Auxiliar 11

Previo C3-Estadísticos e intervalos de confianza.

Profesor: Vicente Acuña
Auxiliares: Sebastián López, Bruno Hernández

P1. Un emisor emite una señal de tamaño θ (consante) de tal forma que el receptor recibe $Y = \theta + X$ con $X \sim \mathcal{N}(0, 4)$.

a) Como θ es desconocido para el receptor, éste toma una M.A.S. $Y_1, \cdots, Y_{15}$ obteniendo $\bar{Y} = 15$. Construya un intervalo de confianza del 90 % para θ.

b) ¿Con qué confianza está construido el intervalo dado por $(14{,}5, 15, 5)$?

c) El emisor afirma que su señal θ es mayor a 14,5. ¿Qué puede concluir?

d) El emisor envía la señal dos veces (de forma independiente). Calcule la probabilidad de que las señales recibidas difieran en menos de una unidad.

P2. Se le ha encargado hacer un estudio acerca de la concentración de metano en un cierto depósito. La variable a estudiar se modela como una $\mathbb{N}(\mu, \sigma^2)$.

a) El dueño del depósito le proporciona σ^2. Usando TCL y 50 muestras, calcule la probabilidad de que la concentración media sea mayor a un valor C.

b) Por un despiste, usted pierde el dato que le proporciono el dueño. Por esto, usted toma 20 nuevas muestras y obtiene un promedio de 100 y una desviación estandar de 4. Obtenga intervalos de confianza para μ y σ^2 al 95 %.

P3. En un laboratorio se desea estudiar la variabilidad de las mediciones tomadas en un complejo experimento.
Se tomaron 6 mediciones:

$$9,54 \quad 9,61 \quad 9,32 \quad 9,48 \quad 9,70 \quad 9,26$$

Suponiendo que ellos provienen de una distribución normal, obtenga un intervalo de confianza de la varianza σ^2 al 90 %.

- Sean N y n el número de elementos en una población y una muestra rspectivamente. Si el muestreo se realiza de manera que cada una de las $\binom{N}{n}$ posibles muestras tiene la misma probabilidad de ser elegida, entonces decimos que es un **muestreo aleatorio simple**.

  Además definimos un **Estadístico** a cualquier función $T(X)$ que dependa sólamente de una muestra dada.

- Dada $y_1, y_2, \cdots, y_n$ una muestra de tamaño n, definimos la media muestral como:
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
  También definimos la varianza muestral S^2 como:
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- Sean $Y_1, \cdots, Y_n$ una muestra aleatoria de tamaño n, de una distribución normal con media μ y varianza σ^2. Entonces $\bar{Y}$ está distribuída normalmente con media $\mu_{\bar{Y}=\mu}$ y varianza $\sigma_{\bar{Y}^2=\sigma^2/n}$.

- **T.C.L: Teorema central del límite.** Sean $Y_1, , Y_n$ v.a.i.i.d. con $\mathbb{E}(Y_i) = \mu$ y $Var(Y_i) = \sigma^2 < \infty$. Definamos:
$$U_n = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$
  Entonces la función de distribución de U_n converge a la distribución normal estándar $\mathcal{N}(0,1)$ cuando $n \to \infty$, independiete de la distribución inicial que puedan tener Y_i.

- **Funciones importantes:**

  1. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ con μ desconocido y σ conocido, entonces:
  $$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$
     Es una buena función para un intervalo de confianza para μ.
  2. Si ahora $X \sim \mu, \sigma^{\in}$, pero, ni μ ni σ son conocidos, y sea S^2 la varianza muestral de una M.A.S. de X de tamaño n. Entonces:
  $$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$
     Es una buena función para un intervalo de confianza para μ. Donde T_{n-1} es la distribución *t-Student* con $n-1$ *grados de libertad*.
  3. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, con μ y σ desconocidos, y sea S^2 la varianza muestral de una M.A.S. de X de tamaño n. Entonces:
  $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$
     Es una buena función para un intervalo de confianza para σ^2. Donde χ^2_{n-1} es la distribución de *Chi cuadrado* con $n-1$ *grados de libertad*.

Soluciones:

P1. *a*) El objetivo de esta parte es tratar de tener una idea del valor del parámetro θ, como nosotros sólo podemos ver las observaciones Y_i (señal con cierto ruido que nos hace perder precisión sobre $theta$), trataremos de dar un intervalo en que θ estará con alta probabilidad. Notemos lo siguiente: como $(Y_i)_{i=1}^n$ es una M.A.S., cada una de las observaciones es independiente de la otra. Además:

$$\mathbb{E}(Y_i) = \mathbb{E}(\theta + X) = \mathbb{E}(\theta) + \mathbb{E}(X) = \theta + 0 = \theta < \infty$$

$$\sigma^2 = Var(Y_i) = Var(\theta) + Var(X) = 0 + 4 = 4 < \infty$$

Como los parámetros de distribución de Y_i son finitos para cada observación, podemos concluir, por el **T.C.L** (ver resumen) que, con $n = 15$, $\mathbb{E}(Y_i) = \theta$ y $\sigma^2 = 4$:

$$U = \frac{\bar{Y} - \theta}{2/\sqrt{15}} \sim \mathcal{N}(0, 1)$$

Por lo tanto, sabiendo la distribución de U, podemos encontrar un intervalo de gran probabilidad (poque ahora podemos calcular esa probabilidad) para U. Como U depende de θ, encontraremos un intervalo para θ.

Como la distribución de U, la normal estándar, es una distribución simétrica centrada en el valor esperado 0, que tiene la forma:

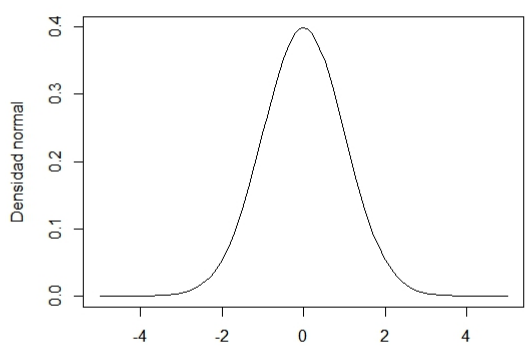

Figura 1: Densidad de la distribución normal estándar.

Entonces un 'buen intervalo' es uno que abarque aquellos valores que poseen la mayor cantidad de valor en la densidad, un intervalo centrado en la media.

Es decir, dos valores Z_1 y Z_2 tal que $U \in [Z_1, Z_2]$ con 90 % de probabilidad:

$$\mathbb{P}(Z_1 < U < Z_2) = 0{,}9 \tag{1}$$

Ocupando la simetría respecto al 0 de la distribución normal, podemos asumir que $Z_1 = -Z_2$, además sabemos que estos zeta dependerán de la prbabilidad pedida:

$$\mathbb{P}(-Z_2 < U < Z_2) = 0{,}9 \Rightarrow \mathbb{P}(U < -Z_2) + \mathbb{P}(Z_2 < U) = 1 - 0{,}9 = \alpha$$

Pero, nuevamente por la simetría, tenemos que $\mathbb{P}(U < -Z_2) = \mathbb{P}(Z_2 < U)$, entonces tenemos que:

$$\mathbb{P}(Z_2 < U) = \frac{\alpha}{2} \tag{2}$$

La intuición detrás de esto es: tengo un intervalo sobre una distribución simétrica, por lo tanto, la probabilidad que dejo afuera del intervalo la distribuyo 'hacia las colas' en la mimsma cantidad, $\alpha/2$ para cada lado.

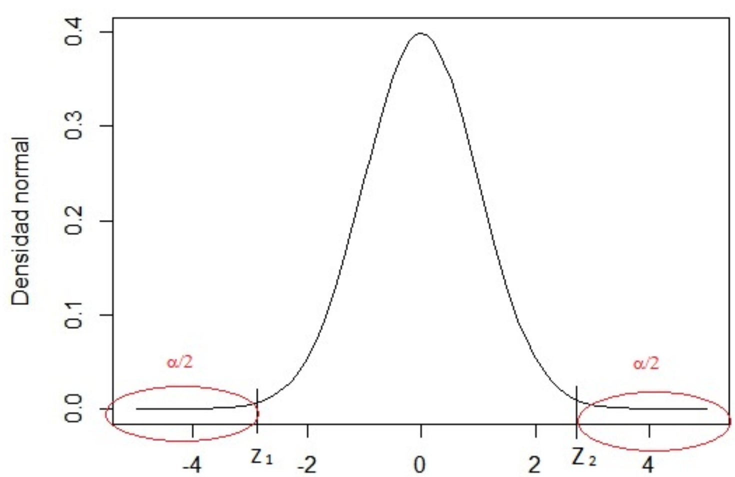

Figura 2: Densidad normal, con colas de $\alpha/2$.

Ahora busquemos un Z que haga lo que estamos pidiendo, a este valor le llamaremos $Z_{\frac{\alpha}{2}}$ pues ahora depende de esta probabilidad. Para esto utilizaremos la *Tabla de distribución normal estándar*.

Cómo ocupo la tabla?

En el centro de la tabla aparecen las probabilidades (Figura 3) y en la primera columna y primera fila de la tabla aparece una forma aproximada del valor que acota por debajo al estadístico con tal probabilidad (Figura 4).

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294

Figura 3: Probabilidades indicadas en la tabla normal.

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294

Figura 4: Valores de Z en la tabla.

La relación entre los valores mostrados en la Figura 4 y la figura 3 es la siguiente: La suma de la coordenada horizontal y vertical de la tabla (Figura 4) es el valor Z_a que acota al estadístico con probabilidad a (mostrada en la Figura 3).

$$U \sim \mathcal{N}(0,1) \;\Rightarrow\; \mathbb{P}(Z_a < U) = a \tag{3}$$

Por ejemplo, en la Figura 5, sale marcada la probabilidad 0,2877, sobre ella aparece el valor 0,06 y hacia la derecha, el valor 0,5, luego como $0{,}5 + 0{,}06 = 0{,}56$, entonces el estadístico U es acotado por 0,56 con probabilidad 0,2877.

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611

Figura 5: Ejemplo

$$\Rightarrow\; \mathbb{P}(0{,}56 < U) = 0{,}2877$$

Luego concluímos que $Z_{0,2877} = 0{,}56$. Por lo tanto, para encontrar los valores Z sólo basta encontrar la probabilidad pedida en la tabla.

Qué ocurre cuando no encuentro el valor que busco?

Como las probabilidades son valores contínuos entre 0 y 1, es posible que querramos encontrar valores que no aparecerán en la tabla, en ese caso podemos seguir de dos formas:

- Elijo el valor de la tabla qué se encuentre más cerca de la probabilidad que busco.
- Si hay 2 valores que se encuentran a la misma distancia de la probabilidad buscada, promedio los Z´s asociados a esas dos probabilidades.

Volvamos a nuestro problema, teníamos que buscar un $Z_{\frac{\alpha}{2}}$ que cumpliera la ecuación (2):

$$\mathbb{P}(Z_{\frac{\alpha}{2}} < U) = \frac{\alpha}{2}$$

Recordando que $\alpha = 1 - 0{,}9 = 0{,}1$, entonces $\alpha/2 = 0{,}05$. Así, con la tabla, buscamos la probabilidad 0,05 y su $Z_{0,05}$ asociado. El problema es que esta probabilidad no aparece en la tabla, por lo tanto proseguimos con buscar las probabilidades que más se acerquen.

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294

Figura 6: Valor 0.05 en la tabla normal.

Vemos que los Z's más cercanos al valor 0.05 son 1.64 y 1.65, por lo tanto:

$$Z_{\frac{\alpha}{2}} = Z_{0,05} = \frac{1{,}64 + 1{,}65}{2} = 1{,}645$$

Reemplazando en la ecuación inicial, ecuación (1):

$$\mathbb{P}(-Z_{\frac{\alpha}{2}} < U < Z_{\frac{\alpha}{2}}) = 0{,}9$$

$$\Rightarrow \mathbb{P}(-1{,}645 < U < 1{,}645) = 0{,}9$$

$$\Rightarrow \mathbb{P}(-1{,}645 < \frac{\bar{Y} - \theta}{2/\sqrt{15}} < 1{,}645) = 0{,}9$$

$$\Rightarrow \mathbb{P}(-\frac{2 \cdot 1{,}645}{\sqrt{15}} - \bar{Y} < -\theta < \frac{2 \cdot 1{,}645}{\sqrt{15}} - \bar{Y}) = 0{,}9$$

$$\Rightarrow \mathbb{P}(\bar{Y} - \frac{2 \cdot 1{,}645}{\sqrt{15}} < \theta < \bar{Y} + \frac{2 \cdot 1{,}645}{\sqrt{15}}) = 0{,}9$$

Entoces ganamos el intervalo para θ de la forma $[\bar{Y} - \frac{2 \cdot 1{,}645}{\sqrt{15}}, \bar{Y} + \frac{21{,}645}{\sqrt{15}}]$, reemplazando el valor encontrado de $\bar{Y} = 15$.

$$\theta \in [14{,}15\,,\, 15{,}85] \quad \text{Con } 90\,\% \text{ de probabilidad.}$$

b) Para conocer la confianza[1] utilizada en este intervalo es necesario calcular la siguiente probabilidad:

$$\mathbb{P}(\theta \in [14,5; 15,5]) = p$$

[1] La confianza es la probabilidad de estar seguro de encontrar el parámetro en ese intervalo. En la parte a), el nivel de confianza que nos pedían de tal intervalo era de 90 %.

Entonces p será la cantidad de confianza del intervalo.
Notemos que el intervalo dado por $[14,5; 15,5]$ está en función de $\bar{Y}$ cuando toma el valor 15:

$$\theta \in [14,5; 15,5] \Rightarrow \theta \in [15 - 0,5; 15 + 0,5]$$
$$\Rightarrow \theta \in [\bar{Y} - 0,5; \bar{Y} + 0,5]$$

Como ya tenemos una noción de la distribución de $\bar{Y}$ podemos calcular:

$$\mathbb{P}(\bar{Y} - 0{,}5 < \theta < \bar{Y} + 0{,}5) = \mathbb{P}(-0{,}5 < \bar{Y} - \theta < 0{,}5) = \mathbb{P}(\frac{-0{,}5 \cdot \sqrt{15}}{2} < \frac{\bar{Y} - \theta}{2/\sqrt{15}} < \frac{0{,}5\sqrt{15}}{2})$$

$$= \mathbb{P}(-0{,}97 < U < 0{,}97)$$

Si tomamos el valor $Z_\beta = 0{,}97$, podemos escribir la probabilidad anterior como:

$$\mathbb{P}(-Z_\beta < U < Z_\beta) = 1 - 2\beta$$

Como $U \sim \mathcal{N}(0,1)$, ocupamos la tabla nuevamente para encontrar la probabilidad β asociada a 0,97.

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611

Figura 7: Probabilidad asociada a 0.97.

Entonces $\beta = 0{,}1666$.

$$\mathbb{P}(\theta \in [14,5; 15,5]) = \mathbb{P}(-0{,}97 < U < 0{,}97) = 1 - 2 \cdot 0{,}1666 = 0{,}6668$$

La confianza del intervalo $[14,5; 15,5]$ es del 66,68 %.

c) Busquemos la probabilidad de que la señal θ sea mayor a 14,5:

$$\mathbb{P}(14,5 < \theta) = \mathbb{P}(15 - 0{,}5 < \theta) = \mathbb{P}(\bar{Y} - 0{,}5 < \theta) = \mathbb{P}(-0{,}5 < \theta - \bar{Y})$$

$$= \mathbb{P}(\bar{Y} - \theta < 0{,}5) = \mathbb{P}(\frac{\bar{Y} - \theta}{2/\sqrt{15}} < \frac{0{,}5 \cdot \sqrt{15}}{2}) = \mathbb{P}(U < 0{,}97) = 1 - \mathbb{P}(0{,}97 < U)$$

Sólo hace falta buscar en la tabla de distribución normal, la probabilidad asociada al valor 0,97. Tal probabilidad ya la habíamos encontrado en la parte *b*) y había sido nombrada con la variable β, con valor $\beta = 0,1666$. Finalmente:

$$\mathbb{P}(14,5 < \theta) = \mathbb{P}(U < 0{,}97) = 1 - \mathbb{P}(0{,}97 < U) = 1 - \beta = 1 - 0{,}1666 = 0{,}8334$$

La probabilidad de que θ sea mayor a 14,5 es del 83,34 %, por lo tanto el emisor está entregando información con alta probabilidad de acierto.

d) Como Y depende de una constante y una variable normal, podemos asumir que la distribución de Y es normal (por ser una traslación de una normal), con media $\mathbb{E}(Y) = \theta$ y varianza $Var(Y) = 4$ (visto en la parte a). Entonces el nuevo caso se trata de dos variables Y_1 e $Y_2 \sim \mathcal{N}(\theta, 4)$. Como sabemos que la suma y ponderación de variables normales, es tora variable normal, calculemos los parámetros de la variable $T = Y_1 - Y_2$:

$$\mathbb{E}(T) = \mathbb{E}(Y_1 - Y_2) = \mathbb{E}(Y_1) - \mathbb{E}(Y_2) = \theta - \theta = 0$$

$$Var(T) = Var(Y_1 - Y_2) = Var(Y_1) + Var(Y_2) = 4 + 4 = 8$$

Entonces $T = Y_1 - Y_2 \sim \mathcal{N}(0, 8)$.

La condición que nos piden es que estas dos señales difieran en menos de una unidad, de forma más formal, esto es $|Y_1 - Y_2| < 1$:

$$\mathbb{P}(|Y_1 - Y_2| < 1) = \mathbb{P}(-1 < Y_1 - Y_2 < 1)$$

Para seguir con este ejercicio ocuparemos el siguiente resultado que facilitará el cálculo:

$$\text{Si: } X \sim \mathcal{N}(\mu, \sigma^2) \;\; \Rightarrow \;\; \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Así, si $T \sim \mathcal{N}(0, 8)$ entonces tenemos que $T/\sqrt{8} \sim \mathcal{N}(0, 1)$. Con esto, podemos hacer lo siguiente:

$$\mathbb{P}(|Y_1 - Y_2| < 1) = \mathbb{P}(-1 < Y_1 - Y_2 < 1) = \mathbb{P}(\frac{-1}{\sqrt{8}} < \frac{Y_1 - Y_2}{\sqrt{8}} < \frac{1}{\sqrt{8}})$$

$$= \mathbb{P}(\frac{-1}{\sqrt{8}} < \frac{T}{\sqrt{8}} < \frac{1}{\sqrt{8}}) = \mathbb{P}(-0{,}35 < U < 0{,}35)$$

Donde U es la variable normal estándar que hemos estado ocupando en los pasos anteriores. Nuevamente acá podemos ocupar la tabla de distribución normal para calcular la probabilidad asociada. Si $Z_\gamma = 0{,}35$, entonces la probabilidad buscada buede ser resuelta por:

$$\mathbb{P}(-Z_\gamma < U < Z_\gamma) = 1 - 2\gamma$$

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

Figura 8: Probabilidad asociada al valor normal 0.35

Obtenemos el valor $\gamma = 0{,}3632$, finalmente calculamos la probabilidad pedida.

$$\mathbb{P}(|Y_1 - Y_2| < 1) = \mathbb{P}(-0{,}35 < U < 0{,}35) = 1 - 2 \cdot 0{,}3632 = 0{,}2736$$

P2. *a*) Sea $(X_i)_{i=1}^n$ una M.A.S. de tamaño n de la concentración de metano en el depósito. Como σ^2 y μ son valores finitos, por **T.C.L.**:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Entonces, buscar la prpobabilidad de que la media sea mayor que un valor C es equivalente a:

$$\mathbb{P}(C < \mu) = \mathbb{P}(\bar{X} - \mu < \bar{X} - C) = \mathbb{P}(\frac{\bar{X} - C}{\sigma/\sqrt{n}} < \frac{\bar{X} - C}{\sigma/\sqrt{n}}) = \mathbb{P}(U < \frac{\bar{X} - C}{\sigma/\sqrt{n}}) = 1 - \mathbb{P}(U > \frac{\bar{X} - C}{\sigma/\sqrt{n}})$$

Donde $U \sim \mathcal{N}(0,1)$. Dado que $n = 50$, σ es un valor dado (enunciado), $\bar{X}$ es un valor calculable a través de la muestra y C es una constante conocida, el valor cota de esa probabilidad es calculable y puede ser encontrado en la tabla de distribución normal estándar. Supongamos que $\beta_{\bar{X},C}$ es ese valor, entonces:

$$\mathbb{P}(C < \mu) = 1 - \mathbb{P}(U > \frac{\bar{X} - C}{\sigma/\sqrt{50}}) = 1 - \beta_{\bar{X},C}$$

b) Como ya no poseemos el parámetro σ, tendremos que reemplazarlo por su estimador de la muestra, es decir, S^2. Por perder este parámetro, nuestra estimación tendrá que recurrir a otra distribución que modele mejor el problema considerando el dato faltante.

Sabemos, por la parte anterior que :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Pero al momento de reemplazar el dato faltante por su estimador, es decir, $\sigma^2 \to S^2$, la distribución cambia por una T de student:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$

Con los nuevos datos entregados, tenemos una muestra de $n = 20$ datos, cuyo promedio muestral es $\bar{X} = 100$ y varianza muestral $S^2 = 4$. Busquemos un intervalo para μ mediante la T_{n-1} de Student.

Al igual la distribución normal, la distribución T_{n-1} de student es simétrica[2] respecto al cero, bajo estos supuestos, podemos encontrar un α tal que el intervalo de confianza $1 - \alpha$ sea $[-t_{\frac{\alpha}{2}}, t_{\frac{\alpha}{2}}]$, con $t_{\frac{\alpha}{2}}$ el valor asociado a $\frac{\alpha}{2}$ en la tabla t-student. Con esto tenemos

[2] Como es simética, distribuímos $\alpha/2$ a ambas colas de la distribución.

que:

$$\mathbb{P}(-t_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}}) = 1-\alpha$$

Como nos piden que el intervalo tenga confianza de 95 %, haremos que $1-\alpha = 0{,}95$, es decir, $\alpha = 0{,}05$. Como $\alpha/2 = 0{,}025$, busquemos $t_{0{,}025}$ en la tabla de t-student de $n-1 = 20-1 = 19$ grados de libertad[3].

Probabilidad α

Grados de libertad n-1

α / r	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Figura 9: Valor de la t de student para $\alpha = 0{,}025$ y 19 grados de libertad.

Entonces $t_{0{,}025} = 2,093$, con esto calculamos nuestro intervalo de confianza.

$$\mathbb{P}(-2{,}093 < \frac{\bar{X}-\mu}{S/\sqrt{n}} < 2{,}0{,}93) = 0{,}95 \Rightarrow \mathbb{P}(-2{,}093 < \frac{100-\mu}{2/\sqrt{20}} < 2{,}093) = 0{,}95$$

$$\Rightarrow \mathbb{P}(-\frac{2\cdot 2{,}093}{\sqrt{20}} < \bar{X}-\mu < \frac{2\cdot 2{,}093}{\sqrt{20}})0{,}95 \Rightarrow \mathbb{P}(100-\frac{2\cdot 2{,}093}{\sqrt{20}} < \mu < 100+\frac{2\cdot 2{,}093}{\sqrt{20}}) = 0{,}95$$

En conclusión, el intervalo $[99,06; 100,93]$ contiene a μ con 95 % de confianza.

Qué ocurre con la estimación de σ^2?

Como nunca pudimos obtener un valor fijo para μ, el estadístico anterior no nos sirve para despejar σ^2. Esta vez ocuparemos la distribución χ^2 ('Chi cuadrado') para crear el intervalo en que encontraremos a σ^2.

[3] A diferencia de la tabla para la normal estandar, la tabla de t de student tiene los valores T al centro de la tabla, no las probabilidades (ver ejercicio 1).

Ocuparemos la siguiente función que distribuye como Chi:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

En este caso no tenemos una distribución simétrica, por lo tanto aculuramos toda la probabilidad en la de las colas de la distribución (cola derecha). Si $Y \sim \chi^2_{n-1}$ entonces la tabla de chi-cuadrado nos entrega el siguiente valor:

$$\mathbb{P}(Y > y_\alpha) = \alpha$$

Para un α a elección, como nuestra función distribuye χ^2_{n-1} y la probabilidad del intervalo nos pide 95 % de confianza, tomamos $\alpha = 0{,}95$

$$\mathbb{P}(\frac{(n-1)S^2}{\sigma^2} > y_\alpha) = \alpha = 0{,}95$$

Entonces busquemos en la tabla[4] de chi-cuadrado aquel dato que entrega una probabilidad de 0.95 con $n-1 = 19$ grados de libertad.

El resultado es $y_{0{,}95} = 10{,}12$, y lo reemplazamos en nuestra probabilidad:

$$\mathbb{P}(\frac{(n-1)S^2}{\sigma^2} > 10{,}12) = 0{,}95 \Rightarrow \mathbb{P}(\frac{19 \cdot 4}{\sigma^2} > 10{,}12 = 0{,}95$$

$$\Rightarrow \mathbb{P}(\frac{76}{10{,}12} > \sigma^2) = 0{,}95 \Rightarrow \mathbb{P}(7{,}5 > \sigma^2) = 0{,}95$$

Por último, como la varianza es un valor positivo entonces, gratuitamente, σ^2 está acotado inferiormente por 0. Finalmente:

$$\sigma^2 \in (0; 7{,}5) \quad \text{con } 95\,\% \text{ de confianza.}$$

P3. Ocuparemos la distribución de chi-cuadrado para crear un intervalo de alta confianza para el parámetro σ^2. la función a ocupar será:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Donde n es la cantidad de datos de la muestra ($n = 6$ en este caso) y S^2 es la varianza muestral, calculada de la forma:

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

Por lo tanto, para un valor α dado, la tabla entrega otro valor, y_α tal que cumpla que:

$$\mathbb{P}(\frac{(n-1)S^2}{\sigma 2} > y_\alpha) = \alpha$$

Como el intervalo requerido es de un 90 %, haremos que $\alpha = 0{,}9$. Así, revisando la tabla de

[4] La tabla de chi-cuadrado tiene el mismo funcionamiento que la tabla t de student.

chi-cuadrado con $n-1=5$ grados de liberad, obtenemos que $y_\alpha = 1{,}61$.

Reemplazamos este valor para calcular nuestro intervalo:

$$\mathbb{P}(\frac{(n-1)S^2}{\sigma^2} > 1{,}61) = 0{,}9$$

Finalmente, con los datos podemos verificar que $\bar{x} = 9{,}485$ y $S^2 = 0{,}02855$.

Así, $\mathbb{P}(5 \cdot 0{,}02855/1{,}61 > \sigma^2) = 0{,}9$, como la varianza es siempre positiva, tenemos que 0 es cota inferior de σ^2:

$$\sigma^2 \in (0; 0,089) \quad \text{Con } 90\,\% \text{ de confianza.}$$