

12

Teoría de la decisión estadística

En este capítulo se presenta la Teoría de la Decisión Estadística como herramienta básica para la toma de decisiones, basadas en evidencia científica. La manera de hacerlo es plantear las hipótesis posibles y luego efectuarle una prueba o *test estadístico*. Llamada en algunas obras: la docimasia estadística. Cuando una conclusión se valida con un test estadístico se la llama de tipo *cuantitativo*, en caso contrario la decisión adoptada es de tipo *cualitativo*, o sea, una decisión tomada en forma subjetiva. El método consiste en definir una probabilidad de aceptación del orden del 95% (o rechazo) de una hipótesis de trabajo planteada, que permite calcular los valores críticos (o límites de aceptación) de un estadígrafo calculado a partir de los valores medidos. La importancia de este tema es muy grande. Basta decir que el objeto final de la Estadística es la toma de decisiones. En este capítulo se presenta el modelo de las grandes muestras usando la función de Gauss, no solo por razones didácticas sino por el hecho de que muchos modelos que se verán más adelante se pueden aproximar con este cuando las muestras sean lo suficientemente grandes. La manera de trabajar y razonar es la misma para todos los otros modelos estadísticos que se irán mostrando a lo largo de este trabajo. Esta herramienta permite la toma de decisiones en muchos casos prácticos, como ver si un método clínico está calibrado, testear precisión y exactitud, plantear métodos de Control de Calidad, comparar entre sí dos técnicas de laboratorio, o entre dos medicamentos, y otros más.

12.1 Hipótesis estadísticas

Las conjeturas que se realizan sobre los parámetros desconocidos de las poblaciones tienen asociadas una cierta probabilidad de ocurrencia. Conviene hacer planteos o supuestos que luego puedan ser verificados en el ámbito probabilístico. La idea es que estas hipótesis pueden ser aceptadas o rechazadas con un *test de validación*. Pero se debe remarcar que nunca se tendrá la certeza acerca de sus resultados. Cuando un test acepta o rechaza una hipótesis estadística, lo hace con el nivel de confianza deseado por el investigador, tan alto como guste, pero nunca del 100%, o sea, de la certeza.

El primer paso es formular hipótesis estadísticas. Esto es, plantear los supuestos realizados en términos matemáticos cubriendo todos los casos posibles. Por ejemplo, se desea testear si la vida útil de un remedio en condiciones normales de almacenamiento es por lo menos de un año. Se pueden plantear dos alternativas: a) la vida útil es de un año o más, lo cual escrito matemáticamente se expresa con: $T \geq 365$ días; y b) todos los demás casos posibles, es decir: la vida útil es menor a un año, descrito con: $T < 365$ días. Otro ejemplo: se quiere ver si no hay diferencias entre un nuevo método de medición clínico y el viejo que se venía usando hasta ahora.. Las

dos hipótesis serían (a) $\mu_{\text{viejo}} = \mu_{\text{nuevo}}$ y (b) $\mu_{\text{viejo}} \neq \mu_{\text{nuevo}}$. O sea, en la primera hipótesis, se suponen equivalentes, mientras que en la segunda se los toma como diferentes. Se denominan:

Hipótesis nula H_0 : es la hipótesis estadística planteada para ser testada;

Hipótesis Alternativa H_1 : es la hipótesis complementaria de la anterior.

Esto significa que la unión de ambas hipótesis cubre el universo de casos posibles. Como se vio en los ejemplos anteriores, siempre se plantean dos alternativas que son complementarias. Se puede elegir a cualquiera de ellas para ser la H_0 . Conviene usar a la que tiene el signo igual dentro de sus alternativas, pues facilita el cálculo de estadígrafo de prueba. De todos modos es lo mismo, pues muchas veces se plantea una hipótesis para ser rechazada antes que aceptada, y viceversa. Por ejemplo, si se quiere testear un dado para ver si está bien, se plantea: $H_0: p = 1/6$ y $H_1: p \neq 1/6$, donde la hipótesis es que la probabilidad de ocurrencia de cualquiera de sus caras es de $1/6$. Si es una moneda lo que se quiere ensayar, el planteo es similar con una probabilidad de $1/2$. En las comparaciones de dos técnicas clínicas lo que interesa es probar que ambas miden igual el valor de la magnitud en el paciente y por ello se emplea la media aritmética. Sería equivalente a pensar que ambas tienen la misma exactitud. Pero también se puede verificar si ambos métodos tienen similar dispersión usando la varianza como estadígrafo de prueba. En estos casos de comparaciones de dos estadígrafos conviene plantear como H_0 la que contiene la igualdad. El motivo para hacer esto es poder simplificar los valores poblacionales desconocidos, en el cálculo de la tipificada de las diferencias. Así, el valor: $\mu_{\text{viejo}} - \mu_{\text{nuevo}} = 0$ y se evita el tener que usar un suero control (con μ conocido) para hacer las comparaciones.

12.2 Validaciones estadísticas

Las pruebas que se realizan para plantear las hipótesis se conocen con el nombre de *ensayos de validación estadística*. El problema básico es determinar si las diferencias observadas entre el valor obtenido y el valor esperado se deben al azar, o si realmente son diferentes. Por ejemplo, para testear una moneda se arroja al aire 100 veces y se obtienen 55 caras. Entonces, el valor de frecuencia relativa obtenido es de 0,55, mientras que el valor teórico esperado es de 0,5. El problema es poder determinar, en *forma científica*, si esa diferencia observada de 0,05 es porque la moneda está algo cargada o simplemente por casualidad. Notar que si se toma esta decisión en forma subjetiva (*cualitativa*), diferentes personas podrían opinar de manera diferente. Esto es, para algunos la diferencia es tan chica que no vale la pena tenerla en cuenta, pero otros pueden opinar que es demasiado grande para olvidarla, y finalmente un tercer grupo puede creer que los datos son insuficientes para poder tomar una decisión. Este tipo de divergencia de opinión generó un sin número de discusiones estériles entre los científicos del pasado, donde unos apoyaban fervientemente una teoría, mientras los de la escuela contraria se encontraban en franca oposición. Se necesitó un árbitro creíble para ambas partes. Uno que tomase las decisiones en forma *objetiva (cuantitativa)*, sin ser influenciado por nada ni por nadie y creíble para todos. El método de la prueba estadística surgió entonces como el único aceptable, el método *científico* para decidir. Tanto es así que hoy ningún trabajo de investigación es aceptado sino se realizó una validación estadística de sus conclusiones. Trabajos no validados no se aceptan más en los congresos, reuniones, coloquios o seminarios de discusión. Al poder disponer de una herramienta tan poderosa para probar hipótesis, el trabajo de investigación se aceleró en forma notable. Además, ahora se

dispone de técnicas para el diseño de experimentos, como se podrá ver más adelante en el capítulo del ANOVA (Análisis de la varianza).

Como se vio en el Gráfico 4.1 hay dos maneras de acertar al tomar una decisión, y también hay dos maneras de equivocarse. En estadística se denominan:

Error de tipo I: ocurre cuando se rechaza una hipótesis, que debería haber sido aceptada.

Error de tipo II: ocurre cuando se acepta una hipótesis, que debería haber sido rechazada.

En el comercio, al error de tipo I se lo denomina *riesgo del vendedor*. Ocurre cuando es devuelta una mercadería por no pasar el control de calidad, y en realidad era buena. Por ejemplo, en un lote de 2000 artículos se revisan 5 de ellos para decidir si se lo acepta, y justo se eligen para hacer el control los únicos fallados; entonces se le devuelven al vendedor quien tiene que correr con todos los gastos que demanda una reposición de mercadería, si no es que pierde la venta injustamente. A veces, la mercadería viene de otro continente y los costos son muy altos sin contar con la pérdida de imagen.

Por otro parte está el *riesgo del comprador*, que puede comprar algo de mala calidad por haber revisado una parte pequeña del lote que no puso en evidencia tal cosa. Es el caso de un error estadístico de tipo II. El problema de una compraventa justa es minimizar el riesgo de ambas partes. Para eso, se emplean las Normas IRAM respectivas donde se especifican el riesgo de cada uno, el tamaño del lote de prueba y el número aceptable de piezas falladas. Los contratos de compraventa detallan claramente todas esas condiciones negociadas por ambas partes.

Lo ideal sería minimizar ambos tipos de errores a la vez, lo cual puede hacerse aumentando el número de pruebas. Sin embargo, en la práctica, para un tamaño muestral dado, cuando se achica uno, inevitablemente crece el otro. A veces, un tipo de error puede tener más importancia que el otro, como el caso visto de sensibilidad y especificidad de los índices clínicos donde según la enfermedad conviene minimizar uno u otro. Otras veces, no es posible aumentar el número de pruebas, ya sea por su costo, por falta de material o por el tiempo insumido. Pero la regla de oro de los estadísticos es:

Para minimizar los errores, hay que aumentar el número de pruebas.

12.3 Test de hipótesis

El primer paso para contrastar una hipótesis estadística cualquiera es fijar el *nivel de significación* (α) de la prueba, de acuerdo con el nivel de confianza deseado. En Física, Química y otras ciencias exactas, se suele adoptar un valor crítico $Z_{\alpha} = 1$, considerado suficiente. Sin embargo, en otras disciplinas como Biología, Medicina, etc., que trabajan con materiales menos estables, los niveles de confianza usuales son del 95 %, 99 % y del 99,9 %. En efecto, un trozo de metal tiene menos variabilidad durante el proceso de medición que un material biológico. Cuando se trabaja con seres vivos las magnitudes clínicas como glucosa, colesterol, bilirrubina y otras,

varían casi en forma continua dentro de ciertos límites. La ingesta de medicamentos, el ejercicio, la posición adoptada por el individuo y el tiempo de ayuno, modifican los valores de dichas magnitudes. Esto obliga a los investigadores a tomar límites de tolerancia más amplios.

Cuando se rechaza una hipótesis usando un nivel de significación $\alpha = 0,05$, para un nivel de confianza del 95 %, se dice en el informe que los resultados fueron *significativos*. Esto se denota de varias formas. Algunos autores ponen el estadígrafo hallado y el valor de la significación, otros el valor de la probabilidad correspondiente al estadígrafo medido; otra es usar un asterisco para ello. En Medicina, la recomendación actual es poner todo, en especial el intervalo de confianza (CI), junto con el valor del estadígrafo. Son ejemplos de estos modos:

- ... se obtuvieron resultados significativos ($p < 0,05$)
- ... esta hipótesis no es aceptable $Z_x = 2,07$ ($p = 0,0468$)
- ... se rechazó la hipótesis planteada ($Z = 2,07$) al 95% de confianza.
- ... se rechazó la hipótesis planteada ($Z = 2,07^*$)
- ... se rechaza pues $Z = 2,07$ cae fuera de 95% CI (-1,6 ; 1,6)

Son cinco maneras diferentes de decir lo mismo; sin embargo, la última presenta mayor información al lector. Al poner el estadígrafo Z se le está diciendo que se ha empleado la función de Gauss (o la teoría de las grandes muestras) para realizar la validación, cosa que no es clara en la primera al usar la probabilidad p . Además 95% CI dice que se trata de un intervalo al 95% y da los límites superior e inferior del mismo. En síntesis, la notación más recomendable es la última pues es la que mejor resume todos los datos hallados.

Cuando se rechaza la hipótesis al nivel del 99%, se dice que los resultados fueron *muy significativos* y se denota con dos asteriscos en lugar de usar uno solo o con 99% CI. Por ejemplo, $Z = 2,7^{**}$ o bien $p < 0,01$, también $p = 0,003$, o $Z = 2,7$ cae afuera de 99% CI (-2,58 ; 2,58).

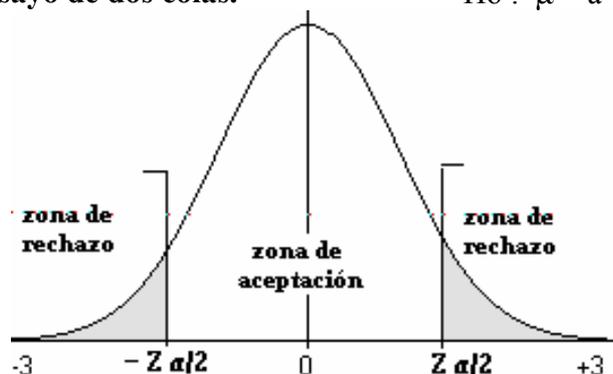
Si se rechaza la hipótesis al nivel del 99,9%, entonces se habla de resultados *altamente significativos* y esto se denota empleando tres asteriscos, o bien con la notación de intervalo 99,9% CI. Como ser, $Z = 3,8^{***}$, o puede ser con $Z = 3,8$ ($p < 0,001$), y también $p = 0,001$, o $Z = 3,8$ cae afuera de 99,9% CI (-3,29 ; 3,29).

En cambio, cuando la hipótesis nula no puede ser rechazada se dice que los resultados *no fueron significativos* y no se coloca ningún asterisco. Por ejemplo, $Z = 1,2$ o $p > 0,05$, o también $Z = 1,2$ cae adentro de 95% CI (-1,96 ; 1,96)

El siguiente paso, una vez definido el nivel de confianza, es determinar las *zonas de aceptación y de rechazo* de la hipótesis planteada. Primero se define el tipo de ensayo a usar. Hay dos tipos: ensayo de *una* o de *dos* colas, de acuerdo a si en la hipótesis nula hay una desigualdad o una igualdad respectivamente. Luego se calcula el valor crítico del ensayo que delimita las zonas buscadas. Cuando la hipótesis nula sea de igualdad, el ensayo será de *dos colas* porque interesa testear los casos de mayor y de menor a la vez. Eso implica calcular dos límites dentro de los cuales estará la zona de aceptación de la (H_0) y a cada lado dos zonas de rechazo - ver Figura 12.1. Cuando en la hipótesis nula figure alguno de los signos de desigualdad, el ensayo tendrá *una cola*, es decir, habrá un solo valor crítico que separa las zonas de aceptación y de rechazo - ver Figura 12.2.

Figura 12.1 : (a) Ensayo de dos colas.

$$H_0 : \mu = a \quad H_1 : \mu \neq a$$

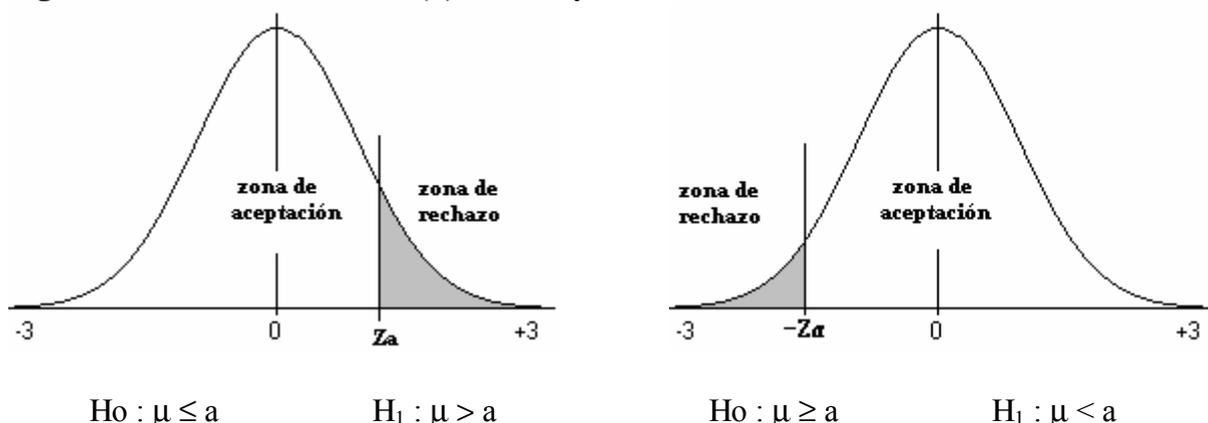


Suponiendo que la hipótesis nula dada $H_0 : \mu = a$, donde a es un valor cualquiera, referente al estadígrafo e que tiene una distribución muestral de tipo gaussiano, definida unívocamente con los parámetros $(\mu_e ; \sigma_e)$. Si se tipifica con $Z = (a - \mu_e) / \sigma_e$, la distribución de la variable tipificada Z será la función normal estandarizada, de parámetros $(\mu = 0 ; \sigma = 1)$ que está tabulada. Si este estadígrafo Z cae dentro de la zona de aceptación, entonces la hipótesis nula será aceptada. Caso contrario, si cae en la zona de rechazo, se considerará válida a la hipótesis alternativa; se rechazará la H_0 con la significación correspondiente. El área total sombreada es el nivel de significación α , por ello se debe repartir en cantidades iguales en ambos extremos de la curva. Eso define el valor crítico $z_{\alpha/2}$ a buscar en tablas. La zona de rechazo es simétrica respecto al origen, cada cola tiene bajo la curva la mitad del nivel de significación adoptado $\alpha/2$. La zona de aceptación de la hipótesis nula tiene bajo la curva una cantidad $1 - \alpha$. Por ejemplo, si se elige un nivel de confianza del 95%, el nivel de significación será $\alpha = 0,05$. Esto significa que se debe dejar un 2,5% en cada cola y el valor crítico será: $Z_{\alpha/2} = 1,96$ que corresponde a la probabilidad buscada.

Si en la hipótesis nula se define una situación de mayor o menor, entonces se trata de un ensayo de una cola. Los dos casos posibles se muestran en la Figura 12.1 (b) siguiente:

Figura 12.1:

(b) Ensayo de una cola.



El área total sombreada es el nivel de significación α y queda a la derecha del valor crítico Z_α , pues en la hipótesis nula se postula que el valor del estadígrafo testado es menor o igual que cierto valor límite dado por a . En cambio, cuando en la hipótesis nula se postula que el parámetro es mayor o igual al límite dado, el área sombreada quedará a la izquierda del valor crítico Z_α y

definirá la zona de rechazo. Por ejemplo, si el nivel de significación es del 95 %, el área total será del 5 %, y así queda $Z_{0,05} = 1,645$.

En síntesis, los pasos a seguir en un ensayo de hipótesis son:

- 1) se formulan las hipótesis de trabajo: nula y su alternativa;
- 2) el investigador define el nivel de confianza y calcula el valor crítico de tablas;
- 3) con este valor, quedan establecidas las zonas de rechazo y aceptación de hipótesis;
- 4) se calcula el valor tipificado del estadígrafo usando μ_e ; σ_e y el valor a de la H_0 ;
- 5) se compara este valor con el crítico para aceptar o rechazar la H_0 como se vio antes.
- 6) para informar los resultados se usa principalmente el intervalo de confianza.

Estos pasos son comunes a todos los modelos estadísticos. La diferencia entre ellos reside en la manera de calcular el valor de comparación y en la Tabla estadística que se empleará para el ensayo. Pero la estrategia es similar para todos ellos. Una manera sencilla y práctica de hacer esta comparación es dibujando las zonas de aceptación y rechazo a mano alzada. Luego, se coloca el valor calculado en el paso 4 en el gráfico, y así ver en cuál zona cae, para proceder en consecuencia. Hay otra manera que se está volviendo más popular en la bibliografía médica. Consiste en calcular el intervalo de confianza al nivel fijado por el investigador. El estadígrafo Z que se obtiene a través de los datos experimentales debería caer dentro de tal intervalo si se cumple la H_0 planteada. En cambio, si el mismo cae fuera, eso significa que se puede rechazar la H_0 al nivel de confianza del intervalo. En resumen:

Forma clásica de hacer el test: comparar Z versus $Z\alpha$ para tomar la decisión.

Forma recomendada en clínica: fijarse si μ_e cae dentro o fuera del 95% CI para decidir.

O bien, fijarse si e cae dentro o fuera del 95% CI poblacional

Se presentan tres casos posibles, a saber, que el estadígrafo Z caiga en una de las zonas de aceptación / rechazo, o bien que coincida justo con el límite de ambas zonas. Esto es:

- Si Z cae dentro de la zona de aceptación: los resultados no son significativos y no se rechaza H_0 . Entonces, $Z < Z_{\text{crítico}}$ O bien, μ_e cae dentro del 95% CI.

- Si Z cae dentro de la zona de rechazo: los resultados son significativos y se rechaza la hipótesis H_0 . Se deben verificar los otros niveles de confianza. En este caso será $Z > Z_{\text{crítico}}$ O μ_e cae fuera del 95% CI

- Si Z cae en el límite: los resultados son dudosos. Por lo tanto, se deberán hacer más pruebas para poder llegar a una decisión. Es cuando μ_e coincide con uno de los valores límites del 95% CI. Acá es $Z \equiv Z_{\text{crítico}}$ (no es necesario que sean exactamente iguales puede ocurrir que a juicio del investigador estén “sospechosamente cercanos”).

Conviene remarcar el hecho siguiente: que una hipótesis nula sea aceptada, no quiere decir que se tenga prueba científica de su validez. Puede haber otro modelo científico más sensible que el de Gauss que detecte diferencias cuando este no lo haga. Solo cuando una hipótesis no es aceptada se puede decir: se ha encontrado evidencia científica para rechazar la hipótesis. Es decir, que se valida el rechazo, pero no la aceptación. Por ello muchos investigadores prefieren plantear hipótesis para ser rechazadas. Aprovechando el hecho de que son ellos quienes deciden

cual es la nula y cual es la alternativa. La sensibilidad de un modelo para detectar diferencias en los ensayos es la *robustez* del mismo. *Cuando se pueda se debe elegir el modelo más robusto.* Esto se verá mejor en los capítulos siguientes cuando se estudie la forma de resolver un mismo problema con más un modelo estadístico.

12.4 Modelo de Gauss para una muestra

El modelo de Gauss sirve para validar hipótesis estadísticas cuando se toma una muestra a los efectos de estimar los parámetros de una población. También se lo puede emplear para comparar dos muestras entre sí, bajo el supuesto de que ambas fueron extraídas de la misma población (como se verá en el punto 12.5 siguiente). En este punto se comenzará trabajando con una sola muestra, de tamaño lo suficientemente grande como para que este modelo pueda ser aplicado. De acuerdo con el estadígrafo con el cual se va a trabajar, se puede encontrar en la Tabla 4 el tamaño mínimo para poder emplear a este modelo. Como las muestras son grandes, es más sencillo y barato usar muestras pequeñas (como se verá en el capítulo siguiente). Pero lo que se busca aquí es ilustrar la manera de razonar y proceder en estos casos.

Cabe destacar que en la última década se comienza a hacer una distinción entre resultados *estadísticamente* significativos y *clínicamente* significativos. Por ejemplo, una diferencia de porcentajes $Z = 13\%$ con 95% CI (7% ; 12%) es estadísticamente significativa, pero si solo importan las diferencias de 15% o más, entonces $Z = 13\%$ no es clínicamente significativa.

En teoría, se pueden ensayar todos los estadígrafos muestrales como los presentados en la Tabla 4. Con las correcciones correspondientes a los casos de muestreo sin reposición en poblaciones finitas. Sin embargo, los estadígrafos de uso más frecuentes son los siguientes:

12.4.1 Medias

El modelo de Gauss para las distribuciones de medias muestrales se puede aplicar para muestras de tamaño treinta o mayores ($n \geq 30$) donde $\mu_e = \mu_{\bar{X}} = \bar{X}$ y $\sigma_e = \sigma / \sqrt{n}$; por lo tanto el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{donde } \sigma \text{ es desconocido y se lo aproxima con DS.}$$

Ejemplo de calibración: Se realizaron 36 mediciones repetidas de una misma magnitud clínica: glucosa; usando una solución calibrada con $\mu = 90$ mg/dl, los resultados obtenidos arrojan una media de 95 mg/dl y un desvío estándar de 8 mg/dl. Ensayar la hipótesis de que el sistema de medición está calibrado (en exactitud) con un nivel de confianza adecuado.

Aquí conviene plantear la igualdad como una hipótesis nula.

$H_0 : \mu = 90$ mg/dl y el sistema está calibrado en exactitud.

$H_1 : \mu \neq 90$ mg/dl y el sistema no está calibrado, tiene un error de tipo sistemático.

Por lo tanto se trata de un ensayo de dos colas. El estadígrafo de comparación es:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{95 - 90}{8/\sqrt{36}} = 3,75 \text{ ***}$$

Los valores críticos de tablas son:

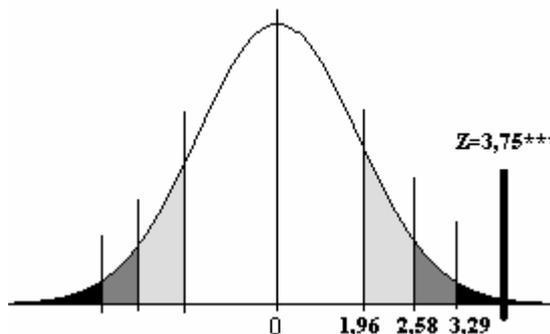
$$Z_{0,95} = 1,96 ; Z_{0,99} = 2,58 \text{ y } Z_{0,999} = 3,29$$

Como $Z \gg Z\alpha$ se rechaza la H_0 . O bien:

$Z = 3,75$ cae fuera de 99,9% CI (-3,29 ; +3,29)

$\mu = 90$ cae fuera de 99,9 % CI (90,6; 99,4)

$\bar{X} = 95$ cae fuera de 99,9% CI (85,6 ; 94,4)



La exactitud del método no es aceptable

Las zonas sombreadas en la figura anterior indican las tres zonas de rechazo de la hipótesis nula. A medida que se oscurece el rechazo es mayor. La primera zona delimitada por $Z_{0,95} = 1,96$ es cuando hay resultados significativos (95%). Como el estadígrafo Z es mayor y cae dentro de esta zona, hasta ahí se dice que se obtuvieron resultados *significativos* como para rechazar H_0 , y se lo denota con $Z = 3,75^*$. En estos casos, se debe continuar comparando con el siguiente nivel del 99%; como $Z_{0,99} = 2,58$ es todavía menor que Z , ahora se puede decir que se obtuvieron resultados *muy significativos* que se denotan con $Z = 3,75^{**}$, para rechazar H_0 . Entonces, se debe continuar con el último nivel acostumbrado del 99,9%. Nuevamente, $Z_{0,999} = 3,29 < Z$; por lo tanto la conclusión final es que se tienen resultados *altamente significativos* $Z = 3,75^{***}$ que prueban la existencia de un error sistemático que ha descalibrado al sistema de medición, en cuanto a su exactitud. Esto significa una *validación estadística* de la conclusión adoptada. La decisión consecuente será calibrar el sistema. La estadística es una herramienta para probar la existencia de tal error, pero no explica la forma de corregirlo. Esto lo debe hacer el investigador con sus conocimientos profesionales y su experiencia técnica. Todavía se puede estimar el módulo de este error sistemático (ES) con: $ES = \text{valor promedio} - \text{valor control} = 5 \text{ mg/dl}$.

También se puede efectuar la validación con los CI. Se pueden plantear tres tipos de CI: El CI de la variable tipificada Z es 99,9% CI (-3,29 ; 3,29) como $Z = 3,75$ cae fuera se rechaza H_0 . Otro tipo de CI es el muestral visto en el capítulo anterior: 99,9 % CI (90,6; 99,4) y como el $\mu = 90$ cae fuera del mismo se rechaza H_0 . El último tipo de CI es el que se forma con los valores poblacionales armando el intervalo con $(\mu \pm Z_{\alpha} \cdot \sigma_e)$ y viendo si $\mu_e = \bar{X}$ cae dentro del mismo. En este caso será: $(90 \pm 4,4)$ para un 99,9% de confianza, como $\bar{X} = 95$ cae fuera se rechaza H_0 . Estos tres tipos de CI son totalmente equivalentes y se pueden usar según convenga, por ejemplo para las diferencias de medias o proporciones conviene el último pues el valor esperado es $\mu = 0$.

Se calculan las probabilidades para el valor $Z = 3,75$ y $Z = -3,75$ con la ecuación gaussiana usando el programa Excel, y se obtiene una probabilidad de rechazo $p = 0,00018$ (esto es la probabilidad dentro de la zona de aceptación es de 0,99982). Como $p < \alpha$ se puede rechazar la hipótesis nula (cuando $Z > 3,5$ ya se puede rechazar H_0 sin ir a tablas). Algunos investigadores prefieren informar de esta manera, diciendo por ejemplo: *se encontró un error sistemático* ($p = 0,00018$). Comparando ambas maneras de informar, se concluye que es más corto y completo informar así: *se encontró un error sistemático de 5 mg/dl; $\mu_e = 95$ con 99,9% CI (90,6 ; 99,4)*. Al ver la Z el lector deduce que se ha empleado el modelo de Gauss y como el valor cae fuera del intervalo de confianza del 99,9%, deduce que se tiene validación estadística de lo afirmado.

También se puede abreviar con: *mediante Gauss se obtuvieron resultados altamente significativos: el sistema está descalibrado en exactitud.*

Para poder calibrar el sistema desde el punto de vista de la precisión habrá que estudiar como se comporta el desvío estándar si se lo compara con uno “patrón”, esto es contra un desvío máximo admisible (σ_{\max}) que se obtiene de los textos de análisis clínicos.

12.4.2 Varianzas

El modelo de Gauss para las distribuciones de varianzas muestrales se puede aplicar para muestras de tamaño cien o mayores ($N \geq 100$), donde $\mu_e = \mu_{\sigma^2} = DS^2$ y $SE(\sigma^2) = \sigma_{\sigma^2} = \sigma^2 \sqrt{2/n}$. Por lo tanto, el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{DS^2 - \sigma^2}{\sigma^2 \sqrt{2/n}} \quad \text{donde } \sigma \text{ es desconocido y se lo postula con } H_0.$$

Ejemplo de calibración: Continuando con el caso anterior, el investigador completa las 100 muestras obteniendo una nueva media de 94,6 mg/dl con un desvío de 9,65 mg/dl. De los libros de texto surge que la dispersión aceptada para la determinación de Glucosa viene dada por un coeficiente de variación del 10%. Se desea analizar si la dispersión encontrada en el experimento es menor, o por lo menos igual que la aceptable.

El CV% = 10% significa: $0,1 = \sigma_{\max} / \mu = \sigma_{\max} / 90$. De allí surge $\sigma_{\max} = 9$ mg/dl. Se plantean:

H_0 : $\sigma \leq 9$ mg/dl y el sistema tiene una dispersión aceptable.

H_1 : $\sigma > 9$ mg/dl y el error casual es muy grande. El sistema no está calibrado en precisión.

Por lo tanto, se trata de un ensayo de una sola cola y el estadígrafo de comparación se saca con:

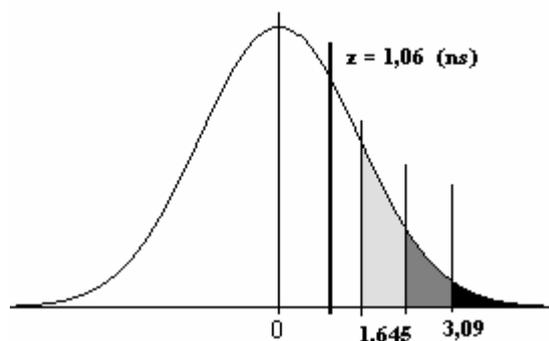
$$Z = \frac{DS^2 - \sigma^2}{\sigma^2 \sqrt{2/n}} = \frac{(9,65)^2 - (9)^2}{(9)^2 \sqrt{2/100}} = 1,06$$

Los valores críticos de tablas son:

$$Z_{0,95} = 1,645 ; Z_{0,99} = 2,33 \text{ y } Z_{0,999} = 3,09$$

Como $Z < Z_{\alpha}$ no se rechaza la H_0 .

$$\sigma^2 \leq 81 \in 95\% \text{ CI } (-\infty ; +87,2)$$



Por lo tanto, los resultados no fueron significativos. Esto es, no hay evidencia suficiente como para pensar que la dispersión no sea aceptable. El error casual no es significativo. Dicho en otros términos, la precisión del sistema de medición es menor que la máxima admisible recomendada en los libros de texto. La forma de expresarlo puede ser:

- no se rechaza la hipótesis nula: $Z = 1,06$ (ns) y también con 95% CI $(-\infty ; +87,2)$;
- no hay evidencia que pruebe una dispersión no aceptable.

12.4.3 Proporciones

El modelo de Gauss para las distribuciones de la proporciones se puede aplicar para muestras de tamaño $n > 25$ o mayores donde $\mu_e = \mu_p = \pi$ y $\sigma_e = SE(\pi) = \sqrt{\pi \cdot (1 - \pi) / n}$. Por lo tanto el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{p - \pi}{\sqrt{\pi \cdot (1 - \pi) / n}}$$

donde p es el valor experimental de la proporción.

O bien,

$$Z = \frac{r - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$$

si se trabaja con número de éxitos en vez de proporciones.

Notar que la distribución de las proporciones, o el número de éxitos r, es en realidad una distribución binomial, por lo tanto para que la aproximación sea buena, hay que efectuarle la corrección por continuidad de Yates, o lo que sería mejor la aproximación de Molenaar vista en el punto 9.8.1 anterior.

Ejemplo 1: En un experimento sobre PES (percepción extrasensorial), el sujeto investigado debía adivinar el símbolo de las cartas Zenner que el investigador iba sacando en otro cuarto separado. Cada vez que sacaba una carta tocaba un timbre y el sujeto anotaba el símbolo que creía ver. Se realizó esta prueba 400 veces y acertó 160 veces. Decidir si el sujeto tiene PES.

$H_0 : \mu = \pi = 0,2$ El sujeto está adivinando, $p = 1/5$ (hay 5 símbolos diferentes en las cartas).

$H_1 : \mu > 0,2$ El sujeto tiene capacidad extrasensorial pues acierta más de lo esperado.

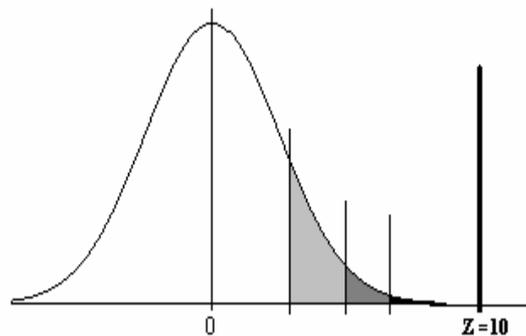
Por lo tanto, se trata de un ensayo de una cola. El porcentaje de aciertos medido en este experimento fue de $p = 160 / 400 = 0,4$. Además $\mu_p = 0,2 = \pi$ y $\sigma_{\pi} = 0,02$. Reemplazando en la primera ecuación

$$Z = (0,4 - 0,2) / 0,02 = 10***$$

Como $Z_{0,95} = 1,645$; $Z_{0,99} = 2,33$ y $Z_{0,999} = 3,09$.

El resultado obtenido es altamente significativo.

Se tiene una fuerte evidencia que este sujeto tiene poder extrasensorial. $p = 0,4$ no cae en 99,9% CI $(-\infty ; +0,26)$



Notar que en realidad se debería hacer la corrección por continuidad, porque se está usando una distribución continua para validar una variable discreta. Pero cuando la evidencia es tan fuerte como en este caso no es necesaria. Para ver esto se puede calcular $Z_{Yates} = 9,94***$ y con la de Molenaar es $Z_M = 8,83***$. Por su parte la probabilidad exacta con la binomial es casi 1.

Ejemplo 2: El fabricante de un medicamento asegura que el mismo tiene un 85% de efectividad para aliviar las neuralgias en medio día. Se tomó una muestra de 500 individuos con neuralgia, a

los cuales se les suministró el medicamento y 400 de ellos se aliviaron en el plazo previsto. Decidir si lo que asegura el fabricante es cierto.

Se trata de un ensayo de una sola cola se pueden postular dos tipos de H_0 : una usando las proporciones de éxitos como se vio en el ejemplo 1, y otra empleando directamente las frecuencias.

$H_0 : \mu \geq 500 (0,85) = 425$ pacientes aliviados y lo que dice el fabricante es cierto.

$H_1 : \mu < 425$ y no es cierta la afirmación realizada.

$$\mu_r = n \cdot \pi = 425 \text{ y } \sigma_r = [(n \cdot \pi (1-\pi))]^{1/2} = 7,98 \approx 8$$

Reemplazando y efectuando las correcciones por continuidad resulta:

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|400 - 425| - 1/2] / 8 = 3,19^{***}$$

$$Z_M = [(4 \cdot 400 + 3)0,15]^{1/2} - [(4 \cdot 500 - 4 \cdot 400 - 1)0,85]^{1/2} = 2,91^{**}$$

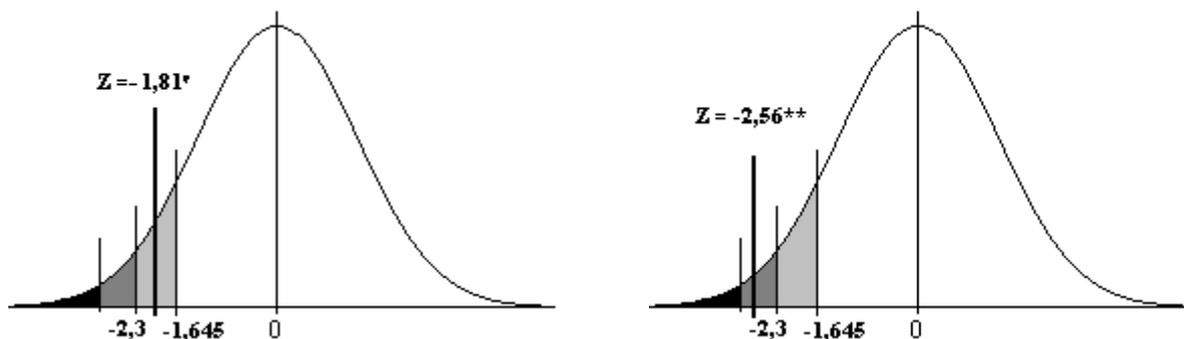
La probabilidad binomial es 0,998407 la cual corresponde a un valor $Z = 2,96$

En conclusión el resultado obtenido es muy significativo. Se encontró evidencia como para rechazar la afirmación del fabricante.

Conviene recordar que la H_0 define el lugar donde se coloca la cola: si en la H_0 se postula mayor o igual, corresponde la cola de la izquierda como en este caso. Si en la H_0 se postula menor o igual, es al revés. Tanto Yates como Molenaar siempre dan valores positivos porque lo que se busca es la probabilidad, y para determinar de cual cola se trata hay que estudiar la H_0 .

Ejemplo 3: El mismo caso anterior, solo que esta vez se curaron 411 individuos. Es todo similar solo que el nuevo valor de Z es:

$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|411 - 425| - 1/2] / 8 = 1,81^*$ Como se trata de la cola izquierda de la distribución el valor de Z debería ser negativo y graficarse como:



Ahora pasan dos cosas: este nuevo valor de Z cae en la zona de rechazo para el 95% de confianza, pero en la zona de aceptación para el 99% de confianza. Estos resultados no son contradictorios como puede parecer a primera vista. Ocurre que se tiene evidencia como para rechazar la H_0 al nivel del 95%, pero no es tan fuerte como para rechazarla al nivel del 99%. Y es por eso que se le colocó un solo asterisco. En cambio, si por ejemplo se hubieran curado 405 individuos, resultaría $Z = -2,56^{**}$. O sea, se rechaza a los niveles del 95% y del 99% pero no alcanza

para rechazar la H_0 al nivel del 99,9% ($Z_{0,999} = -3,09$). En las figuras de más arriba se han graficado estas dos situaciones. Las mismas se deben expresar como sigue:

$$r = 411 \notin 95\% \text{ CI } (412 ; + \infty) \text{ y con eso se deduce que cae dentro del } 99\% \text{ CI}$$

$$r = 405 \notin 99\% \text{ CI } (406 ; + \infty) \text{ y con eso se deduce que cae dentro del } 99,9\% \text{ CI}$$

Notar que si se hubiese usado la aproximación de Molenaar los valores hubieran sido:

$$Z_M = [(4.411 + 3)0,15]^{1/2} - [(4.500 - 4.411 - 1)0,85]^{1/2} = 1,682^*$$

$$Z_M = [(4.405 + 3)0,15]^{1/2} - [(4.500 - 4.405 - 1)0,85]^{1/2} = 2,346^{**}$$

Los valores anteriores corresponden a la cola izquierda de la distribución y por lo tanto deben ser negativos. Las conclusiones son análogas pero los valores dan más cercanos a los límites. Por otra parte se pueden calcular las probabilidades exactas con la probabilidad Binomial que son:

$$P_{Bi} (r \leq 411) = 0,04784 \text{ lo que implica un valor } Z = -1,666$$

$$P_{Bi} (r \leq 405) = 0,04784 \text{ lo que implica un valor } Z = -2,378$$

Una vez más se puede apreciar que la aproximación con el método de Molenaar es mejor que el procedimiento clásico de Yates.

12.5 Comparaciones de dos muestras

El modelo de Gauss sirve para comparar dos muestras entre sí y validar la hipótesis estadística de que ambas provienen de dos poblaciones con la misma media. Esto se puede hacer comparando dos medias muestrales para el caso de variables continuas, o dos proporciones en el caso de variable discreta. Si además se pudiese probar que las varianzas son iguales, entonces se podría validar el hecho de que ambas muestras provienen de la misma población normal. Ya que se está suponiendo que tienen igual media y varianzas, que son los dos parámetros que definen en forma unívoca a una curva de Gauss.

La hipótesis nula más conveniente es que ambas muestras tienen igual media (o igual proporción), porque de esa forma no se necesitan sus valores en la fórmula del cálculo del estadígrafo Z . Se puede trabajar con tamaños de muestras diferentes, aunque si fueran iguales las fórmulas son más sencillas. Los dos casos principales son:

12.5.1 Comparaciones de dos medias muestrales

Para este caso se pueden emplear las fórmulas vistas en el capítulo anterior para la definición de los intervalos de confianza, y así tipificar la diferencia de medias. Esto es postular:

$$H_0 : \mu_1 = \mu_2 \quad \text{O sea, } \mu(\bar{x}_1 - \bar{x}_2) = 0 : \text{ no hay diferencia entre las medias.}$$

$H_1 : \mu_1 \neq \mu_2$ O sea, $\mu(\bar{x}_1 - \bar{x}_2) \neq 0$: hay diferencia entre las medias

Entonces siempre será un ensayo de dos colas. Y el estadígrafo Z de comparación es:

$$Z = [(\bar{x}_1 - \bar{x}_2) - (\mu \bar{x}_1 - \bar{x}_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Aplicando la H_0 a la ecuación anterior queda: $Z = [(\bar{x}_1 - \bar{x}_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Y si las muestras son de igual tamaño y desvío: $Z = [(\bar{x}_1 - \bar{x}_2)] / [\sigma \sqrt{\frac{2}{n}}]$

Ejemplo 1) Se realizaron 36 mediciones repetidas de glucosa, usando una solución calibrada con $\mu = 90$ mg/dl, con dos marcas comerciales diferentes. Los resultados obtenidos para cada caso son: Marca A (media 95 mg/dl y desvío 8 mg/dl), Marca B (media 93 y desvío 7 mg/dl). Ensayar la hipótesis de que ambas marcas son equivalentes.

$H_0 : \mu_A = \mu_B$ El uso de una marca u otra es indistinto

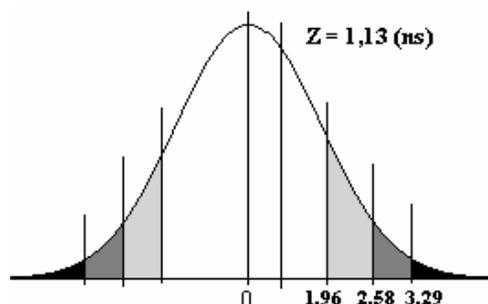
$H_1 : \mu_A \neq \mu_B$ Hay diferencia entre ambas marcas

Donde $(\bar{x}_1 - \bar{x}_2) = 95 - 93 = 2$ y $\sigma_{1-2} = 1,772$

Entonces resulta:

$$Z = [(\bar{x}_1 - \bar{x}_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 2 / 1,772 = 1,13 \text{ (ns)}$$

El valor 2 cae dentro del CI 95% (- 3,5 ; + 3,5)



Como $Z < 1,96 = z_{\alpha/2}$ no se puede rechazar la hipótesis nula. Este test no arroja evidencia significativa, como para pensar que hay diferencias entre las marcas A y B.

Ejemplo 2) En una industria farmacéutica se cambió un dosificador líquido por uno nuevo. El anterior tenía un promedio histórico de 5,01 ml con un desvío de 0,08 ml en 100 pruebas de calidad. Al actual se le realizaron 50 pruebas arrojando una media de 5,048 ml y un desvío de 0,05 ml. Con esta información se debe decidir si ambos dosificadores son equivalentes.

$H_0 : \mu_N = \mu_V$ Son equivalentes.

$H_1 : \mu_N \neq \mu_V$ Hay diferencia entre ambos.

Usando la fórmula general donde el supuesto principal es que las varianzas son diferentes, pero los valores esperados son equivalentes entonces resulta:

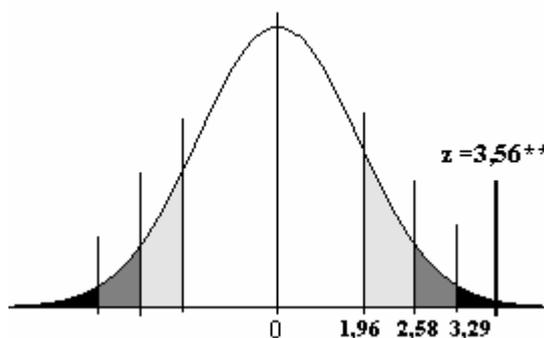
$\sigma_1 \approx DS_1 = 0,08$ y $\sigma_2 \approx DS_2 = 0,05$ con sus tamaños muestrales de 100 y 50 respectivamente:

$$Z = [(5,048 - 5,01)] / \sqrt{\frac{0,05^2}{50} + \frac{0,08^2}{100}}$$

$$Z = 3,56^{***} \notin \text{CI } 99,9\% (-3,29 ; +3,29)$$

$$\mu_N - \mu_V = 0,038 \notin \text{CI } 99,9\% (-0,035 ; +0,035)$$

Hay fuerte evidencia de que no son equivalentes.



Notar que se ha probado que hay una diferencia del 1,2% entre ambos. Lo que significa que cada medicamento fabricado tendrá, en promedio, una cantidad mayor que la histórica lo que implicará un aumento de los costos de producción. Esta evidencia aconseja no usar el nuevo dosificador.

12.6 Comparaciones de dos proporciones

Para este caso se pueden emplear las fórmulas vistas en el capítulo anterior para la definición de los intervalos de confianza de la diferencia de dos proporciones y así tipificarla. O sea:

$$Z = (p_1 - p_2) - (\mu p_1 - p_2) / \sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n_1} + \frac{\pi_2 \cdot (1 - \pi_2)}{n_2}}$$

Nuevamente conviene hacer el supuesto de igualdad de proporciones en la hipótesis nula para simplificar la fórmula cuando se pueda. Esto es:

$H_0 : \mu p_1 - p_2 = 0$: No hay diferencia entre ambas proporciones.

$H_1 : \mu p_1 - p_2 \neq 0$: Hay diferencia entre ambas.

También se puede trabajar con las frecuencias en lugar de las respectivas proporciones.

Ejemplo 1) Dos grupos A y B formados por 100 pacientes sufren un cierto tipo de infección. Se administra un nuevo remedio a los del grupo A y un placebo a los del grupo B que es el grupo testigo o control. En todo lo demás, se trata a los 200 pacientes de la misma forma. Transcurrido el tiempo de prueba se verifica que se han curado 75 individuos del grupo A y 65 del B. Decidir si el nuevo remedio ayuda a curar la enfermedad.

$H_0 : \pi_A \leq \pi_B$ Las diferencias detectadas se deben al azar, el remedio no es efectivo.

$H_1 : \pi_A > \pi_B$ Hay diferencia entre ambas y el remedio es efectivo.

Acá conviene hacer un ensayo de una sola cola porque lo que interesa es ver si el remedio sirve, esto es, si el número de curados es mayor en el grupo A que en el grupo B. Los valores muestrales de proporciones se usan para estimar los poblacionales. Entonces:

$$\mu_{1-2} = 0,75 - 0,65 = 0,10 \quad \text{y} \quad \sigma_{1-2} = \sqrt{\frac{0,75 \cdot 0,25}{100} + \frac{0,65 \cdot 0,35}{100}} = 0,0644$$

Con estos valores se puede calcular:

$$Z = 0,10 / 0,0644 = 1,55 \text{ (no significativo)} \quad \mu_{1-2} = 0,10 \text{ cae dentro de CI 95\% } (-\infty ; +0,13)$$

Esto significa que no se tiene prueba científica como para poder afirmar que el suero es efectivo. Si se le realiza la corrección por continuidad el valor de Z es todavía menor ($Z = 1,48$).

Otra forma de resolver este problema es haciendo los supuestos siguientes:

a) Ambas muestras provienen de la misma población, esto es $\pi = \pi_A = \pi_B$

b) La mejor estimación del valor desconocido π es con la media ponderada de las proporciones observadas. Esto es: $\pi \approx [100 (0,75) + 100 (0,65)] / (100 + 100) = 0,7$

Y la varianza se calcula ahora con este valor estimado de 0,7. Esto es:

$$\sigma^2 = \pi (1 - \pi) [(1/n_1) + (1/n_2)] = 0,7 \cdot 0,3 [(1/100) + (1/100)] = 0,0042 \quad \text{O sea, } \sigma = 0,06481$$

c) Se formula la H_0 lo que significa que el valor observado de la diferencia se debe al azar.

$$Z = (p_1 - p_2) / \sigma = 0,10 / 0,06481 = 1,543 \text{ (no significativo)}$$

Notar que los dos procedimientos son muy similares y puede adoptarse cualquiera de ambos en casi todos los casos. La excepción es cuando las diferencias entre los tamaños muestrales es muy grande y cuando las proporciones difieren substancialmente. En todo otro caso, los valores del desvío estándar son muy parecidos como en el ejemplo anterior $\sigma_{1-2} = 0,0644$ y $\sigma = 0,06481$.

Ejemplo 2) Una muestra de 300 clientes de la Farmacia A demostró que el 58% paga de contado. En cambio en otra muestra de 200 clientes de la Farmacia B solo lo hace el 44%. Se desea saber si esa diferencia se debe al azar o si existe alguna causa como su ubicación en la ciudad.

$H_0 : \mu p_1 - p_2 = 0$: No hay diferencia entre ambas.

$H_1 : \mu p_1 - p_2 \neq 0$: Hay diferencia entre ambas.

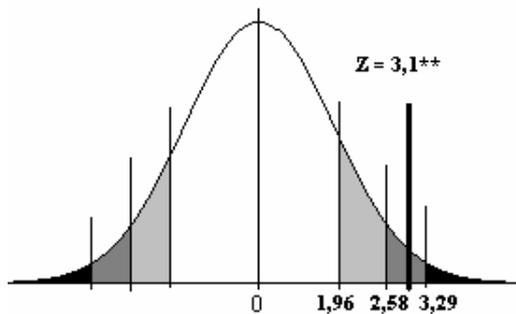
Acá se hace un ensayo de dos colas porque lo que interesa es ver si hay diferencia entre ambas Farmacias. Por lo tanto se calculan:

$$\mu_{1-2} = 0,56 - 0,48 = 0,14 \quad \text{y} \quad \sigma_{1-2} = \sqrt{\frac{0,58 \cdot 0,42}{300} + \frac{0,44 \cdot 0,56}{200}} = 0,0452$$

Con estos valores se puede calcular:

$$Z = 0,14 / 0,0452 = 3,1^{**} \quad \text{pues } Z_{0,95} = 1,96 ; Z_{0,99} = 2,58 \text{ y } Z_{0,999} = 3,29$$

En el caso de desvío ponderado es $p = 0,524$ y $\sigma = 0,0456$ valor muy parecido al anterior σ_{1-2} . Cualquiera de los dos procedimientos dará valores muy parecidos de Z y pueden ser usados indistintamente.



Esto significa que se tienen resultados muy significativos que prueban la diferencia entre la clientela de ambas farmacias, respecto al pago contado.

$$2,58 < Z < 3,29$$

$\mu_{1-2} = 0,14 \notin 99\% \text{ CI } (-0,12 ; +0,12)$ como cae fuera del intervalo se rechaza la H_0

12.7 Intervalos de confianza versus test de hipótesis

En los comienzos del uso de la validación estadística de los resultados experimentales, la costumbre era informar la probabilidad asociada al evento estudiado. Por ejemplo, se informaba que $p < 0,05$ lo que implicaba decir que los resultados fueron significativos, o bien $p = 0,039$ que significa lo mismo, pero usando el valor hallado de probabilidad. Esta costumbre fue criticada porque el lector del trabajo de investigación no podía saber el tipo de test estadístico usado, ni la manera de emplear los datos del estudio. El primer cambio fue introducir el valor del estadígrafo hallado junto con la probabilidad del mismo. Por ejemplo, informar que se encontró un $Z = 2,64$ con $p = 0,0047$, o bien $Z = 2,64^{**}$, de donde el lector podía deducir que se había aplicado la función normal y se habían encontrado valores muy significativos ($p < 0,01$). A esta manera de informar se la denominó: test de hipótesis o valores de probabilidad. En un trabajo de Gardner y Altman publicado en 1986, los autores presentan un estudio comparativo entre el uso de los intervalos de confianza y los test de hipótesis en Medicina.

Según estos autores desde 1950 se produjo un incremento inusitado en el uso de estadística en Medicina, lo que se tradujo en consecuencias desafortunadas porque en vez de analizar los resultados obtenidos clínicamente, lo reducían a una dicotomía del tipo H_0 vs. H_1 . El excesivo uso del test de hipótesis a expensas de otras formas de análisis llegó a tal grado, que en los trabajos de investigación solo se hacía referencia a los valores de probabilidad, o si se acepta o rechaza la H_0 , sin mencionar cosas tales como las concentraciones halladas, las proporciones, sus diferencias etc. La alternativa propuesta fue usar los intervalos de confianza para el enfoque estadístico en los trabajos de investigación, porque entregan más información al lector. Esta es la recomendación para trabajos futuros a publicarse en las revistas médicas y sus asociadas.

El problema básico es que diferencias pequeñas encontradas en una magnitud clínica pueden ser estadísticamente significativas en grandes muestras, mientras que en clínica son irrelevantes. O bien, importantes efectos clínicos pueden aparecer como estadísticamente no significativo solo porque el número de casos estudiados era pequeño. La idea básica es que: *cuando un resultado es estadísticamente significativo, no significa necesariamente que clínicamente lo sea*. Otro problema es que una magnitud clínica no tiene porque poseer una distribución normal en poblaciones humanas, pero si se le toman muestras aleatorias e independientes lo suficientemente

grandes, la distribución muestral va a ser normal independientemente de la población de la que fue extraída. Esto hecho ya fue denotado en los casos de las aproximaciones normales a poblaciones binomiales o de Poisson. Entonces hay que tener cuidado de no suponer a la población como normal, solo porque las muestras extraídas lo sean.

En conclusión la propuesta es incluir toda la información estadística posible en un informe para que tanto el lector como el autor eviten confusiones. Por ejemplo, conviene informar los resultados así: “La diferencia entre las medias muestrales de la presión sistólica entre diabéticos y no diabéticos encontrada fue de 6,0 mm Hg., con un intervalo de confianza del 95% entre 1,1 y 10,9 mm Hg. y el estadístico hallado fue $Z = 2,9$ con una probabilidad asociada de $p = 0,002$.” O en forma sintética: “ $\Delta \bar{x} = 6$ mm Hg, 95% CI (1,1 ; 1,9), $Z = 2,4^{**}$ ($p = 0,002$)”

12.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) La H_0 es la hipótesis que se plantea para ser testeada. | V | F |
| 2) Las H_0 y la H_1 son complementarias. | V | F |
| 3) Explicar que son los riesgos del comprador y del vendedor | | |
| 4) La mejor forma de minimizar los errores es aumentando el número de experimentos. | V | F |
| 5) Explicar los tres niveles de significación usuales en variables biológicas..... | | |
| 6) La zona de aceptación en un ensayo de dos colas es igual a la de una cola. | V | F |
| 7) Explicar los pasos a seguir en un ensayo de hipótesis | | |
| 8) Los resultados son muy significativos si Z cae en la zona de rechazo del nivel 99,9% | V | F |
| 9) Los resultados no son significativos si Z cae en la zona de aceptación. | V | F |
| 10) Si se acepta la H_0 se tiene prueba científica de la misma. | V | F |
| 11) Cuando se rechaza la H_1 se dice que se han validado las conclusiones alcanzadas. | V | F |
| 12) Las pruebas de validación más usuales se hacen con la distribución muestral de | | |

2) La vida útil de un medicamento A se midió con 200 muestras y dio una media de 1400 días, con un desvío de 120 días. El medicamento B se determinó con 100 muestras y se obtuvo una media de 1200 días con un desvío de 80 días. Decidir si hay diferencia entre ambos.

3) Suponiendo que se usa una muestra de sangre control y al suero obtenido se lo fracciona en 30 alícuotas, luego a cada una se le determina la creatinina, y con los valores medidos se obtienen un promedio de 10 mg/dl y un desvío de 2,2 mg/dl. Si el valor real es de 9,5 mg/dl. Decidir si el método clínico está calibrado.

4) Se toman 40 muestras de cierto antibiótico y se les mide su contenido de Sulfato de Neomicina, resultando un promedio de 87 mg con un desvío de 4 mg. Decidir si el contenido del Sulfato está por debajo del límite de control de producción, fijado en 85 mg.

5) Se realizaron 100 exámenes de Bioestadística. Si se toman 30 de ellos al azar entre los varones y se obtiene una media de 6,5 puntos con un desvío de 1,5 puntos, mientras que se eligen otros 30 entre las mujeres obteniendo una media de 7 con un desvío de 1 punto. Decidir si hay diferencias entre las notas de ambos sexos.

6) Se lanza una moneda al aire 100 veces y se obtienen 53 caras. Decidir si la moneda está bien hecha. ¿Qué ocurre si se sacan 58 caras?

7) Se tomaron 200 muestras aleatorias de presión sistólica a niños cuyos padres son hipertensos, obteniéndose una media de 100 y un desvío de 5. Luego se tomaron 100 muestras de niños cuyos padres tienen la presión sanguínea normal, y se obtuvo una media de 95 con un desvío de 4. Decidir si hay diferencia entre ambos grupos.

8) El peso promedio de 50 estudiantes de la Facultad elegidos al azar entre los que toman parte de pruebas atléticas es de 60,5 Kg. con un desvío de 5 Kg. Mientras que otros tantos alumnos, elegidos entre los que no demuestran interés, arrojó una media de 64,5 Kg. con un desvío de 3 Kg. Testear la hipótesis de que los alumnos que participan en pruebas atléticas son más flacos que los otros.

9) Para hacer un test de embarazo se emplea la técnica 1 en 400 pacientes y se encuentran 220 pacientes que dieron un resultado (+). Los aciertos en (+) fueron 180 y 160 en (-). Con estos datos calcular la Sensibilidad y Especificidad de esta técnica. A las mismas pacientes se les aplicó la técnica 2 y resultaron 240 pacientes con (+), mientras el número de aciertos en (+) fue de 190 y en (-) 150. Decidir cuál de ambas técnicas es la mejor para adoptar en nuestro Laboratorio (a un nivel del 95%).

10) Se sabe que la concentración de coloides en el agua de un río es del 5,4 /ml. Se tomó una muestra del mismo a la que se le agregó un reactivo. Luego de un rato se extrajo una pequeña cantidad para analizarla en una cámara de recuento. Los resultados fueron:

Nº de coloides: 2 3 4 5 6 7 (nº/ml)

Frecuencia: 10 20 170 100 90 10

Decidir si hubo algún cambio significativo en la concentración del agua (a un 95% y 99%)

11) Se midieron 50 alícuotas de un suero con el Kit de Glucosa marca X, y otros 50 con la marca Y. Los resultados fueron:

Marca X: entre (80-80,9) 8 veces, (81-81,9) 22 veces, (82-82,9) 15 veces y (83-83,9) 5 veces.

Marca Y: entre (80-80,9) 6 veces, (81-81,9) 30 veces, (83-83,9) 10 veces y (83-83,9) 4 veces

Decidir si hay diferencia en la exactitud y en la precisión de ambos métodos, comparando las medias y desvíos estándar al 95% de confianza. Plantear los intervalos de confianza para cada uno al 95%, 99% y 99,9%.

12) Para hacer un test VDRL se emplea el *Kit A* en 200 pacientes y se encuentran 120 pacientes que dieron un resultado (-). Los aciertos en (+) fueron 70 y 110 en (-). Con estos datos calcular la Sensibilidad y Especificidad de A y B. Si con los mismos pacientes se empleó el *Kit B* y resultaron 140 pacientes con (-), mientras el número de aciertos en (+) fue de 55 y en (-) 130. Decidir cuál de ambas técnicas es la mejor para adoptar en nuestro Laboratorio (a un nivel del 95%).

13) En un recuento de rojos en cámara se espera obtener 4,5 millones por cm³ de cierto paciente. Se le toma una muestra de sangre, y de acuerdo con los resultados hallados decidir (a un nivel de confianza del 95%) si se verifica lo esperado.

Nº de rojos por millón 2 3 4 5 6

Frecuencia 10 90 110 150 40