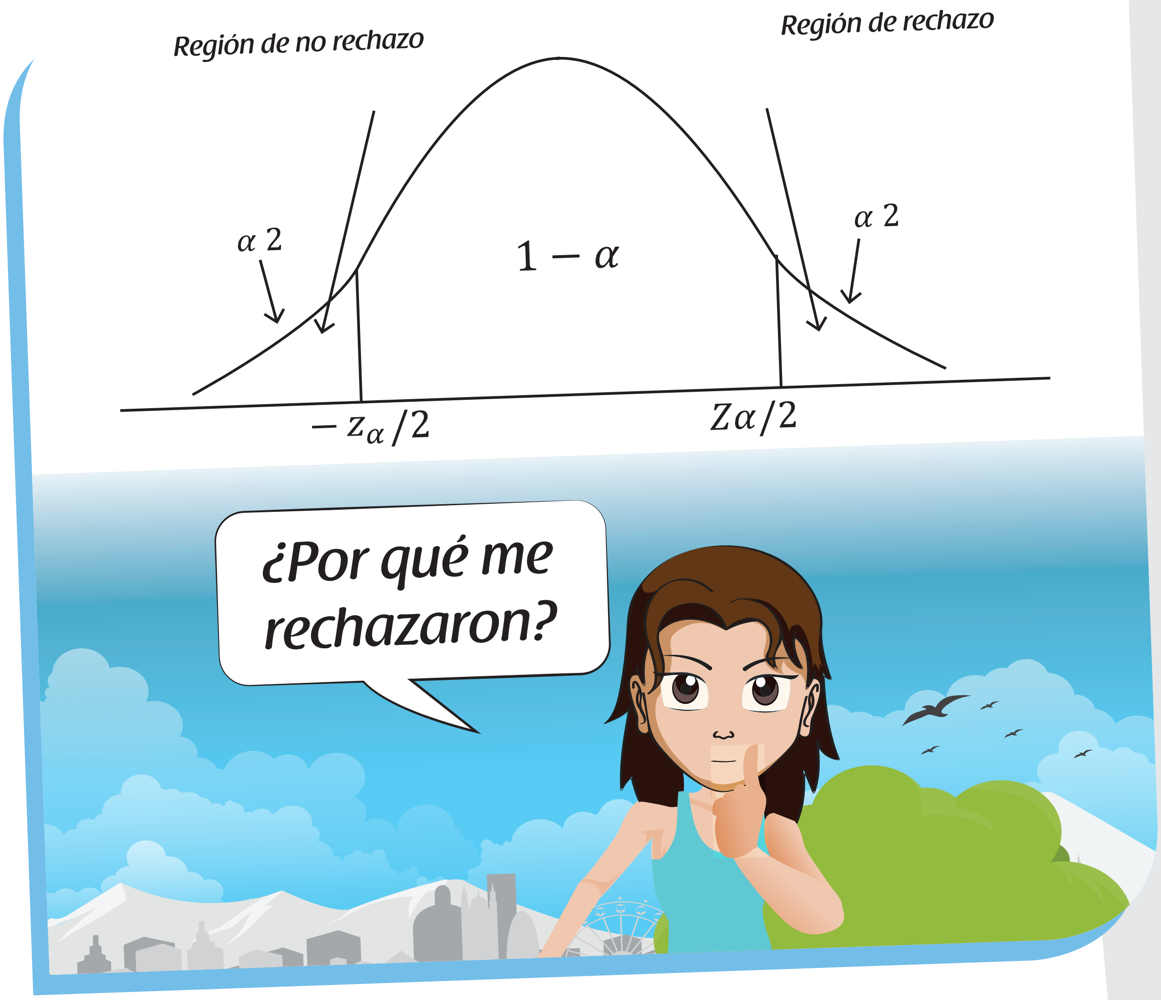
Guía 5

Empleemos la hipótesis estadística



Indicadores de desempeño

Conceptual:

- Define hipótesis en un conjunto de datos.

Procedimental:

- Realiza pruebas de hipótesis en un conjunto de datos.

Actitudinal:

- Muestra responsabilidad en las decisiones que toma con respecto a la información derivada de procedimientos estadísticos.

A Vivencia

TRABAJO EN PAREJAS

1. Contestamos las preguntas según la tabla de datos:

¿Cuáles de las siguientes situaciones ha vivido con su familia?				
	Sí	*No*	*No aplica*	*Total*
Darles preferencia a sus hermanos por ser varones.	39.8 %	58.4 %	1.8%	100.0%
Obligarla a atender a sus hermanos.	33.0 %	65.8 %	1.2%	100.0%
Que las tareas de la casa las hagan sólo las mujeres.	43.7 %	55.7 %	.6%	100.0%
Mandarla a una escuela donde aprenda los roles de madre y esposa.	9.4 %	90.0 %	.6%	100.0%
No permitirle estudiar.	21.8%	77.4%	.8%	100.0%
Darle menos libertad que a sus hermanos.	48.0 %	51.0 %	1.0%	100.0%

a. Determinamos las medidas de tendencia central: Promedio, media y moda.

b. Calculamos las medidas de dispersión: Rango, varianza y desviación estándar.

c. Escribimos una conclusión sobre las situaciones familiares según la información presentada.

2. La siguiente tabla muestra los resultados del fenómeno de bullying en el Colegio Corrientes de Paz. La encuesta se realizó con la pregunta: ¿Cuál de las siguientes acciones realiza para maltratar a alguien con más frecuencia?:

Acciones	Número de personas
Ignoro	67
No dejo participar	47
Insulto	42
Hago notas ofensivas	33
Hablo mal de alguien	35

Escondo cosas	12
Rompo cosas	19
Robo cosas	13
Pego	4
Amenazo para meter miedo	15
Obligo con amenazas	16

a. Elaboramos las tablas de frecuencia.

b. Determinamos las medidas de tendencia central y dispersión.

c. Escribimos una hipótesis de los datos que indican los comportamientos de los estudiantes para maltratar a otros.

TRABAJO EN EQUIPO

3. Escribimos por qué es importante determinar una conclusión de un conjunto de datos y su importancia para tomar decisiones.

4. Discutimos con nuestros compañeros y profesor el trabajo realizado y establecemos acuerdos sobre cuáles son las respuestas.

BC Fundamentación Científica y Ejercitación

TRABAJO EN EQUIPO

1. Asignamos los roles correspondientes al interior del equipo y le solicitamos respetuosamente al compañero indicado realizar la siguiente lectura y elaboramos un mapa conceptual con las ideas más importantes:

Hipótesis

En estadística, una **hipótesis** es una afirmación acerca de una propiedad de una población. Esta se prueba a través de procedimientos estadísticos para identificar el nivel de significancia de la propiedad de una población.

La idea que se tiene para realizar una hipótesis es: Si, bajo un supuesto dado, la probabilidad de un suceso observado particular es excepcionalmente pequeña, concluimos que el supuesto probablemente es incorrecto. En ese sentido, una hipótesis de un conjunto de datos es un intento por distinguir resultados que pueden ocurrir más fácilmente que otros por el azar.

Tipos de hipótesis

Las hipótesis que se pueden construir son la nula y la alternativa.

La hipótesis nula (denotada por H_0) es la afirmación de que el valor de un parámetro de población (como una proporción, media o desviación estándar) es igual a un valor aseverado.

La hipótesis alternativa (denotada por H_1 o H_a o H_A) es la afirmación de que el parámetro tiene un valor que, de alguna manera, difiere de la hipótesis nula.

Pasos para la prueba de hipótesis

Paso 1: Datos
Se debe entender la naturaleza de los datos que forman la base de los procedimientos de prueba ya que estos determinan el tipo de prueba a utilizar. Por ejemplo, se debe determinar si los datos constan de conteo o medidas.

Paso 2: Restricciones o supuestos
Las diferentes suposiciones son las que modifican los procedimientos. Por ejemplo, las suposiciones respecto a la normalidad de la distribución de la población o la igualdad de varianzas e independencia de las muestras.

Paso 3: Hipótesis
En la prueba de hipótesis se trabaja con dos hipótesis estadísticas que deben anunciarse explícitamente. La primera es la hipótesis que debe probarse, es decir, la hipótesis nula. Este enunciado está relacionado con las condiciones que se suponen ciertas en la población de interés. En general, la hipótesis nula se establece con el propósito expreso de ser rechazada. En caso que esta no se rechace, indica que los datos no proporcionan evidencia suficiente que cause el rechazo y si conduce a este, señala que los datos disponibles no son compatibles con la hipótesis nula pero sirven para generar otra llamada alternativa, la cual es una proposición que se creerá cierta si los datos de la muestra llevan al rechazo de la hipótesis nula.

Para construir la hipótesis nula se utiliza el indicador de igualdad ($=, \leq o \geq$). Por ejemplo, la media de una poblacion es igual a 50 y se representa: $H_0: \mu = 50$. Y para la hipótesis alternativa se utiliza el indicador de diferente ($\neq, < o >$). Por ejemplo:

$$H_a: \mu \neq 50$$

$$H_a: \mu < 50$$

$$H_a: \mu > 50$$

Es posible establecer las siguientes reglas para determinar si una proposición se puede utilizar como hipótesis nula o alternativa:

1. La conclusión a la que se desea o espera llegar como resultado de la prueba generalmente se usa como hipótesis alternativa.
2. La hipótesis nula debe contener una proposición de igualdad, ya sea $=, \leq o \geq$.
3. La hipótesis nula es la que se debe comprobar.
4. Las hipótesis nula y alternativa son complementarias. Es decir, las dos contemplan de manera exhaustiva todos los valores posibles que los parámetros de suposición pueden asumir.

Paso 4: Estadística de prueba

En una forma para calcular la estadística de la prueba, existen muchos valores posibles que pueden asumirse y dependen del valor particular observado en la muestra.

Las siguientes son las fórmulas que existen para realizar la estadística de prueba:

Estadístico de prueba para proporciones $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$

Estadístico de prueba para medias $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ o $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$

Estadístico de prueba para desviaciones estándar $X^2 = \dfrac{(n-1)s^2}{\sigma^2}$

Paso 5: Regla de decisión

Todos los valores posibles que la estadística de prueba puede asumir son puntos sobre el eje horizontal de la gráfica de su distribución, los cuales se dividen en dos grupos: Uno de ellos constituye lo que se conoce como región de rechazo y el otro forma la región de no rechazo. Los valores de la estadística de prueba que forman la región de rechazo son aquellos que tienen la menor probabilidad de ocurrir, mientras que los que forman la región de no rechazo tienen la mayor probabilidad de ocurrir, si la hipótesis nula es verdadera para ambas regiones.

La regla de decisión señala que se debe rechazar la hipótesis nula si el valor de la estadística de prueba que se calcula a partir de la muestra es uno de los valores de la región de rechazo, y que no se debe rechazar la hipótesis nula si el valor calculado de la estadística de prueba es uno de los valores de la región de no rechazo.

La decisión en cuanto a qué valores van hacia la región de rechazo y cuáles van hacia la de no rechazo se toma con base en el nivel de significación deseado, indicado por α. El término de significación refleja el hecho que algunas veces la prueba hipótesis recibe el nombre de prueba de significación, la cual designa el área bajo la curva de la distribución de la estadística que está por encima de los valores sobre el eje horizontal que constituyen la región de rechazo.

Ejemplo 1:

Las puntuaciones en un examen que mide la variable de resolución de problemas siguen una distribución normal de media 11.5 en la población general de adolescentes.
En la Escuela Rojas Pinilla se ha implantado un programa de estimulación a una muestra de 30 estudiantes y ha proporcionado los siguientes resultados: 11, 9, 12, 17, 8, 11, 9, 4, 5, 9, 14, 9, 17, 24, 19, 10, 17, 17, 8,23, 8, 6, 14, 16, 6, 7, 15, 20, 14, 15; a un nivel de confianza del 95% ¿Puede afirmarse que el programa es efectivo?

Solución:
La hipótesis nula es $H_0: \mu = 11.5$ y la alternativa es $H_a: \mu > 11.5$.
La forma de calcular el estadístico es $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$. Donde $\bar{x} = 12.47$ es la media y la desviación estándar es $s = 5.22$. Sustituyendo tenemos:

$$t == \frac{12.47 - 11.5}{\frac{5.22}{\sqrt{30}}} = \frac{0.97}{\frac{5.22}{5.47}} = 1.01088$$

Buscamos en la tabla de la $t\text{-}student$ con 30 grados de libertad, el valor que deja por debajo de sí una probabilidad de 0.95 que resulta ser 1.697. El valor del estadístico es menor que el crítico, por consiguiente se acepta la hipótesis nula. Entonces, no hay evidencia de que el programa sea efectivo.

La tabla $t\text{-}student$ se utiliza para determinar con sus valores si una hipótesis se rechaza o no. La muestra de la tabla es para 30 datos:

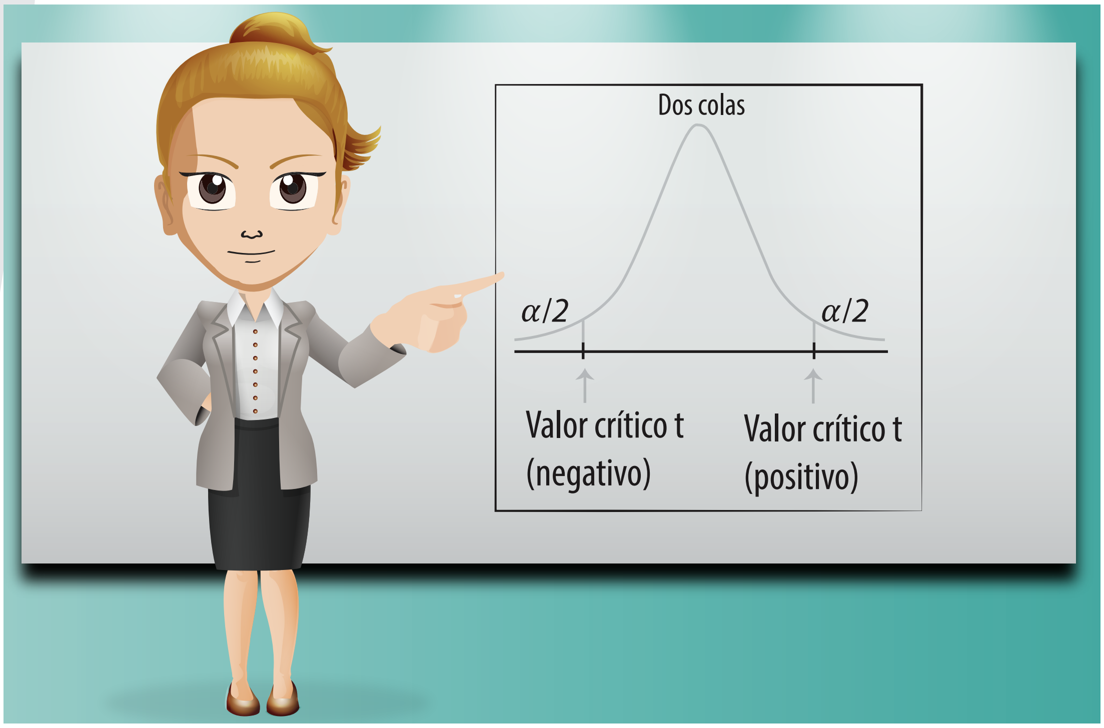
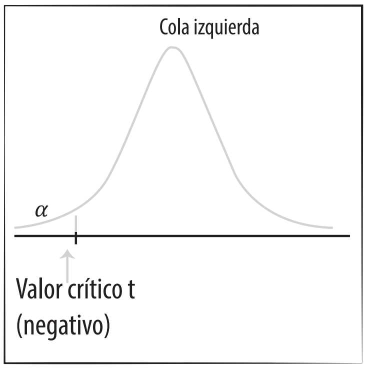
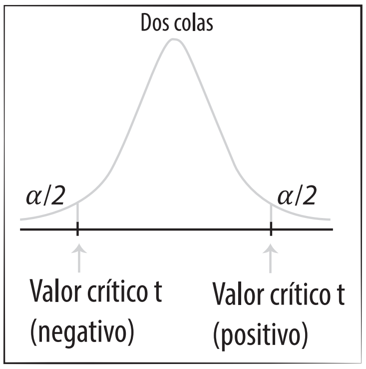


Tabla A-3	Distribución *t:* Valores críticos *t*				
	Área en una cola				
	0.005	0.01	0.025	0.05	0.10
Grados de libertad	Área en dos colas				
	0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.694	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.967	1.310





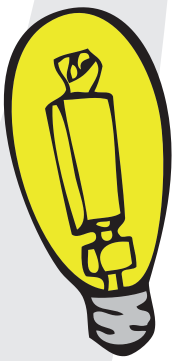
2. Resolvemos las siguientes situaciones utilizando la información de la tabla anterior que indica el valor de ***t-student*** con relación al número de grados de libertad y determinamos si se rechaza o no la hipotesis nula:

   a. Se desea determinar si una clase de 16 estudiantes puede desempeñarse bien tanto en español como en matemáticas. A continuación se presentan las calificaciones que se dieron en el primer período:

ESTUDIANTE	ESPAÑOL	MATEMÁTICAS
Andrea	8.4	8.4
Beto	5.5	5.7
Carolina	8.5	9
Diego	8	9.7
Enrique	5.5	7.4
Federico	8	5.4
Guillermo	6.4	7
Holman	9.5	7.5
Inírida	8.6	9
Jairo	9.1	9.3
Luis	7.5	7.7
Mariana	9.2	8.6
Nancy	8.5	8.8
Olivia	7.5	8.7
Pedro	9.1	8.5
Rosa	9.2	8.6

Considerando que las calificaciones de la prueba se distribuyen normalmente, se debe probar que la hipótesis de la puntuación media de la población en español es la misma que en matemáticas contra la hipótesis alternativa de que son diferentes para $\alpha = 0.05$.

b. El ingreso medio de los trabajadores de una industria petrolera es de \$ 1 580 000 con una desviación de 300 000 y la empresa afirma que los ingresos se han incrementado. Comprobamos esta afirmación con una muestra de 19 empleados, encontrando un ingreso medio de 1 650 000 pesos. A un nivel del 5% de significación, ¿apoyaría la afirmación de la empresa?



c. La vida útil promedio de una muestra aleatoria de 10 bombillos Metal Halide (luz para escenarios deportivos) es de 7 000 horas, con una desviación estándar muestral de 200 horas. Se supone que la afirmación dice que la vida útil de los bombillos es superior a 4 000 horas. A un nivel del 5% de significación, ¿apoyaría esta afirmación?

3. De las anteriores hipótesis, determinamos qué decisión es la más acorde con relación al conjunto de datos.

4. Invitamos a nuestro docente a que revise nuestro trabajo y le solicitamos amablemente que nos aclare las dudas.

D Aplicación

TRABAJO EN EQUIPO

1. Elaboramos una encuesta entre 10 a 30 personas sobre los ingresos mensuales que tienen contrastados con sus gastos. Si es posible nos apoyamos en ayudas ofimáticas para poder realizar las encuestas. Elaboramos una hipótesis nula y la comprobamos.

2. Realizamos una encuesta entre 5 a 15 personas de nuestro curso sobre sus rendimientos en las materias de matemáticas, español, ciencias sociales y ciencias naturales. Elaboramos una hipótesis y la comprobamos.

3. Los siguientes datos muestran algunas denuncias sobre exclusión social. A partir de ellos, elaboramos una hipótesis y la comprobamos. Escribimos una de las decisiones que se pueden tomar para evitar este tipo de acciones:

TABLA 3.1. **Estimación de la población y del número de hogares en situación de exclusión social, 2007 - 2013**

	Población			Hogares		
Total exclusión social	Mediados 2007	Finales 2009	Mediados 2013	Mediados 2007	Finales 2009	Mediados 2013
Total (miles)	44.874	45.983	46.425	16.329	17.121	17.441
Proporción excluidos (%)	16,3	18,7	25,1	15,8	17,2	21,9
Estimación excluidos (miles)	7.314	8.599	11.746	2.580	2.945	3.820
Crecimiento respecto de 2007 (%)	–	17,6	60,6	–	14,1	48,0
	Población			**Hogares**		
Exclusión social severa	Mediados 2007	Finales 2009	Mediados 2013	Mediados 2007	Finales 2009	Mediados 2013
Total (miles)	44.874	45.983	46.610	16.329	17.121	17.441
Proporción excluidos (%)	6,2	7,5	10,9	5,6	6,7	8,9
Estimación excluidos (miles)	2.782	3.449	5.080	914	1.147	1.552
Crecimiento respecto de 2007 (%)	–	24,0	82,6	–	25,5	69,8

Fuente: EINSFOESSA 2007,2009 Y 2013.

4. En la actividad de conjunto, socializamos las decisiones que se pueden tomar a partir de los análisis de datos.

5. Solicitamos a nuestro profesor que nos aclare todas nuestras inquietudes al respecto y revise el desarrollo de nuestro trabajo.

E Complementación

TRABAJO EN PAREJAS

1. Realizamos la lectura sobre la prueba *t-student* para aceptar o rechazar:

La distribución T

Es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

La prueba *t-student* fue desarrollada en 1899 por el químico inglés William Sealy Gosset (1876-1937), mientras trabajaba en técnicas de control de calidad para las destilerías Guiness en Dublín. Debido a que en la destilería, su puesto de trabajo no era inicialmente de estadístico y su dedicación debía estar exclusivamente encaminada a mejorar los costos de producción, publicó sus hallazgos anónimamente firmando sus artículos con el nombre de "Student".

2. Buscamos la biografía y los aportes a la estadística de William Sealy Gosset y cómo su contribución se generalizó a todos los estudios de datos.

3. Continuamos con la lectura:

La prueba de *t-student* es un método de análisis estadístico que compara las medias de dos grupos diferentes. Esta es paramétrica, es decir, sólo sirve para comparar variables numéricas de distribución normal.

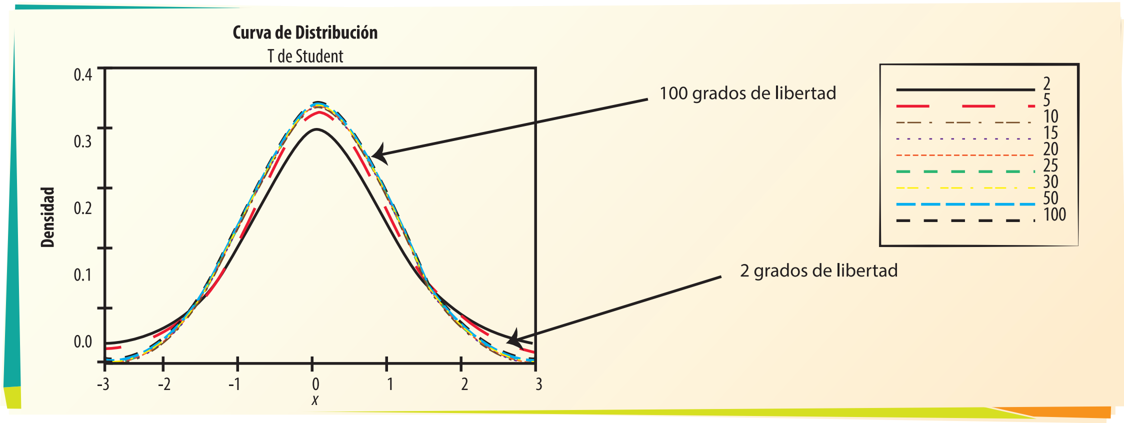
La prueba *t-student* arroja el valor del estadístico t. Según sea el valor de t, corresponderá un valor de significación estadístico determinado. Para calcularlo, su fórmula es:

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

Diferencia a probar

Desviación estándar de la diferencia o Error Estándar

La forma de la distribución *t-student* depende del parámetro, llamado el número de grados de libertad, que es igual al tamaño de la muestra:



Aunque parece una distribución normal, la distribución t tiene un poco más de área en los extremos y menos en el centro cuando los grados de libertad son pocos. Además, la distribución t es más bien una colección de distribuciones, una para cada número de grados de libertad.

4. Complementamos la información buscando en páginas de internet o libros de estadística sobre la prueba *t-student*.

Evaluación por competencias

1. Si tengo las hipótesis nula (H_0) y alternativa (H_a) de un conjunto de datos, el nivel de significación se define como:

A. Probabilidad de rechazar H_0 cuando H_0 es verdadera.

B. Probabilidad de aceptar H_1 cuando H_1 es verdadera.

C. Probabilidad de aceptar H_0 cuando H_0 es falsa.

D. Probabilidad de rechazar H_0 cuando H_1 es verdadera.

1

2. Un laboratorio introduce un medicamento para mejorar el estado de atención de los estudiantes. Al finalizar el año, 125 estudiantes lo tomaron obteniendo un rendimiento de 90 con una desviación estándar de 35. En años anteriores su rendimiento promedio fue de 78. Si se quiere probar el mejoramiento por el medicamento a un nivel de significación del 5%, las hipótesis a probar son:

A. $H_0: \mu = 90. H_a: \mu > 90$

B. $H_0: \mu = 78. H_a: \mu < 78$

C. $H_0: \mu: 90. H_a: \mu < 90$

D. $H_0: \mu = 78. H_a: \mu > 78$

2

INFORMACIÓN PARA CONTESTAR LAS PREGUNTAS 3 y 4

En los pasados seis meses algunos de los empleados nuevos han asistido a un seminario de entrenamiento dado por la empresa A y la empresa B. Al final del seminario, todos los empleados tomaron un examen para obtener el certificado. El seminario en la empresa A es más caro, pero en general se cree que el entrenamiento que se ofrece en ella es mejor que el de la empresa B. Se han recibido los resultados de las calificaciones de 15 empleados que estudiaron en la empresa A y de 15 empleados que estudiaron en la empresa B. Basado en estas calificaciones:

Persona	Empresa A	Empresa B
1	99	98
2	99	96
3	98	96
4	97	95
5	90	85
6	85	80
7	84	79
8	82	78
9	81	75
10	79	73
11	79	72
12	68	69
13	61	67
14	60	62
15	56	60

3. ¿Cómo puedo comprobar que el programa de la empresa A es mejor que el de la empresa B?:

A. Si el valor es menor de 0.05, sí podemos decir que el entrenamiento en la empresa A es significativamente mejor que el entrenamiento de la empresa B.

B. Si el valor no es menor de 0.05, no podemos decir que el entrenamiento en la empresa B es significativamente mejor que el entrenamiento de la empresa A.

C. Si el valor no es menor de 0.05, no podemos decir que el entrenamiento en la empresa A es significativamente mejor que el entrenamiento de la empresa B.

D. Si el valor es menor de 0.05, sí podemos decir que el entrenamiento en la empresa B es significativamente mejor que el entrenamiento de la empresa A.

3

4. La fórmula *t-student* aplicada a la empresa A es:

A. $t = \dfrac{81.2 - 0.05}{\dfrac{14.5}{\sqrt{15}}}$

B. $t = \dfrac{79 - 0.05}{\dfrac{12.6}{\sqrt{15}}}$

C. $t = \dfrac{81.2 - 79}{\dfrac{14.5}{\sqrt{15}} - 12.6}$

D. $t = 0.667$

4

5. Para comprobar la utilidad de una técnica de motivación, un investigador pasa una prueba de rendimiento académico a una muestra de 9 personas. Después, aplica su técnica motivacional y tras ello, vuelve a pasar la prueba de rendimiento. Los resultados fueron los siguientes:

Persona	Primer test	Segundo test
Andrés	8	12
Carolina	14	11
Alejandra	11	9
Willin	10	10
Elber	19	12
Tania	17	10
Olga	11	12
Alejandro	19	16
Carlos	14	8

A un nivel de confianza del 95%, ¿podemos rechazar que los rendimientos académicos son iguales antes y después, frente a la alternativa de que se produzca una mejora?:

A. SÍ

B. NO

Glosario

- **Desviación estándar:** Es una medida del grado de dispersión de los datos con respecto al valor promedio. Es decir, esta es simplemente el "promedio" o variación esperada con respecto a la media aritmética:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x}\,)^2}$$

- **Hipótesis:** Es una aseveración, afirmación o proposición acerca de la propiedad de una población.

- **Mediana:** Representa el valor de la variable de posición central en un conjunto de datos ordenados.

- **Moda:** Es el de mayor frecuencia absoluta.

- **Nivel de significancia:** Es la probabilidad de que el estadístico de prueba caiga en la región crítica, cuando la hipótesis nula es verdadera.

- **Promedio:** Es la división entre la suma de los valores y el número de datos involucrados. Simbólicamente:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

- **Prueba de hipótesis:** Es un procedimiento estándar para probar una afirmación acerca de una población. También se conoce como prueba de significancia.

- **Rango**: Intervalo entre el valor máximo y el mínimo; por ello, comparte unidades con los datos.

- **Región crítica:** Es un área donde se encuentran todos los valores del estadístico de prueba que provocan que se rechace la hipótesis nula.

- **Varianza:** Una variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2$$

Siendo X_i cada dato, n el número de datos y $\bar{X}$ la media de los datos.

Bibliografía

Ángel, A (2008). Álgebra intermedia. Séptima edición. México: Pearson Educación.

Arriza, B (2006). Guía práctica para análisis de datos. España: Consejería de innovación, ciencia y empresa.

Swokowski, E. y Cole, J (2009). Álgebra y trigonometría con geometría analítica. Doceava edición. México: Pearson Educación.

Camargo, L. y otros (1999). ALFA 9. Serie de matemáticas para educación básica secundaria y media. Editorial NORMA.

Paz, M (s.f.). Números complejos. Etsi Industriales: Departamento de matemáticas.

Triola, M.F (2009) Estadística. Décima Edición. México: Pearson Educación.

Walpole, R.; Myers, R. y Myers, S (2009). Probabilidad y estadística para ingenieros. México: Pearson Educación.