Métodos estadísticos para pruebas sensoriales en la industria

Enrique Villa Diharce
villadi@cimat.mx

CIMAT, Guanajuato, Gto. Julio 14-18 de 2014

Programa

1.-Estudios sensoriales
2.-Métodos de discriminación
3.- Análisis de potencia para pruebas sensoriales
4.- Pruebas de vida de anaquel

Referencias y bibliografía

Meilgaard, M. Civille, G.V. y Carr, B. T. 1999. “Sensory Evaluation Techniques” Third Edition. CRC Press, Boca Raton, FL.

Gacula, M.C. Jr., Sing, J., Bi, J. y Altan, S. 2009. “Statistical Methods in Food and Consumer Research”, Second Edition. Academic Press, New York, NY.

Stevens, S.S. 1946. On the theory of scales of measurement. Science **103**, 677-680.

Hough, G. 2010. “Sensory Shelf Life Estimation of Food Products” CRC Press, Boca Raton, FL.

Jian, Bi. 2006. “Sensory discrimination Tests and Measurements, Statistical Principles, Procedures and Tables” Blackwell Publishing.

Lawless, H.T. and Hildegrade H. 2010. “Sensory Evaluation of Food, Principles and Practices”, Second Edition. Springer, New York.

Evaluación sensorial

Evaluación sensorial es una disciplina científica usada para evocar, medir, analizar e interpretar reacciones hacia aquellas características de alimentos y materiales como son percibidos por los sentidos de vista, olfato, gusto, tacto y oído.

División de Evaluación Sensorial del Instituto de Tecnólogos de Alimentos, USA, 1975.

Evaluación sensorial

De los diferentes sectores de la industria de productos de consumo (alimentos, bebidas, cosméticos, productos de cuidado personal, telas, farmacéuticos, etc.), los sectores de alimentos y bebidas tuvieron mayor interés en la evaluación sensorial.

Durante los 1940´s y la primera mitad de los 1950´s, la evaluación sensorial tuvo un impulso importante por parte del ejercito de EUA, desarrolló programas de investigación sobre la aceptación de alimentos que se distribuían entre las miembros del ejercito.

Después del impulso que dio el ejercito a la evaluación sensorial, la industria empezó a reconocer la importancia de la evaluación sensorial en la formulación y evaluación de los productos.

Introducción

El análisis sensorial es una ciencia multidisciplinaria en la que se utilizan panelistas humanos que utilizan los sentidos de la vista, olfato, gusto, tacto y oído para medir las características sensoriales y la aceptabilidad de los productos alimenticios y de otros.

No existe ningún otro instrumento que pueda reproducir o reemplazar la respuesta humana; por lo tanto la evaluación sensorial resulta un factor esencial en cualquier estudio sobre alimentos. El análisis sensorial es aplicable a muchos sectores, tales como desarrollo y mejoramiento de productos, control de calidad, etc.

Si se desea obtener resultados confiables y válidos en los estudios sensoriales, el panel debe ser tratado como un instrumento científico.

Introducción

Toda prueba que incluya paneles sensoriales debe llevarse a cabo en condiciones controladas, utilizando diseños experimentales, métodos de prueba y análisis estadísticos apropiados. Solo así, el análisis sensorial podrá producir resultados consistentes y reproducibles.

La información sobre los gustos, las preferencias y requisitos de aceptabilidad se obtiene empleando métodos de análisis adaptados a las necesidades del consumidor y evaluaciones sensoriales con panelistas no entrenados.

La información sobre las características sensoriales específicas de un alimento requiere pruebas orientadas al producto.

Introducción

La identificación y medición de las propiedades sensoriales es factor esencial para el desarrollo de nuevos productos alimenticios, reformulación de productos ya existentes, identificación de cambios causados por los métodos de procesamiento, almacenamiento y uso de nuevos ingredientes, así como para el mantenimiento de las normas de calidad.

Para obtener este tipo de información, se llevan a cabo evaluaciones sensoriales en el laboratorio con paneles entrenados. Cuando hay un cambio en la formulación de un producto o una nueva fórmula , las pruebas orientadas al producto preceden a las pruebas orientadas al consumidor.

Introducción

Pruebas orientadas al consumidor. En estas pruebas se toma una muestra aleatoria grande, compuesta de personas representativas de la población de posibles usuarios, con el fin de obtener información sobre las actitudes o preferencias de los consumidores.

En las pruebas con consumidores no se emplean panelistas entrenados, ni seleccionados por su agudeza sensorial; sin embargo los panelistas deben ser usuarios del producto. Por lo general, para este tipo de pruebas se entrevistan de 100 a 500 personas. Los resultados se utilizan para predecir actitudes de una población determinada.

Las entrevistas o pruebas pueden realizarse en un lugar central tal como un mercado, una escuela, centro comercial, centro comunitario o también en los hogares de los consumidores. Una verdadera prueba orientada al consumidor, requiere seleccionar un panel representativo de la población

Introducción

Pruebas orientadas al producto. Estas pruebas emplean pequeños paneles entrenados que funcionan como instrumentos de medición. Los paneles entrenados se utilizan para identificar diferencias entre los productos alimenticios similares o para medir la intensidad de características tales como el sabor (olor y gusto), textura o apariencia.

Por lo general, estos paneles constan de 5 a 15 panelistas seleccionados por su agudeza sensorial, los que han sido especialmente entrenados para la tarea que se realizará.

Los panelistas entrenados no deben utilizarse para evaluar aceptabilidad de alimentos, ya que debido a su entrenamiento especial, no solo son mas sensibles a las pequeñas diferencias que lo que es el consumidor promedio, sino que también pueden poner a un lado sus preferencias y aversiones cuando están midiendo parámetros sensoriales.

Escalas de medición

Los datos de los estudios sensoriales pueden clasificarse de acuerdo a alguna de las siguientes escalas (Stevens, 1946, 1951):

Nominal, ordinal, de escala y de razón.

El tipo de análisis estadístico de los datos, dependerá del tipo de escala elegido. La escala de medición deberá seleccionarse después de analizar y determinar los objetivos del estudio.

Escala Nominal. Esta escala es la más sencilla, ya que aquí los números no tienen un valor numérico, ya que se emplea para clasificar unidades o características. Por ejemplo:

- En estudios de aceptación rechazo usamos:

  0=rechazo, 1=aceptación

- En la identificación de características olfativas de salsas de tomate, los panelistas pueden usar una escala nominal:

  1=a fruta, 2=dulce, 3=picante y 4=acre.

Escalas de medición

Escala Ordinal. En esta escala, los números representan posiciones y las muestras se ordenan de acuerdo a la magnitud. El orden no dice nada sobre la diferencia entre las muestras. En los paneles de consumidores, las muestras se ordenan en base a su preferencia o aceptabilidad.

Por ejemplo si se tienen tres formulaciones de un tipo de pan dulce, estas pueden ordenarse de la siguiente forma:
1: se prefiere más , 2: de preferencia media y 3: de menor preferencia.

En las pruebas orientadas al producto, el ordenamiento se basa en las intensidades de una característica específica del producto. Así, una serie de 5 muestras de sopa de pollo podría ordenarse por su contenido de sal, asignando el numero 1 a la sopa más salada, 5 a la menos y las otras muestras quedan ubicadas entre estas dos.

Escalas de medición

Escala de Intervalo. Esta escala permite ordenar las muestras de acuerdo a la magnitud de una sola característica del producto o de acuerdo a la aceptabilidad o preferencia . Además, las escalas de intervalo permiten indicar el grado de diferencia entre muestras, ya que esta escala posee una unidad de medición.

Por ejemplo, si utilizamos una escala de intervalo para evaluar las sopas de pollo, no solo se puede identificar la muestra más salada, sino que también se sabe el número de intervalos que separan la sopa más salada de la menos salada.

La mayoría de los datos estadísticos que encontramos en la práctica, se encuentran en una escala de intervalo. Ejemplos de escalas de intervalo son las escalas hedónicas y escalas de temperatura Celsius y Fahrenheit.

Escalas de medición

Escala de Razón. Esta escala es similar a la escala de intervalo, pero además tiene un verdadero punto cero, es decir tiene un origen natural.

En una escala de intervalo, el valor cero elegido arbitrariamente no indica necesariamente la ausencia de la característica que se mide. En una escala de razón, el punto cero indica la ausencia completa de la característica.

Si tenemos X y Y medidos en esta escala, no solo podemos decir cuantas unidades es mayor Y que X, sino también, cuantas veces es mayor. Por ejemplo, si dos muestras A y B reciben las calificaciones 3 y 6 en una escala de razón que mide la intensidad de salado, entonces la muestra B es tres unidades más salada que A y también podemos decir que la muestra B es el doble de salada que la muestra A.

Este tipo de escala raramente se utiliza en pruebas orientadas al consumidor, ya que para poder emplearla adecuadamente se requiere entrenamiento.

Escalas de medición

Pruebas estadísticas para datos sensoriales.
Los datos de las escalas nominales y ordinales se analizan empleando métodos estadísticos no paramétricos, mientras que los datos de las escalas de intervalo y de razón, se analizan empleando métodos estadísticos paramétricos.

Los métodos no paramétricos permiten un grado de discriminación menor que los métodos paramétricos, pero no requieren que los datos tengan una distribución determinada, como ocurre con las pruebas paramétricas.

Variabilidad en los procesos

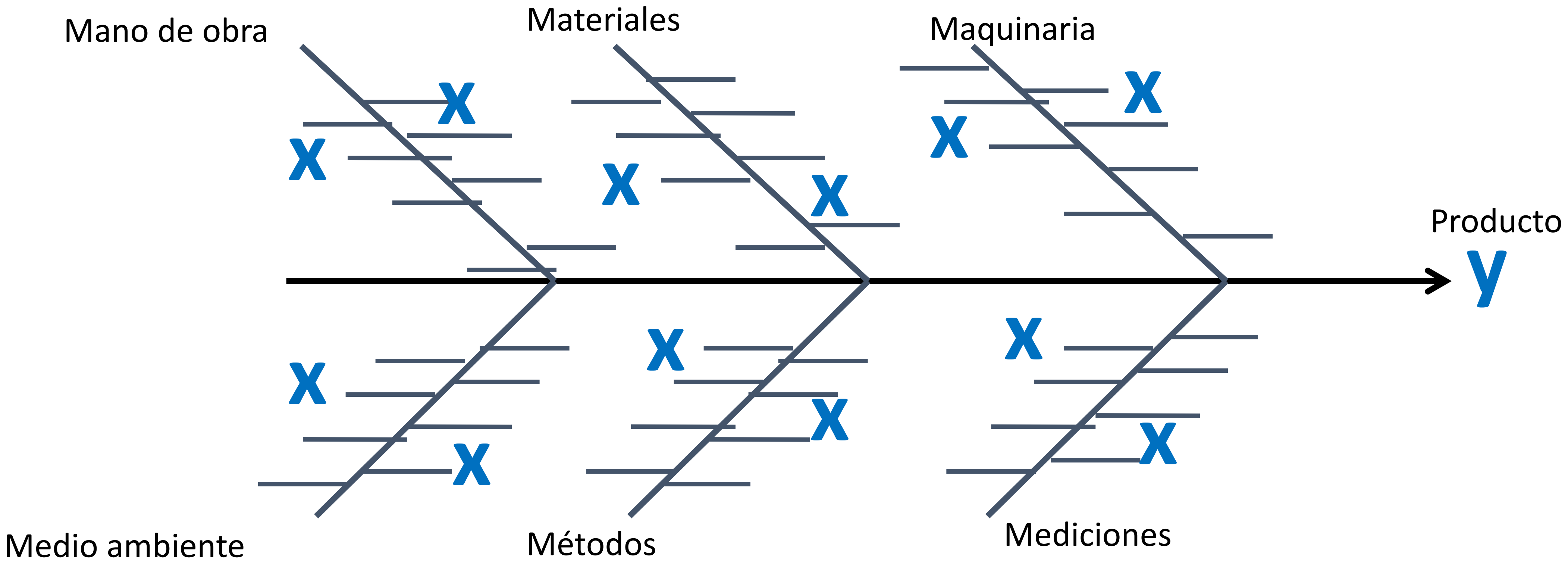


Diagrama de causas y efectos. Causas de variación

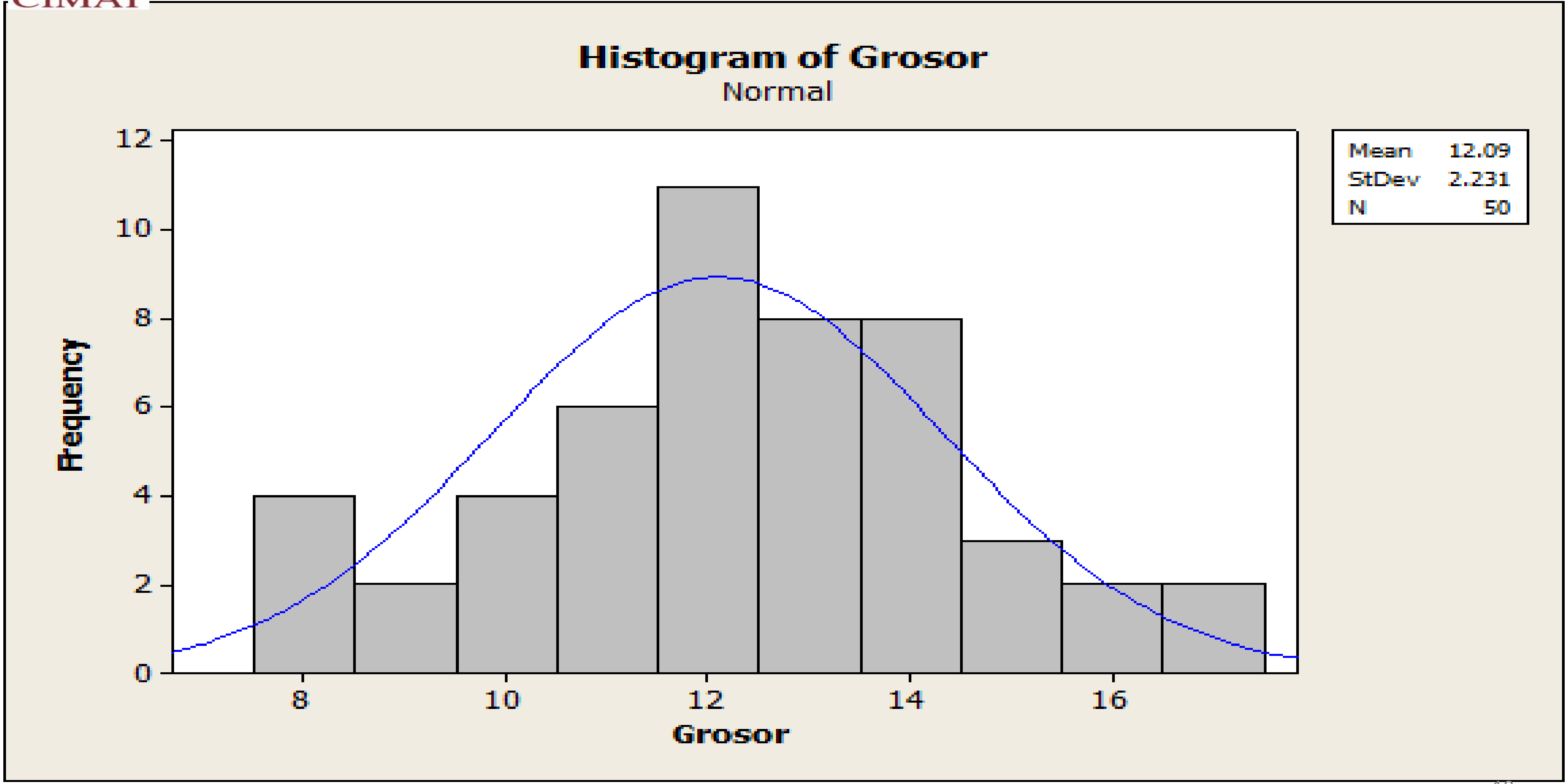
La variabilidad del producto final tiene un gran número de contribuyentes.
Por el Teorema Central del Límite, la distribución de la característica de interés es Normal

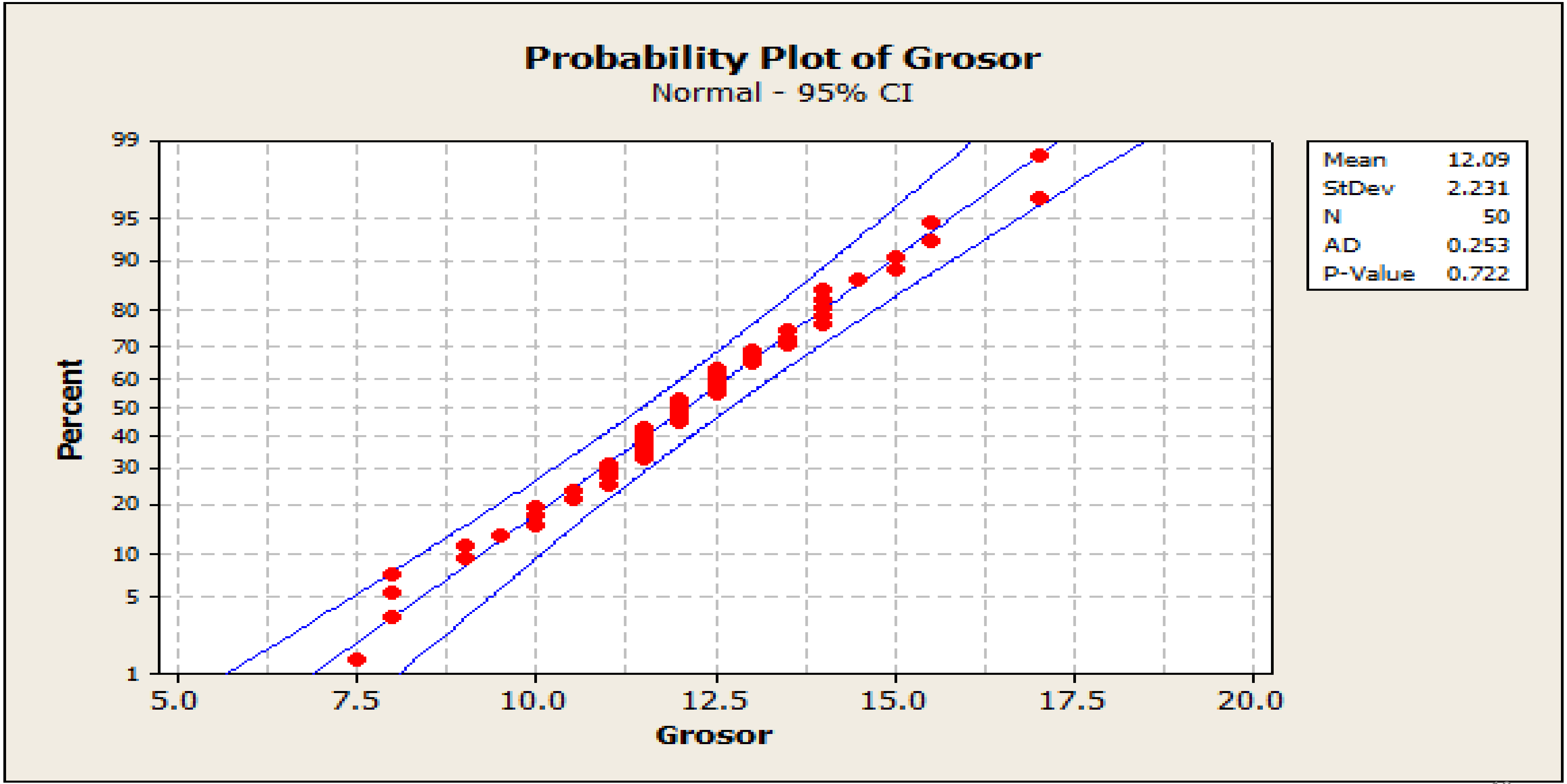
Ejemplo

Una máquina produce piezas de mica. Se toman 50 piezas consecutivas y se mide el grosor (en mm):

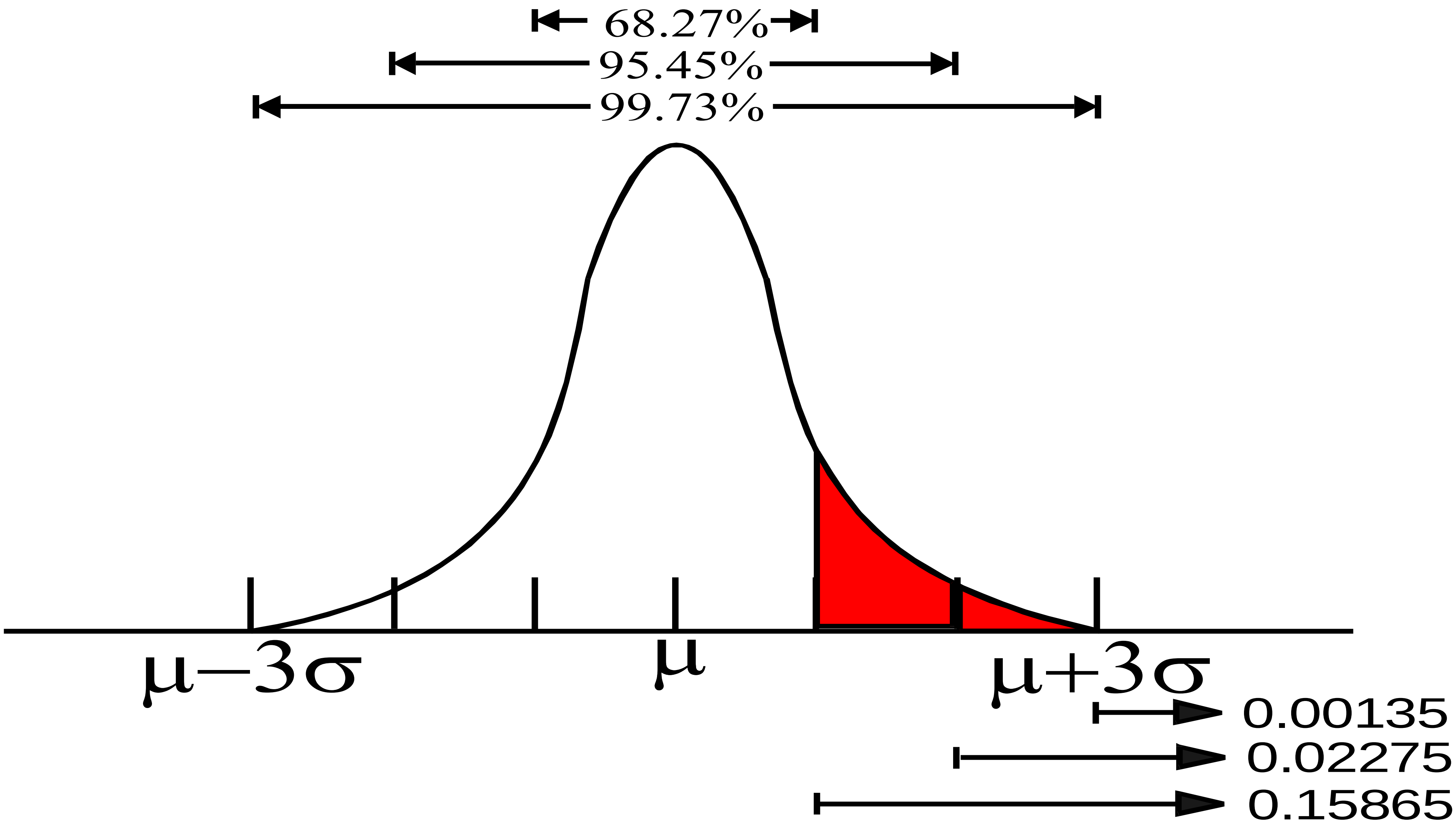
8.0	12.5	12.5	14.0	13.5	12.0	14.0	12.0	10.0	14.5
10.0	10.5	8.0	15.0	9.0	13.0	11.0	10.0	14.0	11.0
12.0	10.5	13.5	11.5	12.0	15.5	14.0	7.5	11.5	11.0
12.0	12.5	15.5	13.5	12.5	17.0	8.0	11.0	11.5	17.0
11.5	9.0	9.5	11.5	12.5	14.0	11.5	13.0	13.0	15.0

¿ Estos datos tienen distribución normal?

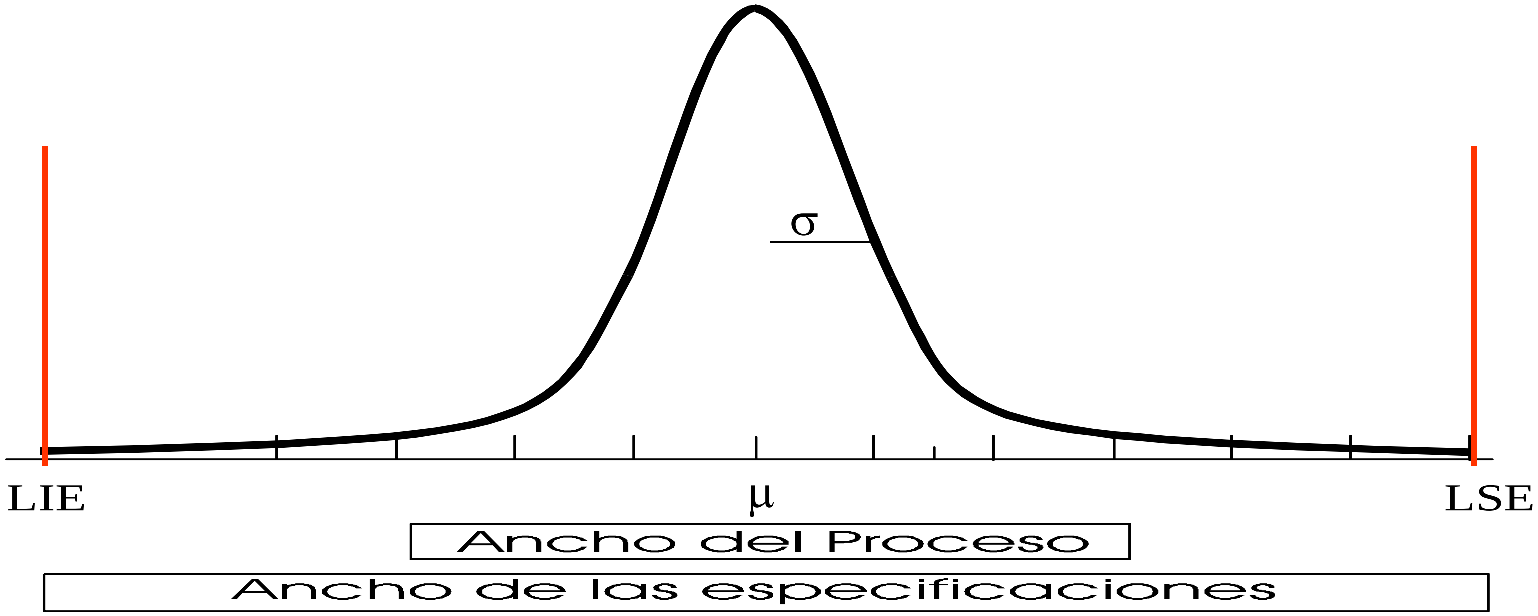

Histogram of Grosor
Normal
Mean 12.09
StDev 2.231
N 50
Frequency
Grosor
0
2
4
6
8
10
12
8
10
12
14
16


Probability Plot of Grosor
Normal - 95% CI
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Grosor
Mean 12.09
StDev 2.231
N 50
AD 0.253
P-Value 0.722

Áreas Bajo la Curva Normal



ESPECIFICACIONES Y ANCHO DEL PROCESO



El ancho del proceso, suponiendo normalidad, se define como 6σ. El ancho de las especificaciones puede medir “cualquier cantidad” de sigmas.

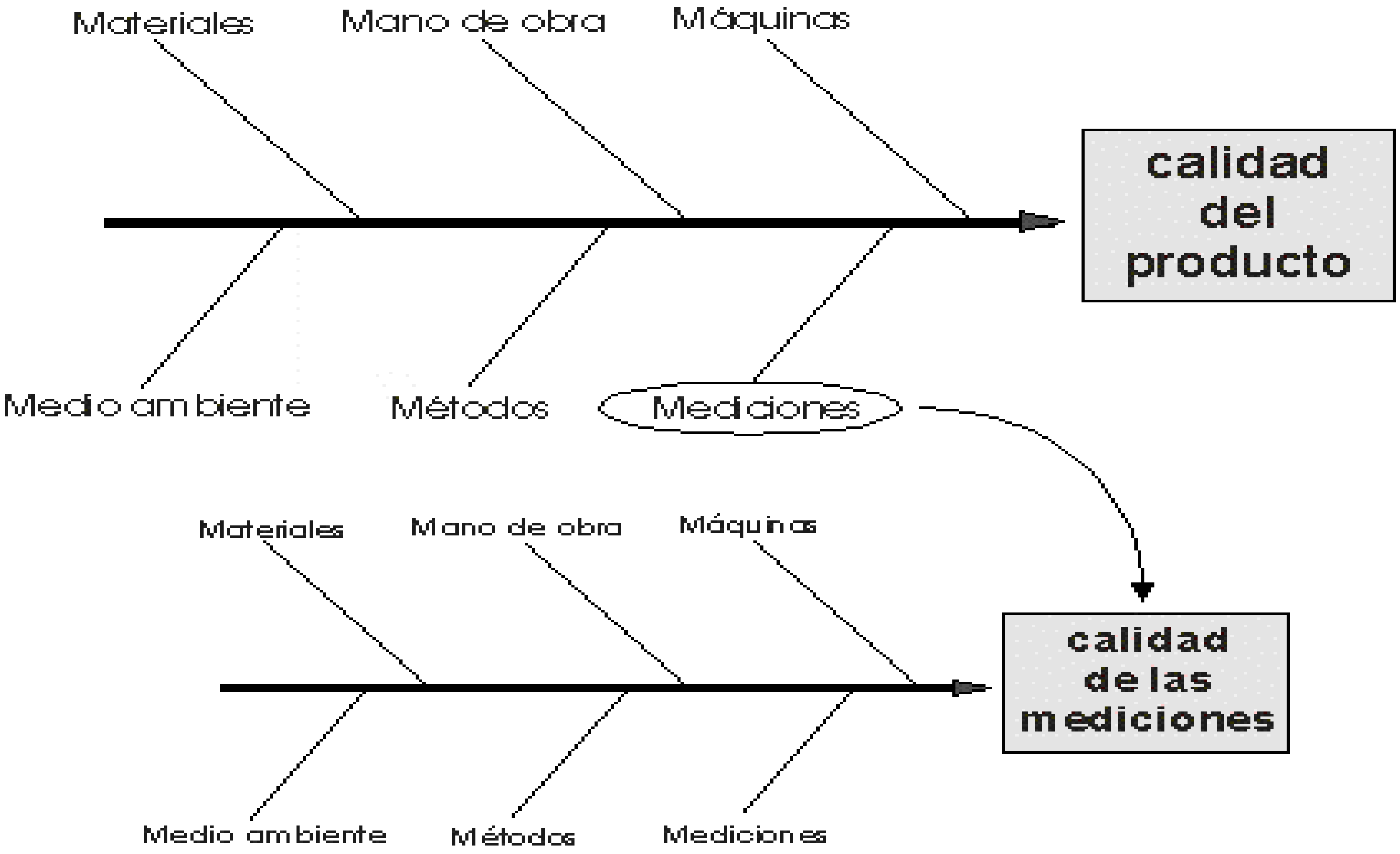


Figura. Las 6M del proceso de medición

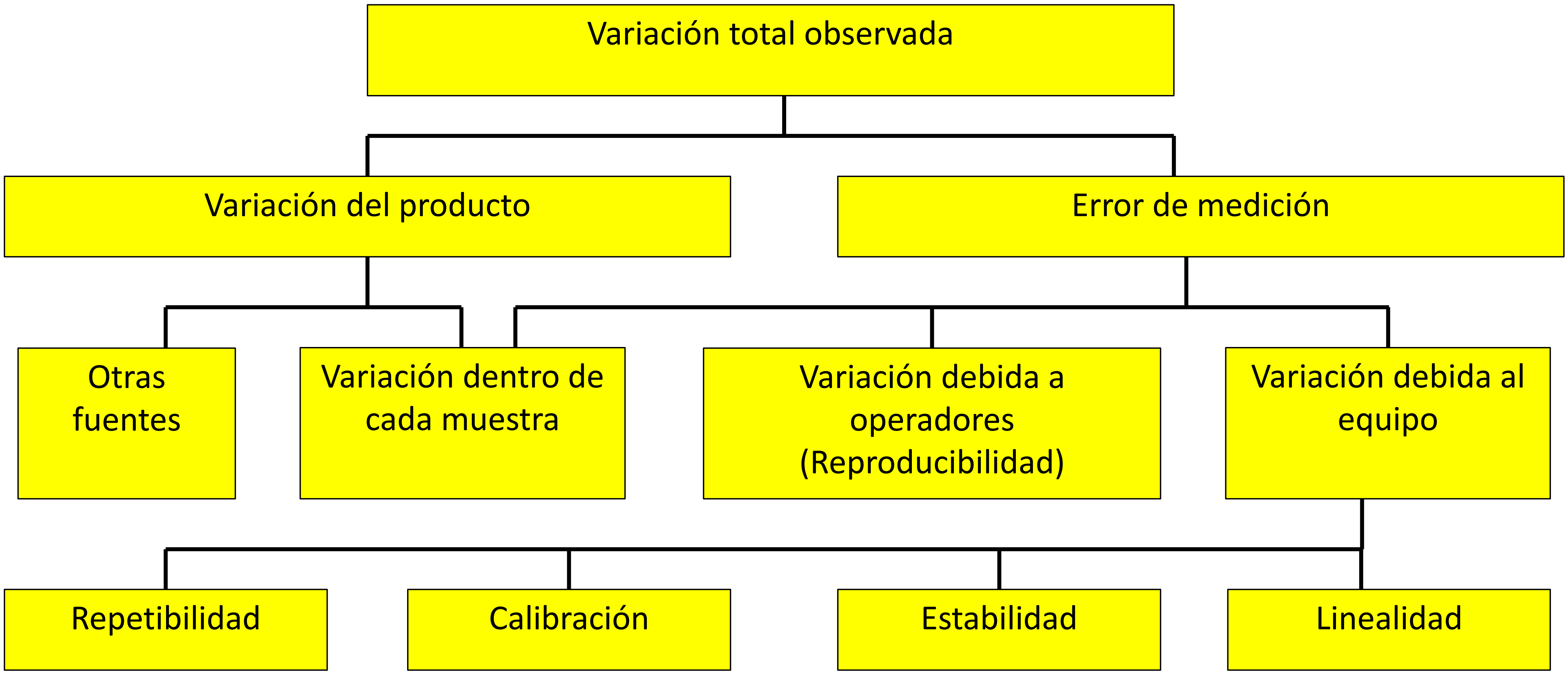
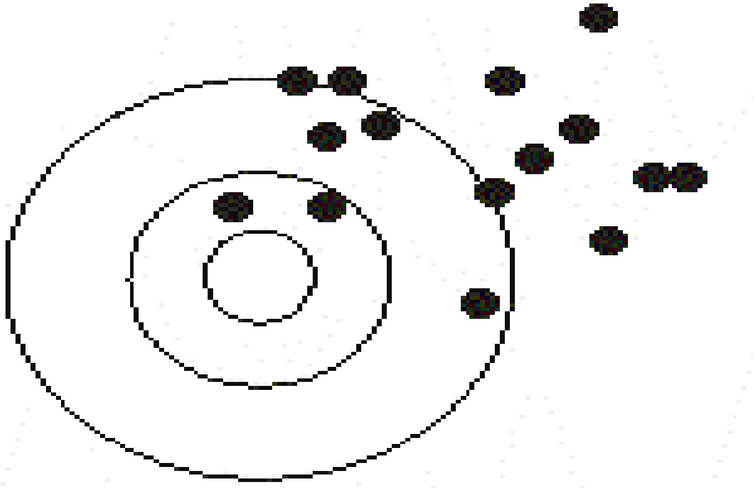
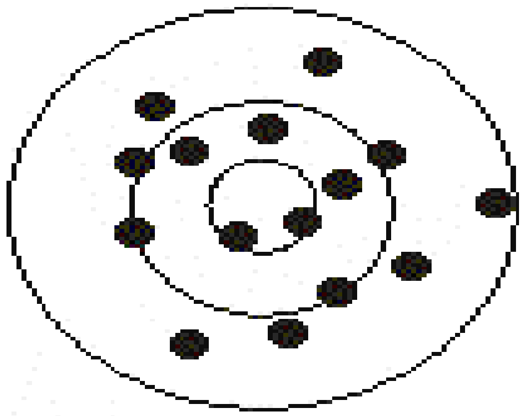


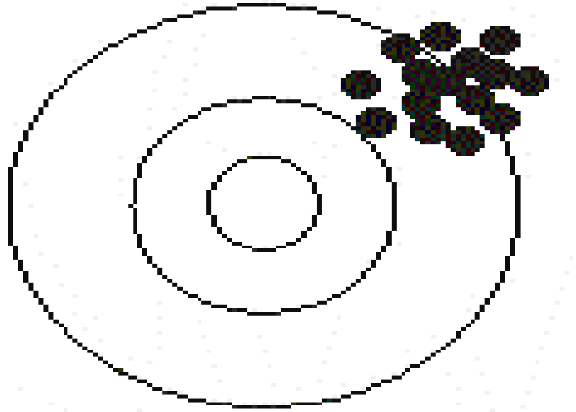
Figura. Fuentes de variabilidad en las mediciones



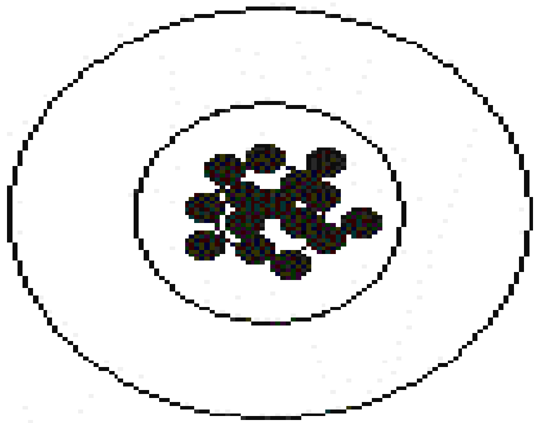
a) Impreciso e inexacto



b) Impreciso y exacto



c) Preciso e inexacto



d) Preciso y exacto

Figura. Precisión y exactitud (como tiro al blanco)

¿Cómo es la calidad de medición en las evaluaciones sensoriales?

Análisis sensorial

Análisis sensorial:
Tiene como funciones principales conducir pruebas validas y hacer mediciones sensoriales confiables.
La inferencia estadística es la base teórica de las pruebas sensoriales.

Análisis sensorial de laboratorio
Se emplea un panel entrenado como un instrumento analítico para medir las propiedades sensoriales de los productos.

Análisis sensorial de consumidores
Se emplea una muestra de la población de consumidores para probar y pronosticar las respuestas de los consumidores respecto a los productos.

Análisis sensorial

Metodología:
La metodología básica para el análisis sensorial de laboratorio y los consumidores son el análisis discriminativo y el análisis descriptivo.

Análisis discriminativo
Incluye mediciones y pruebas discriminativas. Las pruebas discriminativas se usan para determinar cuando existe **una diferencia** de tratamientos para propiedades sensoriales confundibles de los productos.

Análisis descriptivo
Determina, usando una escala de calificación, que tanto difiere entre productos, una característica específica. Es un procedimiento cuantitativo que caracteriza atributos sensoriales de un producto.

Análisis sensorial

Pruebas de aceptación:
Es el análisis descriptivo cuantitativo de las preferencias de los productos.

Las pruebas de aceptación o preferencia para un panel de laboratorio tiene un valor muy limitado, porque no reflejan las preferencias de la población de consumidores. En cambio estas pruebas son de gran utilidad cuando se aplican a muestras de la población de consumidores de un producto.

Pruebas de similaridad
Su objetivo no es demostrar diferencia sino similaridad o equivalencia entre dos productos. Esto es de interés cuando el fabricante modifica la formulación de un producto y quiere saber si los consumidores perciben el producto original y el modificado como similares.

Pruebas de diferencia sensorial

Las pruebas de diferencia sensorial, también conocidas como pruebas de diferencias discriminantes, se utilizan para juzgar y establecer la diferencia entre productos, sin reportar la magnitud o dirección de la diferencia.

Las pruebas de discriminación pueden usarse con diferentes objetivos prácticos. Algunas veces los investigadores están interesados en demostrar que dos muestras son perceptiblemente diferentes. En otros casos están interesados en determinar si dos muestras son suficientemente similares para utilizarse de forma intercambiable.

Existen varias pruebas discriminantes, aquí veremos algunas pruebas para investigar si existe una diferencia sensorial entre muestras. Estas pruebas son, la prueba Triangular, la prueba Dúo-trio y la prueba de diferencias pareadas, que se diseñan para mostrar cuando los sujetos pueden detectar diferencias entre las muestras.

Pruebas de diferencia sensorial

Las pruebas sensoriales más simples se utilizan con la intención de saber si existe una diferencia perceptible entre dos tipos se productos. El análisis de los datos usualmente se basa en la estadística de frecuencias y proporciones.

Un ejemplo clásico de este tipo de pruebas fue el procedimiento triangular que se utilizó en una fábrica de cerveza (Carlsberg), (Helm y Trolle,1946) y en las destilerías de Seagrams, (Peryam y Swartz, 1950).

En la prueba Triangular se toman dos productos de un lote y otro de un lote diferente. Los jueces prueban los tres productos y se les pide que identifiquen el producto diferente.

La capacidad para discriminar diferencias se podría inferir después de una serie consistente de elecciones correctas, que este por encima de lo que se podría esperar por azar.

Prueba Triangular

Prueba triangular

A cada juez se le proporcionan 3 muestras (dos iguales y una diferente) y este tiene que indicar cual es la muestra diferente.

La probabilidad de señalar correctamente cual es la muestra diferente, solo por azar, es $P_T = \frac{1}{3}$.

Si los jueces pueden discriminar, entonces la probabilidad de selección correcta es significativamente mayor que 1/3. En este caso, las hipótesis nula y alternativa se establecen como,

$$H_0\colon P_T = 1/3,\ \ H_1\colon P_T > 1/3.$$

Si la hipótesis nula no puede rechazarse, entonces:

1. La muestra impar, que se considera diferente, realmente no es diferente, o
2. Los jueces no son suficientemente discriminantes.

También podemos argumentar que los jueces dejan de ser discriminantes cuando las muestras son casi iguales

Prueba triangular

Diseño experimental para la prueba triangular

Sean, A: muestra estándar o control y
B: la muestra de prueba o tratamiento.
Se usan 6 arreglos de las tres muestran que forman el diseño básico.

El diseño básico se repite para tener replicas adicionales y balance en el efecto orden.

Para balancear las muestras A y B se designan como impares en el diseño en un número igual de veces.

La muestra A aparece en una posición particular el mismo número de veces que la muestra B.

Los panelistas se asignan aleatoriamente a los diferentes ordenes de presentación de las muestras.

Prueba triangular

Existe evidencia experimental de que en las pruebas triangulares, los resultados son sesgados por la designación de la muestra impar.

Grim y Goldblyth (1965). Usaron datos de perdida de sabor en un estudio de irradiación de un producto y observaron que los panelistas que recibieron dos muestras irradiadas y la muestra control como la impar, fueron quienes menos pudieron identificar correctamente la muestra impar, que aquellos que recibieron la muestra irradiada como la impar.

Wasserman y Talley (1969). Notaron que un juez podía determinar una diferencia significativa entre salchichas ahumadas y no ahumadas, cuando la salchicha ahumada se designaba como la muestra impar. En cambio era menos probable detectar diferencias significativas para un juez cuando la salchicha no ahumada era la que se designaba como impar.

Prueba triangular

El siguiente diseño toma en cuenta estos sesgos, así que, es balanceado con respecto a la designación de las muestras pares e impares.

Tabla 1. Diseño básico para una prueba triangular.

Panelistas	Muestra diferente	Ordenes	Codigo de tres dígitos
1	A	ABB	101, 475, 178
2		BAB	946, 557, 721
3		BBA	453, 412, 965
4	B	BAA	876, 633, 798
5		ABA	321, 613, 197
6		AAB	154, 995, 668

Prueba Triangular

Nombre______________________________________**Fecha**_____ **Tipo de muestra**____ ____		
Instrucciones: • Probar las muestras de izquierda a derecha. Dos son idénticas; determinar cual es la impar. • Si no observa diferencias proceda al azar.		
Conjunto de tres muestras	Cuál es la muestra impar?	Comentarios
___ ___ ___		
___ ___ ___		
___ ___ ___		

Figura 3. Formato para la prueba triangular.

Prueba triangular

Consideraciones al usar la prueba triangular en evaluación sensorial

- Es importante que los miembros del panel tengan claro entendimiento de los criterios de la prueba.
- El juicio debería basarse solo en la característica o respuesta estudiada. También, las muestras iguales deberían ser homogéneas.
- Si en una prueba triangular, las diferencias entre muestras iguales es mayor que la diferencia a ser detectada entre las muestras impares e iguales, entonces la hipótesis nula de no diferencia tiende a ser aceptada erróneamente (falsamente).
- Sean X_1 y X_2 las respuestas psicológicas generadas por las muestras iguales y Y la respuesta evocada por la muestra impar. Se obtiene una respuesta correcta en la prueba triangular si tenemos las siguientes respuestas:

$$|X_1 - X_2| < |X_1 - Y| \text{ y } |X_1 - X_2| < |X_2 - Y|.$$

Prueba triangular

Cuando los productos a ser evaluados tienen una gran variación de muestra a muestra, las relaciones anteriores pueden no cumplirse, y la prueba triangular puede ya no aplicar y debería usarse con discreción (reservas).

Los supuestos estadísticos importantes en que se basa la prueba triangular son, que los panelistas hacen sus selecciones independientemente uno de otro.

Prueba triangular

Prueba de hipótesis

Para investigar la existencia de una diferencia sensorial en dos tipos de productos utilizamos pruebas de hipótesis.

Las hipótesis nula y alternativa son:

$$H_0: p_c = 1/3 \text{ contra } H_1: p_c > 1/3$$

La hipótesis nula establece que no se percibe una diferencia sensorial entre los productos y por lo tanto, la selección correcta del producto diferente es aleatoria y con esto tenemos una probabilidad de selección correcta $p_c = 1/3$.

En la hipótesis alternativa tenemos que si se percibe una diferencia sensorial entre los productos y por lo tanto la probabilidad de selección correcta es $p_c > 1/3$.

Prueba triangular

Prueba de hipótesis

En esta prueba tenemos un experimento binomial:

a. El experimento consiste en n selecciones del producto diferente.
b. Cada respuesta es binaria, ya que es correcta o incorrecta.
c. Los ensayos (selecciones) son independientes.
d. La probabilidad de selección correcta p_c, permanece constante durante todas las selecciones.

La estadística de prueba es el número de respuestas correctas, X.

La distribución de X es una distribución Binomial(n, p_c).

Tenemos evidencia en contra de H_0 cuando el número de respuestas correctas es grande.

Tomamos una probabilidad α de rechazar erróneamente H_0.

Prueba triangular

El valor crítico c para esta prueba satisface lo siguiente,

Tenemos evidencia a en contra de H_0 (a favor de H_1) cuando el número de respuestas correctas x satisface,

$$P(X \geq x) = \sum_{i=x}^{n} \binom{n}{i} {p_0}^i (1 - p_0)^{n-i} = 1 - \sum_{i=0}^{(x-1)} \binom{n}{i} {p_0}^i (1 - p_0)^{n-i} \leq \alpha$$

La probabilidad de respuesta correcta en esta prueba es, $p_c = 1/3$.

Esta probabilidad se obtiene fácilmente utilizando el lenguaje de computo estadístico R:

$$P(X \geq x) = 1 - pbinom(x - 1, n, p_0)$$

Prueba triangular

Ejemplo.
Se llevó a cabo un estudio para evaluar el sabor de salchichas preparadas con y sin nitrato de sodio. Para eliminar los diferentes incidentes debidos al color, a los panelistas se les cubrieron los ojos. En la Tabla 2 se muestran los resultados.

¿Existe una diferencia de sabor detectable entre los dos tipos de salchicha?

En este caso, $x = 4,\ n = 12$ y $p_0 = 1/3$.

Utilizamos el software R.

```
> 1-pbinom(3,12,(1/3))
[1] 0.6069253
```

Como esta probabilidad es grande, no tenemos evidencia en contra de H_0 con niveles de significancia $\alpha = .01, .05, .10$. Concluimos así, que no hay una diferencia de sabor detectable entre los dos tipos de salchicha.

Tabla 2. Diseño de una prueba triangular para evaluar el sabor de salchichas con y sin nitrito de sodio.

Panelista	Muestra impar	Orden	Selección correcta	Selección incorrecta
1	A (nitrito)	ABB	A	--
2		BAB	--	B
3		BBA	A	--
4	B (no nitrito)	BAA	--	A
5		ABA	--	A
6		AAB	B	--
7	A (nitrito)	ABB	--	B
8		BAB	--	B
9		BBA	A	--
10	B (no nitrito)	BAA	--	A
11		ABA	--	A
12		AAB	--	A
Total			x=4	8

Prueba triangular-Ejemplos

Ejemplo. **Percepción de la diferencia de dos tipos de pan**. El responsable del empaque de un cierto tipo de pan, desea probar la efectividad de un nuevo material de empaque contra el plástico que actualmente se usa para elaborar los bolsas del pan.

El responsable del proceso de empaque piensa que si puede probar una diferencia significativa a las tres semanas después del empaque, el puede justificar el cambio de material para el empaque del producto.

La muestra de pan tipo A corresponde al que tiene la envoltura tradicional y el tipo B es el que tiene la envoltura nueva.

En esta prueba triangular, el panel consta de 30 jueces, que se programan en grupos de seis para para asegurar una aleatorización completa dentro de los grupos.

La distribución de los conjuntos de tres muestras de pan se hace de acuerdo a la tabla siguiente:

Panelista	Orden
1, 7, 13, 19, 25, 31	ABB
2, 8, 14, 20, 26, 32	BAB
3, 9, 15, 21, 27, 33	BBA
4, 10, 16, 22, 28, 34	BAA
5, 11, 17, 23, 29, 35	ABA
6, 12, 18, 24, 30, 36	AAB

Prueba triangular-Ejemplos

Ejemplo. Percepción de la diferencia de dos tipos de pan. (continuación)

Se lleva a cabo la prueba y al final, después de tres semanas que dura el producto almacenado, se tiene que de los 30 participantes, 17 identificaron correctamente la muestra diferente.

Para calcular la significancia de este resultado, calculamos la probabilidad (utilizando R) de tener un resultado igual o mas extremo, esto es,

$$P(X \geq 17) = 1 - P(X \leq 16) = 1 - pbinom(16{,}30, (1/3)) \ = 0.00722$$

Como esta probabilidad es menor que .01, decimos que la diferencia es significativa con un riesgo α del 1%. Concluimos que hay una diferencia sensorial detectable entre los dos productos.

Prueba Dúo-Trío

Prueba Dúo-trío

La prueba Dúo-trío desarrollada por Peryam & Swartz (1950) es similar a la Triangular, excepto que una de las dos muestras idénticas es conocida por el panelista.

Procedimiento:

Se preparan tres muestras que se codifican adecuadamente, dos de las muestras son idénticas y la otra diferente (impar).

Una de las muestras idénticas se presenta inicialmente para su evaluación sensorial, y se define como la muestra de referencia.

Las dos muestras restantes se presentan en orden balanceado y se pide al panelista que elija la muestra idéntica a la referencia.

Al igual que la prueba triangular esta prueba permite identificar si hay diferencia sensorial entre dos productos, pero no indica en qué atributo difieren.

Planteamiento de la hipótesis

Bajo el supuesto de que los sujetos están homogéneamente entrenados y que no hay diferencia entre los dos tipos de productos comparados (es decir, que el sujeto termina eligiendo al azar la respuesta), se espera que la probabilidad de que el sujeto elija de forma correcta la respuesta sea 1/2, mientras que si los productos muestran una diferencia lo esperado es que tal probabilidad sea mayor que 1/2:

$$H_0\colon p_c = 1/2 \text{ contra } H_0\colon p_c > 1/2$$

En este caso, las hipótesis nula y alternativa se pueden entender de la siguiente forma:

H_0: Los dos productos no son perceptiblemente diferentes.

H_1: Los dos productos son perceptiblemente diferentes.

Análisis de datos

En esta prueba el análisis de datos es similar al realizado en el caso de la prueba Triangular, solo que ahora la probabilidad de respuesta correcta debida solo al azar es 1/2 .

Dos modalidades para la prueba

- Referencia constante
  - La misma muestra, usualmente de la producción habitual, es siempre la muestra referencia.
  - Se usa con sujetos entrenados y cuando uno de los productos es bien conocido por estos el cual se usa como referencia.
- Referencia balanceada
  - Ambas muestras que están siendo comparadas se usan aleatoriamente como referencia.
  - Se usa cuando ambas muestras son desconocidos o cuando se tienen sujetos no entrenados.

Procedimiento de la prueba Dúo-trío, considerando la modalidad

- Preparar igual número de muestras de cada una de las posibles combinaciones.
  a) Con referencia constante todos los sujetos reciben la misma muestra referencia R_A (usando como referencia el producto A) dando así dos combinaciones posibles (R_AAB, R_ABA).
  b) Con referencia balanceada la mitad de los sujetos reciben una muestra como referencia y la otra mitad reciben la otra, obteniendo así cuatro secuencias de presentación (R_AAB, R_ABA, R_BAB, R_BBA, donde R_A y R_B en 4 combinaciones anteriores denotan que las referencias son A y B respectivamente).

- Asignar en forma aleatoria estas combinaciones a los sujetos (una muestra identificada como referencia y dos muestras codificadas, una de las cuales es idéntica a la muestra de referencia).
- A cada sujeto se le pide que prueben (sientan, examinen) cada muestra de izquierda a derecha e indique cuál muestra codificada es similar a la muestra referencia.
- Contar el número de respuestas correctas, analizar e interpretar el resultado.

Prueba Dúo-Trío
Nombre:_______________________________________ **Fecha:**_______   **Tipo de muestra:**____
Instrucciones: Pruebe (sienta, examine) las muestras de izquierda a derecha. La muestra de la izquierda es la referencia. Determine cual de las dos muestras coincide con la de referencia e indique colocando una X.   Si no hay diferencia clara entre las dos muestras desconocidas debe adivinar.
Referencia Código _______ Código _______
Comentarios:____   ____

Figura 4. Hoja de respuesta para la prueba Dúo-trío

Ejemplo (Fragancia para cajas de pañuelos faciales)

- Se quiere determinar si dos métodos de administración de fragancia producen diferencias en la fragancia percibida de dos pañuelos después de haber estado almacenados por un periodo de tiempo comparable a la edad normal del producto al momento de uso.
- El primer método consiste en administrar la fragancia directamente a los pañuelos, el segundo en administrar la fragancia al interior de la caja.
- Se usará la prueba de referencia balanceada.
- La prueba se realiza con 40 sujetos, quienes tienen alguna experiencia en evaluación de olor.

Diseño de la prueba

- Cada uno de los dos tipos de pañuelos (uno correspondiente al método 1 y el otro al 2) se usaron como como referencia en la mitad (20) de las 40 evaluaciones hechas.
- Después de asignar aleatoriamente las cuarenta combinaciones a los 40 sujetos, se obtuvo 21 de éstos eligieron de forma correcta la referencia designada.

Cálculos con R:

```
> 1-pbinom(20,40,(1/2))
[1] 0.4373147
```

Como $P(X \geq 21) = 1 - P(X \leq 20) = 0.437$

Los datos no muestran evidencia en contra de la hipótesis nula.

Se concluye que los dos métodos de administración de fragancias no son perceptiblemente diferentes en cuanto a su fragancia, con un 5% de significancia.

Pruebas de Elección Forzada

Métodos de elección forzada

En las pruebas anteriores se busca establecer si hay o no diferencia entre dos muestras, sin considerar la razón por la cual existe tal diferencia.

Cuando se identifica un atributo o característica como la fuente de posible diferencia (como dulzor, amargor, acidez, etc.), podemos realizar una prueba de diferencia, en base a la característica que origina la diferencia.

Las pruebas de diferencia que indican los atributos a evaluar, como la 2-AFC y 3-AFC, son más potentes que aquellas que no los especifican, como la Dúo-Trio y la Triangular.

Cuando se trabaja con jueces no entrenados en detectar atributos específicos, como consumidores, es comprensible el uso de pruebas más simples.

Método de elección forzada con dos alternativas (2-AFC)

Este método que también denotamos por (2-AFC), por sus siglas en inglés (2-Alternative Forced-Choice), (Green and Swets,1966) se emplea cuando una característica sensorial particular difiere entre dos muestras, (por ejemplo, cual muestra es más dulce, o más amarga)

En este método, el panelista recibe un par de muestras codificadas, A y B, para su comparación en base a una característica sensorial especificada. Las posibles parejas son: AB y BA.

Al panelista se le pide que seleccione la muestra con la característica sensorial más fuerte (o más débil). El panelista debe señalar una muestra, aún si no puede detectar la diferencia, en este caso lo hace al azar.
Las normas ISO 5495-1983 y ASTM E2164-2001describen esta prueba.

A diferencia de la prueba Dúo-Trío, esta prueba no tiene una muestra de referencia.

Método de elección forzada con dos alternativas (2-AFC)

En esta prueba, la probabilidad de selección correcta, solo por azar, es 1/2, por lo cual, la hipótesis nula es,

$$H_0\colon p_c = 1/2.$$

En este caso podemos tener una hipótesis alternativa unilateral o también bilateral. El tipo de alternativa depende del propósito del experimento.

Por ejemplo, si en una prueba de dulzura de dos productos (productos tradicional y nuevo) sabemos de antemano que el producto nuevo contiene más azúcar. Entonces elegimos la alternativa unilateral, porque solo es de interés una diferencia en un sentido. En este caso tenemos,

$$H_1\colon p_c > 1/2.$$

En cambio, si en esta prueba, a los panelistas se les pide que expresen su preferencia por una muestra sobre otra, no sabemos previamente cual producto es más popular. Aquí debemos elegir una alternativa bilateral,

$$H_1\colon p_c \neq 1/2.$$

Método de elección forzada con dos alternativas (2-AFC)

La decisión de utilizar una alternativa unilateral o bilateral debe hacerse antes del experimento.

Método de elección forzada con dos alternativas (2-AFC)

Ejemplo. Alternativa bilateral (Diferencia direccional)

Una investigación de consumidores de un tipo de limonada, indica que estos tienen preferencia por un sabor de "limón natural" . El productor de la bebida, ha desarrollado dos tipos de jarabe A y B para preparar la limonada y desea saber cual de los dos jarabes tiene un sabor más parecido a limón natural.

Como diferentes personas pueden tener diferente una percepción diferente del sabor a limón natural, entonces se elige un panel grande. Se lleva a cabo una prueba de comparación pareada (2-AFC) con 40 participantes y un error α del 5%. La hipótesis nula es H_0: Sabor natural de A= Sabor natural de B y la hipótesis alternativa es, H_1: Sabor natural de A$\neq$ Sabor natural de B. Ambos resultados son de interés, así que la prueba es de dos lados.

De los 40 participantes, 26 eligen la muestra B como la que tiene una sabor más natural. Cuatro sujetos reportan "No hay diferencia" y estos cuatro se dividen en ambas muestras, así que concluimos que 28 eligen la muestra B.

Método de elección forzada con dos alternativas (2-AFC)

Ejemplo. Alternativa bilateral (Diferencia direccional)
Tomamos como estadística de prueba X el número de participantes que eligen uno de los productos. Si el producto considerado es la muestra B, entonces $x = 28$.

Como en el experimento se tienen 40 participantes, calculamos la probabilidad de tener un resultado como $X = 28$ o más extremo, esto es,

$$P(X \geq 28) = 1 - P(X \leq 27)$$

Utilizando R tenemos:

```
> 1-pbinom(27,40,(1/2))
[1] 0.008294502
>
```

Como esta probabilidad es menor que $\alpha/2 = .025$, decimos que los datos tienen evidencia en contra de la hipótesis nula y concluimos que la formulación B debería emplearse, ya que de acuerdo al panel, es la que tiene un sabor más natural.

Método de elección forzada con dos alternativas (2-AFC)

Ejemplo. Alternativa unilateral
Un fabricante de cerveza recibe el reporte de que su cerveza (A) le perece poco amarga a los consumidores, así que elabora otro tipo de cerveza (B), con un sabor ligeramente más amargo.

Para comparar los dos tipos de cerveza se realiza una prueba de diferencia (2-AFC), para saber si se ha logrado un incremento en el sabor amargo. Se elige un riesgo α de 1%.

Se lleva a cabo la prueba con un panel de 30 jueces que tienen una capacidad probada para detectar cambios pequeños en el sabor amargo.

La hipótesis nula es, H_0: Amargor de A=Amargor de B, y la alternativa en este caso es, H_1: Amargor de B>Amargor de A, así que la prueba es de un lado.

De los 30 jueces que prueban las muestras, 22 eligen la muestra B como la más amarga.

Método de elección forzada con dos alternativas (2-AFC)

Ejemplo. Alternativa unilateral
Calculamos la probabilidad de tener un resultado como $X = 22$ o más extremo, esto es,

$$P(X \geq 22) = 1 - P(X \leq 21)$$

Con R obtenemos,

```
> 1-pbinom(21,30,(1/2))
[1] 0.008062401
>
```

Como esta probabilidad es menor que $\alpha = 0.01$, podemos decir que la diferencia del sabor amargo de la nueva cerveza se percibió con una significancia $\alpha = .01$.

Método de elección forzada con dos alternativas (2-AFC)

Comentario: Para decidir si una prueba de comparación pareada (2-AFC) es de uno o dos lados, nos fijamos en el tipo de hipótesis alternativa, esto es, si es de uno o dos lados. Las pruebas unilaterales ocurren cuando el objetivo de la prueba es confirmar una mejora específica o un efecto de tratamiento.

Algunos ejemplos de pruebas de uno y dos lados son los siguientes:

Objetivos en pruebas unilaterales:

Confirmar que la cerveza de prueba es más amarga.
Confirmar que el producto de prueba es el preferido.

Objetivos en pruebas bilaterales:

Decidir cual cerveza es más amarga.
Decidir Cual producto es el preferido.
En situaciones de pruebas donde la hipótesis alternativa es que las muestras son diferentes, en lugar de "Una es más que la otra"

Método de elección forzada con tres alternativas (3-AFC)

En este método, a cada panelista se presentan tres muestras de dos productos A y B. Dos de ellas son iguales, los posibles conjuntos de muestras son AAB, ABA, BAA, ABB, BAB, BBA.

El estimulo de A es más fuerte que el de B. Al panelista se le presenta uno de los conjuntos ABB, BAB o BBA con la instrucción de seleccionar el estimulo más fuerte de los tres. O también ocurre que al panelista se le presentan los conjuntos AAB, ABA o BAA con la instrucción de seleccionar el estímulo más débil de los tres.

El panelista tiene que seleccionar una muestra aunque no pueda identificar una que tenga el valor mas alto o bajo.

El diseño de la prueba 3-AFC es igual al de la prueba Triangular usual, pero las instrucciones son diferentes (Ver la siguiente figura). En un caso la instrucción es selecciona la muestra “diferente” y en el otro selecciona la muestra “con el nivel más alto”.

Método de elección forzada con tres alternativas (3-AFC)

Nombre____ Fecha________________
Esta es una Prueba 3-AFC. Debe hacer una selección.
Una de las tres muestras es la más dulce (amarga, etc.).
Escriba el número (código) de la muestra.___________________________

Comentarios: ____
____

Figura. Cuestionario para la prueba 3-AFC.

Las diferentes instrucciones pueden afectar profundamente el desempeño del panelista y la potencia de la prueba. Para una intensidad sensorial, esto es una magnitud percibida de un estimulo, la proporción de respuestas correctas en una prueba 3-AFC será mayor que en una Triangular.

Método de elección forzada con tres alternativas (3-AFC)

La estadística de prueba en la prueba 3-AFC es el número X de selecciones correctas en las n comparaciones. Si todas las selecciones se hacen en forma independiente, entonces la estadística de prueba sigue una distribución binomial, con parámetros p_c y n. La hipótesis nula es,

$$H_0: p_c = 1/3,$$

y la hipótesis alternativa es,

$$H_1: p_c > 1/3.$$

Método de elección forzada con tres alternativas (3-AFC)

La probabilidad de respuestas correctas en una hipótesis nula es 1/3 porque, si las muestras A y B tienen la misma intensidad sensorial, la probabilidad de seleccionar A es “mas fuerte” para los conjuntos ABB, BAB y BBA o seleccionar B es “mas debil” para los conjuntos BAA, ABA y AAB, es simplemente la probabilidad de respuesta correcta por azar, 1/3.

La prueba siempre es unilateral porque la probabilidad de respuestas correctas no puede ser menor que la probabilidad de selección correcta por azar 1/3.

Método de elección forzada con tres alternativas (3-AFC)

Ejemplo.
El fabricante de una bebida con azúcar, recientemente ha producido una nueva bebida reducida en azúcar y quiere investigar si los consumidores podrían detectar la diferencia en lo dulce entre las dos bebidas . Se conduce una prueba 3-AFC con 100 panelistas elegidos al azar y con un nivel de significancia de .05.

Cuarenta y seis panelistas de los 100 seleccionaron correctamente el nuevo producto como la bebida menos dulce.

Calculamos en R la probabilidad de tener un resultado como $X = 46$ o más extremo,

```
> 1-pbinom(45,100,(1/3))
[1] 0.005709371
>
```

Como esta probabilidad es menor que $\alpha = .05$, concluimos con un nivel de significancia $\alpha = .05$, que los consumidores perciben la nueva bebida como menos dulce que la original.

Fin de Primera Parte

Potencia de las pruebas

Pruebas de hipótesis estadísticas

Hipótesis sobre la proporción de una distribución binomial

Hipótesis nula, $H_0 \colon p = p_0$

Hipótesis alternativa: (a) $H_0 \colon p \neq p_0$, (b) $H_1 \colon p > p_0$, : (c) $H_1 \colon p < \mu_0$

Estadística de prueba: X= numero de aciertos (o fallas)

Región de rechazo: $\mathrm{X} \geq x_R$

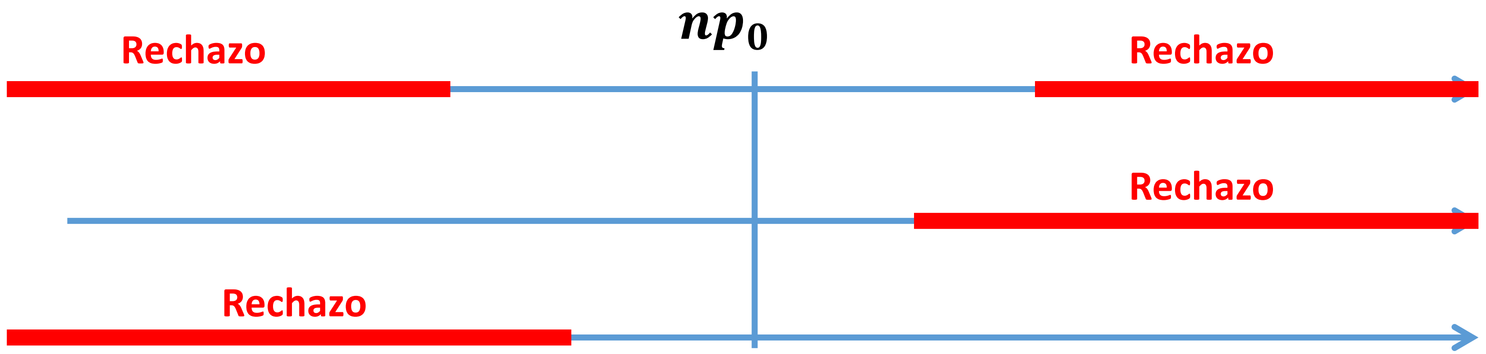
Bajo la hipótesis H_0, la estadística de prueba tiene una distribución Binomial(n, p_0).

La región de rechazo es diferente en las distintas hipótesis alternativas, como se muestra enseguida:

Con la alternativa (a)

Con la alternativa (b)

Con la alternativa (c)



Pruebas de hipótesis estadísticas

Después de que se han formulado las hipótesis nula y alternativa, se construye una estadística de prueba (muestral), que se usa de acuerdo a un criterio de decisión para rechazar o no rechazar la hipótesis nula.

Con frecuencia la estadística de prueba es una expresión que mide la discrepancia entre la hipótesis nula y los datos. En este caso un valor grande de la estadística muestra evidencia en contra de la hipótesis nula.

¿Podemos tomar otra estadística de prueba?

Como la regla de decisión es estadística, esto es, se basa en una muestra, esta debe reflejar los errores y la variación muestral. Por lo tanto, el rechazo o la aceptación de la hipótesis nula están sujetos a error. Tenemos dos tipos de error:

- Error tipo I. Se tiene cuando rechazamos un hipótesis nula que es correcta.
- Error Tipo II. Se tiene cuando aceptamos una hipótesis nula que es falsa.

Pruebas de hipótesis estadísticas

Las probabilidades de estos errores son las siguientes:

α =Probabilidad de rechazar H_0 cuando es cierta
=$P(Rechazar\ H_0|H_0\ es\ cierta)$

β =Probabilidad de aceptar H_0 cuando es falsa
= $P(Aceptar\ H_0|H_0\ es\ falsa)$

Nos gustaría tener pruebas de hipótesis con probabilidades de error α y β pequeños. En general esto se puede lograr con muestras grandes.

Si un error Tipo I es mas serio que un error Tipo II, queremos una estadística de prueba con un valor de α pequeño y controlamos β tanto como sea posible.

Si ambos tipos de error son igualmente serios, entonces queremos tener valores de α y β muy cercanos y lo más pequeños que se pueda.

Pruebas de hipótesis estadísticas

Usualmente el tamaño de α se fija por el investigador durante la planeación del experimento. Luego se propone una estadística de prueba con el nivel especificado de α y al mismo tiempo, mantenga la probabilidad del error Tipo II a un nivel mínimo para todos los puntos de la hipótesis alternativa.

Si H_0 es cierta, la probabilidad de selección correcta (que en este caso es elegir H_0) es el nivel de confianza $(1-\alpha)$. De manera similar, si H_0 es falsa, la probabilidad de elección correcta (rechazar H_0) esta dada por **la potencia de la prueba** $(1-\beta)$.

La potencia de una prueba de hipótesis determina la capacidad de la prueba para detectar diferencias cuando las hay. Una prueba con una potencia alta puede detectar diferencias, aunque estas sean pequeñas. En cambio una prueba con una potencia reducida detectara diferencias, solo si son diferencias muy evidentes por su magnitud.

Pruebas de hipótesis estadísticas

Una herramienta de gran utilidad para juzgar la calidad de una prueba de hipótesis es la función potencia. Esta función se evalúa en cada punto del espacio paramétrico y su valor es la probabilidad de rechazar la hipótesis nula, cuando el valor del parámetro tiene un valor θ determinado, esto es,

$$p(\theta) = P(\text{Rechazar } H_0 | \text{el parametro es } \theta).$$

Cuando tenemos una prueba en donde las hipótesis nula y alternativa se refieren a la media de una distribución normal, entonces el espacio paramétrico es todo el conjunto de los números reales.

En cambio si tenemos una prueba relacionada con el número de aciertos de un panel de jueces entrenados, el parámetro es la proporción aciertos del panel. En este caso el espacio paramétrico es el intervalo [0,1], ya que una proporción no puede ser mayor que 1 ni menor que 0.

En la siguiente figura, se muestran tres funciones potencia (para n=5, 10 y 100) de la prueba de hipótesis sobre la proporción de una distribución binomial, donde las hipótesis nula y alternativa son respectivamente $H_0 : p = p_0$ y $H_1 : p > p_0$. En este ejemplo tomamos una proporción $p = .6$.

Pruebas de hipótesis estadísticas

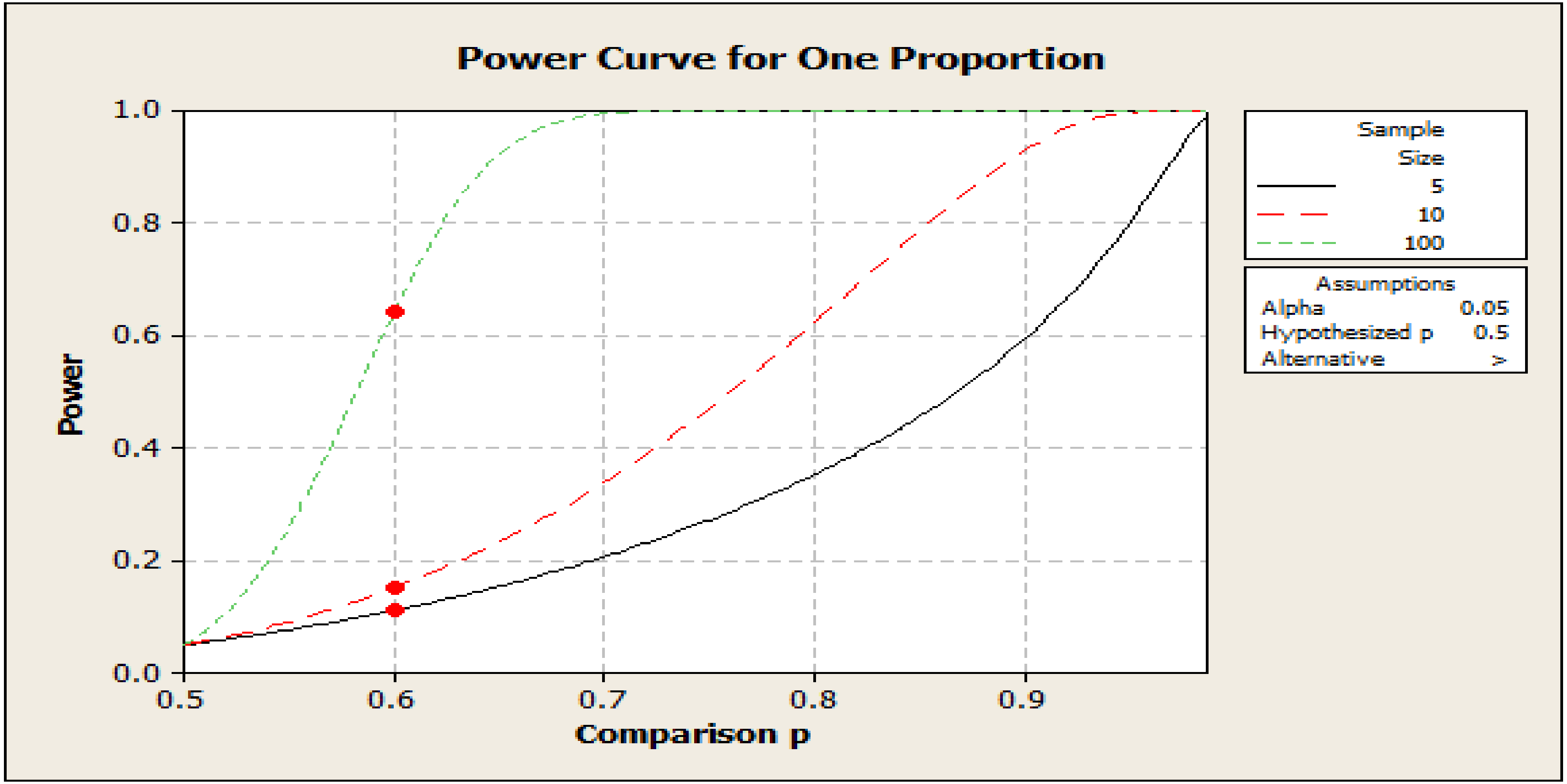


Figura 2. Funciones potencia de una prueba de hipótesis sobre la proporción de una distribución binomial, donde la hipótesis alternativa es unilateral derecha, $H_0\colon p > p_0$. Las 3 funciones corresponden a diferentes tamaños de muestra.

Pruebas de hipótesis estadísticas

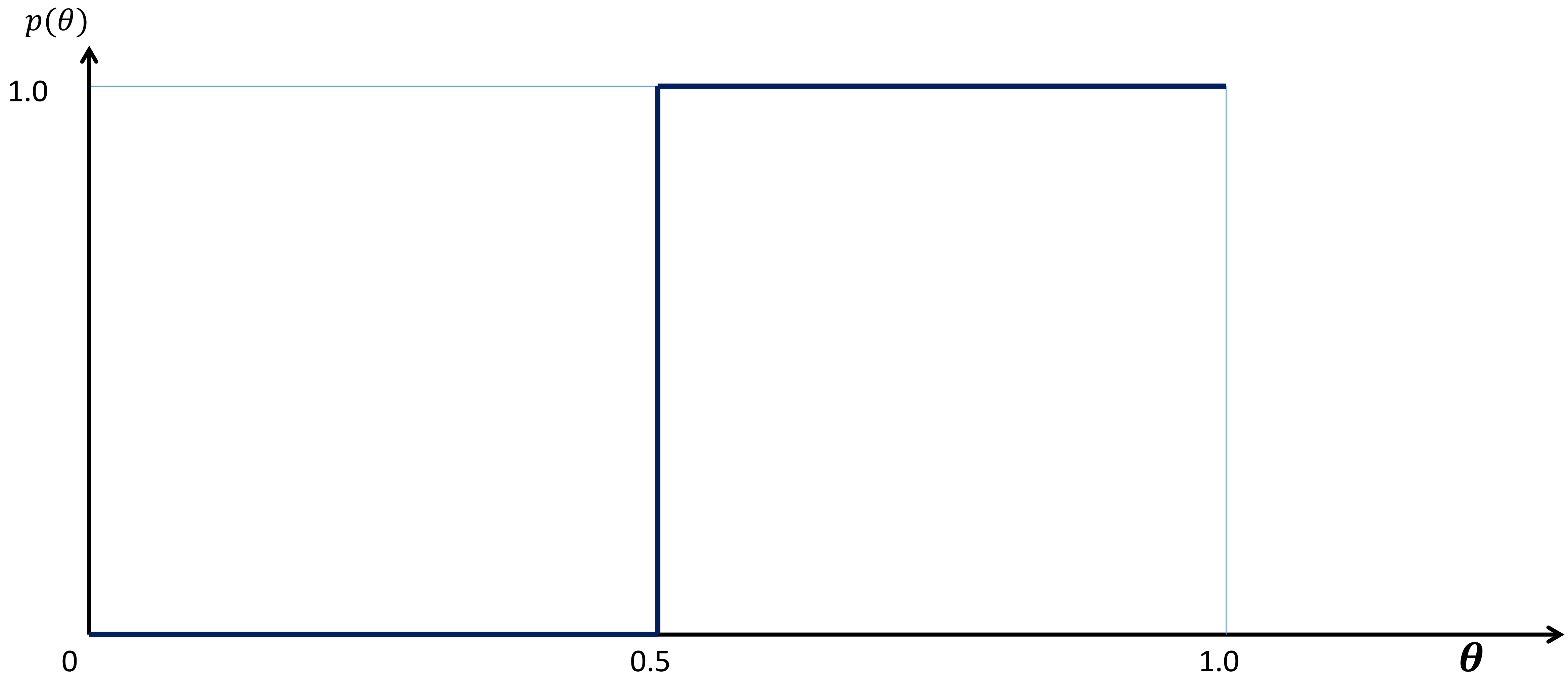


Figura. Función potencia ideal, cuando la hipótesis nula a probar es que la proporción de aciertos es .5 (o menos), contra la alternativa que plantea una proporción de aciertos mayor que .5.

Análisis de la potencia de las pruebas

Potencia de prueba:

$$Potencia = 1 - \beta = P\left(\frac{\hat{p} - p_0}{\frac{\sigma_0}{\sqrt{n}}} > z_{1-\alpha} | H_1 es\ cierta\right),$$

donde p_0 es la probabilidad de selección correcta en la hipótesis nula H_0 y $\sigma_0 = \sqrt{p_0(1 - p_0)}$.

Si $p = p_1$ cuando H_1 es cierta, entonces:

$$Potencia = P(\hat{p} > z_{1-\alpha}\, \sigma_0/\sqrt{n} + p_0 | p = p_1)$$

$$= P\left(\frac{\hat{p} - p_1}{\sigma_1/\sqrt{n}} > \frac{z_{1-\alpha}\sigma_0\ /\ \sqrt{\mathrm{n}} + p_0 - p_1}{\sigma_1/\sqrt{n}}\, | p = p_1\right)$$

$$= P(Z > \lambda) = 1 - \Phi(\lambda),$$

donde, $\lambda = \frac{z_{1-\alpha}\sigma_0/\sqrt{\mathrm{n}} + p_0 - p_1}{\sigma_1/\sqrt{n}}$ y $\sigma_1 = \sqrt{p_1(1 - p_1)}$.

Análisis de la potencia de las pruebas

Tamaño muestral:

Si $1-\beta = P(Z>\lambda) = 1-\Phi(\lambda)$, entonces, $\beta = \Phi(\lambda)$, así, $z_\beta = \lambda$.

Sustituyendo λ tenemos, $z_\beta = \frac{z_{1-\alpha}\sigma_0/\sqrt{\mathrm{n}}+p_0-p_1}{\sigma_1/\sqrt{n}}$

Despejamos el tamaño muestral,

$$n = \left(\frac{z_{1-\alpha}\sigma_0 - z_\beta\sigma_1}{p_1-p_0}\right)^2 = \left(\frac{z_{1-\alpha}\sigma_0 + z_{1-\beta}\sigma_1}{p_1-p_0}\right)^2.$$

Cuando crece la discrepancia p_1-p_0 entre las probabilidades de selección correcta, tenemos que:

- Aumenta la potencia $1-\beta$
- Disminuye el tamaño muestral n.

Análisis de la potencia de las pruebas

Ejemplo.
Determinar la potencia de una prueba 2-AFC (con $\alpha = .05$) cuando se tiene un panel de 25 participantes y se quiere detectar una probabilidad de respuesta correcta $p_1 = .75$.

$$Potencia = P(Z > \lambda) = 1 - \Phi(\lambda), \qquad \lambda = \frac{z_{1-\alpha}\sigma_0/\sqrt{\mathrm{n}} + p_0 - p_1}{\sigma_1/\sqrt{n}}$$

$$z_{.05} = 1.64\ , p_1 = .75.\ , \quad \sigma_1 = \sqrt{.75(1-.5)} = .433,$$

$$\lambda = \frac{z_{1-\alpha}\sigma_0/\sqrt{\mathrm{n}} + p_0 - p_1}{\sigma_1/\sqrt{n}} = \frac{1.64\times\frac{.5}{\sqrt{25}} + .5 - .75}{.433/\sqrt{25}} = -.993$$

Potencia=$1 - \Phi(3.88) = 1 - .18 = .84$

Si la probabilidad real de respuesta correcta fuera $p_1 = .75$, esta prueba tendría una probabilidad de .84 de detectarla. Cuando crece la discrepancia.

Esta potencia corresponde tanto a la prueba 2-AFC, como a la prueba Dúo-Trío.

Análisis de la potencia de las pruebas

Ejemplo.
Determinar el número necesario de jueces para llevar a cabo una prueba Dúo-Trío, con una probabilidad de error tipo I, $\alpha = .05$, y además requerimos que si la probabilidad de respuesta correcta es $p_c = .70$, esta sea detectada con una probabilidad de 90%.

En este problema tenemos: $z_{.95} = 1.645,\ z_{.10} = -1.282, \sigma_0 = .5,\ \sigma_1 = \sqrt{.7 \times .3} = .4583$

$$n = \left(\frac{z_{1-\alpha}\sigma_0 - z_\beta \sigma_1}{p_1 - p_0}\right)^2 = \left(\frac{z_{1-\alpha}\sigma_0 + z_{1-\beta}\sigma_1}{p_1 - p_0}\right)^2 =$$

$$= \left(\frac{1.645 \times .5 - (-1.282).4583}{.70 - .50}\right)^2 = (7.05)^2 = 49.7 \cong 50.$$

En esta prueba se requiere un panel formado por 50 jueces.

Pregunta: ¿Si solo se cuenta con 40 jueces, con que probabilidad se detectará la probabilidad de respuesta correcta, $p_c = .70$?

Análisis de la potencia de las pruebas

Ejercicio.
Si tenemos una prueba Dúo-Trío con una significancia $\alpha = .10$, con que probabilidad la prueba detectaría la probabilidad de una respuesta correcta $p_1 = .70$. Considere los siguientes tres casos:
Caso 1) El panel consta de 10 participantes.
Caso 2) El panel consta de 20 participantes.
Caso 3) El panel consta de 50 participantes.

Comparación de métodos

Algunos métodos (como, 2-AFC y Dúo-Trio) tienen el mismo modelo de prueba estadística, y en consecuencia, la misma potencia para la misma proporción (p_1) en la hipótesis alternativa, bajo las mismas condiciones para α y n. Esto también ocurre para la pareja 3-AFC y Triangular.

Sin embargo, la misma proporción refleja diferentes diferencias sensoriales en diferentes métodos de discriminación.

La proporción de respuestas correctas en una prueba de diferencias usando un método de elección forzada es un estadístico de prueba adecuado. Sin embargo no es un buen índice para medir para medir diferencias sensoriales, ya que no es independiente del método utilizado.

Para un mismo par de estímulos, las proporciones de respuesta correcta utilizando distintos métodos son diferentes.

La misma probabilidad de selección correcta usando diferentes métodos, refleja diferentes diferencias sensoriales.

Comparación de métodos

Un índice de diferencia sensorial (d) que ha resultado de gran utilidad en la comparación de pruebas de diferencia es el índice (d o δ) de Thurstone.

Modelo de Thurstone

Este modelo se basa en dos suposiciones:

- La variabilidad de la percepción sensorial
- El establecimiento de la estrategia cognitiva propia de cada prueba sensorial

Variabilidad de la percepción
Para ilustrar el significado, imaginamos la degustación de estímulos.

Al recibir un estimulo determinado, por la variabilidad en la respuesta de los sentidos, la intensidad de percepción tiene un rango y un patrón de variación, de tal forma que la percepción del estímulo puede representarse por una distribución normal.

Comparación de métodos

Según el modelo de Thurstone, las percepciones de dos estímulos diferentes se distribuyen como dos distribuciones normales con varianza común.

Entre más separadas estén las medias de las distribuciones de percepción, más fácil será distinguirlas.

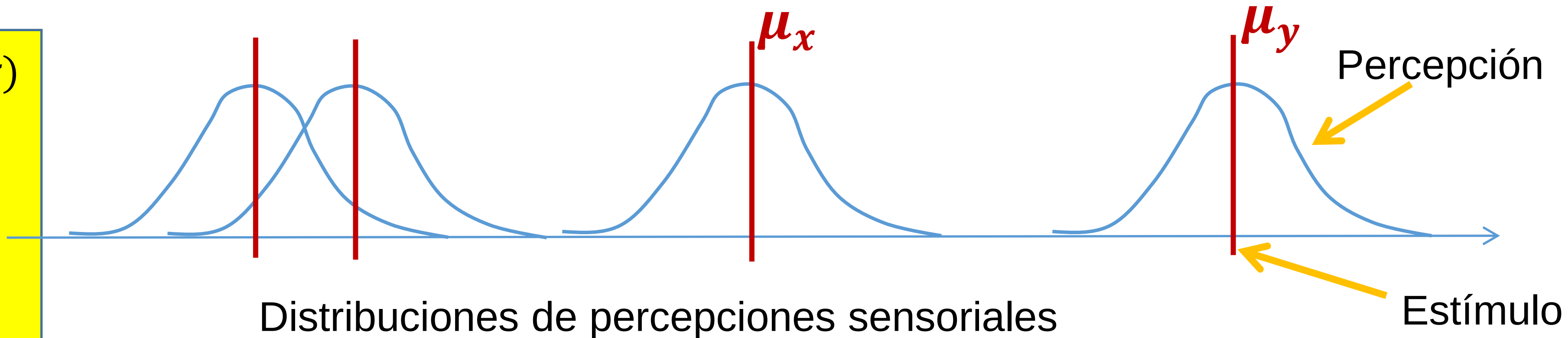
El nivel de separación de las distribuciones puede utilizarse como una medida de que tan bien los jueces discriminan entre los dos estímulos.

El parámetro d (o δ) se define como la distancia entre las medias medida en unidades de desviación estándar. Este parámetro es un índice de diferencia sensorial.

Si $\boldsymbol{X \sim N(\mu_x, \sigma)}$ **y** $\boldsymbol{Y \sim N(\mu_y, \sigma)}$

El índice de diferencia sensorial es:

$$\boldsymbol{d = \frac{\mu_y - \mu_x}{\sigma}}$$



Distribuciones de percepciones sensoriales

Análisis de la potencia de las pruebas

Potencia de prueba:

d = diferencia sensorial

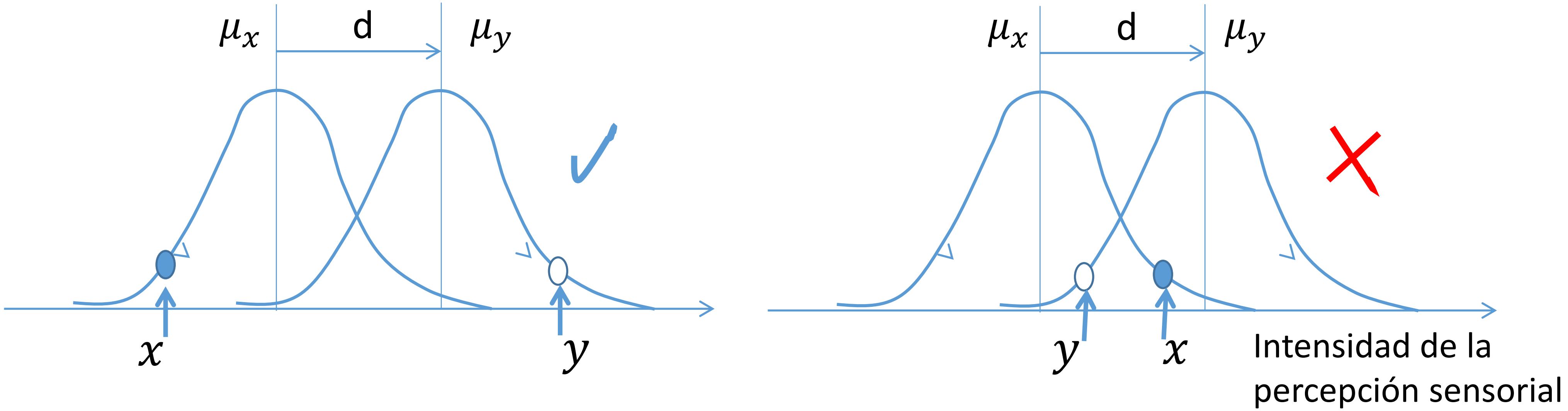


Figura. Las distribuciones representan los estímulos. Se ilustra la prueba de comparación por pares.

Probabilidad de respuesta correcta: $p_c = P(Y > X)$

Análisis de la potencia de las pruebas

Probabilidad de respuestas correctas (p_c) con la prueba 2-AFC:

Se tiene una respuesta correcta cuando Y>X y consideramos que $X{\sim}N(\mu_x, \sigma_x)$ y $Y{\sim}N(\mu_y, \sigma_y)$

Probabilidad de respuesta correcta:

$$p_c = P(Y > X) = \Phi\left((\mu_y - \mu_x)/\sqrt{\sigma_y{}^2 + \sigma_x{}^2}\right)$$

Tenemos que $\sigma_x = \sigma_y = \sigma$, así que:

$$p_c = \Phi\left((\mu_y - \mu_x)/\sqrt{2\sigma^2}\right) = \Phi\left([(\mu_y - \mu_x)/\sigma]/\sqrt{2}\right) = \Phi\left(d/\sqrt{2}\right)$$

En el caso de la prueba 2-AFC de diferencias, la proporción p_c de respuestas correctas y la diferencia sensorial d están relacionadas por la función:

$$p_c = \Phi\left(d/\sqrt{2}\right)$$

Análisis de la potencia de las pruebas

Calculo de la probabilidad de selección correcta por simulación.

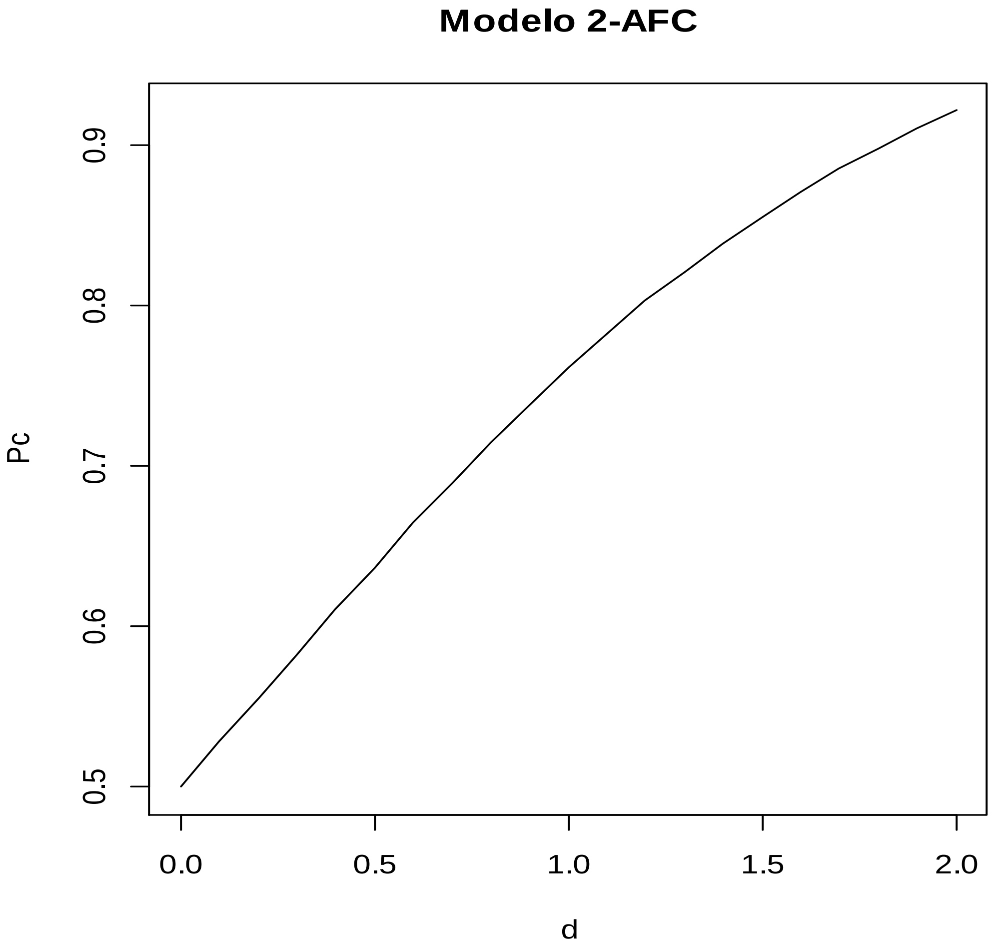
$$\hat{p}_c = \frac{Num.\, de\ respuestas\ correctas}{Num.\, de\ ensayos}$$

Consideramos dos estímulos X y Y que presentan un grado de discriminación d.

Simulamos un número grande de veces observaciones X y Y, con distribuciones N(0,1) y N($d, 1$) y calculamos el número promedio de respuestas correctas.

Calculo de la probabilidad de selección correcta por simulación.

```
#Calculo de la probabilidad de selección correcta
#Método 2-AFC
#Simulación de dos estimulos X y Y
#X con distribucion N(0,1)
#Y con distribución N(d,1)
#d=diferencia sensorial
#N= Numero de simulaciones
#
d<-1
N<-100000
CA<-0
i<-1
while(i<=N){
		x<-rnorm(1,0,1)
		y<-rnorm(1,d,1)
		if(x<y){CA<-CA+1}
i<-i+1
}
Pc<-CA/N
Pc
```



Relación entre la diferencia sensorial d y la probabilidad de selección correcta p_c para el modelo 2-AFC.

Análisis de la potencia de las pruebas

Cuando detectamos una diferencia sensorial, puede ser de interés poder estimar su magnitud. Un índice que mide la diferencia sensorial es la d de Thrustone. Este índice es no observable, pero puede estimarse a partir de los datos.

De la relación entre p_c y d tenemos que, $d = \sqrt{2}\Phi^{-1}(p_c)$

Un estimador de d es:

$$\boxed{\hat{d} = \sqrt{2}\Phi^{-1}(\hat{p}_c),}$$

donde $\hat{p}_c$ es un estimador de la probabilidad de respuestas correctas,

$$\hat{p}_c = \frac{Num.\, de\ respuestas\ correctas}{Num.\, de\ ensayos}$$

Para hacer inferencia sobre d es necesario estimar la varianza de $\hat{d}$.

Análisis de la potencia de las pruebas

Varianza de $\hat{d}$

Consideramos una aproximación de Taylor para $f(\hat{d}) = \Phi(\hat{d}/\sqrt{2})$,

$$f(\hat{d}) \approx f(d) + f^{'}(d)(\hat{d} - d)$$

De aquí tenemos que,

$$\hat{p}_c - p_c = f(\hat{d}) - f(d) = f^{'}(d)(\hat{d} - d)$$

$$V(\hat{p}_c) = E(\hat{p}_c - p_c)^2 = [f^{'}(d)]^2 E(\hat{d} - d)^2 = [f^{'}(d)]^2 V(\hat{d})$$

Y como, $f^{'}(d) = \Phi^{'}(d/\sqrt{2})(1/\sqrt{2}) = \phi(d/\sqrt{2})(1/\sqrt{2})$

Entonces la varianza estimada de $\hat{d}$ es, $\boxed{\hat{V}(\hat{d}) = \frac{2\hat{p}_c(1-\hat{p}_c)}{n\phi^2(\hat{d}/\sqrt{2})} = \frac{2\hat{p}_c(1-\hat{p}_c)}{n\phi^2(\sqrt{2}\Phi^{-1}(\hat{p}_c)/\sqrt{2})}}$

Análisis de la potencia de las pruebas

Ejemplo. En una prueba 2-AFC, donde participaron 30 panelistas, se obtuvieron 19 respuestas correctas.

1. Encontrar la estimación $\hat{p}_c$ de la probabilidad de respuestas correctas p_c, así como también la varianza de la estimación.
2. Encontrar la estimación $\hat{d}$ del índice de diferencia sensorial d, así como también su varianza.

La proporción de respuestas correctas es una estimación de la probabilidad de respuestas correctas, $\hat{p}_c = \frac{21}{36} = 75$.

Estimamos la varianza de $\hat{p}_c$ por, $V(\hat{p}_c) = \frac{\hat{p}_c(1-\hat{p}_c)}{n} = \frac{.75(1-.75)}{36} = .0052$

Estimamos el índice d como, $\hat{d} = \sqrt{2}\Phi^{-1}(\hat{p}_c) = \sqrt{2}\Phi^{-1}(.75) = .954,$

Entonces la varianza estimada de $\hat{d}$ es,

$$\hat{V}(\hat{d}) = \frac{2\hat{p}_c(1-\hat{p}_c)}{n\phi^2(\hat{d}/\sqrt{2})} = \frac{2\hat{p}_c(1-\hat{p}_c)}{n\phi^2(\sqrt{2}\Phi^{-1}(\hat{p}_c)/\sqrt{2})} = \frac{2(.75)(.25)}{36\phi^2(\frac{.954}{\sqrt{2}})} = \frac{.375}{3.636} = .103.$$

Análisis de la potencia de las pruebas

d = diferencia sensorial

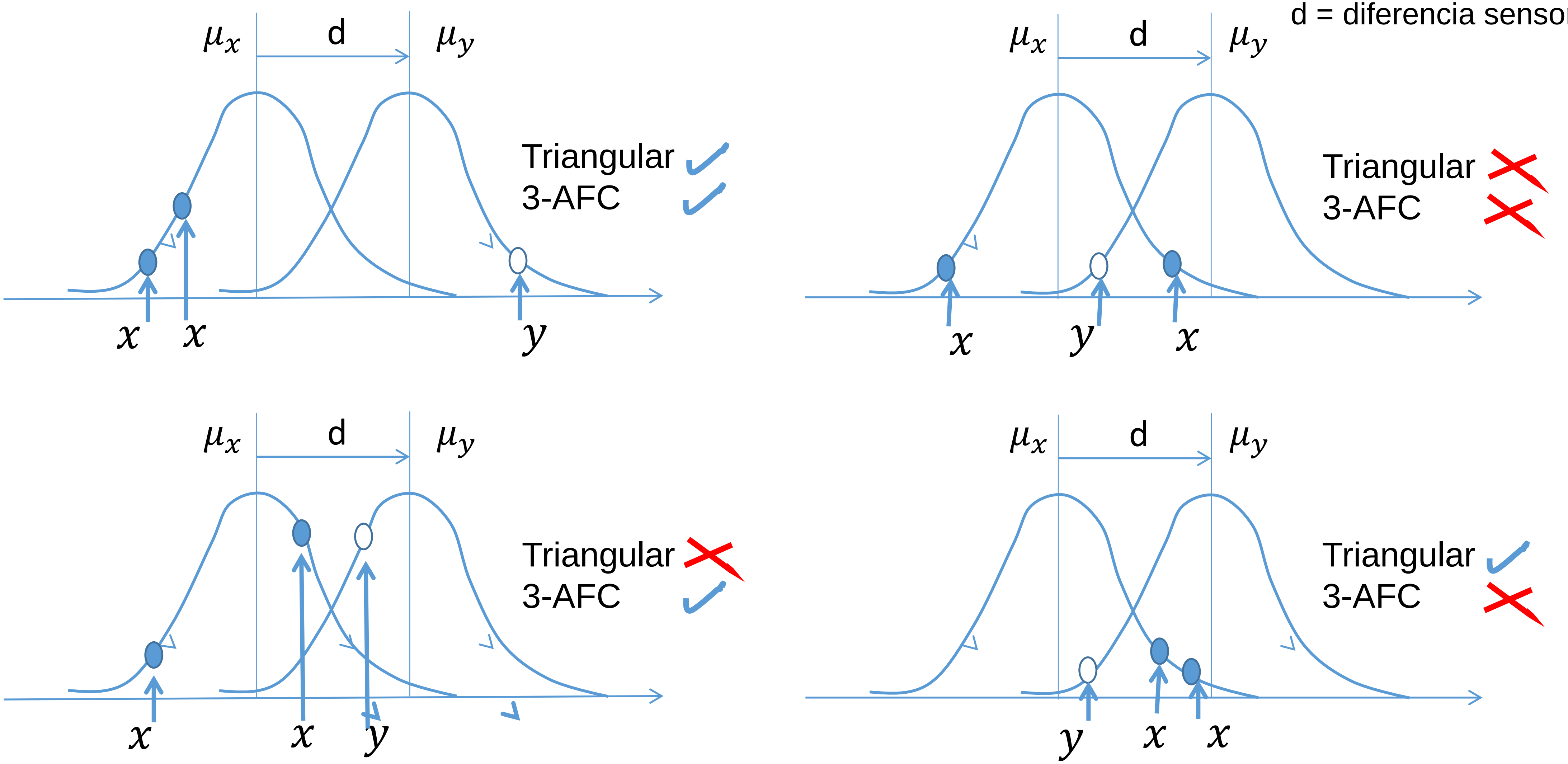


Figura. Aplicación del modelo Thurstone a las pruebas Triangular y 3-AFC

Análisis de la potencia de las pruebas

Reglas de decisión y funciones psicométricas para diferentes métodos

Método 2-AFC

$$p_c = \Phi\left(d/\sqrt{2}\right)$$

Método 3-AFC

$$p_c = \int_{-\infty}^{\infty} \Phi^2(u)\,\phi(u-d)du$$

Método Dúo-Trio

$$p_c = 1 - \Phi\left(d/\sqrt{2}\right) - \Phi\left(d/\sqrt{6}\right) + 2\Phi\left(d/\sqrt{2}\right)\Phi\left(d/\sqrt{6}\right).$$

Método Triangular

$$p_c = \int_{0}^{\infty} \left[\Phi\left(-u\sqrt{3} + d\sqrt{2/3}\right) + \Phi\left(-u\sqrt{3} - d\sqrt{2/3}\right)\right] \exp\left(-\frac{u^2}{2}\right) / \sqrt{2\pi}du$$

p_c (%)	2-AFC	3-AFC	Dúo-Trío	Triangular
50	0	0.56	0	1.47
55	0.18	0.72	0.76	1.72
60	0.36	0.89	1.12	1.98
65	0.55	1.06	1.42	2.23
70	0.74	1.24	1.72	2.50
75	0.95	1.43	2.02	2.80
80	1.19	1.65	2.36	3.13
85	1.47	1.91	2.75	3.52
90	1.82	2.23	3.26	4.03
95	2.33	2.71	4.06	4.80
99	3.29	3.62	5.70	6.09

Tabla. Valores de d en función de la proporción de respuestas correctas y del tipo de prueba de diferencia.

Tabla. Valores de p_c para diferentes valores del índice d de Thurstone

d	2-AFC	DT	3-AFC	TRI	d	2-AFC	DT	3-AFC	TRI
0.3	.58	.51	.42	.34	**1.2**	.80	.61	.69	.45
0.4	.61	.51	.45	.35	**1.3**	.82	.63	.72	.47
0.5	.64	.52	.48	.36	**1.4**	.84	.65	.74	.49
0.6	.66	.53	.51	.37	**1.5**	.86	.66	.77	.51
0.7	.69	.54	.54	.38	**1.6**	.87	.68	.79	.53
0.8	.71	.55	.57	.39	**1.7**	.89	.70	.81	.55
0.9	.74	.57	.60	.40	**1.8**	.90	.71	.83	.57
1.0	.76	.58	.63	.42	**1.9**	.91	.73	.85	.59
1.1	.78	.60	.66	.43	**2.0**	.92	.75	.87	.60

Tabla. Tamaños muestrales requeridos para las pruebas de diferencia, para una significancia $\alpha = .05$ y una potencia 0.8.

d	2-AFC	DT	3-AFC	TRI	d	2-AFC	DT	3-AFC	TRI
0.3	300	23229	206	20928	**1.2**	26	136	17	120
0.4	174	7591	119	6749	**1.3**	23	106	15	93
0.5	115	3196	78	2840	**1.4**	21	84	13	74
0.6	83	1590	56	1419	**1.5**	19	68	12	60
0.7	63	893	42	795	**1.6**	17	57	11	49
0.8	50	550	33	488	**1.7**	16	48	10	42
0.9	41	360	27	320	**1.8**	15	41	9	36
1.0	35	250	23	222	**1.9**	14	36	9	31
1.1	30	181	19	160	**2.0**	13	32	8	27

La paradoja de las pruebas discriminatorias

Los resultados de la aplicación de las pruebas Triangular y 3-AFC reportan mejor desempeño de los jueces cuando utilizan la prueba 3-AFC.

Esta diferencia en el desempeño de los jueces se reportó por primera vez en1937 por Abrahams y colaboradores y durante muchos años no se pudo explicar. Años más tarde, se le conoció como la paradoja de las pruebas de diferencia (discriminatorias y no discriminatorias)

No fue sino hasta 1979 que Frijters encontró la solución a la paradoja , a través de los conceptos del modelo de Thurstone

Frijters, JER. The paradox of discriminatory nondiscriminators resolved. Chem. Senses Flavor, **4**: 355-358; 1979.

Estrategias cognitivas

¿Por qué las pruebas 3-AFC y 2-AFC permiten un desempeño superior a los jueces? La explicación se resume a través de la estrategias cognitivas que utiliza el juez en las pruebas triangular y 3-AFC (Siguiente figura)

En la prueba Triangular, el juez ubica la muestra diferente como la que más se aleja de las otras dos muestras. Para esto, compara las distancias entre las distancias entre los estímulos en el eje de intensidad de la percepción.

En la prueba 3-AFC, el juez elige la muestra de mayor intensidad de percepción. Esta estrategia se conoce como descremado porque el estimulo de mayor intensidad es eliminado de la misma manera que se elimina la espuma de la cerveza.

En el análisis de la siguiente figura se observa que la estrategia de descremado es más eficiente que la de comparación de distancias.

Estrategias cognitivas

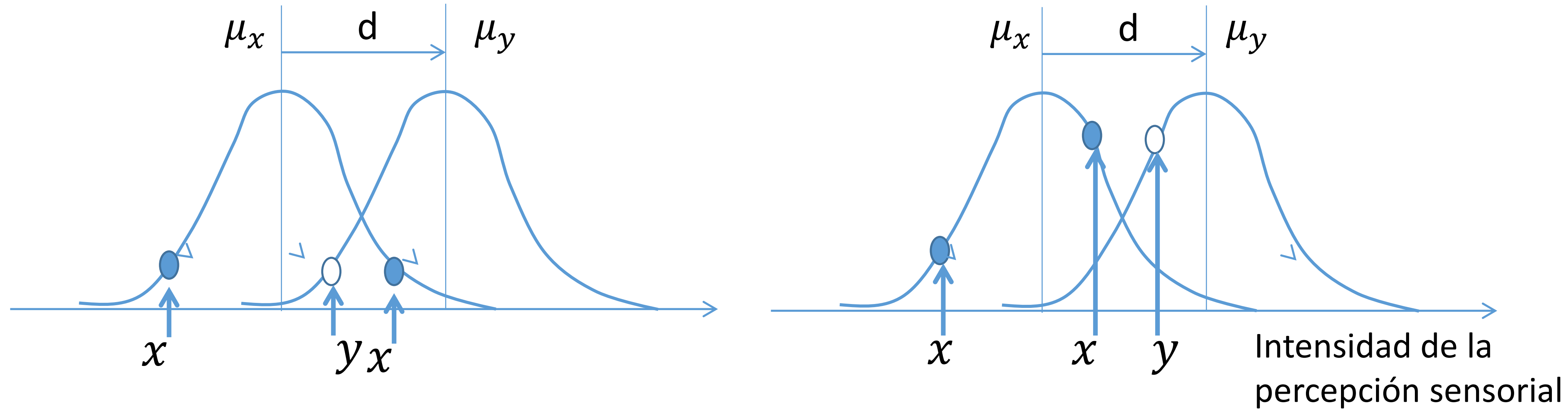


Figura. Ilustración de dos reglas de decisión usadas durante pruebas de diferencia. Estrategia de comparación de distancias (izq.). Estrategia de descremado (der.)

Comparación de pruebas a través de la función potencia.

Potencia de las pruebas 2-AFC y Dúo-Trío, como funciones del indice d.

Para construir estas funciones, tomamos, $\alpha = .05$ y $n = 30$.