Apunte sobre TEORIA DEL ERROR en Test de Hipótesis

(en particular para el parámetro μ con σ conocido)

El siguiente cuadro explica la Teoría del error en los test de hipótesis y las probabilidades involucradas

Ho	**V**	**F**
Rechazar	P(Error Tipo I) = α	(1-β)
No Rechazar	(1-α)	P(Error Tipo II) = β

Se llama α (o probabilidad de cometer error de tipo I) a la probabilidad de rechazar la hipótesis nula cuando es verdadera; y se llama β (o probabilidad de cometer error de tipo II) a la probabilidad de no rechazar la hipótesis nula cuando es falsa.
Por la definición clásica de probabilidades, aparecen entonces (1-α), como la probabilidad de no rechazar la hipótesis nula cuando es verdadera y (1-β), como la probabilidad de rechazar la hipótesis nula cuando es falsa.
Se define así a α como nivel de significación y a (1-β) como potencia del test.

El valor de α, es fijado arbitrariamente por el investigador a priori de la realización del test de hipótesis; mientras que el valor de β se obtiene a posteriori de realizado el test y únicamente ante el caso de No rechazar la hipótesis nula. El valor de β varía junto con los distintos y verdaderos valores del parámetro poblacional testeado.

Veamos lo antedicho a través del siguiente ejemplo:

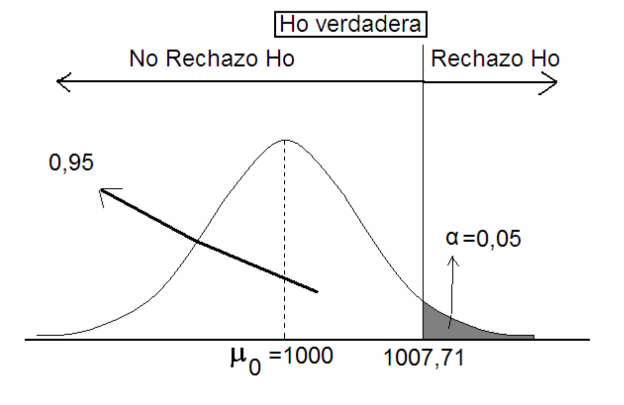
Una empresa envasa desodorante en aerosol. El volumen de aerosol colocado en cada cilindro se controla por muestreo, porque seria muy costoso (tanto en tiempo como en dinero) hacer el recuento de todos los cilindros.
Cuando el proceso productivo funciona correctamente, el volumen x de aerosol colocado en cada cilindro debe ser en promedio de a lo sumo 1000 cm3. Asimismo se conoce que dicha variable se distribuye normalmente en la población con un σ = 37,5 cm3
La empresa selecciona a diario y en forma aleatoria una muestra de 64 cilindros con un nivel de significación del 5%. En una de esas muestras obtuvo un promedio de 1010 cm3 por cilindro.

Es decir: Se desea testear Ho: $\mu_o \leq 1000$ (continuar con el llenado de las cilindros)
H1: $\mu_o > 1000$ (detener el proceso inmediatamente)

α = 0,05 significa la probabilidad de detener el proceso cuando realmente haya 1000 o menos cm3 de aerosol en promedio en los cilindros y constituye la probabilidad de cometer error de tipo I.
Busquemos el valor crítico de la media muestral:

$$\overline{x}_{crit} = 1000 + 1{,}645 * \frac{37{,}5}{\sqrt{64}} = 1007{,}71 cm3$$

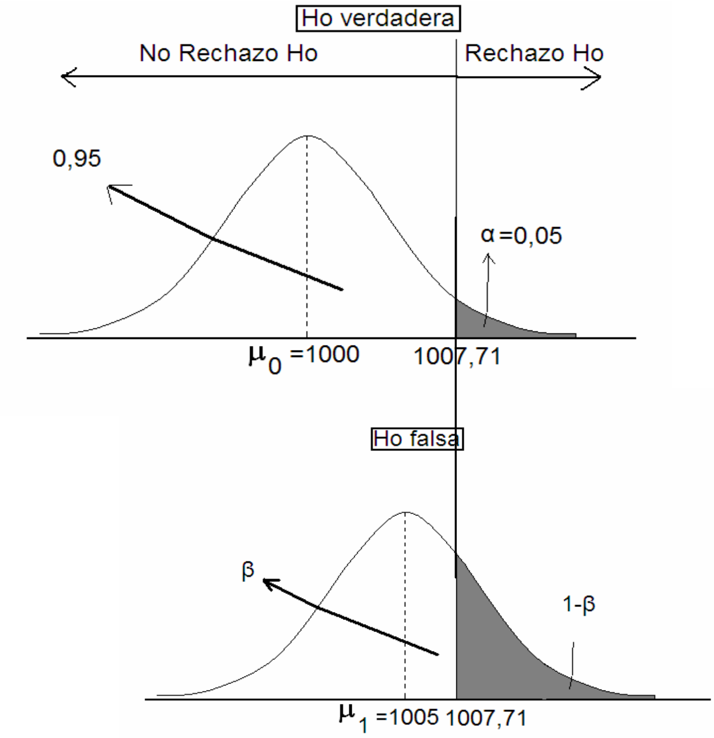
Como la condición de rechazo es (C.R.: si $\overline{x}$ > 1007,71 entonces rechazo Ho) y en la muestra hemos obtenido
$\overline{x}$ = 1010 entonces rechazamos Ho y detenemos el proceso inmediatamente.



La muestra ha arrojado evidencia suficiente en contra de que el proceso está funcionando bien.
Tras esta conclusión sería ilógico suponer que apareciera un error de tipo II, pues hemos rechazado la Ho.

El nuevo ingeniero de planta nos indica que en un proceso idéntico el verificó que en promedio los cilindros se llenan con 1005 cm3 y que no tiene razones para creer lo contrario, por lo tanto y dada su amplia experiencia lo tomará como valor poblacional ¿Cuál será la probabilidad de detener el proceso si esta afirmación es verdadera?

Esto significa que si el gráfico anterior estaba referido a un valor poblacional μ_o = 1000 que suponíamos verdadero, ahora graficaremos una nueva curva pero centrada en μ_1 = 1005 que en caso de ser cierta tornaría falsa nuestra Ho
Veamos



Allí podemos observar claramente que la zona grisada de la curva centrada en 1005, se corresponde con el concepto de potencia del test (1-β) pues en esta curva, la Ho es falsa y en dicha zona grisada se debe rechazar la misma, por lo tanto representa la probabilidad de rechazar Ho cuando en realidad es falsa.
Calculemos entonces el valor de (1-β).
Simplemente se trata de calcular un área bajo la curva normal.
Entonces haremos:

$$1-F\left(\frac{1007{,}71-1005}{37{,}5\Big/\sqrt{64}}\right)=1-F(0{,}58)=1-0{,}719=0{,}281$$

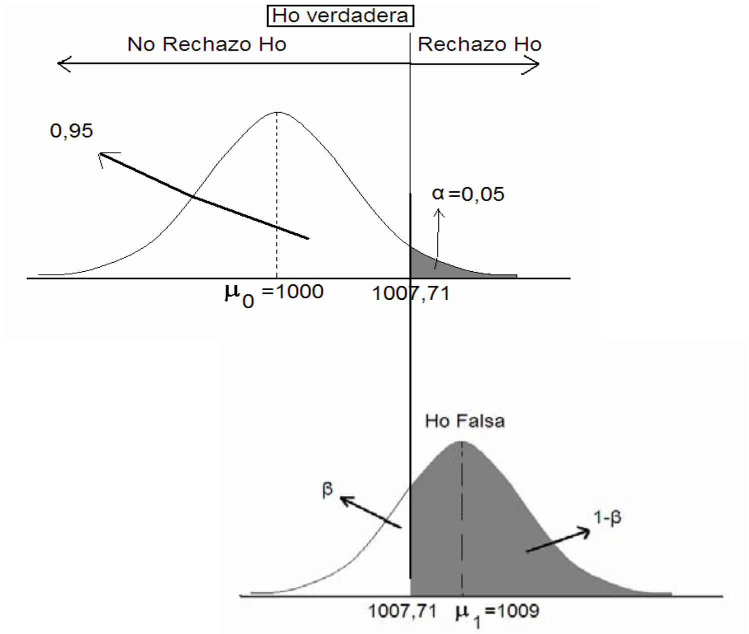
Potencia que es muy baja. Razonemos que significan en este caso el valor de β y de (1-β).

Ho	**V** **EL PROCESO FUNCIONA BIEN**	**F** **EL PROCESO FUNCIONA MAL**
Rechazar (Detener el proceso)	P(Error Tipo I) = α = 0,05	(1-β) = 0,281
No Rechazar (No detener el	(1-α) = 0,95	P(Error Tipo II) = = β = 0,719

proceso)		

Supongamos ahora que $\mu_1 = 1009$

Veamos los gráficos de las curvas



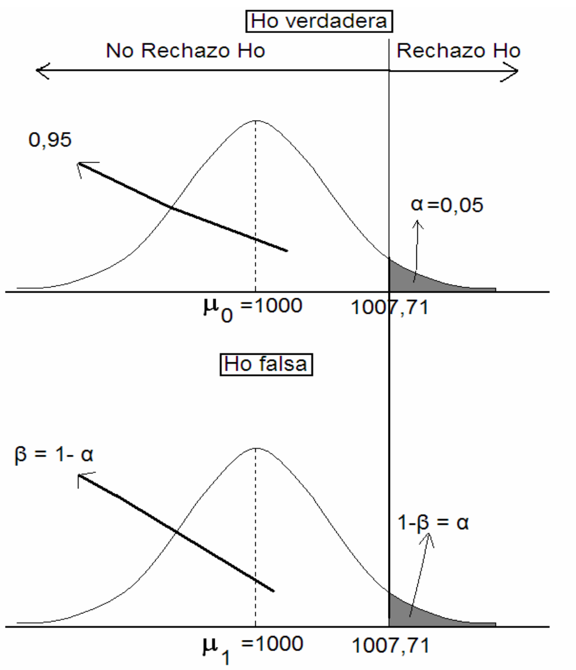
A simple vista vemos que la potencia del test es otra. Calculemos entonces cuánto vale.

$$1-F\left(\frac{1007{,}71-1009}{37{,}5\big/\sqrt{64}}\right)=1-F(-0{,}28)=1-0{,}3897=0{,}6103$$

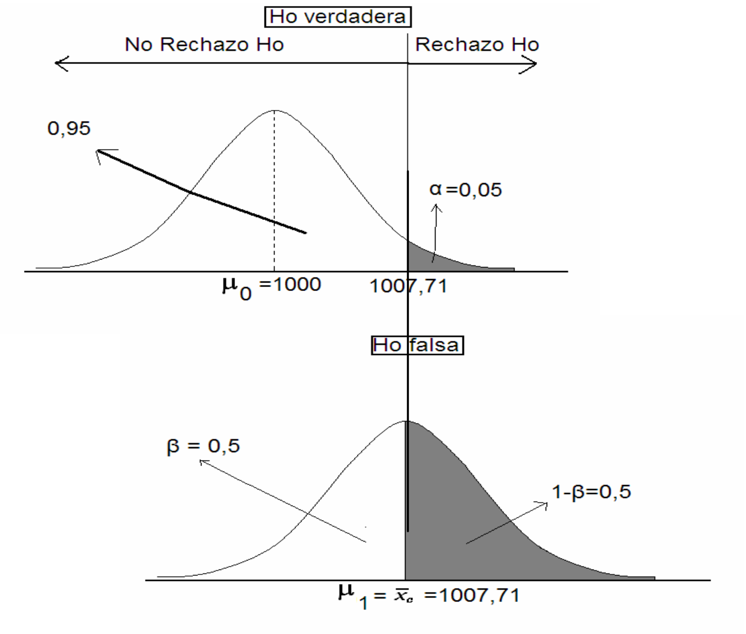
El test es ahora mas potente. Al alejarse la realidad $\mu_1 = 1009$ de nuestro planteo (que suponemos verdadero) μ_o=1000 aumenta la probabilidad de detener el proceso cuando está funcionando mal.

Vemos ahora un par de casos particulares.

Caso particular en que $\mu_1 = \mu_o$



Caso particular en que $\mu_1 = \bar{x}_{crit}$



De esta manera se puede construir una tabla para los valores de β y de (1-β) para los distintos valores de μ_1

μ_1	β	**(1-β)**
1000	**(1-α)**	α
1005	0,7190	0,2810
1007,71	**0,5**	**0,5**
1009	0,3897	0,6103
1010	0.3126	0.6874
1020	0.0043	0.9957

Con esto se podrían graficar las curvas de β y de (1-β) en función de los distintos valores de μ_1
A las curvas de β se las denomina habitualmente curvas características de operación (C.O.) y se las suele utilizar en gráficos de control estadístico. A las curvas de (1-β) se las llama curvas de potencia. Puede encontrar ambos gráficos en la bibliografía obligatoria.

Ahora la idea fundamental en estos tests es mantener en un nivel bajo tanto el error de tipo I como el error de tipo II, por lo tanto vamos a ver dos formas de disminuir el error de tipo II que pueden resultarnos útiles.

Relación entre α y β:

Si en el test original aumentamos α=0,10, la región de rechazo se agranda, pues el crítico se desplaza hacia la izquierda tomando el valor 1006 cm3.
Como μ_1 = 1005 recalculamos…
Y consecuentemente β va a disminuir a 0,5822 (antes era de 0,7190) y por lo tanto la potencia del test aumentará a 0,4178 (antes era de 0,2810)

Decimos entonces que α y β están inversamente relacionados. Cuando uno de ellos aumenta el otro disminuye.

Aumentando el valor de α (recuerde que este valor es fijado de manera arbitraria por el investigador) se logra entonces un incremento en la potencia del test. Por ello esta forma no es recomendable para controlar el error de tipo II, pues aumenta el error de tipo I.

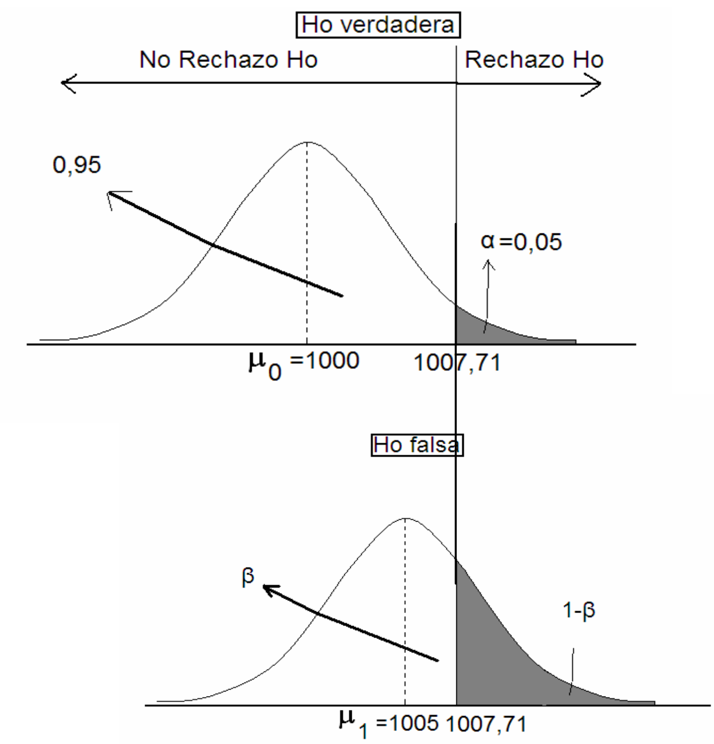
Ahora veremos la segunda y última manera de reducir β y consiguientemente conseguir un aumento de la potencia del test.

Esto es aumentando el tamaño de la muestra (n)
(Nótese que el inconveniente principal está en el tiempo y recursos necesarios que deberán invertirse en tal tarea)

Como el cálculo de β depende del desvío del estimador y este se calcula $\sigma(\bar{x}) = \sigma / \sqrt{n}$, resulta claro que al aumentar (n) disminuye el valor del desvío y esto resulta en un apuntamiento de la curva normal.
Esto provoca que (si se deja el valor de α constante) se reduzca el valor de β en la segunda curva.
Esto es realmente complicado de demostrar desde lo gráfico, pero espero que con la explicación analítica previa lo logre comprender.

Concluyendo entonces, la manera aconsejable para mantener bajo control los valores de α y de β es aumentando el tamaño de la muestra.

Por lo tanto deduciremos la fórmula de cálculo correspondiente.
Si observamos nuevamente los gráficos del planteo original vemos que



En la curva donde Ho es verdadera el $\bar{x}_{crit} = \mu_o + z_{1-\alpha} * \sigma / \sqrt{n}$ y

En la curva donde Ho es falsa el $\bar{x}_{crit} = \mu_1 + z_{\beta} * \sigma / \sqrt{n}$

Por lo tanto, por igualación debe ser

$$\mu_o + z_{1-\alpha} * \sigma / \sqrt{n} = \mu_1 + z_\beta * \sigma / \sqrt{n}$$

la cuál al despejar algebraicamente (n) se obtiene

$$n = \left[\frac{\left(z_{1-\alpha} + z_{1-\beta} \right) * \sigma}{\mu_0 - \mu_1} \right]^2$$

que sirve para los tests unilaterales.

Si el test fuera bilateral debemos cambiar $z_{1-\alpha}$ por $z_{1-\alpha/2}$