Generalidades de Prueba de Hipótesis

En estadística, se denomina *población* al universo de casos posibles. Este término, como muchos otros, tienen sus orígenes en la aplicación de la estadística al estudio de características demográficas. Sin embargo, no se aplica sólo a individuos, sino que se utiliza en forma mucho más amplia.
Lamentablemente, en general no conocemos exactamente la población. De hecho, la población puede ser infinita. Por lo tanto, se suele *inferir* información acerca de la población a partir de la observación de una *muestra* de tamaño finito de la misma. Ya hemos visto, por ejemplo, algunas formas de estimar parámetros de la población a partir de una muestra.

Una *hipótesis estadística* es una afirmación sobre parámetros de una población. Una *prueba de hipótesis* consiste en una forma de contrastar varias hipótesis (mutuamente excluyentes) entre sí, evaluando su posible veracidad. Dado que, en general, no conocemos completamente la población, nunca podemos tener certidumbre total sobre la veracidad de un hipótesis estadística.

Aquí nos concentraremos en pruebas de hipótesis *binarias*, es decir, en las cuales se contrastan sólo dos hipótesis. En particular, en las aplicaciones se tiene interés en poner a prueba la veracidad de cierta hipótesis particular, la cual se denomina *hipótesis nula* (H_0). Para ello, se la compara con una hipótesis complementaria o *hipótesis alternativa* (H_1). El análisis de la veracidad de ambas hipótesis se realiza sobre la base de los datos de una o varias muestras. Nosotros nos enfocaremos en los casos en que se utiliza una sola muestra. Para facilitar el análisis, los datos de la muestra se suelen resumir en un solo valor denominado *estadístico de prueba* (Λ). Basados en en el valor del estadístico de prueba para una muestra particular, se toma una decisión: aceptar o rechazar la veracidad de la hipótesis nula. La hipótesis se rechaza cuando $\Lambda \in R$, siendo R una región pre-determinada de rechazo conocida como la *región crítica*.

Dado que nunca se puede tener una certeza total, se pueden cometer distintos tipos de errores:

	H_0 verdadera	H_0 falsa
Se acepta H_0	OK	Error Tipo II
Se rechaza H_0	Error Tipo I	OK

Una forma de plantear la prueba de hipótesis es fijando un valor α (pequeño) para la máxima probabilidad admisible de cometer un Error Tipo I. A este valor se lo denomina el *nivel de significación* de la prueba. Dado α, se puede determinar la región crítica $R^{(\alpha)}$ tal que

$$P_{H_0}\left(\Lambda \in R^{(\alpha)}\right) \leq \alpha,$$

donde la probabilidad se evalúa en el caso en que la hipótesis nula sea verdadera. Fijada la región crítica, se puede cometer un Error Tipo II. La probabilidad de cometer este error se suele denominar con la letra β. Se conoce como *potencia de la prueba de hipótesis* a $1-\beta$.

Dada una muestra, se puede calcular un valor observado λ_{obs} del estadístico Λ. Se define el *valor p* de la prueba como la probabilidad de que Λ sea "peor" que λ_{obs} cuando H_0 es verdadera. Por "peor" queremos decir que hace más sospechoso que la hipótesis nula sea verdadera. La hipótesis nula se rechaza si el valor p es menor que el nivel de significación α.

Prueba de hipótesis para la media - Generalidades

Sea X una variable aleatoria con media μ. Existen tres tipos de pruebas de hipótesis:

1. De dos colas:
   - Prueba:
   $$\begin{aligned} H_0 &: \ \mu = \mu_0 \qquad \mu_0 \text{ es un valor dado} \\ H_1 &: \ \mu \neq \mu_0 \end{aligned}$$
   - Condición de rechazo: $\Lambda \notin (\lambda_l, \lambda_u) \subset \mathbb{R}$. Si $\mu = \mu_0$ (la hipótesis nula es cierta), $E_{\mu_0}[\Lambda] \in (\lambda_l, \lambda_u)$.
   - Probabilidad de Error Tipo II: $\beta(\mu_1) = P_{\mu_1}(\Lambda \in (\lambda_l, \lambda_u))$, donde $P_{\mu_1}(\cdot)$ denota la probabilidad cuando $\mu = \mu_1$. Hay cinco valores característicos:
     - $\beta(\mu_0) = 1 - \alpha$.
     - $\beta(\mu_l) = \beta(\mu_u) \approx 0.5$ si la distribución de Λ es simétrica.
     - $\beta(-\infty) = \beta(+\infty) = 0$.

     La gráfica de $\beta(\mu_1)$ vs. μ_1 se conoce como la *curva característica*. Por otro lado, se llama *curva de potencia* a la de $1 - \beta(\mu_1)$ vs. μ_1.
   - Valor p: $p(\lambda_{\text{obs}}) = P_{\mu_0}(|\Lambda - E_{\mu_0}[\Lambda]| > |\lambda_{\text{obs}} - E_{\mu_0}[\Lambda]|)$.
2. De cola derecha:
   - Prueba:
   $$\begin{aligned} H_0 &: \ \mu \leq \mu_0 \qquad \mu_0 \text{ es un valor dado} \\ H_1 &: \ \mu > \mu_0 \end{aligned}$$
   - Condición de rechazo: $\Lambda > \lambda_u$.
   - Probabilidad de Error Tipo II: $\beta(\mu_1) = P_{\mu_1}(\Lambda \leq \lambda_u)$. Hay tres valores característicos:
     - $\beta(\mu_0) = 1 - \alpha$.
     - $\beta(\lambda_u) = 0.5$ si la distribución de Λ es simétrica.
     - $\beta(+\infty) = 0$.
   - Valor p: $p(\lambda_{\text{obs}}) = P_{\mu_0}(\Lambda > \lambda_{\text{obs}})$.
3. De cola izquierda:
   - Prueba:
   $$\begin{aligned} H_0 &: \ \mu \geq \mu_0 \qquad \mu_0 \text{ es un valor dado} \\ H_1 &: \ \mu < \mu_0 \end{aligned}$$
   - Condición de rechazo: $\Lambda < \lambda_l$.
   - Probabilidad de Error Tipo II: $\beta(\mu_1) = P_{\mu_1}(\Lambda \geq \lambda_l)$. Hay tres valores característicos:
     - $\beta(\mu_0) = 1 - \alpha$.
     - $\beta(\lambda_l) = 0.5$ si la distribución de Λ es simétrica.
     - $\beta(-\infty) = 0$.
   - Valor p: $p(\lambda_{\text{obs}}) = P_{\mu_0}(\Lambda < \lambda_{\text{obs}})$.

Prueba de hipótesis para la media con varianza conocida

Asumamos que la muestra $X_1, X_2, \cdots, X_n$ de una población puede considerarse un conjunto de n variables aleatorias i.i.d. *normales* con media μ **desconocida** y varianza σ^2 **conocida**. Dos estadísticos de prueba habituales son la media muestral $\bar{X}$ y

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

Dado que es fácil pasar de uno estadístico a otro, aquí sólo presentaremos los resultados usando Z. Si α es el nivel de significación, las pruebas de hipótesis quedan:

Prueba de Hipótesis	**Condición de rechazo**	**Prob. Error Tipo II** $\beta(\mu_1)$	**Valor** p $p(z_{\text{obs}})$
$H_0:\ \mu = \mu_0$ $H_1:\ \mu \neq \mu_0$	$Z < -z_{1-\frac{\alpha}{2}}$ o $Z > +z_{1-\frac{\alpha}{2}}$	$\Phi\left(+z_{1-\frac{\alpha}{2}} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$ $-$ $\Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$2\left(1 - \Phi\left(z_{\text{obs}}\right)\right)$
$H_0:\ \mu \leq \mu_0$ $H_1:\ \mu > \mu_0$	$Z > +z_{1-\alpha}$	$\Phi\left(+z_{1-\alpha} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$1 - \Phi\left(z_{\text{obs}}\right)$
$H_0:\ \mu \geq \mu_0$ $H_1:\ \mu < \mu_0$	$Z < -z_{1-\alpha}$	$1 - \Phi\left(-z_{1-\alpha} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$\Phi\left(z_{\text{obs}}\right)$

Gracias al Teorema Central del Límite, **estas ecuaciones son válidas aunque las variables aleatorias no sean normales si n es grande**.

Sean o no normales las variables aleatorias, **si n es grande, las mismas condiciones de rechazo y fórmulas para el valor p se pueden utilizar aún en el caso que la varianza sea desconocida**. En este caso, el estadístico de prueba que se utiliza es

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}.$$

Prueba de hipótesis para una proporción

Sea q la probabilidad **desconocida** de un evento dado. Otra forma de entender q es como la proporción de la población que satisface cierta propiedad. Sea X el número de individuos con la propiedad indicada de una muestra de n casos independientes. Si $n \ll N$, siendo N el tamaño de la población, $X \sim \text{Binomial}(n, q)$. Si n es grande, por el Teorema Central del Límite, la distribución de X es aproximadamente normal con media $\mu = nq$ y varianza $\sigma^2 = nq(1-q)$.
Tres estadísticos de prueba posibles son X, $\hat{q} = X/n$ y

$$Z = \frac{\hat{q} - q_0}{\sqrt{\frac{q_0(1-q_0)}{n}}}.$$

Dado que es sencillo pasar de un estadístico a otro, sólo presentaremos resultados correspondientes a Z. Si n es grande (> 100), gracias al Teorema Central del Límite las pruebas de hipótesis para un nivel de significación α quedan:

Prueba de Hipótesis	**Condición de rechazo**	**Prob. Error Tipo II** $\beta(q_1)$	**Valor** p $p(z_{\text{obs}})$
$H_0: \ q = q_0$ $H_1: \ q \neq q_0$	$Z < -z_{1-\frac{\alpha}{2}}$ o $Z > +z_{1-\frac{\alpha}{2}}$	$\Phi\left(+z_{1-\frac{\alpha}{2}}\sqrt{\frac{q_0(1-q_0)}{q_1(1-q_1)}} + \frac{q_0-q_1}{\sqrt{\frac{q_1(1-q_1)}{n}}}\right) - \Phi\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{q_0(1-q_0)}{q_1(1-q_1)}} + \frac{q_0-q_1}{\sqrt{\frac{q_1(1-q_1)}{n}}}\right)$	$2\left(1-\Phi\left(z_{\text{obs}}\right)\right)$
$H_0: \ q \leq q_0$ $H_1: \ q > q_0$	$Z > +z_{1-\alpha}$	$\Phi\left(+z_{1-\alpha}\sqrt{\frac{q_0(1-q_0)}{q_1(1-q_1)}} + \frac{q_0-q_1}{\sqrt{\frac{q_1(1-q_1)}{n}}}\right)$	$1-\Phi\left(z_{\text{obs}}\right)$
$H_0: \ q \geq q_0$ $H_1: \ q < q_0$	$Z < -z_{1-\alpha}$	$1-\Phi\left(-z_{1-\alpha}\sqrt{\frac{q_0(1-q_0)}{q_1(1-q_1)}} + \frac{q_0-q_1}{\sqrt{\frac{q_1(1-q_1)}{n}}}\right)$	$\Phi\left(z_{\text{obs}}\right)$

Ejercicio 16 de la guía 9

La proporción de defectuosos en un proceso es 2 %. Como se desea establecer un control sobre el mismo se inspeccionan con una frecuencia de una vez por hora muestras de tamaño n y se revisa el proceso si se encuentran c o más defectuosas entre las n. Para determinar n y c se han fijado un riesgo del 10 % (0.1 es la probabilidad de detener el proceso cuando en realidad no habría que hacerlo) y una probabilidad 0.2 de cometer el error de no detener el proceso en el caso en que el porcentaje de defectuosos sea de 6 %.

1. Especificar claramente la prueba de hipótesis planteada (hipótesis nula, alternativa, nivel de significación, estadístico de prueba y zona de rechazo).
2. Determinar los valores de n y c.
3. Representar gráficamente en forma aproximada las curvas de operación característica (curva OC) y de potencia indicando claramente las abscisas y ordenadas de al menos tres puntos.

Respuesta:
Dado que se quiere "mantener a raya" el porcentaje de defectuosos, lo que importa es vigilar que el mismo no aumente respecto del 2 % aceptable. Por lo tanto, la prueba de hipótesis debe ser

$$H_0 : \; q \leq 0.02 \tag{1}$$
$$H_1 : \; q > 0.02 \tag{2}$$

El proceso se detiene si los resultados sugieren que la hipótesis nula es falsa. Entonces, detener el proceso cuando no habría que hacerlo es cometer un Error de Tipo I. Ergo, el problema requiere un nivel significación $\alpha = 0.1$. Por otro lado, no detener el proceso cuando el porcentaje de defectuosos es superior al 2 % es un Error de Tipo II: el problema pide que $\beta(0.06) \leq 0.2$. Si n es grande,

$$\beta(0.06) = \Phi\left(+z_{1-0.1}\sqrt{\frac{0.02(1-0.02)}{0.06(1-0.06)}} + \frac{0.02-0.06}{\sqrt{\frac{0.06(1-0.06)}{n}}} \right) \leq 0.2 \tag{3}$$

$$\Phi\left(+z_{0.9}\sqrt{\frac{49}{141}} - \frac{2\sqrt{n}}{\sqrt{141}} \right) \leq 0.2 \tag{4}$$

$$z_{0.9}\sqrt{\frac{49}{141}} - \frac{2\sqrt{n}}{\sqrt{141}} \leq z_{0.2} \tag{5}$$

$$\frac{2\sqrt{n}}{\sqrt{141}} \geq z_{0.9}\sqrt{\frac{49}{141}} - z_{0.2} \tag{6}$$

$$\sqrt{n} \geq \frac{7}{2}z_{0.9} - \frac{\sqrt{141}}{2}z_{0.2} \tag{7}$$

$$n \geq \left(\frac{7}{2}z_{0.9} - \frac{\sqrt{141}}{2}z_{0.2}\right)^2 \approx \left(\frac{7}{2}1.2816 - \frac{\sqrt{141}}{2}(-0.8416)\right)^2 \tag{8}$$

$$n \geq 89.9136 \tag{9}$$

$$n \geq 90 \tag{10}$$

Podemos verificar:

$$\Phi\left(+z_{0.9}\sqrt{\frac{49}{141}}-\frac{2\sqrt{89}}{\sqrt{141}}\right)\approx 0.2023 \tag{11}$$

$$\Phi\left(+z_{0.9}\sqrt{\frac{49}{141}}-\frac{2\sqrt{90}}{\sqrt{141}}\right)\approx 0.1998 \tag{12}$$

Por lo tanto, el mínimo n es 90. La condición de rechazo es

$$\frac{\hat{q}-0.02}{\sqrt{\frac{0.02(1-0.02)}{90}}}>z_{0.9} \tag{13}$$

$$\hat{q}>0.02+\frac{7}{300}\sqrt{\frac{2}{5}}\,z_{0.9}\approx 0.0389\approx 0.04 \tag{14}$$

Definiendo $X=n\hat{q}=90\hat{q}$,

$$X>1.8+\frac{21}{10}\sqrt{\frac{2}{5}}\,z_{0.9}\approx 3.5021. \tag{15}$$

Es decir, el proceso se revisará (es decir, la hipótesis nula se rechazará) cuando el número de defectuosos de una muestra de $n=90$ sea mayor o igual a $c=4$.

La curva de operación característica se puede graficar a partir de

$$\beta(q_1)=\Phi\left(+z_{0.9}\sqrt{\frac{0.02(1-0.02)}{q_1(1-q_1)}}+\frac{0.02-q_1}{\sqrt{\frac{q_1(1-q_1)}{90}}}\right) \tag{16}$$

Es fácil ver que

$$\beta(0.0200)=0.9, \tag{17}$$
$$\beta(1.0000)=0.0. \tag{18}$$

A partir de las ecuaciones (13)-(14), tampoco es difícil mostrar

$$\beta(0.0389)=0.5. \tag{19}$$

Si bien $\beta(1)=0$, $\beta(q_1)$ es muy pequeño para valores de q_1 bastante más chicos. En efecto,

$$\beta(0.1)=0.0267, \tag{20}$$
$$\beta(0.2)\approx 7\times 10^{-5}. \tag{21}$$

La Fig. 1 muestra la curva de operación característica.

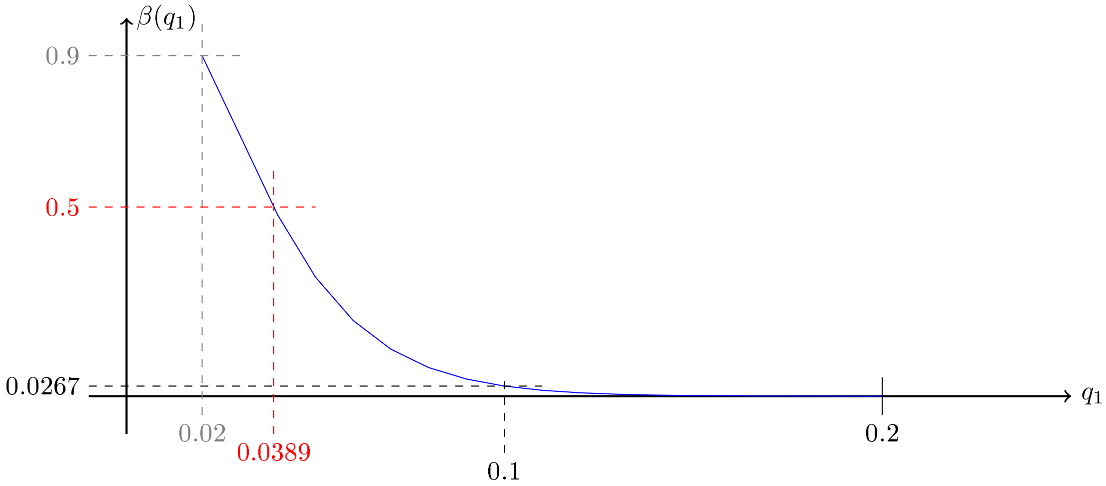

Figura 1: Curva de operación característica del ejercicio 16 de la guía 9.

Prueba de hipótesis para la media con varianza desconocida

Asumamos que la muestra $X_1, X_2, \cdots, X_n$ de una población puede considerarse un conjunto de n variables aleatorias i.i.d. *normales* con media μ y varianza σ^2 **desconocida**. El estadístico de prueba es

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}.$$

Si α es el nivel de significación, las pruebas de hipótesis quedan:

Prueba de Hipótesis	**Condición de rechazo**	**Prob. Error Tipo II** $\beta(\mu_1)$	**Valor p** $p(t_{\text{obs}})$
$H_0: \ \mu = \mu_0$ $H_1: \ \mu \neq \mu_0$	$T < -t_{n-1,1-\frac{\alpha}{2}}$ o $T > +t_{n-1,1-\frac{\alpha}{2}}$	$\Xi_{n-1}\left(+t_{n-1,1-\frac{\alpha}{2}} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$ $-$ $\Xi_{n-1}\left(-t_{n-1,1-\frac{\alpha}{2}} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$2\left(1 - \Xi_{n-1}\left(t_{n-1,\text{obs}}\right)\right)$
$H_0: \ \mu \leq \mu_0$ $H_1: \ \mu > \mu_0$	$T > +t_{n-1,1-\alpha}$	$\Xi_{n-1}\left(+t_{n-1,1-\alpha} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$1 - \Xi_{n-1}\left(t_{\text{obs}}\right)$
$H_0: \ \mu \geq \mu_0$ $H_1: \ \mu < \mu_0$	$T < -t_{n-1,1-\alpha}$	$1 - \Xi_{n-1}\left(-t_{n-1,1-\alpha} + \frac{\mu_0-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$	$\Xi_{n-1}\left(t_{\text{obs}}\right)$

Ξ_{n-1} es la función de distribución (acumulada) correspondiente a una variable aleatoria t de Student con $n-1$ grados de libertad. Cuando n es grande (> 200), $\Xi_{n-1}(\cdot) \approx \Phi(\cdot)$ y $t_{n-1,\delta} \approx z_\delta$.

Ejercicio 8 de la guía 9

Según un fabricante, la resistencia media de cierto alambre aleado de Al es de 250 MN/m^2. Un contratista adquiere un lote de alambre y pone a prueba una muestra de tamaño 35. El valor medio y la desviación estándar obtenidos a partir de esa muestra son 274.4 MN/m^2 y 11.2 MN/m^2, respectivamente. ¿Es justificable que el contratista concluya que la remesa tiene una resistencia significativamente diferente a lo especificado por el fabricante? Tomar un nivel de significación del 5 %.

Respuesta:
Al contratista le preocupa que si la afirmación del fabricante es cierta o no. La prueba de hipótesis adecuada es entonces

$$H_0: \ \mu = 250 \tag{22}$$
$$H_1: \ \mu \neq 250 \tag{23}$$

Siendo $n = 35$, el valor crítico de la prueba de hipótesis es $t_{34,0.975} \approx 2.0322$. El estadístico observado es

$$t_{\text{obs}} = \frac{274.4 - 250}{\frac{11.2}{\sqrt{35}}} \approx 12.8886. \tag{24}$$

Dado que $t_{\text{obs}} > t_{34,0.975}$, existe suficiente evidencia para rechazar la hipótesis nula a un nivel de significación del 5 %.