

PRUEBAS DE HIPOTESIS PARA **MEDIA**

Introducción

En este capítulo ya estamos en el final del ciclo del método científico, tenemos todas las bases para poder hacer inferencia estadística, es decir usar la información de la muestra para generalizar y llegar a conclusiones acerca de la población de interés. Partiremos con lo más básico que es la inferencia para medias en una población y veremos que los pasos para hacer inferencia se repiten, lo que cambia es el parámetro de interés.

EL LENGUAJE DE TOMA DE DECISIONES EN ESTADISTICA

Definiciones:

La hipótesis nula, denotada por H_0 , es el “status quo”, lo convencional, lo que sabemos de la población, lo aceptado hasta el momento.

La hipótesis alternativa, denotada por H_1 , es una alternativa a la hipótesis nula, implica cambio, es lo que el investigador espera que sea cierto.

EL LENGUAJE DE TOMA DE DECISIONES EN ESTADISTICA

Durante el curso revisaremos los pasos para tomar una decisión en estadística, por ahora los podemos resumir en:

1. Tenemos una hipótesis, que la asumimos cierta.
2. Obtenemos datos de un experimento relativo a la hipótesis.
3. Tomamos una decisión acerca de la hipótesis a partir de cuán probable son esos datos proviniendo de la hipótesis.
4. Al tomar una decisión se cometerán 2 tipos de errores.

EL LENGUAJE DE TOMA DE DECISIONES EN ESTADISTICA

En términos estadísticos estos errores tienen nombres especiales.

Definición:

Error Tipo I: es el error que se comete cuando rechazamos la hipótesis nula (H_0) en circunstancia que es la hipótesis verdadera.

Error tipo II: es el error que se comete cuando aceptamos la hipótesis nula (H_0) en circunstancia que es la hipótesis falsa.

EL LENGUAJE DE TOMA DE DECISIONES EN ESTADISTICA

Conceptos claves:

En estadística, la probabilidad de cometer el error tipo I se llama **nivel de significancia** o **nivel de significación** y se denota por la letra griega alfa, α .

La probabilidad de cometer el error tipo II se denota por la letra griega beta, β .

α = nivel de significancia.

α = probabilidad de cometer el error tipo I.

¿De qué otra manera podemos escribir la definición de α ?

β = probabilidad de cometer el error tipo II.

¿De qué otra manera podemos escribir la definición de β ?

Pruebas de Hipótesis

Pasos en una prueba de hipótesis:

1. Establecer la hipótesis.
2. Definir el nivel de significación.
3. Obtener los datos.
4. Definir test estadístico y verificar los supuestos.
5. Calcular el test estadístico observado bajo H_0 .
6. Calcular el valor p .
7. Tomar la decisión con respecto a H_0 .
8. Conclusión del investigador.

Pruebas de Hipótesis para Media

Ahora veremos el caso donde lo que nos interesa es investigar sobre la media de una población.

Escriba las hipótesis nula y alternativa que usaría para probar las siguientes aseveraciones. Las hipótesis deben ser expresadas en términos del parámetro μ , la media de la población de interés.

- a) La media de la edad de los trabajadores en una empresa es de más de 60 años.
- b) La media del contenido de cafeína de una taza regular de café es de menos de 110 mg.
- c) El promedio del peso de carga de camionetas difiere de 300 kilos.

Inferencia para media con desviación estándar desconocida

Varianza de la población (σ^2) desconocida

El cálculo del error estándar de la media muestral incluye a σ , pero casi nunca vamos a conocer la variabilidad en la población en estudio. Cuando se desconoce el σ del universo, el error estándar del promedio debe calcularse a partir de la desviación estándar de la muestra:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

En este caso ya no es lícito trabajar con la distribución normal y la variable normal estándar,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

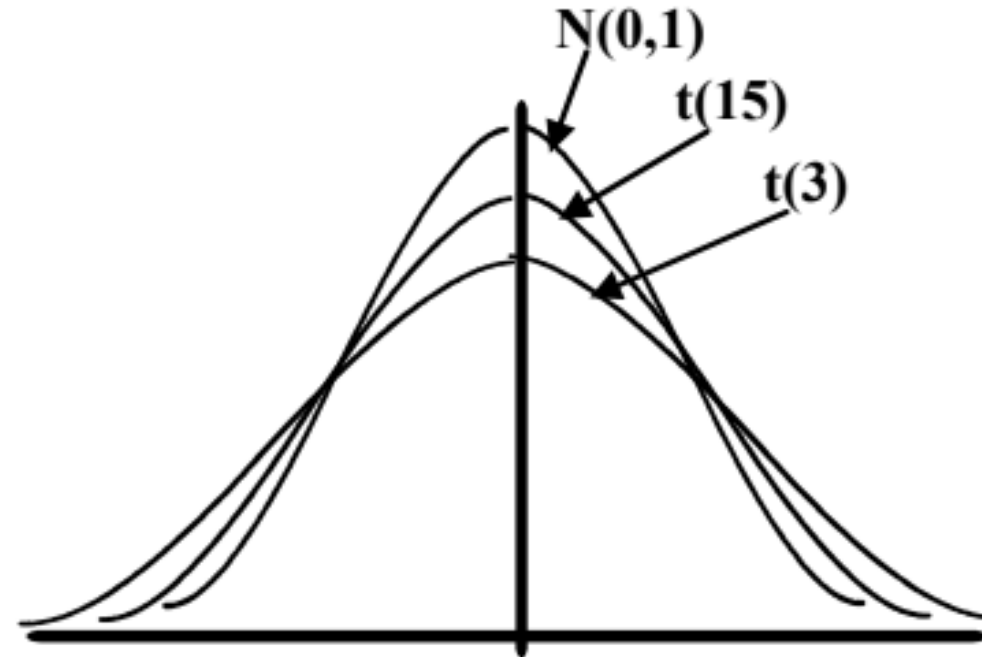
sino que se trabajará con la variable t de Student:

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

Inferencia para media con desviación estándar desconocida

Esta variable t sigue una distribución t de Student con $(n-1)$ grados de libertad.

Propiedades de la distribución t de Student



- Los valores de t dependen del número de grados de libertad, los que se determinan a partir del número usado en el denominador para el cálculo de la desviación estándar (s) es decir $(n-1)$.
- La función de densidad de la distribución t de Student tiene forma de campana simétrica, similar a la distribución normal $N(0,1)$.
- Es un poco más “chata” y tiene “colas más pesadas” que la $N(0,1)$.
- Cuando el tamaño de la muestra aumenta, la distribución t de Student se aproxima a la $N(0,1)$.

Resumen

Resumen del test t simple para la media de una población

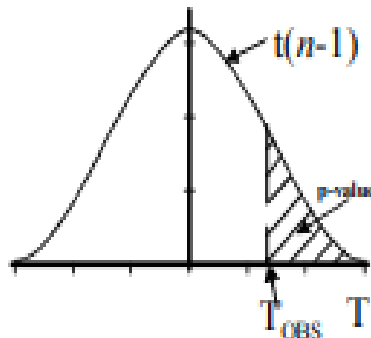
- Estamos interesados en **docimar la hipótesis** acerca de la **media de una población** μ . La hipótesis nula es $H_0: \mu = \mu_0$ donde μ_0 es un **valor hipotético** para μ . La hipótesis alternativa da la dirección del test.
- Se asume que los datos provienen de una **muestra aleatoria** de tamaño n de una población con **distribución Normal** con desviación estándar σ **desconocida**. El supuesto de normalidad no es crucial si el tamaño de la muestra es grande.
- Nuestra decisión acerca de μ , será en base al **valor de la media muestral estandarizada** \bar{x} , el cual es $t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$.
Este es el **test estadístico** y su distribución bajo H_0 , es una distribución t con $n-1$ grados de libertad.

Resumen

- El **valor-p** del test, depende de la hipótesis alternativa:

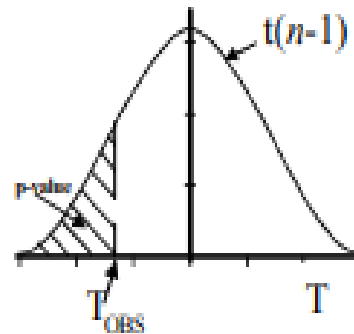
Test Unilateral, cola superior

Si $H_1 : \mu > \mu_0$, entonces el valor-p es $P(t > t_{obs})$ es el área a la derecha del test estadístico observado bajo H_0 .



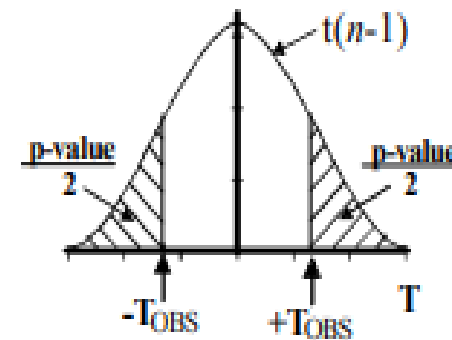
Test Unilateral, cola inferior

Si $H_1 : \mu < \mu_0$, entonces el valor-p es $P(t < t_{obs})$ es el área a la izquierda del test estadístico observado bajo H_0 .



Test Bilateral

Si $H_1 : \mu \neq \mu_0$, entonces el valor-p es $2P(t > |t_{obs}|)$ el área afuera de las dos colas del test estadístico observado bajo H_0 .



- Decisión:** Si el valor-p es menor que el nivel de significancia entonces rechazamos H_0 .

Tabla Distribución t de Student

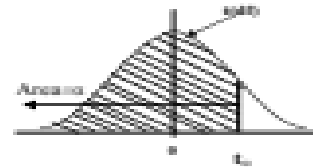


Tabla: Percentiles de la distribución t de Student

gl	$t_{0.40}$	$t_{0.30}$	$t_{0.20}$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.885	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.657
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.525
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.341
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576
gl	$-t_{0.40}$	$-t_{0.30}$	$-t_{0.20}$	$-t_{0.10}$	$-t_{0.05}$	$-t_{0.025}$	$-t_{0.01}$	$-t_{0.005}$

Ejemplo :

Un profesor universitario lee un artículo que discute sobre el tiempo que dedican los alumnos para estudiar cada semana. El artículo dice que los alumnos estudian en promedio 7,06 horas por semana. El profesor cree que los alumnos estudian más de 7,06 horas por semana en promedio. El profesor elige una muestra aleatoria simple de 15 estudiantes de su universidad y encuentra que el promedio muestral de estudio es $\bar{x} = 8,43$ horas y la desviación estándar es $s = 4,32$ horas . Asuma que el tiempo de estudio semanal sigue una distribución normal.

- a) Establezca la hipótesis a docimar.
- b) Calcule el test estadístico.
- c) Encuentre el valor p del test.
- d) Los resultados, ¿son estadísticamente significativos al 5%?
- e) Escriba su conclusión al profesor.