0,975}\frac{S_n}{\sqrt{n-1}}$$

Como conocemos la desviación típica de la muestra, utilizamos la segunda expresión para obtener el error máximo de estimación:

$$E_{\max} = t_{n-1;1-\alpha/2}\sigma_{\bar{Y}} = t_{n-1;0,975}\frac{S_n}{\sqrt{n-1}} = 2,060\frac{15}{\sqrt{26-1}} = 6,18$$

El intervalo de confianza con un nivel de confianza del 95% es:

$$IC = \bar{Y} \pm E_{\max} = 126 \pm 6,18 \rightarrow \begin{cases} 132,18 \\ 119,82 \end{cases}$$

Que indica que el tiempo medio que los empleados dedican a otras actividades es un valor comprendido entre 119,82 y 132,18 minutos de su jornada laboral semanal con un nivel de confianza del 95%.

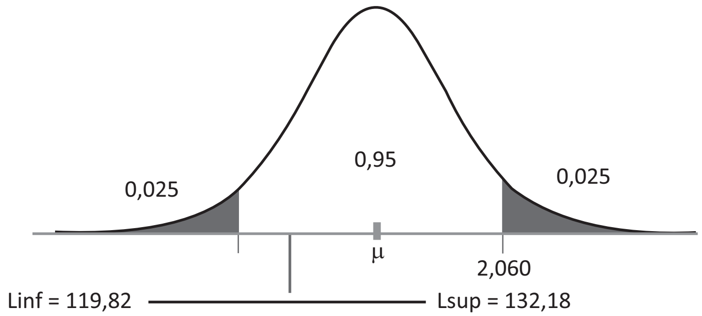


Figura 18: *Representación gráfica del intervalo de confianza de la media.*

Para el contraste de hipótesis seguiremos los siguientes pasos:

CONDICIONES Y SUPUESTOS: El investigador diseña una investigación en la que utiliza una muestra pequeña (n < 30). La variable de estudio «tiempo» dedicado a otras actividades es de naturaleza cuantitativa y medida con escala de razón. Aunque la muestra es pequeña se aplicará un contraste paramétrico ya que variable se distribuye normalmente en la población lo que asegura que la distribución muestral de la media sea una distribución t de Student.

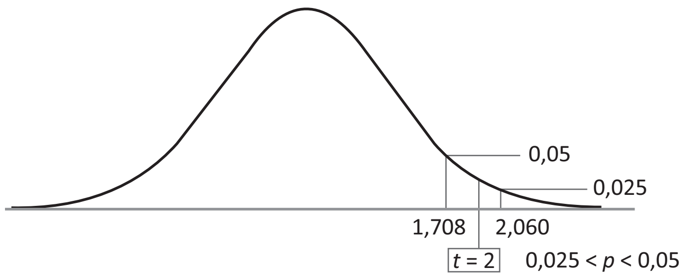
HIPÓTESIS: Se formula un contraste unilateral en la que el investigador quiere comprobar que el tiempo medio de dedicado a otras actividades es mayor de 120 minutos.

$$H_0 : \mu \leq 120 \qquad H_1 : \mu > 120$$

ESTADÍSTICO DE CONTRASTE: Aunque la muestra es pequeña, la distribución muestral de la media es la distribución t de Student porque proviene de una población con distribución normal.

$$T = \frac{\overline{Y} - \mu_0}{\frac{S_n}{\sqrt{n-1}}} = \frac{126-120}{\frac{15}{\sqrt{26-1}}} = \frac{6}{3} = 2$$

CONCLUSIÓN: Con un nivel de confianza del 95% en un contraste unilateral derecho, el valor crítico (1,708) hay que buscarlo en la distribución *t* de Student con grados de libertad. Como el estadístico de contraste supera esta máxima diferencia que cabe esperar por simple azar, rechazamos la H_0 con un nivel de confianza del 95%.



g.l.	Probabilidad											
	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950	0,975	0,990	0,995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	$\boxed{1,708}$	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771

Tabla 10: *Reproducción parcial de la tabla de la distribución t y representación del valor crítico y el nivel crítico p para el estadístico de contraste* **T = 2**.

Por otra parte, el estadístico de contraste $T = 2$ se encuentra entre los valores 1,708 y 2,060 que dejan por encima probabilidades de 0,05 y 0,025, respectivamente. Por tanto, el valor crítico asociado al estadístico de contraste $T = 2$ es $p < 0{,}05$ pero $p > 0{,}025$. De forma que no podríamos rechazar la Ho con un nivel de confianza del 99%

INTERPRETACIÓN. El investigador ha podido comprobar que, con un nivel de confianza del 95% el tiempo que dedican los empleados a actividades como navegar por internet o enviar e-mail privados es superior a los 120 minutos durante una jornada laboral semanal. Además el tiempo medio que dedican los empleados a esta actividad es un valor comprendido entre 119,82 y 132,18 minutos a la semana.

SITUACIÓN 13

SITUACIÓN 13. Chaves y Noguera (2008) estudiaron la inserción laboral de los titulados en Psicología de cinco promociones consecutivas (desde 2002 hasta 2006). Entre aquellos sujetos que estaban trabajando 6 meses después de finalizar sus estudios, el 81% son mujeres, el 31% accedieron al trabajo mediante contactos personales y el 54% trabajaban a jornada completa. Si el número de titulados que trabajaban 6 meses después de finalizar sus estudios es igual a 400, determine:

13.1. El intervalo de confianza para la proporción de mujeres tituladas en psicología que estaban trabajando 6 meses después de finalizar sus estudios, con un nivel de confianza del 99%.

13.2. Con un nivel de confianza del 99%, ¿cuál debería ser el tamaño de la muestra para estimar con un error máximo de estimación del 2% la proporción de titulados que accedieron al trabajo mediante contactos personales?

13.3. Si sospecha que la proporción de titulados que trabajan a jornada completa es distinta de los que trabajan a media jornada, ¿cuál sería su conclusión?

SOLUCIÓN:

13.1. A partir de una muestra de 400 titulados en psicología, el intervalo de confianza de la proporción de mujeres tituladas en Psicología que estaban trabajando 6 meses después de finalizar sus estudios, es:

$$E_{\max} = Z_{1-\alpha/2}\sqrt{\frac{p\cdot(1-p)}{n}} = 2{,}58\cdot\sqrt{\frac{0{,}81\cdot(1-0{,}81)}{400}} = 0{,}05061$$

$$L_{\text{inf}} = p - E_{\max} = 0{,}81 - 0{,}05061 = 0{,}759$$

$$L_{\text{sup}} = p + E_{\max} = 0{,}81 - 0{,}05061 = 0{,}861$$

13.2. Si el error máxima de estimación es del 2% (E_{max} = 0,02)

$$E_{\max} = Z_{1-\alpha/2}\sqrt{\frac{p\cdot(1-p)}{n}} \rightarrow 0,02 = 2,58\sqrt{\frac{0,31\cdot(1-0,31)}{n}}$$

Entonces, despejando el valor de n, el tamaño de la muestra debería ser:

$$n = p(1-p)\frac{z^2_{\alpha/2}}{E^2} = 0,31(1-0,31)\frac{2,58^2}{0,02^2} = 3560 \text{ titulados}$$

13.3. Para el contraste de hipótesis de que la proporción de titulados que trabajan a jornada completa es distinta de los que trabajan a media jornada, seguimos los siguientes pasos:

CONDICIONES Y SUPUESTOS: Se utiliza la información proporcionada por una muestra. La variable de estudio es la «proporción de titulados que trabajan a jornada completa» que es de naturaleza dicotómica y con distribución binomial que se aproxima a la normal cuando la muestra es grande (n > 25).

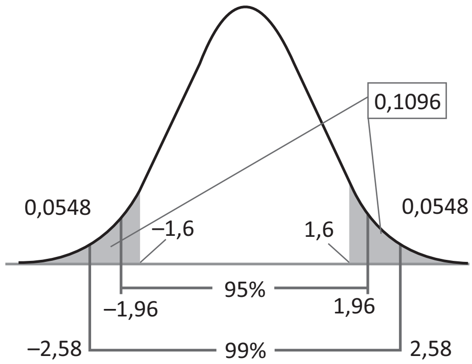
HIPÓTESIS: La hipótesis nula establece que no existen diferencias significativas entre la proporción de titulados que trabajan a jornada completa respecto a los que trabajan a media jornada. Si no existen diferencias, entonces estas proporciones son iguales a 0,50 (50%).

$$H_0 : \pi = 0,50; \qquad H_1 : \pi \neq 0,50$$

ESTADÍSTICO DE CONTRASTE: Se utiliza el estadístico Z por aproximación de la distribución binomial a la normal.

$$Z = \frac{P-\pi_0}{\sigma_p} = \frac{0,54-0,50}{\sqrt{\frac{0,50(1-0,50)}{400}}} = 1,6$$

REGLA DE DECISIÓN: Los valores críticos para un contraste bilateral con un nivel de confianza del 95% son –1,96 y +1,96 y con un nivel de confianza del 99% son –2,58 y +2,58. Por otra parte, en la distribución normal la puntuación típica Z = 1,6 deja por debajo una probabilidad de 0,9452 y por encima 1 – 0,9542 = 0,0548, con lo que el nivel crítico p para un contraste bilateral es 0,0548 + 0,0548 = 0,1096.



z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964

Tabla 11: *Tabla de la distribución normal para localizar los valores críticos y nivel crítico p para un estadístico Z = 1,6 en un contraste bilateral y representación gráfica de estos valores.*

CONCLUSIÓN: Como el nivel critíco *p* es mayor que los niveles de significación del 0,05 y del 0,01, no tenemos evidencia suficiente para rechazar la hipótesis nula. A la misma conclusión se llegaría comparando el estadístico de contraste con los valores críticos ya que el estadístico de contraste no supera ninguno de esos valores.

INTERPRETACIÓN: No hay evidencia suficiente para afirmar que la proporción de titulados que trabajan a jornada completa es distinta de los que trabajan a media jornada.

SITUACIÓN 14. *«PIAAC (Programme for International Assessment of Adult Skills) es un estudio internacional que mide las destrezas cognitivas básicas que per-*

mite a las personas adultas participar en la vida social y económica del siglo XXI y también las habilidades laborales básicas que exige su puesto de trabajo. En su último estudio realizado en el año 2013 comparando los resultados de 23 países de la OCDE muestra que los adultos españoles de entre 16 y 65 obtuvieron en matemáticas el último puesto, con una media de 246 puntos, 23 por debajo de la media de la OCDE y 22 por debajo de la media de la UE. Entre sus conclusiones destaca que los jóvenes españoles puntúan mucho más alto que los mayores y están más cerca de la media de la OCDE y que la tasa de paro de los españoles con competencias matemáticas en los niveles inferiores es tres veces superior de la que están en los niveles superiores». Imagine que usted aplica la misma prueba a una muestra representativa de 121 jóvenes de su localidad con edades comprendidas entre 16 y 34 años obteniendo una media de 252 puntos con una desviación típica insesgada de 77 puntos, siendo la proporción de jóvenes desempleados del 28,5%. Y estableciendo un nivel de confianza del 95%, desea conocer:

14.1. El intervalo de confianza de la media de la población de los jóvenes españoles en competencias matemáticas.

14.2. A partir de los datos de su estudio, y con un nivel de confianza del 95%, ¿entre que valores estimaría la proporción jóvenes desempleados en su localidad?

14.3. El intervalo de confianza de la varianza de las puntuaciones en competencias matemáticas de la población de jóvenes españoles de su localidad es

14.4. Si hubiéramos querido estimar la varianza poblacional con un error máximo de estimación que no supere los 200 puntos, ¿cuál debería haber sido, aproximadamente, el tamaño de la muestra seleccionada?

14.5. Si su interés es comprobar que la media en la prueba de competencias matemáticas de los jóvenes españoles de su localidad es significativamente inferior que la media general de los adultos de la UE, ¿cuál sería su conclusión?

SOLUCIÓN:

De los datos del enunciado extraemos la siguiente tabla resumen de los datos:

	España	OCDE	UE	Sus datos
Media	246	269	268	252
Cuasi-desvTipica				77
Tamaño muestra				121

14.1. Nos encontramos ante un diseño de una muestra en la que la variable de estudio: «las competencias matemáticas» es de tipo cuantitativo y medida con escala de intervalo. Se desconoce la forma de su distribución en la población y su varianza por lo que la distribución muestral de la media es la *t* de Student.

Para obtener el intervalo de confianza de la media de la prueba en la población de los jóvenes españoles en competencias matemáticas con un nivel de confianza del 95%, calculamos primero el error máximo de estimación.

$$E_{\max} = t_{1-\alpha/2} \frac{S_{n-1}}{\sqrt{n}} = 1,96 \frac{77}{\sqrt{121}} = 1,96 \cdot 7 = 13,72$$

Y los límites del intervalo son:

$$L_{\text{inf}} = \overline{Y} - E_{\max} = 252 - 13,72 = 238,28$$
$$L_{\text{sup}} = \overline{Y} + E_{\max} = 252 + 13,72 = 265,72$$

14.2. La proporción es una variable dicotómica con distribución binomial. Como la muestra es grande esta distribución se aproxima a la normal. Para obtener el intervalo de confianza de la proporción, calculamos primero el error máximo de estimación:

$$E_{maz} = Z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} = 1,96 \sqrt{\frac{0,285(1-0,285)}{121}} = 1,96 \cdot 0,041 = 0,08$$

Y los límites del intervalo son:

$$L_{\text{inf}} = p - E_{\max} = 0,285 - 0,08 = 0,205$$
$$L_{\text{sup}} = p + E_{\max} = 0,285 + 0,08 = 0,365$$

14.3. La varianza de una variable cuantitativa se distribuye según chi-cuadrado con $n - 1$ grados de libertad. Cuando la muestra es grande, como en este caso, la distribución chi-cuadrado se aproxima a la normal. Para obtener el intervalo de confianza de la varianza calculamos primero el error máximo de estimación:

$$E_{\max} = Z_{1-\alpha/2} S^2 \sqrt{\frac{2}{n}} = 1{,}96 \cdot 77^2 \sqrt{\frac{2}{121}} = 1494{,}03$$

Y el intervalo es:

$$IC = S^2 \pm E_{\max} = 77^2 \pm 1494{,}03 = \begin{cases} 4434{,}97 \\ 7423{,}03 \end{cases}$$

14.4. Si queremos estimar la varianza poblacional con un error máximo de estimación que no supere los 200 puntos, el tamaño de la muestra, con un nivel de confianza del 95%, debería ser:

$$n = 2S^4 \frac{Z^2_{1-\alpha/2}}{E^2_{\max}} = 2 \cdot 77^4 \cdot \frac{1{,}96^2}{200^2} \approx 6752$$

14.5. Para contrastar su hipótesis de que la media de los jóvenes españoles es significativamente inferior que la media general de los adultos de la UE, partimos de las mismas condiciones ya expuestos en el punto 14.1.

CONDICIONES Y SUPUESTOS: Nos encontramos ante un diseño de una muestra en la que la variable de estudio: «las competencias matemáticas» es de tipo cuantitativo y medida con escala de intervalo. Se desconoce la forma de su distribución en la población y su varianza por lo que la distribución muestral de la media es la *t* de Student.

HIPÓTESIS: Se plantea un contraste unilateral con su hipótesis dirigida a encontrar diferencias significativas entre el valor teórico formulado como parámetro poblacional (268) y el valor obtenido en su investigación:

$$H_0 : \mu \geq 268 \qquad H_1 : \mu < 268$$

ESTADÍSTICO DE CONTRASTE: Se aplica el estadístico *T* porque se trabaja con una muestra grande que proviene de una población de la que se desconoce la forma de la distribución y su varianza.

$$T=\frac{\overline{Y}-\mu_0}{\sigma_{\overline{Y}}}=\frac{\overline{Y}-\mu_0}{S_{n-1}/\sqrt{n}}=\frac{252-268}{77/\sqrt{121}}=-2{,}29$$

REGLA DE DECISIÓN: En una distribución t con 121 – 1 = 120 gl, el valor de t es prácticamente igual al de Z de la distribución $N(0;1)$. Y para un contraste unilateral izquierdo esta probabilidad el nivel crítico p asociado al estadístico de contraste obtenido es p = 0,0113.

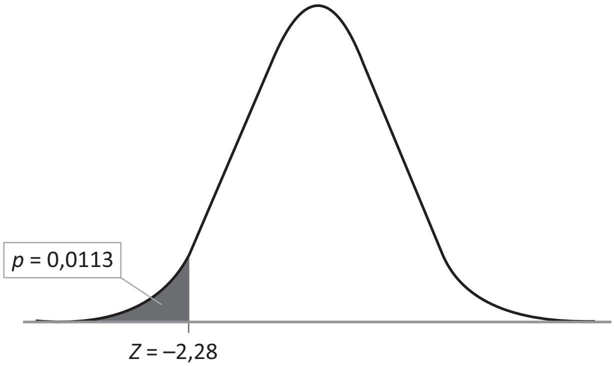


Figura 19: *Representación gráfica de la distribución t, el estadístico de contraste y su nivel crítico p.*

CONCLUSION: Como esta probabilidad es menor que el nivel de significación establecido en el 0,05 (nivel de confianza del 95%) rechazamos la H_0.

INTERPRETACIÓN: Con un nivel de confianza del 95%, las competencias matemáticas de los jóvenes españoles de su localidad significativamente inferior que la media general de los adultos de la UE.

SITUACIÓN 15. *Uno de los mejores indicadores de la salud psicológica es el auto-concepto (Esnaola, Goñi y Madariaga 2008) que hace referencia a la idea u opinión que cada persona tiene de sí misma.* En una investigación realizada por Esnaola y Revuelta (2009) con una muestra jóvenes de entre 12 y 24 años de la Comunidad Autónoma Cántabra y País Vasco se analiza la relación entre la actividad física y el auto-concepto. Por sus respuestas a un cuestionario sobre prácticas físico-deportivas el 31% fueron clasificados como inactivos y el resto como activos. En la escala de auto-concepto físico, la media y desviación típica insesgada de los inactivos fue de 22,41 y 4,1 respectivamente y en el grupo de activos, de 23,7 y 4,83.

15.1. Si con un nivel de confianza del 95% el error máximo para estimar la proporción poblacional de jóvenes activos vale 0,0523 ¿Cuál era el tamaño de la muestra utilizada en el trabajo de Esnaola y Revuelta?:

Suponiendo que el tamaño de la muestra fuera de 132 jóvenes, determine:

15.2. El intervalo de confianza de la varianza poblacional en la escala de auto-concepto físico para los jóvenes activos con edades comprendidas entre 12 y 24 años, con un nivel de confianza del 95%.

15.3. Si el investigador quisiera comprobar que la varianza poblacional de la escala de auto-concepto físico para los sujetos inactivos es significativamente distinta de 25 puntos, ¿Cuál sería su conclusión?

SOLUCIÓN:

Partimos de los siguientes datos resumidos proporcionados en el enunciado:

	Escala autoconcepto
ACTIVOS (69%)	Media = 23,7 Desv. Tip. Inses = 4,83
INACTIVOS (31%)	Media= 22,41 Desv. Tip. Inses = 4,1

15.1. El tamaño de la muestra necesario con un error máximo de estimación fijado en 0,0523 es:

$$E_{\max} = Z\sqrt{\frac{p \cdot q}{n}} \qquad 0,0523 = 1,96\sqrt{\frac{0,69 \cdot 0,31}{n}}$$

Y despejando el valor de *n*:

$$n = 1,96^2 \frac{0,69 \cdot 0,31}{0,0523^2} = 300,41 \cong 300 \text{ jóvenes}$$

15.2. Los jóvenes activos son el 69% de la muestra compuesta por 132 jóvenes. Por tanto, el grupo de jóvenes activos está formado por n = 0,69 × 132 = 91,08 = 91 jóvenes que obtienen en la escala una desviación típica insesgada de 4,83. Los límites del intervalo de confianza son:

$$\frac{(n-1)S_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \leq \sigma^2 \leq \frac{(n-1)S_{n-1}^2}{\chi^2_{\frac{\alpha}{2};n-1}} \qquad \frac{(91-1)\cdot 4,83^2}{118,1359} \leq \sigma^2 \leq \frac{(91-1)\cdot 4,83^2}{65,6466}$$

La varianza poblacional es un valor comprendido entre: $11,77 \leq \sigma^2 \leq 31,98$ con un nivel de confianza del 95%.

15.3. Para el contraste de esta hipótesis seguiremos los siguientes pasos:

CONDICIONES Y SUPUESTOS: La escala de autoconcepto es una variable medida con escala de intervalo aplicada a UNA muestra de jóvenes inactivos que provienen de una población con distribución desconocida. Como la muestra es grande ($n > 30$) podemos aplicar un contraste parámetrico.

HIPÓTESIS: En la muestra se obtiene una varianza de $4,1^2 = 16,81$. El investigador quiere comprobar que la varianza poblacional de la escala de auto-concepto físico para los sujetos inactivos es significativamente distinta de 25 puntos. Como no se marca la dirección (mayor o menor) de la diferencia se formula un contraste bilateral con las siguientes hipótesis:

$$H_0 : \sigma^2 = 25; \qquad H_1 : \sigma^2 \neq 25$$

ESTADÍSTICO DE CONTRASTE: Los sujetos activos son 41 que corresponden al 31% de la muestra de 132 jóvenes. Para estos 41 jóvenes inactivos la desviación típica insesgada en la escala es de 4,1 y el estadístico de contraste es:

$$\chi^2 = \frac{(n-1)S_{n-1}^2}{\sigma_o^2} = \frac{(41-1)4,1^2}{25} = 26,9$$

REGLA DE DECISIÓN: La tabla de la distribución chi-cuadrado proporciona probabilidades por debajo de un valor concreto de la distribución chi-cuadrado. En la tabla de la distribución chi-cuadrado con 40 gl figuran los siguientes valores: 24,433 que deja por debajo una probabilidad de 0,025 y 59,3417 que deja por debajo una probabilidad de 0,975. Estos dos valores corresponden a los valores críticos de una distribución chi-cuadrado con 40 gl para un contraste bilateral con un nivel de confianza del 95%.

Con un nivel de confianza del 99% los valores críticos son 20,7 que deja por debajo una probabilidad de 0,005 y 66,766 que deja por debajo una probabilidad de 0,995. Por tanto, ambos valores representan los valores críticos con un nivel de confianza del 99%.

g.l.	Probabilidad									
	0,005	**0,010**	**0,025**	**0,050**	**0,100**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	12,8382
29	13,1211	14,2565	16,0471	17,7084	19,7677	39,0875	42,5570	45,7223	49,5879	52,3356
30	13,7867	14,9535	16,7908	18,4927	20,5992	40,2560	43,7730	46,9792	50,8922	53,6720
40	20,7065	22,1643	24,4330	26,5093	29,0505	51,8051	55,7585	59,3417	63,6907	66,7660
50	27,9907	29,7067	32,3574	34,7643	37,6886	63,1671	67,5048	71,4202	76,1539	79,4900
60	35,5345	37,4849	40,4817	43,1880	46,4589	74,3970	79,0819	83,2977	88,3794	91,9517

Tabla 12: *Reproducción parcial de la tabla de la distribución chi-cuadrado.*

Por otra parte, para una distribución con 40 gl el valor que aparece en la tabla más próximo al estadístico de contraste obtenido es 26,5093 que deja por debajo una probabilidad de 0,05. Por tanto, el nivel crítico p para un contraste bilateral es $p > 0{,}05$.

CONCLUSIÓN: Como el estadístico de contraste no supera los valores críticos (ni por la derecha ni por la izquierda), no hay evidencia suficiente para rechazar la H_0 con un nivel de confianza del 95% ni del 99%. A la misma conclusión llegamos al comparar el nivel crítico p con el nivel de significación.

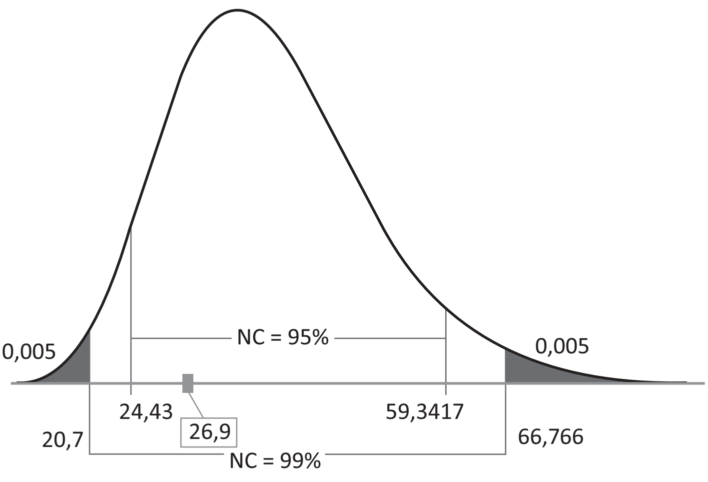


Figura 20: *Representación gráfica de la distribución chi-cuadrado con el estadístico de contraste y los valores críticos para los niveles de confianza del 95% y 99%.*

INTERPRETACIÓN: La varianza poblacional de la escala de auto-concepto físico para los sujetos inactivos no difiere de forma significativa de los 25 puntos, con un nivel de confianza del 95%.

SITUACIÓN 16 (amor de grillo). El artículo «*Well-Fed crickets bowl maidens over*» (*Nature Science Update*, 1999) informó que los grillos de campo hembras son atraídos por machos que tienen una frecuencia de canto (chirridos) elevada e hipotetizaron que la tasa a la que los grillos macho cantan está relacionada con su estado nutricional. La tasa usual del canto del grillo de campo macho es de 60 chirridos por segundo. Para investigar si la tasa de canto está relacionada con el estado nutricional, los investigadores alimentaron a grillos machos con una dieta alta en proteínas durante 8 días, después de lo cual se midió la tasa del canto. Esta tasa de canto en grillos macho con dieta alta en proteínas fue de 109. El tamaño muestral y la desviación típica fueron de $n = 32$ y $S_n = 40$. Utilice un $\alpha = 0{,}01$.

SOLUCIÓN:

Se trata de un estudio con un único grupo de 32 grillos cuya tasa o frecuencia de canto se mide Antes y Después de implementar una dieta alta en proteínas, aunque solo nos proporcionan los datos del post-test (109 Hz) indicándonos indirectamente que la tasa base de grillos no tratados (pre-test) es de 60 Hz. Cada grillo se mide dos veces: antes de alimentarlo con una dieta alta en proteínas y otra después de alimentarlo con esta dieta. Luego la tasa de canto de todos y cada uno de los grillos se ha medido dos veces. Esto significa que las puntuaciones de antes y de después están relacionadas. Es un diseño de dos grupos con medidas dependientes (diseño pre-post). No obstante, no podemos utilizar el contraste de dos muestras relacionadas para la media porque no se proporcionan los datos suficientes (no tenemos la media de las diferencias y su varianza). Pero podemos observar que tenemos datos suficientes para realizar un contraste de una media con varianza poblacional desconocida asumiendo como hipótesis nula que la media es 60 Hz.

Los datos de que disponemos en el enunciado son:

- Tratamiento: dieta alta en proteínas durante 8 días.
- Número de grillos = 32; Desviación típica muestral pos-tratamiento = $S_n = 40$. Tasa de canto después del tratamiento 109 chirridos por segundo (es decir, 109 Hz).
- Tasa usual de canto = 60. Podemos considerar que esta es la tasa de canto previa al tratamiento. Es decir, observamos que el estudio nos indica que se

ha pasado de 60 Hz (línea base supuesta *a priori*) a 109 Hz por efecto del tratamiento. ¿Será significativo este cambio debido al tratamiento o puede considerarse un efecto del azar?

- Variable dependiente: chirridos por segundo (Hz). Debido a que esta variable admite un valor de 0 como ausencia absoluta de canto (silencio), es una variable de razón.

CONDICIONES Y SUPUESTOS:

- La variable dependiente se encuentra medida en una escala de razón.
- No nos indican en el enunciado que la distribución de la variable dependiente sea normal en la población pero se cumple el supuesto de que $n > 30$.
- No conocemos la varianza poblacional.

HIPÓTESIS: Debemos plantear una hipótesis unilateral derecha ya que no resulta lógico en el marco del estudio plantear que incrementar el suplementos de proteínas pueda disminuir la tasa de canto.

$$H_0: \mu \leq 60$$
$$H_1: \mu > 60$$

ESTADÍSTICO DE CONTRASTE: Aplicaremos la prueba *t* de Student.

$$T = \frac{\bar{Y} - \mu_0}{\frac{S_n}{\sqrt{n-1}}} = \frac{109 - 60}{\frac{40}{\sqrt{32-1}}} = \frac{49}{7,18} = 6,82$$

REGLA DE DECISIÓN: Al ser un contraste unilateral derecho, solo existe un valor crítico. Debemos buscar en la Tabla de la *T* con 32 – 1 = 31 grados de libertad al 99% de confianza. Utilizamos 30 grados de libertad porque es el más cercano a 31. El valor obtenido es $t = 2,457$.

g.l.	Probabilidad											
	0,550	**0,600**	**0,650**	**0,700**	**0,750**	**0,800**	**0,850**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660
70	0,126	0,254	0,387	0,527	0,678	0,847	1,044	1,294	1,667	1,994	2,381	2,648
80	0,126	0,254	0,387	0,526	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639

CONCLUSIÓN: Como el estadístico de contraste obtenido (T = 6,82) es mayor que el valor crítico (t = 2,457), es decir, 6,82 se encuentra en la región de rechazo de H_0. Por consiguiente, rechazamos H_0.

INTERPRETACIÓN: Debemos concluir que el tratamiento alimentario ha sido efectivo en el incremento de la frecuencia del chirrido de los grillos macho (T = 6,82; p < 0,005).