Unidad 7. Prueba de hipótesis

1. Introducción

Una prueba o test de hipótesis es un procedimiento inferencial que nos permitirá decidir, con un criterio estadístico, acerca de la verdad o falsedad de cierta conjetura. Por ejemplo

- si un sistema ha sido infectado por un virus,
- si la actualización de un software ha sido efectiva,
- si la velocidad de conexión promedio es mayor a 48 Mbps, como lo afirma un proveedor de servicio de internet,
- si los tiempos de espera asociados a un servidor provienen de una distribución exponencial

para mencionar algunas conjeturas que interesan en computación. La aplicación de la inferencia estadística incluye campos como la computación, la medicina, la ecología de poblaciones, la biología molecular, etc.

Típicamente, en una prueba de hipótesis se contrasta una afirmación a la que se denomina *hipótesis nula* versus otra, excluyente de la primera, que recibe el nombre de *hipótesis alternativa*. Utilizamos la siguiente notación:

$$H_0: \quad \text{Hipótesis nula}$$

$$H_A: \quad \text{Hipótesis alternativa}$$

Un test de hipótesis concluye con el rechazo - de la verdad- de H_0 a favor de H_A o bien con el no rechazo de H_0. Una decisión u otra está basada en la *evidencia* que proveen los datos.

Ejemplo 1.

Un proveedor de servicio de internet afirma que la velocidad de conexión promedio que provee supera 48 Mbps. ¿Será válida esta afirmación?.

$$H_0 : \mu = 48 \text{ versus } H_A : \mu > 48$$

siendo μ la velocidad media poblacional. □

Definición 1. *Sea μ_0 una constante y μ un parámetro poblacional (no necesariamente la media).*

- *Si la hipótesis alternativa es de la forma $H_A : \mu \neq \mu_0$ al test de hipótesis se le denomina de dos colas y a H_A se le llama bilateral.*
- *Si la hipótesis alternativa es de la forma $H_A : \mu > \mu_0$ ó $H_A : \mu < \mu_0$ al test de hipótesis se le llama de una cola y a H_A, unilateral.*

Ejemplo 2.

Para contrastar si el número medio de usuarios de un servicio de red se modificó de un año a otro consideramos

$$H_0 : \mu_2 - \mu_1 = 0 \text{ versus } H_A : \mu_2 - \mu_1 \neq 0,$$

siendo μ_1 el número medio de usuarios en el año 2021 y μ_2 el del año 2022. □

La utilización de una H_A bilateral o unilateral depende del tipo de problema, del contexto.

Ejemplo 3.

Para verificar que la proporción, p, de componentes electrónicos defectuosos en un lote es mayor a 0.3, se contrastan las hipótesis

$$H_0 : p = 0.3 \text{ versus } H_A : p > 0.3 \quad \square$$

Cualquier decisión que tomemos podría o bien ser correcta o ser errónea, con lo cual hay dos errores posibles.

Definición 2. *Al contrastar las hipótesis H_0 versus H_A, pueden ocurrir dos tipos de errores.*

- *Un error de tipo I ocurre cuando rechazamos una hipótesis nula verdadera.*
- *Un error de tipo II ocurre cuando no rechazamos una hipótesis nula falsa.*

La decisión de rechazar o no H_0 la tomaremos cuantificando la probabilidad de cometer error de tipo I.

Definición 3. *Al contrastar las hipótesis H_0 versus H_A, definimos*

$$\begin{aligned} \alpha &= P(\textit{ cometer error de tipo I}) = P(\textit{rechazar} H_0 \mid H_0 \textit{ es verdadera}), \\ \beta &= P(\textit{ cometer error de tipo II}) = P(\textit{aceptar} H_0 \mid H_0 \textit{ es falsa}), \end{aligned}$$

con $\alpha, \beta \in (0, 1)$.

El valor de α se determina a priori, antes de realizar el experimento que proveerá los datos. Típicamente los valores de α son $0.01, 0.05$ y 0.10.

Continuaremos a través del ejemplo 1 y al mismo tiempo formularemos enunciados sobre la estructura general de una prueba o test de hipótesis.

Ejemplo 4. *Ejemplo 1 revisitado*

Las **hipótesis** a contrastar son

$$H_0 : \mu = 48 \text{ versus } H_A : \mu > 48$$

siendo μ la velocidad media poblacional. Supongamos que la variable aleatoria "velocidad de conexión", que denotaremos con X, tiene una distribución $\mathrm{N}(\mu, 72)$. Asumiremos además que contamos con 80 mediciones que arrojan una media muestral de 49.6 Mbps. ¿Portan estos datos evidencia significativa para concluir que la media poblacional, μ, es mayor a 48, asumiendo que el nivel de significación ha sido fijado en 0.05?

Paso 1. Estadístico de la prueba.

Utilizando como estimador de μ a $\overline{X}_n$ definimos

$$T = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}, \tag{1}$$

donde $\mu_0 = 48$. Luego, T satisface

$$T \sim N(0,1) \text{ asumiendo } H_0 \text{ como válida}$$

y, se le denomina estadístico de la prueba.

Paso 2. Regiones de aceptación y rechazo .

Rechazaríamos la hipótesis nula cuando $\overline{X}_n$ esté "muy" por arriba de $\mu_0 = 48$ ó, equivalentemente, cuando el valor del estadistico $T = t$ esté "muy" por arriba de 0. ¿Qué significa "muy" por debajo? y, la respuesta a esta pregunta, dependerá del nivel α.

El razonamiento entonces es el siguiente: si z_o es tal que

$$
\begin{aligned}
\alpha &= P\,(\text{ cometer error de tipo I}) \\
&= P\,(\text{rechazar } H_0 \mid H_0 \text{ es verdadera}) \\
&= P\,(T \geq z_o \mid H_0)\,,
\end{aligned}
$$

entonces cualquier valor de T mayor o igual que z_o nos conduce a rechazar la hipótesis nula con una probabilidad de error (α) pequeña y, por ende, rechazamos H_0.

Definición 4. *Consideremos un test de hipótesis en el que se contrastan H_0 y H_A y sea T el estadístico del test. Al conjunto, $\mathcal{R}$, de valores del estadístico que conducen a rechazar H_0 se le llama zona de rechazo. La zona de aceptación, $\mathcal{A}$, se define como el complemento de $\mathcal{R}$.*

Para nuestro ejemplo, como

$$
\begin{aligned}
\alpha = 0.05 &= P\,(T \geq z_o \mid H_0) \\
&= P\,(T \geq z_o \mid T \sim \mathrm{N}(0,1)) \\
&= 1 - \Phi(z_o)
\end{aligned}
$$

entonces z_0 es el cuantil 0.95 de la distribución normal estándar; es decir, $z_o = z_{1-\alpha}$ satisface $P(T \leq z_o) = 0.95$ y, en consecuencia, $z_o = 1.64$. Luego , la zona de rechazo es

$$\mathcal{R} = [1.64, \infty).$$

En la Figura 1 se grafica en el eje de abscisas la zona de rechazo y se marca el nivel de significación como área.

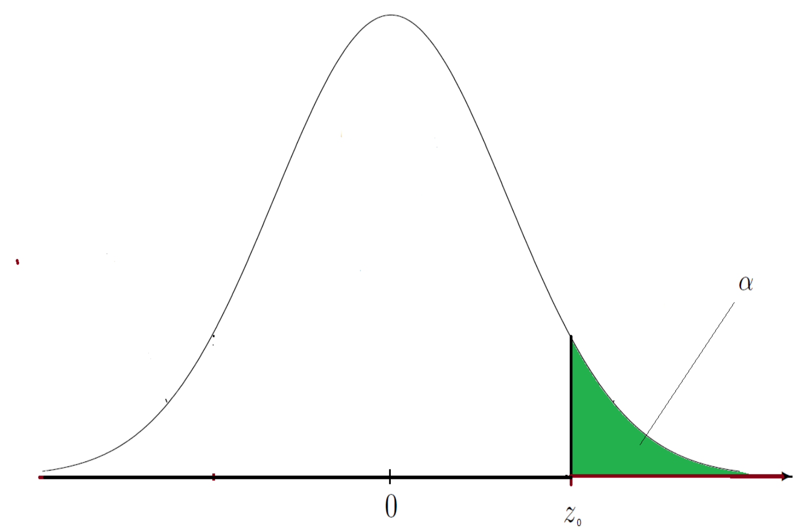

Figura 1: Representación de la zona de rechazo (semirrecta en color marrrón), y el nivel de significación (área en color verde).

Notar que el nivel de significación y, por ende, las zonas de rechazo y aceptación se determinan previamente a la recolección y observación de los datos.

Paso 3. Resultados e interpretación

En este paso incorporamos la evidencia proporcionada por las observaciones. Para los $n = 80$ datos con una media de 49.6 Mbps y para una varianza poblacional de 72 Mbps^2, se observa un valor de T

$$t_{\text{obs}} = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{49.6 - 48}{\sqrt{72}/\sqrt{80}} = 1.69.$$

Dado que (ver Figura 2)

$$t_{\text{obs}} = 1.69 \in \mathcal{R} = [1.64, \infty),$$

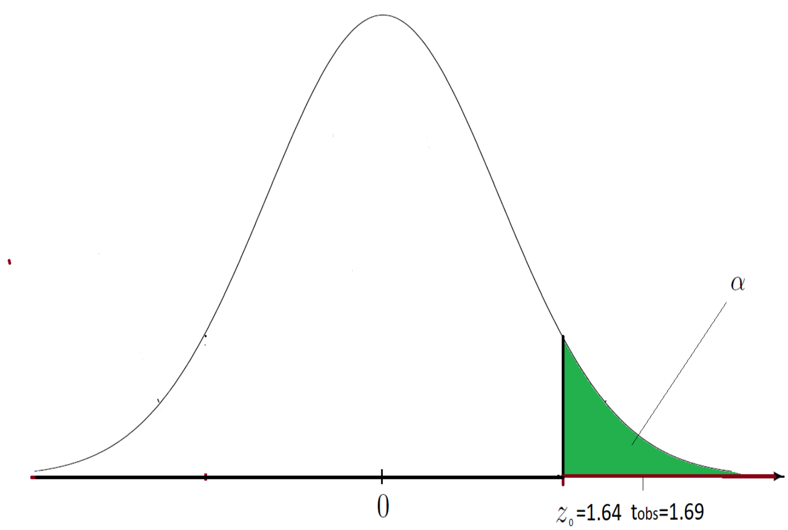

Figura 2: Representación de la zona de rechazo (semirrecta en color marrón), nivel de significación (área en color verde) y t_{obs} (punto marcado en rojo).

rechazamos H_0 y concluimos que "con una probabilidad de error de 0.05, la velocidad media del servicio prestado por el proveedor supera los 48 Mbps." □

Nos interesa, a partir del ejemplo, introducir otro concepto. Supongamos que para otra muestra del mismo tamaño $n = 80$ el valor de la media muestral hubiese sido 51.4 Mbps. De este modo

$$t^*_{\text{obs}} = \frac{51.4 - 48}{\sqrt{72}/\sqrt{80}} = 3.58 \in \mathcal{R} = [1.64, \infty),$$

y también rechazamos H_0. No obstante nos percatamos que esta muestra exhibe una evidencia a favor de H_A más fuerte que la muestra anterior. Una forma de cuantificar esta diferencia es a través del valor p del test dado en la siguiente definición.

Definición 5. *Consideremos un test de hipótesis en el que se contrastan H_0 y H_A y sea t_{obs} el valor del estadístico T basado en una realización muestral $(x_1, \ldots, x_n)$. Asumiendo que H_0 vale, el valor p del test o p-valor se define como la probabilidad de que el estadístico T asuma valores más* *extremos (que lleven a rechazar H_0) que t_{obs}.*

El $p-$valor permite modificar la secuencia de los pasos anteriores y quedarnos, luego del paso 1 sólo con el siguiente paso:

Paso 2*. Resultados basados en el $p-$valor e interpretación. Para un nivel de significación α, si el valor p del test es menor o igual que α entonces rechazamos la H_0 y concluimos la validez de H_A. Si en cambio, el valor p del test es mayor que α entonces no rechazamos H_0. $\square$

Para el ejemplo anterior, en el caso en que la muestra hubiese arrojado una media de 49.6 y $t_{\text{obs}} = 1.69$ entonces (ver Figura 3).

$$\text{p-valor} = P(T \geq 1.69 \mid T \sim \text{N}(0,1)) = 0.046,$$

mientras que para la segunda muestra, dado que la media muestral es 51.4 y $t^*_{\text{obs}} = 3.58$, se tiene que

$$\text{p-valor} = P(T \geq 3.58 \mid T \sim \text{N}(0,1)) = 0.0002$$

(ver Figura 4)

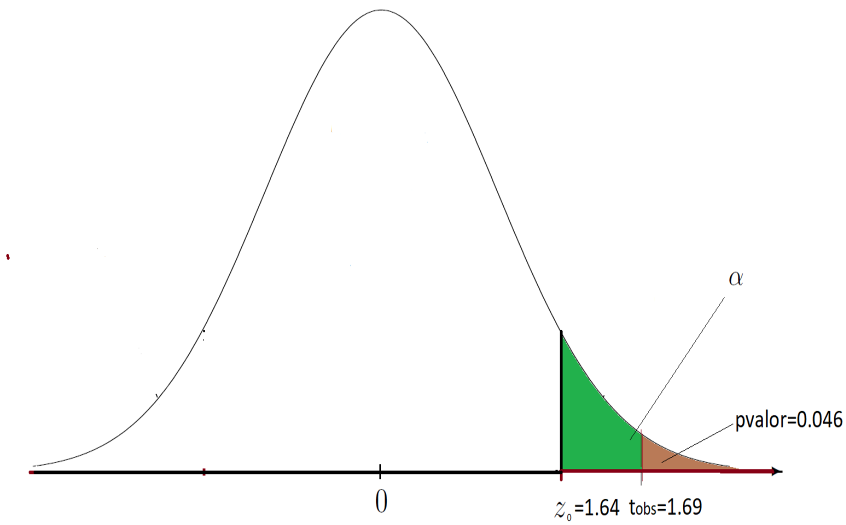

Figura 3: Representación del valor p (área en color marrón) si $t_{obs} = 1.69$ y su relación con el nivel de significación $\alpha = 0.05$ (área en color verde).

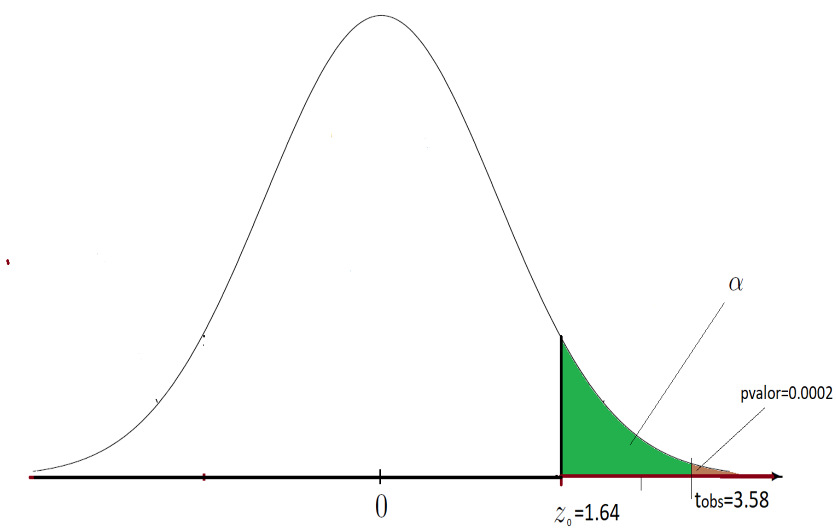

Figura 4: Representación del valor p (área en color marrón) si $t^*_{obs} = 3.58$ y su relación con el nivel de significación $\alpha = 0.05$ (área en color verde).

Así, en ambos casos, para un nivel de significación $\alpha = 0.05$ se concluye que $\mu > 48$.

Pero, si $\alpha = 0.01$ entonces para la primera muestra no se rechaza H_0 por ser el p-valor $= 0.045 > 0.01$ mientras que para basados en la segunda muestra sí rechazamos H_0.

Observación 1.

En una prueba de hipótesis la evidencia proporcionada por los datos permite, en caso de rechazar H_0, afirmar la hipótesis alternativa con una probabilidad de error determinada por el nivel de significación α. Pero, **en caso de no rechazar H_0 sería incorrecto concluir la validez de H_0**. La conclusión correcta es "los datos no exhiben evidencia estadísticamente significativa para afirmar la hipótesis alternativa". La razón de esta diferencia se debe a que la probabilidad de error de tipo II, β, definida arriba depende del valor particular que elijamos en H_A. Una comprensión exhaustiva de la distinción efectuada a través de las frases "hallar evidencias significativas para afirmar H_A" o "no hallar evidencias significativas para afirmar H_A" requiere del concepto de potencia de un prueba que nosotros no abordaremos aquí. □

2. Test de hipótesis para parámetros de distribuciones

En esta sección construimos pruebas de hipótesis para la media de una población con distribución normal, para la diferencia de medias de muestras independientes de poblaciones normales con varianza conocida y para una proporción.

2.1. Test de hipótesis para la media de una población con distribución normal

Analizaremos dos situaciones, cuando conocemos la varianza poblacional (caso 1) y cuando no (caso 2).

Caso 1. Varianza conocida

Sea $X \sim \mathrm{N}(\mu, \sigma^2)$ con varianza σ^2 conocida y asumamos que queremos contrastar las hipótesis $H_0 : \mu = \mu_0$ versus alguna de las alternativas unilaterales $H_A : \mu > \mu_0$ ó $H_A : \mu < \mu_0$ o bien $H_A : \mu > \mu_0$ ó bien la alternativa bilateral $H_A : \mu \neq \mu_0$. Dado los datos $(x_1, \ldots, x_n)$ consideremos t_{obs} el valor del estadístico computado a partir de los datos utilizando (1). A continuación, para cada H_A damos la expresión del valor p. Para contrastar

- $H_0 : \mu = \mu_0$ versus $H_A : \mu > \mu_0$,

$$\text{p-valor} = P(T \geq t_{\text{obs}} \mid T \sim \mathrm{N}(0,1)),$$

- $H_0 : \mu = \mu_0$ versus $H_A : \mu < \mu_0$,

$$\text{p-valor} = P(T \leq t_{\text{obs}} \mid T \sim \mathrm{N}(0,1))$$

  y,

- $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$,

$$\text{p-valor} = 2P(T \geq |t_{\text{obs}}| \mid T \sim \mathrm{N}(0,1)).$$

Caso 2. Varianza desconocida

Cuando la varianza poblacional no se conoce entonces la estimamos con la varianza muestral s_n^2. El estadístico del test se define de modo análogo a (1) como

$$T = \frac{\overline{X}_n - \mu_0}{s_n/\sqrt{n}}, \tag{2}$$

y

$$T \sim t_{n-1} \text{ asumiendo } H_0 \text{ como válida}$$

donde t_{n-1} es la distribución t de Student con $n-1$ grados de libertad.

Ejemplo 5.

En el contexto del Ejemplo 1 supongamos que modificamos la hipótesis alternativa y que las hipótesis a contrastar son ahora

$$H_0 : \mu = 48 \text{ versus } H_A : \mu \neq 48$$

pero ahora desconocemos la varianza poblacional σ^2 y, asumamos que además de contar con la información de que la media muestral es 49.6 conocemos que la varianza muestral es 70.3. ¿A qué conclusión arribamos ?.

Dado que $t_{\text{obs}} = \dfrac{49.6 - 48}{\sqrt{70.3/80}} = 1.71$, entonces

$$\text{p-valor} = 2 \cdot P(T \geq |t_{\text{obs}}|\, |T \sim t_{79}) = 2 \cdot P(T \geq 1.71 | T \sim t_{79}).$$

En R obtenemos

```
> 2*(1-pt(1.71,df=79))
[1] 0.09119093
```

con lo cual para un nivel $\alpha = 0.05$ concluimos que “no hay evidencias estadísticamente significativas para afirmar que la media poblacional sea diferente de 48” $\square$

Notar que el nuevo estadístico observado $t_{\text{obs}} = 1.71$, basado en la misma muestra de antes, está cercano al estadístico observado anterior (cuyo valor es 1.69), pero como la prueba es bilateral se vuelve más exigente para rechazar la hipótesis nula.

2.2. Test de hipótesis para la diferencia de medias de dos poblaciones

En esta subsección abordaremos el problema de comparar las medias de dos poblaciones, basados en muestras independientes. La resolución a este problema depende de la información con la que contemos respecto de las varianzas poblacionales. En esta asignatura sólo estudiamos el caso en que las varianzas sean conocidas. Comencemos con la siguiente situación problemática.

Ejemplo 6.

La grabación de CD's puede ser demandante en términos de consumo energético y en consecuencia podría reducir la vida media de la batería de una laptop. Para investigar este problema se les solicitó a 30 usuarios utilizar sus equipos hasta que la señal de "batería baja" aparezca. Dieciocho usuarios sin grabadora de CD informaron un tiempo promedio de 5.3 horas mientras que el grupo que posee grabadora informó una media de 4.8 horas. Las muestras correspondientes a ambos grupos son independientes. Asumiendo distribuciones normales con una varianza poblacional de 1.8 hs^2 para el grupo sin grabadora y de 2.4 hs^2 para el otro grupo, ¿podemos afirmar que los tiempos medios poblacionales son diferentes? $\square$.

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria de $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $(Y_1, \ldots, Y_m)$ una muestra aleatoria de $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ con ambas muestras independientes. Basados en los datos queremos contrastar la hipótesis

$$H_0 : \mu_X - \mu_Y = 0$$

versus una de las siguientes hipótesis alternativas

$$H_A : \mu_X - \mu_Y > 0 \text{ ó } H_A : \mu_X - \mu_Y < \mu_0 \text{ ó } H_A : \mu_X - \mu_Y \neq 0.$$

Como

$$\overline{X}_n - \overline{Y}_m \sim \mathrm{N}\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

entonces

$$T = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathrm{N}(0,1) \text{ si vale } H_0 \tag{3}$$

y, en consecuencia, T es un estadístico para la prueba. Dependiendo de la H_A, como antes, el test puede ser bilateral o unilateral.

Ejemplo 7. *Ejemplo 6 revisitado.*

De acuerdo al enunciado del problema tiene sentido contrastar

$$H_0 : \mu_X - \mu_Y = 0 \text{ versus } H_A : \mu_X - \mu_Y > 0$$

donde X representa el tiempo para una laptop sin grabadora e Y para una laptop con grabadora. Por los datos obtenidos, para los $n = 18$ usuarios sin grabadora de CD el tiempo promedio es de 5.3 horas y corresponden a una población (X) con varianza $\sigma_X^2 = 1.8$ hs^2 mientras que para los $m = 12$ usuarios que posee grabadora la media muestral es de 4.8 horas, que provienen de una población con varianza $\sigma_X^2 = 2.4$ hs^2. El valor del estadístico es $\dfrac{5.3 - 4.8}{\sqrt{\frac{1.8}{18} + \frac{2.4}{12}}} = 0.91$ y por lo tanto el valor p de la prueba es

$$P(T > 0.91 | T \sim \mathrm{N}(0, 1)) = 0.18.$$

En consecuencia concluimos que, "no hay evidencias estadísticamente significativas para asumir que la vida media de la población de baterías de laptops con grabadora sea menor que aquella correspondiente a la población sin grabadora (valor p=0.18)".

Observación 2.

Notar que no hemos hecho alusión a α ya que el valor p es muy grande. Los valores de α usuales son 0.01, 0.05 y, a lo sumo de, 0.10. $\square$

Observación 3.

Tener en cuenta que si la diferencia entre las medias muestrales fuese $5.3 - 4.8 = 0.5$ pero con tamaños de muestra mayores, por ejemplo $n = 57$ y $m = 60$ (un total de 117 laptops) el valor p hubiese sido 0.03 y, por lo tanto hubiésemos afirmado H_A con un nivel de significación de $\alpha = 0.05$. Esto muestra que la diferencia entre las medias muestrales es pequeña o grande relativa a los tamaños de las muestras. Un análisis similar podría efectuarse en relación al impacto de las varianzas poblacionales. □

2.3. Test de hipótesis para de nivel aproximado (asintótico) para una proporción

Consideremos una muestra aleatoria, $(X_1, \ldots, X_n)$, de una variable X con distribución Bernoulli de parámetro $p \in (0,1)$ y asumamos que queremos contrastar

$$H_0 : p = p_0$$

versus una de las siguientes hipótesis alternativas

$$H_A : p > p_0 \text{ ó } H_A : p < p_0 \text{ ó } H_A : p \neq p_0$$

con p_0 un valor del intervalo $(0,1)$. Por el teorema central del límite, si vale H_0 ($\mathrm{E}(X) = p_0$), para n grande la distribución de $T = \dfrac{\widehat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ es aproximadamente normal estándar y en consecuencia T puede ser utilizado para contrastar las hipótesis nula y alternativa.

Ejemplo 8. *Ejemplo 3 revisitado*

Recordemos que el objetivo es verificar que la proporción, p, de componentes electrónicos defectuosos en un lote es mayor a 0.330 %. Es decir, queremos contrastar las hipótesis

$$H_0 : p = 0.3 \text{ versus } H_A : p > 0.3$$

En 50 componentes inspeccionados la proporción de componentes defectuosos es de 0.41. ¿Provee esta muestra evidencia estadísticamente significativa para rechazar H_0 con un nivel de significación de $\alpha = 0.05$?

Para responder a la pregunta tengamos en cuenta que el valor del estadístico es $t_{\text{obs}} = 1.697$ y en consecuencia

$$\text{valor p} = P(T \geq 1.697 | \text{ vale } H_0) \approx 1 - \Phi(1.697) = 0.04$$

y por lo tanto concluimos que la proporción poblacional de componentes defectuosos supera 0.3 para ese nivel de significación. ¿Qué hubiese concluido si $H_A : p \neq 0.3$?