

CUESTIONES DE INFERENCIAS SOBRE DISTRIBUCIONES NORMALES

Curso 2018/2019

Matemáticas 1º. Grado en Bioquímica.

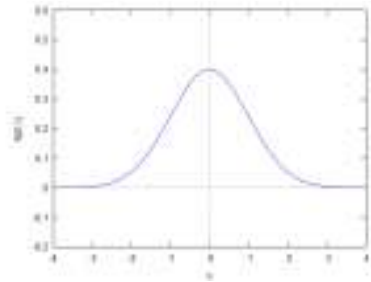
Universidad de Murcia.

I. LA NORMAL $N(0, 1)$. INTERVALOS CARACTERÍSTICOS. TIPIFICACIÓN DE UNA VARIABLE NORMAL.

INTERVALOS CARACTERÍSTICOS

Recordemos que la normal $N(0,1)$ es la distribución continua cuya función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

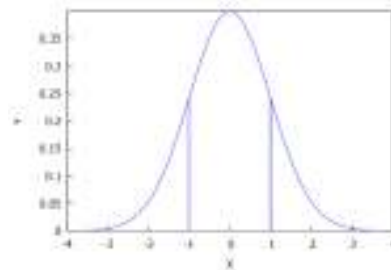


Recordemos también que, dada una variable aleatoria X con distribución normal $N(0,1)$, la probabilidad de que X se encuentre entre dos valores a y b es justamente

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Nos podemos preguntar para qué intervalo, centrado alrededor de la media $\mu=0$, se tiene que la probabilidad de que la variable aleatoria X , de tipo $N(0,1)$, caiga en él es justamente $p \in [0,1]$:

$$P(\mu - k \leq X \leq \mu + k) = p.$$



Si llamamos $\alpha=1-p$, buscamos un valor $k = Z_{\frac{\alpha}{2}}$ tal que $P\left(-Z_{\frac{\alpha}{2}} \leq X \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$. Por las propiedades de simetría de $f(x)$, la condición anterior equivale a decir que dicho valor debe cumplir $P\left(-\infty \leq X \leq -Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} = P\left(Z_{\frac{\alpha}{2}} \leq X \leq +\infty\right)$ o bien que

$$P\left(-\infty \leq x \leq Z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

El intervalo $\left[-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right]$ se conoce como el *intervalo característico* para el cual la probabilidad de que la variable aleatoria X caiga en él es precisamente $1 - \alpha$.

EJEMPLO:

1. Para una probabilidad del 90%, es decir, con $p=0.9$, esto es, $\alpha=0.1$, en una tabla de la normal $N(0,1)$ encontramos que $z_{\frac{\alpha}{2}} = 1.645$, de modo que el intervalo característico, en una normal $N(0,1)$, correspondiente a $p=0.9$ es

$$[-z_{0.05}, z_{0.05}] = [-1.645, 1.645].$$

2. Por otra parte, consultando una tabla de la normal $N(0,1)$, podemos comprobar que a los intervalos $[-1,1], [-2,2], [-3,3]$ (observemos que $\sigma=1$), les corresponden las probabilidades $p=0.6826, p=0.9544, p=0.9974$, respectivamente. Por ejemplo, para $[-1,1]$, en una tabla encontramos que $0.8413 = P(-\infty \leq X \leq 1) = 1 - \frac{\alpha}{2}$, de manera que se deduce que $\alpha = 2(1 - 0.8413) = 0.3174$, luego $1 - \alpha = 0.6826$. Para $[-2,2]$, ahora $0.9772 = P(-\infty \leq X \leq 2) = 1 - \frac{\alpha}{2}$, de modo que $\alpha = 2(1 - 0.9772) = 0.0456$, así que $1 - \alpha = 0.9544$. Para acabar, $0.9987 = P(-\infty \leq X \leq 3) = 1 - \frac{\alpha}{2}$, de manera que $\alpha = 2(1 - 0.9987) = 0.0026$, luego $1 - \alpha = 0.9974$.

De otro modo, podemos decir que

$$z_{\frac{0.3174}{2}} = z_{0.1587} = 1,$$

$$z_{\frac{0.0456}{2}} = z_{0.0228} = 2,$$

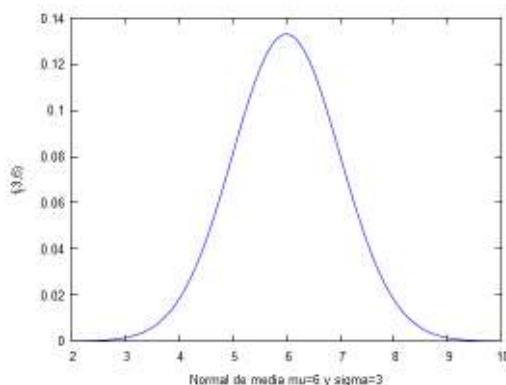
$$z_{\frac{0.0026}{2}} = z_{0.0013} = 3.$$

En la normal $N(0,1)$ (o en cualquier normal $N(\mu, \sigma)$) los datos anteriores se recogen en la llamada regla del 68-95-99, los tres porcentajes que corresponden a los intervalos $[\mu - \sigma, \mu + \sigma], [\mu - 2\sigma, \mu + 2\sigma], [\mu - 3\sigma, \mu + 3\sigma]$, respectivamente.

TIPIFICACIÓN DE UNA VARIABLE NORMAL $N(\mu, \sigma)$.

Cuando la distribución de la variable aleatoria X es una normal $N(\mu, \sigma)$, es decir, con una función de densidad

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$



podemos **“tipificar la variable”** de manera que el cambio de variables $Y = \frac{X-\mu}{\sigma}$ nos transforma la normal $N(\mu, \sigma)$ en una de tipo $N(0,1)$.

De esta forma, por ejemplo, calcular la probabilidad de que una variable normal X , de media $\mu = 6$ y desviación típica $\sigma = 3$, caiga en el intervalo $[5,7]$ se puede calcular relacionando $N(6,3)$ con $N(0,1)$ del modo siguiente (tenemos en cuenta que $X = \mu + \sigma Y$)

$$\int_5^7 f_{6,3}(x)dx = P(5 \leq X \leq 7) = P(5 \leq \mu + \sigma Y \leq 7) = P\left(\frac{5 - \mu}{\sigma} \leq Y \leq \frac{5 + \mu}{\sigma}\right)$$

$$= P\left(\frac{-1}{3} \leq Y \leq \frac{1}{3}\right) = \int_{-\frac{1}{3}}^{\frac{1}{3}} f_{0,1}(x)dx = 2 \int_{-\infty}^{\frac{1}{3}} f_{0,1}(x)dx - 1 = 2(0.6306) - 1$$

$$= 0.2612,$$

es decir, una probabilidad de un 26.12%. (En el cálculo de la última integral se ha hecho uso de las tablas de la normal interpolando las cantidades 0.6293 y 0.6331, correspondientes a los valores 0.33 y 0.34; también se puede usar WxMaxima, dando la instrucción **integrate(1/sqrt(2*pi)*exp(-x^2/2), x, -inf, 1/3)**).

Si, en general, queremos averiguar $\int_{\mu-k}^{\mu+k} f_{\mu,\sigma}(x)dx$, siguiendo el mismo procedimiento que hemos visto en el ejemplo anterior, tendremos que

$$P(\mu - k \leq X \leq \mu + k) = \int_{\mu-k}^{\mu+k} f_{\mu,\sigma}(x)dx = \int_{-\frac{k}{\sigma}}^{\frac{k}{\sigma}} f_{0,1}(x)dx = P\left(\frac{k}{\sigma} \leq Y \leq \frac{k}{\sigma}\right).$$

Así, si nos piden cuál es el intervalo característico $[\mu - k, \mu + k]$ correspondiente a una probabilidad $p = 1 - \alpha$, es decir, para el cual se tiene que

$$P(\mu - k \leq X \leq \mu + k) = \int_{\mu-k}^{\mu+k} f_{\mu,\sigma}(x)dx = 1 - \alpha,$$

deducimos que esto equivale a que

$$\int_{-\frac{k}{\sigma}}^{\frac{k}{\sigma}} f_{0,1}(x)dx = P\left(\frac{-k}{\sigma} \leq Y \leq \frac{k}{\sigma}\right) = 1 - \alpha.$$

Si recordamos ahora cómo se ha definido el intervalo característico $\left[-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right]$ para la normal $N(0,1)$, encontramos que $\frac{k}{\sigma} = Z_{\frac{\alpha}{2}}$, y al despejar resulta que $k = \sigma Z_{\frac{\alpha}{2}}$.

CONCLUSIÓN: El **intervalo característico de una normal** $N(\mu, \sigma)$ correspondiente a $1 - \alpha$ viene dado por

$$\left[\mu - \sigma Z_{\frac{\alpha}{2}}, \mu + \sigma Z_{\frac{\alpha}{2}}\right],$$

en donde $Z_{\frac{\alpha}{2}}$ es el valor para el cual $\int_{-Z_{\frac{\alpha}{2}}}^{Z_{\frac{\alpha}{2}}} f_{0,1}(x)dx = 2 \int_{-\infty}^{Z_{\frac{\alpha}{2}}} f_{0,1}(x)dx - 1 = 1 - \alpha$.

EJERCICIOS:

- 1) Crear una rutina con WxMaxima que calcule el valor de $Z_{\frac{\alpha}{2}}$ tal que

$$P\left(-\infty \leq x \leq Z_{\frac{\alpha}{2}}\right) = \int_{-\infty}^{Z_{\frac{\alpha}{2}}} f_{0,1}(x) dx = 1 - \frac{\alpha}{2}.$$

- 2) Calcular los intervalos característicos correspondientes a una normal $N(0,1)$ y a las probabilidades $p = 1 - \alpha = 0.9$, $p = 0.95$, $p = 0.99$.
- 3) Calcular los intervalos característicos correspondientes a una normal $N(20,2)$ y a las probabilidades $p = 0.9$, $p = 0.95$, $p = 0.99$.

Solución: 1) `l(z):=integrate(1/sqrt(2*pi)*exp(-x^2/2),x,-inf,z);`

`z(a):=find_root(l(z)=1-a, z, 0, 100);` por ejemplo, $z_{0.05}$ será $z(0.05)$.

2) $\alpha = 0.1 \rightarrow z_{0.05} = 1.644853626951474$,

$\alpha = 0.05 \rightarrow z_{0.025} = 1.959963984540056$,

$\alpha = 0.01 \rightarrow z_{0.005} = 2.575829303548914$. 1.959963984540056 ;

3) Para $\alpha = 0.1$,

$$\left[\mu - \sigma z_{\frac{\alpha}{2}}, \mu + \sigma z_{\frac{\alpha}{2}} \right] = 20 \pm 2 \cdot 1.644853626951474 = [16.71029274609705, 23.28970725390295].$$

- 4) Se sabe que la concentración de amoniaco en sangre venosa en individuos sanos se distribuye según una normal $N(\mu, \sigma)$, con $\mu = 110$ microgramos/mm³ y desviación típica desconocida.
- (a) Sabiendo que el 99% de la concentración de amoniaco de los individuos se encuentra en el intervalo [85, 135], calcula la desviación típica de la distribución normal.
- (b) Calcular el intervalo característico que contiene el 70% de los valores de dicha concentración.
- (c) Si un individuo tiene una concentración de 70 microgramos/mm³, ¿qué porcentaje de la población tiene una concentración superior a la de ese individuo?
- (d) ¿Qué concentración verifica que tan sólo un 25% de los individuos tiene una concentración mayor?

II. INFERENCIA PARA DISTRIBUCIONES NORMALES.

Empezaremos detallando algunos resultados teóricos, que daremos sin demostración.

1. Es sencillo demostrar que si tenemos n variables aleatorias X_1, \dots, X_N con la misma media μ y la misma desviación típica σ , entonces la distribución de las medias

$$\bar{X} = \frac{X_1 + \dots + X_N}{N}$$

tiene la misma media μ y desviación típica igual a $\frac{\sigma}{\sqrt{N}}$. Incluso, para N suficientemente grande, se sabe que la distribución de \bar{X} es prácticamente normal. (Teorema central del límite).

De hecho,

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$$

o equivalentemente

$$\sqrt{N} \frac{\bar{X} - \mu}{\sigma} \rightarrow N(0,1).$$

2. Si la distribución de las variables aleatorias X_1, \dots, X_N es la misma, y de tipo normal $N(\mu, \sigma)$, entonces

$$\bar{X} = N\left(\mu, \frac{\sigma}{\sqrt{N}}\right).$$

3. Si la distribución de las variables aleatorias X_1, \dots, X_n es la misma, de tipo normal $N(\mu, \sigma)$, entonces resulta que la distribución de

$$\frac{(n-1)s_{n-1}^2}{\sigma^2}$$

es una χ^2 cuadrado con $n-1$ grados de libertad, en donde $s_{n-1} = \sqrt{\frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1}}$.

Por otra parte, se tiene que

$$\sqrt{n} \frac{(\bar{X} - \mu)}{s_{n-1}}$$

presenta una distribución t de Student con $n-1$ grados de libertad.

A. La distribución t de Student.

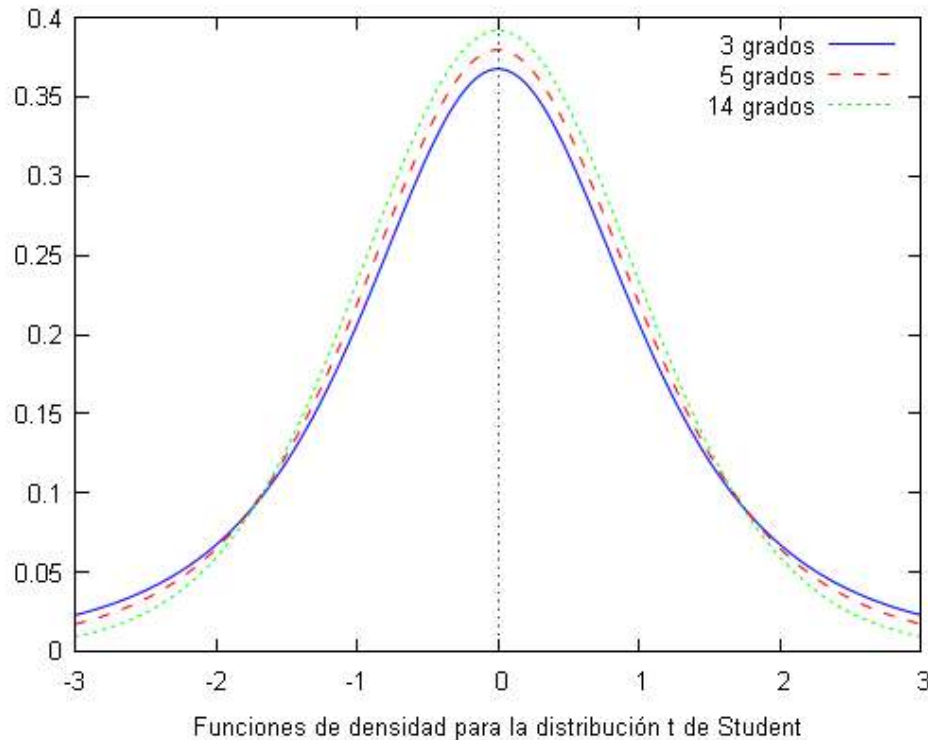
- (1) La distribución t de Student con m grados de libertad corresponde a una variable aleatoria de tipo (absolutamente) continuo, cuya función de densidad viene dada por

$$f_m(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi} \Gamma\left(\frac{m}{2}\right)} \cdot \left(1 + \frac{x^2}{m}\right)^{-\frac{(m+1)}{2}}, \text{ con } x \in \mathbb{R}.$$

La cantidad $\frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi} \Gamma\left(\frac{m}{2}\right)}$ es una constante que envuelve a la función Γ de Euler, función que generaliza el factorial $n! = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$, y que para valores $z > 0$ se define como la integral impropia $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$. Se puede probar que $\Gamma(z+1) = z\Gamma(z)$. Además, $\Gamma(1) = 1$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ y $\Gamma(n) = (n-1)!$ para cada entero positivo n . De manera que para cualquier p

entero positivo se cumple que $\Gamma\left(\frac{p}{2}\right) = \left(\frac{p}{2} - 1\right)!$ cuando p es par, mientras que para p impar se cumple $\Gamma\left(\frac{p}{2}\right) = \frac{p-2}{2} \cdot \frac{p-4}{2} \cdot \dots \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi}$.

- (2) Algunas de las gráficas de la función de densidad $f_m(x)$ de la distribución t de Student son éstas (para hacer las gráficas con WxMaxima, usamos que $f_m(x)$ aparece como pdf_student_t(x,m)):



- (3) Para $\alpha \in (0,1)$, el valor $t_{\frac{\alpha}{2},m}$ es aquel valor $x \in \mathbb{R}$ tal que

$$P\left(|t| > t_{\frac{\alpha}{2},m}\right) = \alpha.$$

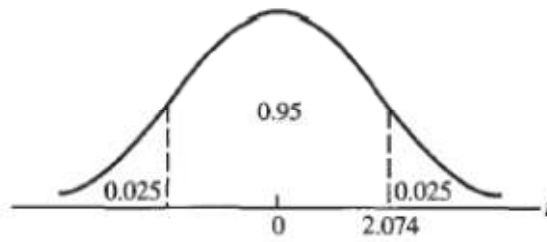
Recuerda que el área que barre la función de densidad es 1, luego nos piden el valor positivo $t_{\frac{\alpha}{2},m}$ de modo que las colas $t < -t_{\frac{\alpha}{2},m}$ y $t > t_{\frac{\alpha}{2},m}$ barren un área total igual a α .

Por la simetría de $f(x)$, cada cola barrerá la misma cantidad de área, a saber, $\alpha/2$.

Por ejemplo, en la siguiente figura hemos obtenido, para una cierta distribución t de Student de m grados de libertad, un valor $t_{\frac{\alpha}{2},m} = t_{0.025,m} = 2.074$ (así que $\alpha =$

0.05), lo que nos indica que

$$\begin{aligned} P\left(|t| > t_{\frac{\alpha}{2},m}\right) = \alpha &\rightarrow P\left(t < -t_{\frac{\alpha}{2},m}\right) + P\left(t > t_{\frac{\alpha}{2},m}\right) = \alpha \\ &\rightarrow P(t < -2.074) + P(t > 2.074) = 0.05 \end{aligned}$$



Para averiguar cuáles son los valores $t_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2}, n-2}$ que corresponden a una t de Student de $n-2$ grados de libertad, tendremos que consultar apropiadas tablas de valores, del tipo de las que vimos en el caso de la normal $N(0,1)$.

(Tabla tomada de [Ross])

TABLE A3 Values of $t_{\alpha, n}$

n	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.474	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

Other t Probabilities:

$P\{T_8 < 2.541\} = .9825$ $P\{T_8 < 2.7\} = .9864$ $P\{T_{11} < .7635\} = .77$ $P\{T_{11} < .934\} = .81$ $P\{T_{11} < 1.66\} = .94$ $P\{T_{12} < 2.8\} = .984$.

Según esta tabla, si $t_{0.025} = 2.074 = t_{0.025, n-2}$, vemos que hay $n-2=22$ grados de libertad.

Con WxMaxima podemos calcular esos valores $t_{a,n}$ usando el comando (hay que ejecutar primero load(distrib))

quantile_student_t(1-a,n);

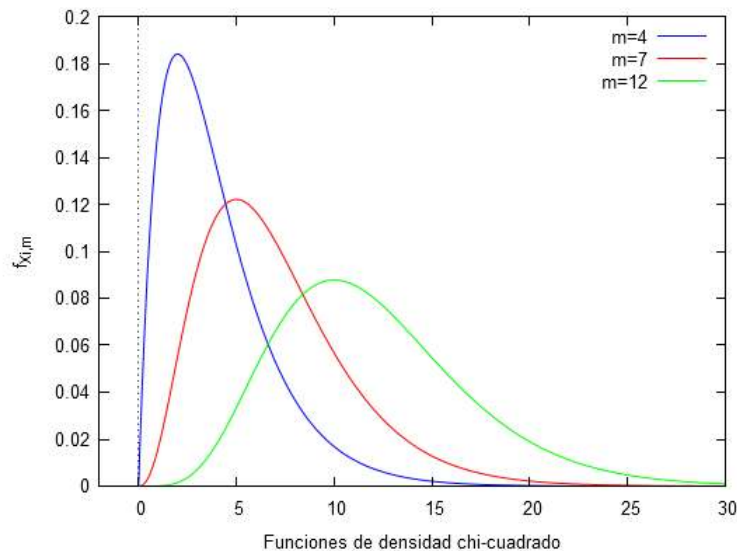
con el que se busca el valor x tal que $P(X \leq x) = \int_{-\infty}^x f_n(s) ds = 1 - a$.

De este modo, $t_{0.025,22} = \text{quantile_student_t}(1 - 0.025, 22) = 2.073873067903899$.

B. En lo que respecta a la distribución χ_i cuadrado con m grados de libertad, digamos que se trata de una variable continua cuya distribución viene dada a través de la función de densidad

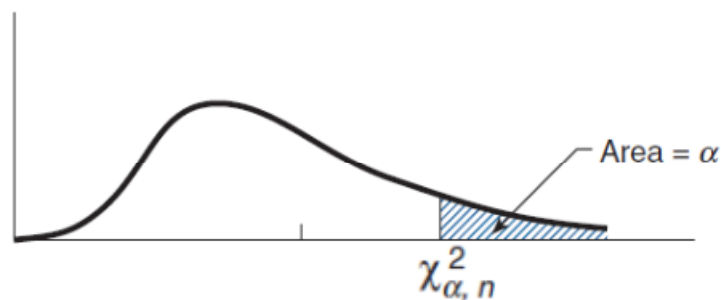
$$f_{\chi,m}(x) = \begin{cases} \frac{2^{-(m/2)}}{\Gamma(\frac{m}{2})} \cdot x^{\frac{m}{2}-1} \cdot e^{-\frac{x}{2}}, & \text{si } x > 0 \\ 0, & \text{si } x \leq 0, \end{cases}$$

en donde $\Gamma(\cdot)$ representa la función gamma de Euler.



Para dibujar con WxMaxima las funciones de densidad $f_{\chi,m}(x)$ de la distribución χ_i -cuadrado, tras cargar el paquete de distribuciones con la orden load(distrib), dibujamos la función mediante la orden wxplot2d([pdf_chi2(x,m)], [x,xmin,xmax]).

Para un valor $\alpha \in (0,1)$, y n grados de libertad, la cantidad $\chi^2_{\alpha,n}$ se define como el valor para el cual se cumple que $P(X \geq \chi^2_{\alpha,n}) = \alpha$.



Si usamos WxMaxima, para obtener esos valores $\chi^2_{\alpha,n}$, tras escribir load(distrib) para cargar el paquete de funciones de distribución, haremos

$$\chi^2_{\alpha,n} : \text{quantile_chi2}(1-\alpha,n).$$

Por ejemplo,

$$\chi^2_{0.01,8} = \text{quantile_chi2}(0.99,8) = 20.09023502966323.$$

También se puede consultar una tabla, como la que aparece a continuación, tomada de [Ross].

TABLE A2 Values of $\chi^2_{\alpha,n}$

<i>n</i>	$\alpha = .995$	$\alpha = .99$	$\alpha = .975$	$\alpha = .95$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597
3	.0717	.115	.216	.352	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	11.070	12.832	13.086	16.750
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.844	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.772	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672

Other Chi-Square Probabilities:

$$\chi^2_{9,9} = 4.2 \quad P[\chi^2_{16} < 14.3] = .425 \quad P[\chi^2_{11} < 17.1875] = .8976.$$

Con los ingredientes anteriores, podemos resolver los siguientes **problemas de inferencia sobre la media y la desviación típica**.

- A) Supongamos que **desconocemos la media μ** de una variable aleatoria continua de una población. No obstante, supongamos que **conocemos la desviación típica σ** de dicha característica. Tomamos una muestra de la población. Averiguamos entonces el valor \bar{x} de la media de la muestra. Si suponemos que la variable aleatoria sigue una distribución normal o que el tamaño de la muestra es suficientemente grande, ¿cuál es entonces el intervalo de confianza para la media μ de la población con un nivel de confianza del $(1-\alpha)\cdot 100\%$? La respuesta es

$$\mu \in \left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right),$$

con un nivel de confianza $1-\alpha$.

- B) Supongamos ahora que **desconocemos tanto la media μ como la desviación típica σ** de una variable aleatoria continua de tipo normal para una cierta población. Tomamos una muestra de n elementos y calculamos su media \bar{x} . Entonces, ¿cuál es el intervalo de confianza para la media μ de la población, con un nivel de confianza del $(1-\alpha)\cdot 100\%$? La respuesta es

$$\mu \in \left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s_{n-1}}{\sqrt{n}}, \quad \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s_{n-1}}{\sqrt{n}} \right),$$

con un nivel de confianza de $1-\alpha$. Recordemos que $t_{\frac{\alpha}{2}, n-1}$ hace referencia al correspondiente valor de la t de Student con $n-1$ grados de libertad y que $s_{n-1} =$

$$\sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}}.$$

- C) En cuanto a los **intervalos de confianza para la varianza (desconocida)** de una distribución normal $N(\mu, \sigma)$, supongamos que tomamos una muestra de tamaño n de la población, de la que calculamos su media \bar{x} y su "cuasi-varianza" $s_{n-1}^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}$. Entonces, el intervalo de confianza para σ^2 viene dado por

$$\left(\frac{(n-1)s_{n-1}^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \quad \frac{(n-1)s_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right)$$

Para calcular $\chi^2_{\frac{\alpha}{2}, n-1}$ y $\chi^2_{1-\frac{\alpha}{2}, n-1}$, necesitamos acudir a unas tablas para la distribución chi-cuadrado o usar el comando apropiado de WxMaxima.

- D) También podemos dar un intervalo de confianza, con un nivel de confianza $(1-\alpha)\cdot 100\%$, para la diferencia $\mu_1 - \mu_2$ de dos medias, si se conoce la media \bar{x}, \bar{y} de dos muestras de poblaciones normales, de tamaños n y m , respectivamente, y **se conocen las desviaciones típicas σ_1, σ_2** de las poblaciones; en este caso,

$$\mu_1 - \mu_2 \in \left(\bar{x} - \bar{y} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \quad \bar{x} - \bar{y} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right).$$

Si, en cambio, **se desconoce tanto las medias de las poblaciones como sus desviaciones típicas**, para dar un intervalo de confianza para la diferencia de las medias de dos poblaciones, con un nivel de confianza del $(1-\alpha)\cdot 100\%$, haremos

$$\mu_1 - \mu_2 \in \left(\bar{x} - \bar{y} - t_{\frac{\alpha}{2}, n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \quad \bar{x} - \bar{y} + t_{\frac{\alpha}{2}, n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right),$$

siendo $S_p = \sqrt{\frac{(n-1)s_{1,n-1}^2 + (m-1)s_{2,m-1}^2}{n+m-2}}$ y $s_{1,n-1}^2, s_{2,m-1}^2$ las cuasi-varianzas de cada una de las muestras de tamaño n y m .

EJERCICIOS:

- Se ha medido la concentración de policlorobifenilos (PCB) de una cierta especie de pez capturado en el Lago Michigan, usando una técnica de la que se sabe que produce un error normalmente distribuido con una desviación estándar σ de 0.08 ppm (partes por millón). Supongamos que los resultados de 10 medidas independientes para la concentración de PCB en este pez son

11.2, 12.4, 10.8, 11.6, 12.5, 10.1, 11.0, 12.2, 12.4, 10.6.

Dar un intervalo para el nivel de PCB de este pez con un nivel de confianza del 95% (Usar el apartado (A)).

- Mediante un método diseñado por el químico Karl Fischer, para determinar el porcentaje de agua en una solución de metanol, se han obtenido los siguientes datos en un conjunto de 10 muestras

.50, .55, .53, .56, .54, .57, .52, .60, .55, .58

Suponiendo que la variable aleatoria es normal, usa esos datos para dar un intervalo de confianza al 95% para el porcentaje real.

- Se han obtenido los siguientes resultados en 24 medidas independientes del punto de fusión del plomo

330°C	322°C	345°C
328.6°C	331°C	342°C
342.4°C	340.4°C	329.7°C
334°C	326.5°C	325.8°C
337.5°C	327.3°C	322.6°C
341°C	340°C	333°C
343.3°C	331°C	341°C
329.5°C	332.3°C	340°C

Suponiendo que las medidas se pueden considerar como una muestra de tipo normal cuya verdadera media μ es el verdadero punto de fusión del plomo, determinar: un intervalo de confianza para este valor del 95%; un intervalo de confianza del 99%. (Usar el apartado (B)).

- Se pretende que un procedimiento para fabricar arandelas lo haga de modo que se consigan pequeñas desviaciones en sus espesores. Se toma una muestra de 10 de tales arandelas y se miden los espesores siguientes (en pulgadas, una pulgada son 25,40 mm):

0.123 0.133 0.124 0.125 0.126 0.128 0.120 0.124 0.130 0.126

¿Cuál es entonces el intervalo de confianza al 90% para la desviación típica del espesor de estas arandelas? (Usar el apartado C).

5. Dos analistas han tomado varias lecturas sobre la dureza del agua de una ciudad. Suponiendo que las lecturas proporcionan una muestra de una población normal con varianzas desconocidas, determinar un intervalo de confianza del 95% para la diferencia de las medias de dichas poblaciones. Las medidas son estas:

Analista 1: 0.46 // 0.62 // 0.37 // 0.40 // 0.44 // 0.58 // 0.48 // 0.53

Analista 2: 0.82 // 0.61 // 0.89 // 0.51 // 0.33 // 0.48 // 0.23 // 0.25 // 0.67 // 0.88

REFERENCIAS

El alumno interesado en las justificaciones de las fórmulas dadas anteriormente en relación a inferencias sobre la normal, así como en otras cuestiones de inferencias, puede consultar los siguientes manuales.

[Milton] J. S. Milton. *Estadística para Biología y Ciencias de la Salud*.

3ª Edición ampliada. McGraw-Hill Interamericana. Madrid, 2007.

[Ross] S. M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*.

3rd Edition. Elsevier Academic Press. San Diego, CA, USA. 2007.

[Walpole] R. E. Walpole, R. H. Myers, S. L. Myers, K. E. Ye.

Probability and Statistics for Engineers and Scientists. 9th Edition.

Pearson. Upper Saddle River, New Jersey. 2012.