4. PRUEBA DE HIPÓTESIS

Con base en: ROWNTREE, Derek. Introducción a la estadística: un enfoque no matemático. Bogotá: Norma, 1984.

PRUEBA DE SIGNIFICACIÓN

Las investigaciones experimentales tienen diseños semejantes a este:

AG_e	X_1	Y	X_2	-
AG_c	X_1	-	X_2	Y

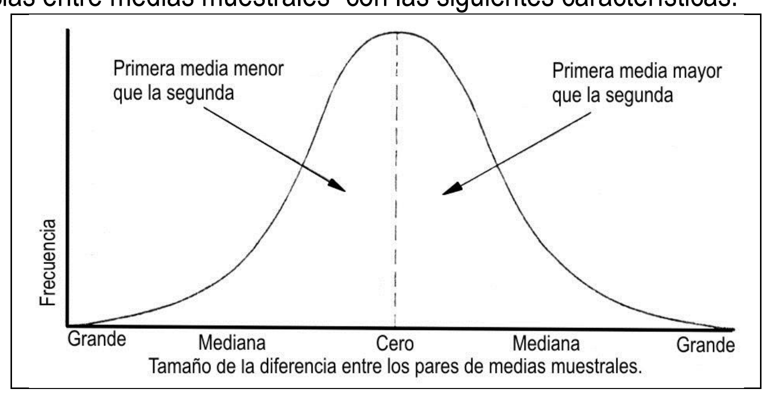
Siendo G_e = Grupo Experimental, G_c = Grupo Control, A = Aleatorización en la división de la muestra en los grupos, X_1 = la preprueba, X^2 = la posprueba, Y = el estímulo experimental, - = la ausencia de estímulo experimental.

La diferencia entre las medias de las prepruebas de los grupos experimental y control debe ser pequeña (cambios debidos al azar), pero entre las pospruebas puede ser o igualmente pequeña (hipótesis nula verdadera) o grande (hipótesis de investigación verdadera). Es decir, las medias de las muestras tienen pequeñas variaciones con el transcurso del tiempo por un sinnúmero de razones que se denominan "azar"; pero cuando la diferencia es demasiado grande, no puede considerarse debida al azar sino al estímulo experimental; cuando esto ocurre se dice que la diferencia es "significativa". Debe definirse entonces un límite a partir del cual la diferencia es significativa, que se denomina "α" (alfa) o "nivel de significancia estadística".

Base teórica:

Si obtuviéramos infinidad de pares de muestras de una población y luego la diferencia entre sus medias, al graficar esta diferencia obtendríamos una curva de "distribución de diferencias entre medias muestrales" con las siguientes características:

- Su media sería cero, a la izquierda están las diferencias negativas (en las cuales la primera muestra fue menor que la segunda) y a la derecha, las diferencias positivas (lo contrario).
- Su desviación estándar se denominada "error estándar de diferencias entre medias" (EE-dif). Y se calcula a partir de los EE de cada muestra: EE-dif = $\sqrt{EE_1^2 + EE_2^2}$



PRUEBA Z

Estadístico que evalúa hipótesis de diferencias entre medias utilizando unidades Z para determinar el punto límite que separa diferencias pequeñas "por azar" de diferencias grandes "significativas", bajo el supuesto de que las observaciones tienen una distribución normal. Los niveles de significancia (α) más utilizados son: 0.05 (5%) y 0.01 (1%), que corresponden a una Z de 1.96 y 2.58 respectivamente. Las diferencias observadas mayores que el límite calculado (Z x EE-dif), se consideran significativas; las menores, debidas al azar.

Ejemplo:

Tenemos la investigación: eficacia de un programa de ejercicio aeróbico en el aumento de la memoria a corto plazo de estudiantes universitarios sedentarios. Se obtiene una muestra de 200 personas, la cual dividimos aleatoriamente en dos grupos: uno recibe el programa y el otro no. El test de posprueba (calificado de 0 a 100) arroja los siguientes datos:

DATOS:

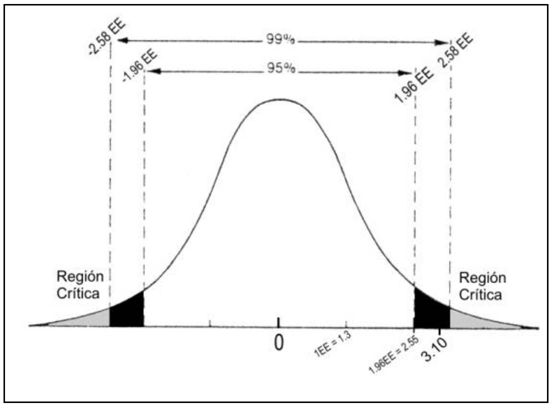
$n_e = 100$ $\quad \bar{x}_e = 63.10$ $\quad DE_e = 10$

$n_c = 100$ $\quad \bar{x}_c = 60.00$ $\quad DE_c = 9$

PRUEBA Z AL 5%:

$$\textbf{EE} = \frac{DE}{\sqrt{n}} \qquad \textbf{EE}_e = \frac{10}{\sqrt{100}} = 1 \qquad \textbf{EE}_c = \frac{9}{\sqrt{100}} = 0.9$$

$$\textbf{EE-dif} = \sqrt{EE_e^2 + EE_c^2} = \sqrt{1^2 + 0.9^2} = 1.3$$



- Límite calculado al 5%: Z x EE-dif = 1.96 x 1.3 = 2.55
- Diferencia observada: 63.10 – 60.00 = 3.10
- La diferencia observada (3.10) **>** Límite calculado (2.55) **==>** Sí hay diferencias significativas, o sea, la probabilidad de que la diferencia observada (3.10) haya sido por azar es menor del 5% (está en la región crítica).

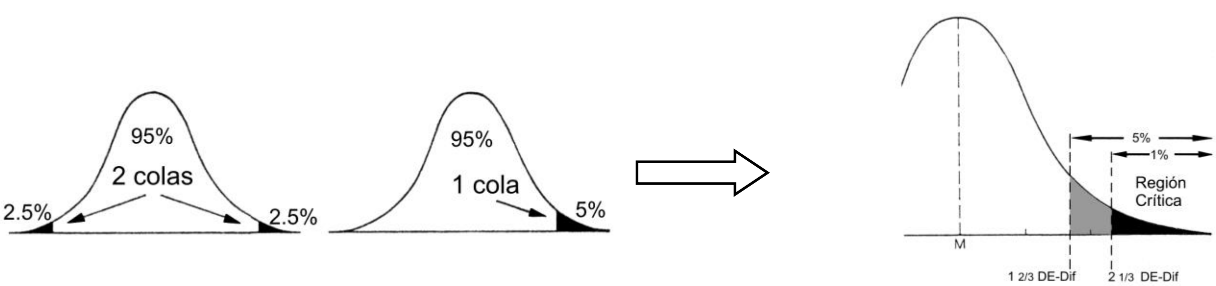
Errores tipo I y tipo II

La hipótesis nula (no hay diferencias significativas) es la que se pone a prueba. Siempre existe la posibilidad de equivocarse: rechazando una hipótesis nula que es verdadera (error tipo I) o aceptando una falsa (error tipo II); como en un juicio en que el acusado puede declarársele culpable siendo inocente, o al revés. Del mismo modo, a más reducimos el riesgo de cometer un error tipo I, más aumentamos el riesgo de cometer un error tipo II, y viceversa. **La ciencia pone el <u>énfasis en evitar errores tipo I</u>: rechazar hipótesis nulas verdaderas** (o sea, aceptar hipótesis de investigación erróneas, que afirmen equívocamente que sí existen diferencias significativas).

Colas

La hipótesis puede plantear:

- Diferencias significativas (sin dirección); en este caso se hacen los cálculos estadísticos teniendo en cuenta ambos extremos de la curva de distribución: **dos colas**.
- Sólo aumento (o sólo disminución) significativa; en este caso se tiene en cuenta sólo uno de los extremos: **una cola**. Por lo tanto, se deja en un solo extremo el 5% (Z de **1.65** ≈ $1^2/_3$) ó 1% (Z de **2.33** ≈ $2^1/_3$).

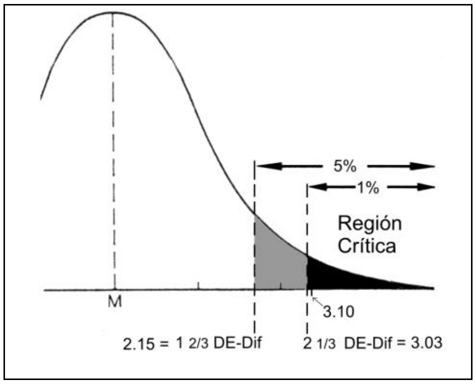


Ejemplo:

Siguiendo el ejemplo anterior, la diferencia no es significativa al 1% (con dos colas):

- Límite calculado al 1% (dos colas): Z x EE-dif = 2.58 x 1.3 = 3.35
- Diferencia observada: 63.10 – 60.00 = 3.10
- La diferencia observada (3.10) **<** Límite calculado (3.35) **==>** NO hay diferencias significativas, o sea, la probabilidad de que la diferencia observada (3.10) haya sido por azar es mayor del 1% (no está en la región crítica).

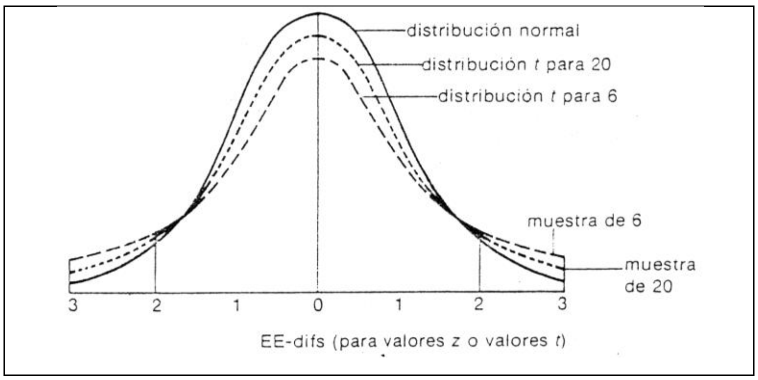
Sin embargo, si de acuerdo a la revisión teórica realizada sólo es posible que el ejercicio aeróbico aumente la memoria (o que no la afecte), pero NO que la disminuya, se puede presentar una hipótesis de una sola cola (antes de recoger la información): "el programa de ejercicio aeróbico aumentará la memoria a corto plazo de estudiantes universitarios sedentarios, a un alfa de 0.01".



La prueba de hipótesis, por consiguiente, será de una sola cola:

- Límite calculado al 1% (una cola): Z x EE-dif = 2.33 x 1.3 = 3.03
- Diferencia observada: 63.10 – 60.00 = 3.10
- La diferencia observada (3.10) **>** Límite calculado (3.03) **==>** SÍ hay diferencias significativas, o sea, la probabilidad de que la diferencia observada (3.1) haya sido por azar es menor del 1% (está en la región crítica).

PRUEBA *t* DE STUDENT



Con **muestras menores de 30**, su desviación estándar no puede considerarse como un estimador confiable de la desviación estándar de la población, sino que es más pequeña. Entonces William Gossett (pseudónimo "Student") elaboró una familia de distribuciones similares a la normal, pero con una mayor dispersión: más bajas y más anchas, a medida que la muestra es menor; de esta forma, el límite fijado es mayor entre menor sea el tamaño muestral. Se calcula como la prueba Z, pero utilizando no la distribución normal sino la **distribución t (más ancha)** correspondiente.

ANÁLISIS DE VARIANZA (PRUEBA DE FISHER o PRUEBA F)

Las pruebas Z y t se utilizan para comparar 2 muestras. Cuando son 3 ó más (por ejemplo, un grupo con ejercicio aeróbico, un grupo con ejercicio con pesas y un grupo en espera), habría que aplicar varias veces la prueba para comparar los pares de muestrales resultantes: para 3 muestras serían tres pruebas (A-B, A-C, B-C); para 4, seis pruebas (A-B, A-C, A-D, B-C, B-D, CD); etc. El problema es que al aumentar el número de pruebas, se incrementa también la probabilidad de un error tipo I: concluir que alguna diferencia es real cuando se debe realmente al azar.

El análisis de varianza permite trabajar simultáneamente varias muestras al **comparar la varianza intragrupal con la intergrupal**. Si la hipótesis nula fuera falsa (esto es, sí existen diferencias significativas entre muestras) la variabilidad intergrupal sería mayor que la intragrupal.

Ejemplo

En el siguiente ejemplo, la variación intragrupal es similar en todos los grupos. En el caso 2, la media del grupo C es claramente superior a la de los grupos A y B, de forma que en este caso, la varianza intergrupal es mayor que la varianza intragrupal. Se rechaza la hipótesis nula y se acepta la hipótesis de investigación: sí hay diferencias significativas entre medias muestrales.

Grupo	Caso 1 A	Caso 1 B	Caso 1 C	Caso 2 A	Caso 2 B	Caso 2 C
	2	3	2	2	3	6
	5	5	3	5	5	9
	4	4	5	4	4	8
	3	3	4	3	3	7
	5	5	4	5	5	8
	3	2	5	3	2	9
	5	4	4	5	4	7
	4	3	3	4	3	8
	4	5	5	4	5	9
	5	3	3	5	3	7
Varianza intragrupal	**1.11**	**1.12**	**1.07**	**1.11**	**1.12**	**1.07**
Media	4.00	3.70	3.80	4.00	3.70	***7.80***
Varianza intergrupal		**0.02**			**2.98**	

MÉTODOS NO PARAMÉTRICOS (VARIABLES CUALITATIVAS)

Desventaja: la desventaja de los métodos no paramétricos es que se pierde información, por lo cual **se requiere una diferencia más grande para que sea significativa**. Ejemplo: el orden "M, F, M, F" da la sensación de que hombres y mujeres son similares, cuando los valores podrían ser, respectivamente: "5, 80, 85, 100", con lo cual la media de los hombres sería "45" y la de las mujeres "90".

Se utilizan cuando las variables son:
(a) Cualitativas nominales: **chi cuadrado**.
(b) Cualitativas ordinales o numéricas sin distribución normal:
- De grupos diferentes: **Mann-Whitney** *(o Kruskall-Wallis si son 3 ó más grupos)*.
- Del mismo grupo: **Wilcoxon** *(o Friedman, si son 3 ó más datos de cada fuente)*.

CHI CUADRADO(χ^2)

Para variables nominales de 2 ó más muestras. χ designa la letra griega chi o ji. En esta prueba **se comparan las FRECUENCIAS observadas con las esperadas por azar**, y luego se busca en la tabla respectiva (correspondiente a una familia de curvas de distribución similar a la de la *t*), si la diferencia encontrada es menor (por azar) o mayor (significativa) del límite calculado para ese número de muestras.

Ejemplo: obesidad vs. género

Frecuencias Observadas

		Sobrepeso Sí	Sobrepeso No	Total
Género	**F**	*70*	*30*	100
	M	*50*	*50*	100
	Total	120	80	200

Frecuencias Esperadas por Azar

		Sobrepeso Sí	Sobrepeso No	Total
Género	**F**	*60*	*40*	100
	M	*60*	*40*	100
	Total	120	80	200

Diferencias

		Sobrepeso Sí	Sobrepeso No
Género	**M**	*10*	*-10*
	F	*-10*	*10*

MANN-WHITNEY *(o Kruskal-Wallis)*

Para variables ordinales y variables numéricas sin distribución normal, de 2 grupos *(de 3 ó más grupos: Kruskal-Wallis)*, por ejemplo, grupo experimental vs. grupo control**. Se compara el ORDEN observado con el esperado por azar**, y luego se busca en la tabla respectiva (correspondiente a una familia de curvas de distribución como la *t*) si la diferencia encontrada es menor (por azar) o mayor (significativa) del límite calculado.

Ejemplo: puesto en carrera de 100 m planos vs. género

Puesto	**1º**	**2º**	**3º**	**4º**	**5º**	**6º**	**7º**	**8º**	**9º**	**10º**	**Total M**	**Total F**	**Diferencia**
Orden Observado	*M*	*M*	*M*	*F*	*M*	*F*	*M*	*F*	*F*	*F*	*18*	*37*	*19*
Orden Esperado	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*25*	*30*	*5*

WILCOXON *(o Friedman)*

Para variables ordinales y variables numéricas sin distribución normal, de un mismo grupo (2 datos de cada fuente) *(3 ó más: Friedman)*, por ejemplo, preprueba vs. posprueba. **Se comparan el ORDEN DE LAS DIFERENCIAS POSITIVAS VS. NEGATIVAS encontradas;** lo esperado por azar es que sean similares, por lo que su suma tienda cero, y luego se busca en la tabla respectiva (correspondiente a una familia de curvas de distribución como la *t*) si la diferencia encontrada es menor (por azar) o mayor (significativa) del límite calculado.

Ejemplo: puesto en carrera de 100 m planos vs. género

Fuente	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**Suma encontrada**	**Suma esperada por azar**
Preprueba	*10*	*9*	*10*	*9*	*8*	*10*	*6*	*10*	*5*	*8*		
Posprueba	*3*	*9*	*4*	*6*	*8*	*2*	*8*	*5*	*6*	*3*		
Diferencia	*7*	*0*	*6*	*3*	*0*	*8*	*-2*	*5*	*-1*	*5*		
Orden diferencias +	*7*		*6*	*3*		*8*		*4.5*		*4.5*	*33*	
Orden diferencias -							*2*		*1*		*-3*	
											30	***0***

Nota: primero se asigna el puesto en orden sin tener cuenta el signo (positivo o negativo). Si la diferencia es 0 no se tiene en cuenta. Si hay varias diferencias iguales, se les asigna la mediana de las posiciones a cada una, por ejemplo, hay dos “5” a los cuales les corresponden los puestos 4º y 5º, entonces a cada uno se les da el puesto 4.5. Luego a la suma de los puestos de las diferencias positivas se le resta la suma de los puestos de las diferencias negativas.