ESTADÍSTICA
GUIA No 8

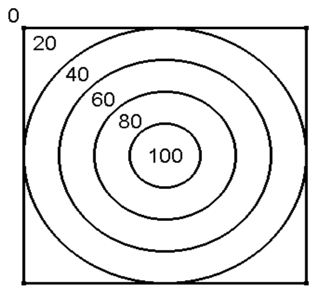
TEORÍA DE LA ESTIMACIÓN

1. Estimación Puntual

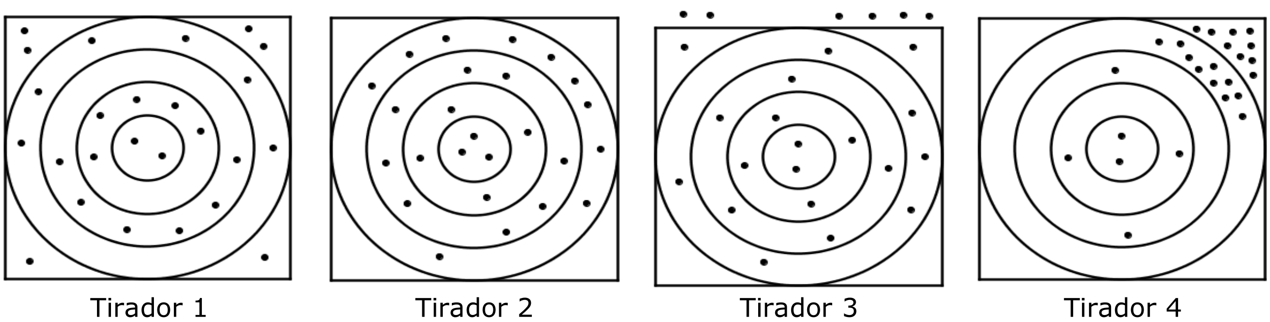
Objetivo:
El objetivo de la estimación puntual es usar una muestra para obtener números (estimaciones puntuales) que sean la mejor representación de los verdaderos parámetros de la población.

1. Sea θ un parámetro (Medida descriptiva en una población).
2. Un estimador $\hat{\theta}$ es un estadístico (Medida descriptiva en una muestra) que se usa para estimar un parámetro, a través de una variable aleatoria.
3. Los valores que toma $\hat{\theta}$ en cada muestra se llaman estimaciones puntuales.

Actividad 1:
Cuatro tiradores desconocidos llegan a una competencia en donde cada uno debe lanzar el dardo 25 veces. Dependiendo de la zona en donde quede el dardo recibirá una puntuación según la siguiente figura:



Al terminar de hacer sus lanzamientos cada tirador, se tomó una foto del objeto de tiro al blanco para ser analizada por los jueces y decidir cuál de los cuatro es el mejor. Las fotos se muestran en la siguiente figura:



Ejercicio 1:
Llene la siguiente tabla, en la primera se colocan los puntajes, y en las demás columnas se colocan la cantidad de dardos que cayó en cada zona para cada tirador.

Puntajes: X_k	Tirador 1: F_k	Tirador 2: F_k	Tirador 3: F_k	Tirador 4: F_k
0				
20				
40				
60				
80				
100				
Total				

Ejercicio 2:
Los jueces, para determinar cuál es el mejor de los cuatro competidores, calcula la media de los puntajes obtenidos por cada tirador. Calcule la media $\overline{X}$ de los puntajes X_k obtenidos por cada tirador, y escríbalos en la siguiente tabla:

Competidor	Media de los puntajes $\overline{X}$
Tirador 1	$\overline{x_1} =$
Tirador 2	$\overline{x_2} =$
Tirador 3	$\overline{x_3} =$
Tirador 4	$\overline{x_4} =$

Preguntas:

1. Desde el inicio de la competencia se podía predecir la media que obtendría cada competidor?

2. ¿Se puede decir que la media $\overline{X}$ es una variable aleatoria?

3. ¿Es $\overline{X}$ una variable aleatoria muestral?

De acuerdo con la última tabla:

4. ¿Cuál es el mejor de los cuatro competidores? (mayor exactitud)

5. ¿Cuál es el peor de los cuatro competidores?

6. ¿Cuál es el más preciso de los cuatro competidores?

Nota:

- La variable $\overline{X}$ es un estimador. $\hat{\theta} = \overline{X}$.
- Cada valor $\overline{x_k}$, es decir, la media de cada jugador es una Estimación puntual.

Actividad 2:

Con motivo de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea conocer el número de hijos que tienen los 200 obreros que ahí laboran. Para esto, se entrevistan a todos los obreros en orden alfabético, como aparecen en la nómina, obteniéndose los resultados que se muestran en la tabla que se muestra en la siguiente página.

Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos
1	8	51	6	101	8	151	4
2	0	52	4	102	0	152	3
3	5	53	7	103	5	153	4
4	3	54	8	104	3	154	5
5	6	55	5	105	4	155	0
6	5	56	7	106	8	156	8
7	0	57	4	107	2	157	4
8	4	58	8	108	0	158	5
9	0	59	8	109	6	159	5
10	8	60	7	110	3	160	8
11	7	61	0	111	0	161	4
12	6	62	2	112	5	162	3
13	1	63	0	113	4	163	8
14	4	64	0	114	2	164	0
15	0	65	2	115	4	165	8
16	7	66	3	116	4	166	0
17	4	67	4	117	2	167	2
18	5	68	8	118	5	168	4
19	8	69	7	119	6	169	8
20	0	70	3	120	8	170	1
21	2	71	5	121	6	171	8
22	8	72	1	122	0	172	2
23	4	73	5	123	7	173	5
24	5	74	3	124	3	174	8
25	6	75	4	125	7	175	2
26	5	76	7	126	0	176	4
27	3	77	6	127	3	177	6
28	8	78	2	128	0	178	3
29	7	79	3	129	5	179	6
30	1	80	0	130	4	180	5
31	4	81	0	131	7	181	3
32	6	82	6	132	3	182	2
33	0	83	2	133	5	183	7
34	5	84	6	134	5	184	2
35	4	85	4	135	6	185	7
36	1	86	4	136	0	186	3
37	1	87	0	137	3	187	2
38	4	88	8	138	5	188	6
39	4	89	8	139	2	189	4
40	3	90	5	140	3	190	6
41	5	91	4	141	3	191	8
42	5	92	4	142	0	192	5
43	7	93	8	143	4	193	3
44	8	94	3	144	8	194	5
45	7	95	3	145	3	195	5
46	8	96	1	146	3	196	3
47	6	97	7	147	1	197	6
48	4	98	8	148	7	198	6
49	3	99	8	149	5	199	4
50	3	100	6	150	4	200	4

Tabla: Obreros y Cantidad de hijos por obrero**.**

Ejercicio 1:
Seleccione una muestra **con reemplazo** de tamaño n=30. Considere la variable X = “Número de hijos por obrero”.

Ejercicio 2:
Elabore una tabla de frecuencias absolutas y relativas para su muestra, clasificando los obreros de acuerdo a la variable X.

Ejercicio 3:
Calcule $\bar{x}$ de su muestra seleccionada

Preguntas finales:

Pregunta 1: ¿Cuál es el estimador?

Pregunta 2: ¿Cuál es la estimación puntual?

Pregunta 3: ¿Puede elaborarse la distribución muestral para $\bar{x}$?

2. Estimación por Intervalos

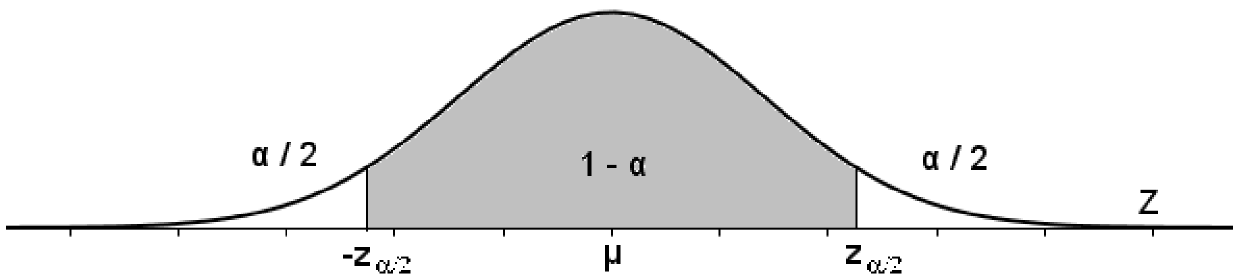
2.1 Estimación de la media por intervalos - Varianza conocida

Objetivo:
El objetivo de la estimación por intervalos es obtener los límites entre los cuales se encuentra el verdadero valor del parámetro de una población con un cierto nivel de confianza.

Teoría:
Un intervalo probabilístico (a,b) para el cual la probabilidad de que el intervalo contenga al parámetro θ (Medida descriptiva para la población) sea igual a 1-α se llama Intervalo de confianza al (1-α)x100% para estimar al parámetro θ. En tal caso se escribe $P(a< \theta <b) = 1- \alpha$.

El valor α se denomina Nivel de significación, y el valor (1-α) se denomina Nivel de confianza.

Si la media μ de la población es desconocida y la varianza σ^2 es conocida, se puede determinar un intervalo (a,b) talque $P(a< \mu <b) = 1- \alpha$. El valor de α se reparte simétricamente entre las dos colas de la curva Normal estándar, quedando (1-α) en la región central y $\alpha/2$ en cada cola de la curva. El problema se reduce a resolver la desigualdad $P\left(\left|\mu - \overline{x_0}\right| < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$, donde $\overline{x_0}$ es una estimación puntual de μ y $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ es el **margen de error**. Los valores $-Z_{\alpha/2}$ y $Z_{\alpha/2}$ se llaman **coeficientes de confianza**.



Actividad 1:

Las calificaciones finales de 250 participantes en unas Olimpiadas Internacionales de Matemáticas y Estadística se presentan en la tabla 1 con sus respectivos códigos. Las pruebas se calificaron de 0 a 10 puntos. Se sabe que la varianza fue de 10.50 puntos cuadrados.

Ejercicio 1:
Seleccione una muestra con reemplazo de tamaño n=40. Considere la variable X = "Nota final del participante".

Ejercicio 2:
Calcule una estimación puntual $\overline{x_0}$ a partir de su muestra seleccionada.

Ejercicio 3:
Calcule un intervalo de confianza para la media poblacional μ con una probabilidad del 95%, tomando la estimación puntual $\overline{x_0}$ calculada por usted en el punto anterior.

Ejercicio 4:

Use la Tabla de la Normal Estándar Z, para encontrar los extremos del intervalo $(-Z_{\alpha/2}, Z_{\alpha/2})$ talque $P\{ (-Z_{\alpha/2}, Z_{\alpha/2}) \} = 0.95$

Explicación:

La desigualdad $\left| \mu - \overline{x_0} \right| < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ es equivalente a $\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. Por lo tanto,

$P\left(\left| \mu - \overline{x_0} \right| < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 0.95$ es equivalente a $P\left(\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 0.95$

Ahora, comparando $P(a < \mu < b) = 0.95$ y $P\left(\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 0.95$ se tiene

$a = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ y $b = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

Ejercicio 5:

¿Cuáles son los valores de $a = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ y $b = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$?

Ejercicio 6:
Finalmente, ¿Cuál es el intervalo de confianza (a,b) ?

Ejercicio 7:
Con base en el desarrollo anterior, complete la siguiente frase:

La probabilidad de que el verdadero valor del parámetro ______ se encuentre en el intervalo ______________ es del __________.

Ejercicio 8:
¿Cuál será el intervalo de confianza, si se toma un nivel de significación del 10% ?

Ejercicio 9:
¿Cuál será el intervalo de confianza si toma un nivel de confianza del 96.6%

Ejercicio 10:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de confianza del 95%, con un margen de error de 1 punto?

Ejercicio 11:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de confianza del 95%, con un margen de error de 2 puntos?

Ejercicio 12:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de significación del 10%, con un margen de error de 2 puntos?

Ejercicio 13:
Si se toma una muestra de tamaño n=20 y se desea obtener un intervalo de confianza con un nivel de confianza del 90%, entonces:
a) ¿Cuáles son los coeficientes de confianza?
b) ¿Cuál es la longitud del intervalo de confianza?

Ejercicio 14:
Ahora, seleccione una muestra sin reemplazo de tamaño n=30. Considere la variable X = "Nota final del participante".
a. Elabore la Tabla de distribución de X.
b. Calcule una estimación puntual de la media poblacional μ.
c. ¿Cuál es el margen de error para un nivel de confianza del 90%,
d. ¿Cuál es el intervalo de confianza?

Ejercicio 15:
Ahora, seleccione una muestra sin reemplazo de tamaño n=40. Considere la variable X = "Nota final del participante".
a. Elabore la Tabla de distribución de X.
b. Calcule una estimación puntual de la media poblacional μ.
c. ¿Cuál es el margen de error para un nivel de significación del 10%,
d. ¿Cuál es el intervalo de confianza?

Código	Nota	Código	Nota	Código	Nota	Código	Nota	Código	Nota
1	3	51	9	101	1	151	8	201	2
2	10	52	5	102	10	152	7	202	1
3	5	53	2	103	7	153	3	203	0
4	10	54	0	104	0	154	3	204	4
5	7	55	10	105	5	155	10	205	4
6	5	56	4	106	6	156	1	206	10
7	9	57	1	107	0	157	6	207	2
8	10	58	7	108	8	158	5	208	3
9	4	59	5	109	7	159	5	209	5
10	9	60	7	110	1	160	7	210	10
11	3	61	4	111	6	161	3	211	1
12	7	62	1	112	7	162	1	212	10
13	2	63	9	113	3	163	4	213	3
14	10	64	8	114	0	164	1	214	8
15	1	65	1	115	4	165	0	215	1
16	9	66	9	116	10	166	8	216	6
17	3	67	2	117	4	167	5	217	10
18	8	68	10	118	10	168	7	218	9
19	8	69	1	119	7	169	9	219	5
20	3	70	0	120	3	170	10	220	7
21	9	71	3	121	7	171	0	221	8
22	9	72	5	122	3	172	10	222	0
23	8	73	0	123	2	173	8	223	3
24	2	74	1	124	9	174	7	224	9
25	5	75	2	125	9	175	3	225	4
26	3	76	8	126	3	176	2	226	0
27	0	77	8	127	9	177	9	227	5
28	0	78	1	128	8	178	0	228	10
29	4	79	6	129	0	179	8	229	8
30	2	80	10	130	7	180	8	230	3
31	5	81	3	131	6	181	1	231	8
32	1	82	0	132	6	182	2	232	1
33	4	83	5	133	7	183	2	233	7
34	8	84	0	134	5	184	8	234	6
35	3	85	3	135	6	185	2	235	5
36	9	86	2	136	6	186	8	236	0
37	9	87	6	137	1	187	6	237	0
38	7	88	7	138	10	188	7	238	2
39	8	89	8	139	4	189	3	239	2
40	2	90	8	140	4	190	7	240	2
41	6	91	3	141	2	191	10	241	2
42	4	92	2	142	4	192	0	242	9
43	6	93	8	143	8	193	10	243	6
44	9	94	7	144	0	194	0	244	6
45	8	95	1	145	8	195	9	245	0
46	7	96	3	146	10	196	10	246	2
47	10	97	3	147	7	197	5	247	6
48	7	98	7	148	9	198	3	248	1
49	1	99	1	149	7	199	3	249	4
50	1	100	9	150	5	200	2	250	1

Tabla 1: Notas finales de 250 participantes.

2.2 Estimación de la media por intervalos - Varianza desconocida

Objetivo:

El objetivo de este taller es realizar la estimación por intervalos de la media poblacional cuando no se conoce la varianza de la población. Se usará la distribución T-Student y se trabajará con la varianza de la muestra. También veremos la utilidad del teorema central del límite.

Distribución T-Student:

La distribución T fue creada por Goset y formalizada por Fisher. Es una familia de distribuciones semejantes a la distribución normal estándar Z. Cada miembro de la familia T está determinado por el tamaño de la muestra n. Para valores pequeños de n, se tiene que $\sigma^2(T)>\sigma^2(Z)$. A medida que aumenta el valor de n, las gráficas de f(T) se acercan a la gráfica de f(Z), esto permite calcular probabilidades de la distribución T-Student con la distribución Z-Estándar.

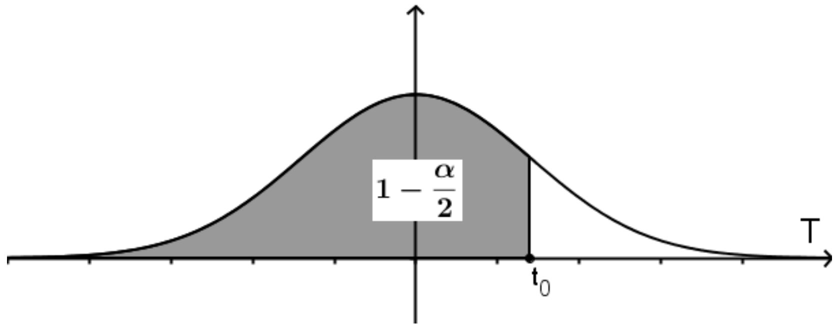
Supóngase que se toma una muestra de tamaño n. Entonces:

1) Cuando la varianza de X se conoce, se usa la expresión $Z=\frac{\overline{X}-\mu_X}{\sigma_X/\sqrt{n}}$ y la tabla de la distribución normal estándar Z.
2) Cuando la varianza de X no se conoce, se usa la expresión $T=\frac{\overline{X}-\mu_X}{S/\sqrt{n}}$ y la tabla de la distribución T-Student. S es la desviación estándar de la muestra $S=\sqrt{\frac{\sum_{k=1}^{k=n}(x_k-\overline{x})^2}{n-1}}$ con (n-1) grados de libertad.
3) La función T-Student para n grados de libertad es $f_n(x)=\frac{\left(\frac{x^2}{n}+1\right)^{-\left(\frac{n+1}{2}\right)}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\;\Gamma\left(\frac{n}{2}\right)}$, donde

$$\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt .$$

Estimación de μ en muestras pequeñas (n<30):

Si X es una variable distribuida normalmente con media μ y varianza desconocidas, S es la desviación estándar muestral de una muestra de tamaño n, entonces, para estimar el intervalo que contiene a la media μ de X con una seguridad del $(1-\alpha)\%$, la tabla T-Student proporciona el valor de t_0 para el cual se acumula el $\left(1-\frac{\alpha}{2}\right)\%$ del área bajo la curva $f_n(x)$. Es decir, $P\left(\left|\mu-\overline{x}\right|<t_0\cdot\frac{S}{\sqrt{n}}\right)=(1-\alpha)\%$. El intervalo de confianza es $\overline{x}-t_0\cdot\frac{S}{\sqrt{n}}<\mu<\overline{x}+t_0\cdot\frac{S}{\sqrt{n}}$.



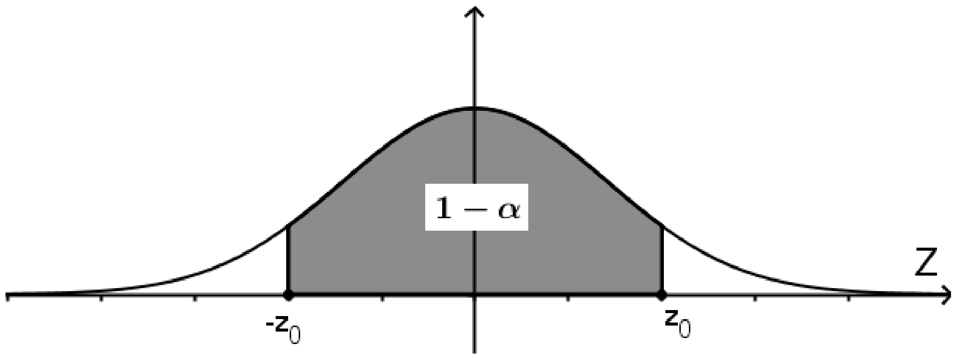
Estimación de μ en muestras grandes (n>30):

Cuando el tamaño de la muestra es n suficientemente grande, el teorema central del límite garantiza que:

1) $\overline{X}$ es aproximadamente normal sin necesidad de que la variable X sea normal.

2) $\mu(\overline{X})=\mu(X)$ y $\sigma^2(\overline{X})=\dfrac{\sigma^2(X)}{n}$.

Por lo tanto se puede utilizar la tabla de la normal estándar Z para estimar el intervalo que contiene a la media μ con una seguridad de $(1-\alpha)\%$ buscando el valor de z_0 para el cual se acumula simétricamente el $(1-\alpha)\%$ del área bajo la normal estándar. Es decir, $P\left(0<\left|\mu-\overline{x}\right|<z_0\cdot\dfrac{S}{\sqrt{n}}\right)=\left(\dfrac{1-\alpha}{2}\right)\%$. El intervalo de confianza es $\overline{x}-z_0\cdot\dfrac{S}{\sqrt{n}}<\mu<\overline{x}+z_0\cdot\dfrac{S}{\sqrt{n}}$.



Ejercicio 1:
Encuentre el valor de t_0, para el cual se cumplen las siguientes igualdades:
Nota: El índice de T corresponde al tamaño de la muestra.

a) $P(T_8<t_0)=0.75$

b) $P(T_{10}<t_0)=0.80$

c) $P(T_{12}<t_0)=0.85$

d) $P(T_{15}<t_0)=0.90$

e) $P(T_{20}<t_0)=0.95$

Ejercicio 2:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Use la tabla T-Student, para completar la siguiente tabla:

n	$\overline{x}$	$1-\alpha$	S^2	t_0
10	8.5	0.80	14.40	
15	7.5	0.90	20.25	
20	9.0	0.95	12.25	
25	10.5	0.90	72.25	
30	11.5	0.99	81.00	

Ejercicio 3:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Use la tabla T-Student, para completar la siguiente tabla:

n	$\overline{x}$	$1-\alpha$	S^2	Intervalo de confianza
10	8.5	0.80	14.40	
15	7.5	0.90	20.25	
20	9.0	0.95	12.25	
25	10.5	0.90	72.25	
30	11.5	0.99	81.00	

Ejercicio 4:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Compare los intervalos de confianza obtenidos usando la tabla T-Student con los obtenidos usando la tabla normal estándar. Complete la siguiente tabla:

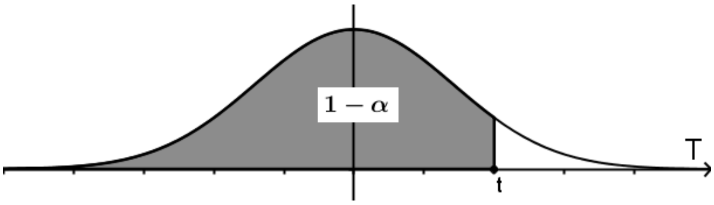
n	$\overline{x}$	$1-\alpha$	S^2	Intervalo de confianza con T	Intervalo de confianza con Z
35	8.5	0.80	14.40		
40	7.5	0.90	20.25		
45	9.0	0.95	12.25		
50	10.5	0.90	72.25		
55	11.5	0.99	81.00		

Ejercicio 5:
Un grupo de 400 estudiantes presentó una prueba de Informática calificada de 0 a 10 puntos. Se toma una muestra de 40 estudiantes. En la tabla se muestran las notas de la muestra.

4	2	10	7	2	4	3	7	1	6
2	9	10	5	9	10	10	5	5	5
1	3	6	8	5	1	4	2	5	5
4	1	0	8	9	6	8	1	4	7

Estime en que intervalo se encuentra el promedio de todo el grupo con una seguridad del 80%, 90% y 95%, utilizando las distribuciones T-Student y Z-Normal estándar.

TABLA DE LA DISTRIBUCIÓN T-STUDENT



	1 - α							
n	**0.75**	**0.80**	**0.85**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576