

Diseño Experimental requiere

1. Asignación aleatoria a tratamientos, a grupos y subgrupos
2. Muestra suficientemente grande
3. Comparación de grupos, subgrupos y factores (tratamientos)

Ejemplo

Un maestro piensa que el método de enseñanza A es mejor que el método B. Enseña dos secciones del curso de Matemáticas Avanzadas con 25 y 22 estudiantes respectivamente. Tiene entonces una **muestra suficientemente grande** de sujetos para hacer el análisis correspondiente. En su análisis **comparará** la puntuación media obtenida por los dos grupos. Los grupos son de aprovechamiento y motivación similares. El maestro:

1. **Diseña los materiales** que utilizará con cada grupo asegurándose que ambos transmiten el mismo conocimiento. La única diferencia entre ellos es la metodología de enseñanza.
2. **Diseña los instrumentos de medición**, una pre-prueba y post-prueba, las cuales en su caso son la misma. La prueba a administrarse tiene una **cantidad suficiente de preguntas relevantes** a aquel conocimiento que deben los estudiantes adquirir, es decir, tiene **validez**. Sabe que si la prueba sólo incluye un número pequeño de preguntas no podrá distinguir el conocimiento adquirido por los estudiantes: habrá muchas puntuaciones similares y hay la posibilidad de que las diferencias entre las puntuaciones no brinden suficiente información. Por ejemplo, si el máximo fuera 5 puntos, un estudiante que obtiene 4 en su pre-prueba, sólo podrá aumentar un punto, hasta un total de 5 en la post-prueba, lo que no ocurre con un estudiante que obtiene puntuación de 1 en la pre-prueba.
3. Ahora **asigna aleatoriamente el método de enseñanza a cada grupo**. Lanza una moneda al aire para decidir cual de los dos grupos recibirá la enseñanza con el método A o B. **Idealmente los estudiantes no deben saber que participan de un experimento, participarían a ciegas**, pero como se requiere permiso de sus padres, eso no es posible. Sin embargo, el maestro no le informa a los grupos, ni hace indicación alguna con su comportamiento, de cual de los métodos participa cada grupo. Tampoco les comunica su creencia de que el método B es mejor que el A.
4. Administra la pre-prueba a cada grupo.
5. Usa los métodos y materiales diseñados para impartir la enseñanza. Como uno de los métodos requiere que los estudiantes sean divididos en subgrupos de trabajo cooperativos, hace la **asignación a los mismos en forma aleatoria**. Para ello usa un generador de números aleatorios (calculadora gráfica o una calculadora Casio, Excel, o números escritos en papelitos y sacados al azar de una urna). Con esta asignación obtiene subgrupos que son estadísticamente homogéneos, pues las diferencias entre los estudiantes, de aprovechamiento, por ejemplo, ha sido distribuida al azar en forma uniforme entre los distintos grupos.
6. Administra la post-prueba, corrige y tabula ambas pruebas. Corrige ambas pruebas al final, así evita prejuiciar su enseñanza con el conocimiento de la puntuación de la primera prueba. Si por ejemplo supiera que los estudiantes que participan del método A no aprendieron bien cierta técnica, como buen maestro, recalcaría esa parte y no

atendería aquellas que los estudiantes ya dominan. Esto sería bueno para la enseñanza, pero sesgaría los resultados del experimento.

El maestro obtuvo los siguientes resultados:

	Grupo A			Grupo B		
	Pre	Post	Dif	Pre	Post	Dif
	38	76	39	53	86	33
	54	79	26	59	89	30
	64	66	2	31	76	45
		80		61		
	34	76	42	53	86	33
	60	81	22		91	
	67	77	10	53		
	33	80	47	61	90	29
	71			47	84	37
	60	77	17	54	87	33
	48	76	28		86	
	48	75	27	50		
	55			38	79	42
	57	66	9	31	76	45
	41	62	22	23	72	49
	47	82	35	65	92	27
	57	70	12	39	80	41
	42	75	33	50	85	35
	64	71	7	42	81	39
	30	81	51	63	91	28
	42	79	38	59	89	30
	39	75	36	50	85	35
				46	83	37
				40	80	40
				78	98	20
Media	49.95	75.31	26.48	48.56	84.33	36.26
Desv Est	11.97	5.53	14.17	11.34	5.52	6.25
n	21	20	19	22	21	19
Max	71	82	51	65	92	49
Min	30	62	2	23	72	27

1. Luego de calcular estadísticas descriptivas y hacer gráficas apropiadas de las distintas variables, el maestro **comparará las diferencias promedio** entre la post-prueba y la pre-prueba para cada grupo. En el grupo A, esa diferencia promedio fue de 26.48 puntos mientras que en el grupo B la diferencia promedio fue de 36.26. Claro, para hacer estadísticamente válida la comparación, deberá examinar e incluir en su análisis la desviación estándar de la media de las diferencias entre los grupos.

2. No incluirá en el análisis las puntuaciones de aquellos estudiantes ausentes a alguna de las dos pruebas, pues este **diseño es pareado**, donde el mismo sujeto es medido en dos ocasiones distintas. Este diseño responde a que la puntuación de cada estudiante en la pre-prueba **no es independiente** de su puntuación en la post-prueba (y viceversa).
3. Mantendrá su pregunta de investigación siempre en mente y comparará las medias de las diferencias de los dos grupos, es decir, comparará la media $\bar{x} = 26.48$ para el grupo A con $\bar{x} = 36.26$ del grupo B.
4. A primera vista estos valores parecen indicar que los estudiantes del grupo (o método) B aprendieron más. Pero el maestro sabe más que eso, pues estos resultados pueden haber sido obtenidos por mero funcionamiento del azar. Para resolver este problema hará una prueba estadística.
5. La prueba que decide usar es una prueba t para dos poblaciones. Esto requiere varias suposiciones:
 - a. Las diferencias mismas tienen distribución normal o el tamaño de cada muestra es lo suficientemente grande para concluir que la distribución de las medias, \bar{x} es aproximadamente normal.
 - b. Las varianzas (o desviaciones estándar) poblacionales de las diferencias son iguales. Si no se puede suponer esto, debe usar otra variante de la prueba t.
 - c. Las puntuaciones del grupo A son independientes del grupo B (no se comunicaron) y las puntuaciones de cada estudiante son independientes de los otros estudiantes de su grupo (no se copiaron).
 - d. Si desea extender estos resultados a otros grupos debe suponer que los grupos estudiados son típicos de los grupos de la población a la cual desea extender los resultados.
6. Establece la hipótesis nula: $H_0 : \mu_A = \mu_B$ Los métodos de enseñanza producen un aprendizaje promedio idéntico ($H_0: \mu_A = \mu_B$) EL valor μ_A es la media poblacional de las diferencias en puntuación usando el método A). La hipótesis alternativa es: $H_a : \mu_B > \mu_A$ El método de enseñanza B produce un mayor aprovechamiento promedio que el método A ($H_a: \mu_B > \mu_A$). Si desea, puede establecer un **nivel de significancia** a priori, tal como 5%, pero es más flexible utilizar el valor p de la prueba.
7. Usa Excel o algún programa estadístico y obtiene:

	Variable 1 Dif Grupo A	Variable 2 Dif grupo B
Mean	26.477	35.462
Variance	200.810	49.825
Observations	19	20
Pooled Variance	123.277	
Hypothesized Mean Difference	0	
df	37	
t Stat	-2.5261	
P(T<=t) one-tail	0.0080	
t Critical one-tail	1.6871	
P(T<=t) two-tail	0.0159	
t Critical two-tail	2.0262	

8. El maestro sabe que el valor importante aquí es el valor p de la prueba dado por $P(T \leq t)$ *one-tail* el cual es igual a .0080, ya que la prueba es de una cola (el aprendizaje promedio bajo el método B es mayor que bajo el método A). La prueba sería de dos colas si la hipótesis alternativa fuera de que las medias son diferentes sin hacer conclusiones sobre cual es mayor.
9. El maestro compara este valor, .0080 con el nivel de significancia previamente seleccionado para la prueba, digamos 5% o .05. Como el valor p es menor que la significancia seleccionada concluye que el método B produce un aprendizaje promedio significativamente mayor que el método A, al nivel de significancia del 5%. Así rechaza la hipótesis nula a favor de la alternativa a ese nivel de significancia.
10. El maestro también sabe que obtendrá la misma conclusión si selecciona cualquier nivel de significancia mayor a .0080. Sin embargo si selecciona un nivel de significancia menor que .0080, tal como el .1% o .001, (es mas conservador a la hora de tomar decisiones, requiriendo evidencia mas contundente) no podrá rechazar la hipótesis nula y su conclusión será necesariamente que el método B no demostró producir un aprendizaje promedio mayor que el método B al nivel de significancia del .1%.