



Diseños experimentales

Apuntes de clase del curso Seminario Investigativo VI

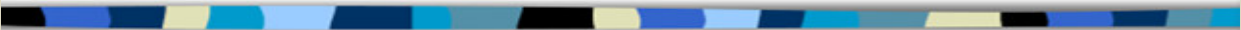
Por:

Gustavo Ramón S.*

* Doctor en *Nuevas Perspectivas en la Investigación en Ciencias de la Actividad Física y el Deporte* (Universidad de Granada).

Docente - Investigador del Instituto Universitario de Educación Física, Universidad de Antioquia (Colombia).

Correo: gusramon2000@yahoo.es



CAPÍTULO VIII

DISEÑOS EXPERIMENTALES

- **Generalidades:**

Las investigaciones y en particular los diseños experimentales intentan establecer básicamente relaciones causa-efecto. Mas específicamente, cuando se desea estudiar como una variable independiente (causa) modifica una variable dependiente (efecto). Un ejemplo podría ser: analizar el efecto que un plan de entrenamiento basado en la técnica Facilitación Neuromuscular Propioceptiva (FNP) (variable independiente) produce sobre la movilidad articular de la cadera (variable dependiente). Sin embargo, un diseño puede contemplar el efecto que producen dos variables independientes tienen sobre una dependiente, por lo que los diseños experimentales poseen una gran variación. En el presente capítulo precisaremos detalles con respecto a este tipo de diseños.

El establecimiento de que, en un experimento, una variable independiente produce una variación específica (que solo puede ser atribuida a la variable independiente) en la variable dependiente depende fundamentalmente de la manera como esté diseñado el experimento. Los procedimientos estadísticos permiten confirmar o validar el procedimiento empleado, pero una cosa debe ser clara: el hecho de aplicar procedimientos estadísticos no permite afirmar que el experimento está bien diseñado, al punto de que se puede afirmar que las variaciones producidas en los sujetos experimentales sólo y exclusivamente pueden ser atribuidas a la variable independiente.

La relación causa-efecto solo puede ser atribuida mediante la aplicación de un pensamiento lógico junto a un diseño bien aplicado. Las siguientes pautas están asociadas a diseños idóneos:

- a) La adecuada selección de un marco teórico
- b) La selección apropiada de los sujetos
- c) El uso de procedimiento estadísticos correctos
- d) El control de la variable independiente
- e) La adecuada medición de la variable dependiente
- f) El control de variables asociadas.

1. **¿ Qué es un Experimento ?**

- ◆ Es un estudio de investigación en el que se manipulan deliberadamente una o mas variables independientes para analizar las consecuencias que la manipulación tiene sobre una o mas variables dependientes, dentro de una situación de control para el investigador.

GRamónS

- ◆ Un experimento se lleva a cabo para analizar si una o mas variables independientes afectan a una o mas variables dependientes y por qué lo hacen.

En cualesquier experimento se presentan tres tipos de variables: la independiente, la dependiente y las intervinientes (contaminadoras o extrañas).

La **variable independiente** representa los tratamientos, factores o condiciones que el investigador controla para probar los efectos sobre la variable dependiente. El término de *tratamiento* se emplea a semejanza de las agentes farmacológicos empleados en la medicina para manejar alguna enfermedad o proceso patológico. En el campo de la actividad física y el deporte, los tratamientos generalmente son los métodos empleados para el desarrollo de las capacidades condicionales o para la enseñanza de los movimientos. En el proceso del experimento, el investigador se limita a **controlar** la variable independiente, es decir, a seguir fielmente el proceso o el método para desarrollar alguna capacidad. Volviendo al campo de la medicina, el tratamiento de una enfermedad implica darle al paciente una dosis de un agente farmacológico en una determinada cantidad, un número de veces por día, durante un intervalo de tiempo definido. Para probar la efectividad de este tratamiento, se debe cumplir con que la dosis sea la indicada, el número de veces que se toma dicha dosis sea el establecido y finalmente, que cumpla con los días estipulados. A este proceso se denomina *controlar* la variable independiente. En el campo de la actividad física y el deporte, el ejemplo específico de control es el hecho de poder dosificar la carga de manera que pueda ser administrada a una determinada intensidad, un número de veces por semana y durante un número fijo de semanas. Si no se puede llegar a este grado de precisión, entonces el experimento no tiene validez porque no tiene control.

Por otra parte, la **variable dependiente** es la que refleja los resultados de un estudio de investigación. Continuando con el ejemplo en el campo de la actividad física y el deporte, en el caso de un estudio sobre la efectividad de un método para enseñar alguna destreza, la variable dependiente es el aprendizaje que logran los sujetos; en el caso de una investigación acerca del efecto que produce el entrenamiento pliométrico en la saltabilidad de los voleibolistas, la variable dependiente es la saltabilidad de los deportistas. Como puede apreciarse, la variable dependiente es la característica que realmente va a variar en un experimento, que va ser modificada, en este caso, por la variable independiente. En el proceso de investigación, la variable dependiente se **mide**. El éxito de un experimento va a depender entonces del grado de precisión con que mida dicha variación. Por lo tanto, tanto el instrumento de medición como el proceso de medición debe ser muy bien definido.

El otro tipo de variable que se presenta en un diseño experimental es la **variable interviniente**, también llamada contaminadora o extraña. Este tipo de variable son variables independientes que, si no se controlan, pueden alterar el resultado de la investigación. Así, en un estudio sobre el desarrollo de la resistencia aeróbica en niños de 8 a 9 años, una variable que produce efectos positivos es el crecimiento y desarrollo. Por lo tanto, en un estudio que desee estudiar el efecto de un determinado

GRamónS

plan de entrenamiento debe tener en cuenta el proceso de crecimiento y desarrollo. En este caso en particular, el diseño debe tener un grupo control y un grupo experimental. El grupo control no estará sometido al plan de entrenamiento pero si lo estará al efecto de crecimiento y desarrollo. Si al final del experimento, tanto en el grupo control como el grupo experimental se producen modificaciones en la resistencia, el investigador no podrá concluir que el plan de entrenamiento fue el que produjo las modificaciones de resistencia en los niños. Por lo tanto, el diseño debe controlar el efecto de las variables intervinientes. De otra manera, la validez externa del experimento se verá seriamente comprometida.

Es posible que en un experimento existan varias variables independientes, varias dependientes y varias intervinientes. Estos tipos de experimentos son complejos y difíciles de llevar a cabo por el control que se debe hacer sobre todas las variables.

Tipología de los Diseños Experimentales

Según Campbell y Stanley (1966)¹, se pueden escoger los siguientes son criterios para la nominación de los diseños experimentales:

- R = Para asignación al azar o aleatorización.
- G = Para determinar el número o grupo de sujetos (G₁ = grupo 1; ...)
- X = Para la presencia de un tratamiento, estímulo, o condición experimental, las cuales pueden ser varias, por lo que puede tener subíndices (X₁, X₂, ..., X_n)
- O = Una medición a los sujetos de un grupo. Puede tener subíndices para determinar el número de mediciones (O₁, O₂ , ..., O_n)
- = Ausencia de un estímulo, indica que se trata de un grupo control.

Según estos criterios, se pueden presentar los siguientes tipos de experimentos:

1°	RG ₁	O ₁	X	O ₂
2°	RG ₁	X	O	
	RG ₂	-	O	
3°	G	O ₁	X	O ₂
4°	RG ₁	X ₁	O ₁	
	RG ₂	X ₂	O ₂	
	RG ₃	X ₃	O ₃	
	RG _k	X _k	O _k	

¹ Campbell, D.T. y Stanley, J.C. 1996. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company.

GRamónS

En el primer experimento, se escoge aleatoriamente (R) un grupo al cual se le realiza una medición previa (O_1), se le aplica un tratamiento (X) y finalmente se vuelve a medir (O_2). En el segundo experimento, aunque similar al primero, varía en que solo se realiza una medición al final del experimento. El tercero, no es seleccionado al azar. El cuarto, es una generalización de K grupos experimentales, con selección al azar (R), sin medición previa, pero con medición al final.

En los próximos apartados veremos cómo se clasifican los experimentos.

2. *¿Qué tipos de experimentos existen?*

En la historia de la investigación, Campbell y Stanley ² (1963) revolucionaron la forma en que se planteaban y se llevaban a cabo las investigaciones. Ellos identificaron tres categorías generales de diseños experimentales: a) experimentos verdaderos, b) cuasi-experimentos, y c) pre-experimentos. En el cuadro 1 se muestran las principales diferencias entre estos tipos de experimentos. La principal diferencia radica en el grado de control que se imponen a las variables que se están estudiando. A continuación estudiaremos cada uno de estos tipos de diseño.

Cuadro 1. Diferencias entre las diferentes tipos de experimentos.

Característica	Pre-experimentos	Cuasi-experimentos	Experimentos
Presencia de grupo control	En algunos casos	Si	Si
Selección aleatoria de sujetos	No	Si, pero restringida	Si
Asignación aleatoria de los sujetos a los grupos	No	No	Si
Asignación aleatoria de tratamientos a grupos	No	Si es posible	Si
Grado de control sobre variables externas	Bajo	Moderado	Alto

2.1. *Experimentos puros o verdaderos*

Un diseño experimental puro es aquel en el cual se manipula una o varias variables independientes para observar sus efectos sobre una o varias variables

² Campbell, D.T., y Stanley, J. Experimental and quasi-experimental design for research on teaching. En Salkind, N.J. Métodos de investigación. 3 ed. Mexico: Prentice Hill.

GRamónS

dependientes en una situación de control. Las siguientes son las condiciones que debe reunir un experimento "puro":

- 2.1.1. Manipulación intencional de una o mas variables independientes.
- 2.1.2. Se debe medir el efecto que la variable independiente tiene en la variable dependiente.
- 2.1.3. Buen control o validez interna de la situación experimental.

2.1.1. Manipulación intencional de una o mas variables independientes.

En un experimento verdadero, se parte de la hipótesis de que la variable independiente producirá modificaciones en la dependiente cuando la primera varíe. Por lo tanto, el investigador debe probar intencionadamente que cuando se producen variaciones en la variable independiente, también lo hará la variable dependiente.

La variable dependiente sólo se mide a diferencia de la variable independiente que se manipula.

¿Cómo se puede manipular la variable independiente?

- ☞ Por la presencia o ausencia de un tratamiento, factor o una característica
 - * Un grupo se expone a la variable independiente o **tratamiento experimental** (grupo experimental) mientras que el otro no (grupo control).
 - * Si en dos grupos de experimentación todo fue "igual", excepto la exposición a la variable independiente, es muy razonable pensar que las diferencias resultantes entre los dos grupos se deban a la presencia o ausencia de la variable independiente.

- ☞ Mas de dos grados o intensidades en la variable independiente.
 - * La variación de la intensidad de la variable independiente puede tener mas de dos grados : alta (X_1), media (X_2) y baja (X_3). Este tipo de estudios responde a la necesidad de observar si la relación entre la variable independiente y la dependiente sigue un comportamiento lineal o no. Por ejemplo, se desea probar el efecto de la intensidad de programas de entrenamiento en resistencia sobre dicha capacidad. Por esta razón, se diseña un plan con una intensidad baja (40%), otro con media intensidad (60%) y otro con alta intensidad (80%). Se espera que el programa de baja intensidad produzca unas pequeñas mejoras sobre la resistencia; el de media intensidad, efectos moderados, y el de alta intensidad, los mayores efectos.

- ☞ Combinación de cantidad y calidad
 - Los grupos experimentales, en el caso de proceso ligados a la actividad física y deportiva, se pueden exponer a procesos de entrenamiento que combinan

variaciones tanto en el volumen de la carga (cantidad) como en la intensidad de la carga (calidad).

- Aspectos a tener en cuenta:
- ✓ En cada nivel o modalidad de la variable independiente se requiere al menos un grupo experimental.
 - ✓ Es útil consultar experimentos previos.
 - ✓ Se debe evaluar la manipulación antes del experimento.
 - ✓ Se deben incluir verificaciones de la manipulación

2.1.2. Se debe medir el efecto que la variable independiente tiene en la variable dependiente.

En primera instancia debemos volver a repasar y precisar los conceptos de medición. Como vimos en la primera parte, la medición es el proceso de asignar valores a resultados. En el caso de diseños experimentales, los valores son numéricos. Dada la variable fuerza muscular, los resultados serán valores numéricos tales como 500 N o 1500 N.

El proceso de medición no es, en la mayoría de las veces, un proceso ciento por ciento exacto puesto que la presencia de factores tales como materiales (precisión de los instrumentos), humanos (grado de concentración o habilidad), ambientales (grado de humedad o temperatura ambiental) hacen que se cometan errores. En el caso citado de la fuerza, se puede afirmar que lo que hemos medido es un valor aproximado puesto que el valor real depende de la ausencia de errores en la medición.

El término error proviene del latín *error-ris*, que significa equivocación, falta, extravío, y se deriva del verbo erro que significa andar errando o vagando de una parte a otra; proviene también del griego *erro* que significa ir equivocado. Por lo tanto, error es un conocimiento equivocado, que no concuerda con la realidad a la que se refiere.

Se pueden distinguir diferentes tipos de error:

- a) de razonamiento: debidos a errores en la deducción o a lógicas indebidas;
- b) de percepción: originados en los engaños que se producen en los órganos de los sentidos o en el proceso de percepción;
- c) de apreciación: derivados de una visión intelectual parcial y subjetiva de la realidad;
- d) de planificación: por concepciones equivocadas de las operaciones o fases de la investigación;

GRamónS

- e) muestral: es un error aleatorio de tipo estadístico que tiene lugar cuando se utilizan y seleccionan las muestras;
- f) sistemático: es el que se comete de manera constante dentro de un proceso de medición.

El error puede proceder del método, es decir, derivado del mismo proceso de medición, ya sea por falta de precisión de los aparatos o instrumentos. Puede proceder del sujeto que realiza la medición (mala visión, mala ubicación, falta de capacitación en el manejo del instrumento). Otro factor de error puede radicar en la falta de colaboración del sujeto que es medido (finge realizar un máximo esfuerzo, falta de habilidad para realizar la prueba, no comprende las instrucciones).

Uno de las premisas fundamentales del éxito en el proceso de investigación es la medición de la variación de la variable dependiente, la cual, para ser válida y confiable debe estar exenta de errores de medición. Si no se pueden eliminar los errores y de esta manera asegurar la medición, los resultados igualmente perderán validez y confiabilidad.

¿Qué es la CONFIABILIDAD ?

En el caso de medir la masa corporal de un grupo de sujetos, en un estudio en el cual la variable dependiente es la masa corporal, se espera que al pesar un sujeto en la primera vez, dicho valor sea igual al ser pesado dos, tres o mas veces, en idénticas circunstancias. Si en una primera medición, el valor es de 65 kg y en una segunda medición el valor es de 66 kg, se puede decidir que esa medición no es confiable.

Una definición de confiabilidad sería entonces el grado en que el proceso de medición elimina los errores. Una operacionalización será:

$$C = \frac{PO}{PO + E} \quad [E-1]$$

donde,

C = Confiabilidad

PO = Puntaje Obtenido

E = Error

Como se puede apreciar, obtener una medición confiable es eliminar los factores de error en el proceso. Si se elimina el factor error, en la ecuación E-1, la confiabilidad sería igual a 1 puesto los valores del numerador y el denominador quedan iguales.

La confiabilidad es el grado en que un proceso de medición elimina los errores. Como se deriva de la E-1. es un cociente, el cual, entre mas se

acerca a 1, mayor grado de confiabilidad tiene la medición. Es el grado en que, al repetir la aplicación de un instrumento o medición al mismo sujeto, se producen iguales resultados.

En el campo de la actividad física y el deporte, la medición de algunas variables es un proceso complejo puesto que la confiabilidad de la medición no depende del instrumento de medición sino de la voluntad del sujeto o del protocolo de medición. Este es el caso de la Cineantropometría en la cual, variables como la talla o los pliegues cutáneos dependen en gran parte del protocolo. Para determinar la talla de un sujeto se deben tener en cuenta la regla de los cinco contactos, el plano de Frankfort y la tracción cervical. Por esto, se deben hacer tres mediciones y algunas escuelas aceptan hasta dos milímetros de diferencias entre las mediciones para aceptarlas como confiables. Algo similar sucede con los pliegues cutáneos, se debe seguir un protocolo y se debe tomar más de dos registros. Para este caso, algunas escuelas toman el dato más bajo de las tres mediciones; otras toman el promedio.

¿Qué procedimientos existen para determinar o medir el grado de confiabilidad?

- ⊗ **Test- Retest** : un instrumento se aplica en dos momentos diferentes a una misma población y sus diferencias no deben ser significativas y su grado de correlación debe ser alto.

El grado de correlación producto-momento de Pearson es un procedimiento estadístico usado en este tipo de pruebas. Los valores varían entre -1 y 1. Los valores positivos expresan una relación directamente proporcional mientras que los valores negativos indican una relación inversamente proporcional. Los siguientes son parámetros de interpretación³:

Las valores entre	Se consideran:	Las valores entre	Se consideran:
- 1.00	Negativa Perfecta	1.00	Positiva Perfecta
- 0.90 y - 0.99	Negativa muy fuerte	0.90 y 0.99	Positiva Muy fuerte
- 0.75 y - 0.89	Negativa fuerte	0.75 y 0.89	Positiva Fuerte
- 0.50 y - 0.74	Negativa moderada o media	0.50 y 0.74	Positiva moderada o media
- 0.10 y - 0.49	Negativa débil	0.10 y 0.49	Positiva débil
0.00 y - 0.09	No existe correlación	0.00 y 0.09	No existe correlación

Para el caso de la confiabilidad de instrumentos, el grado de correlación entre dos mediciones hechas a un grupo de sujetos, el grado de correlación debe ser 0.95 o más. Se debe tener en cuenta que, en el campo de la actividad física y el deporte, el intervalo de tiempo, la hora del día, el nivel de actividad y otras variables intervinientes, pueden afectar las mediciones cambiando el resultado, por lo que las mediciones siempre se deben hacer bajo las mismas circunstancias. Así, la

³ Hernández y col. Metodología de la investigación. 2003. México: MacGraw Hill.

GRamónS

estatura se debe medir inmediatamente se levanta para que el efecto del cansancio acumulado durante el transcurso del día no afecte la medición. Algo similar ocurre con la masa corporal, el consumo de oxígeno, el nivel de fuerza, etc.

- ⊗ Formas alternativas o paralelas: se administran dos instrumentos (que no son iguales pero que son similares en contenido, instrucciones o duración) a un mismo grupo de sujetos en diferentes momentos. Los resultados deben correlacionarse positivamente.

Esta modalidad se emplea para la validación de un instrumento o protocolo. En el caso del consumo de oxígeno, las valoraciones con equipos ergoespirométricos, con medidores de oxígeno y gas carbónico son los más confiables puesto que hacen la medición directamente. Cuando no se dispone de estos equipos, se realizan test de campo, como el test de la milla o los dos mil metros. El resultado de este test se puede correlacionar con el resultado de la medición directa y su grado de correlación no debe ser inferior a 0.95. De hecho, todos los resultados de investigación que pregonan el empleo de métodos de medición indirecto, divulgan el grado de correlación con los métodos directos.

¿Qué es la VALIDEZ ?

El término latino *valere* tiene varios significados: estar sano, ser fuerte, poder, prevalecer. Otras acepciones son: bondad, belleza o toda cualidad positiva. El valor y la validez, derivados de *valere*, se refieren a la cualidad de algo ser sano, fuerte, potente. En el campo de la ciencia se asume que el conocimiento científico debe mostrar o representar la realidad a la que se refiere, es decir, su carácter positivo.

Dado que el conocimiento de la realidad se basa fundamentalmente en ideas de la misma, en el campo científico no se podrán obtener sino ideas aproximadas de la realidad. Por lo tanto, se plantean las siguientes cuestiones:

- a) Se deben tomar las medidas necesarias y disponer las cosas de tal modo que los resultados de investigación sean válidos y no estén viciados o adulterados por la intervención de factores extraños a la investigación, y
- b) Comprobar, una vez realizada la investigación, el grado de exactitud con el que los resultados obtenidos representan a la realidad a la que se refieren.

En resumen, la *validez* se refiere al grado en que un instrumento realmente mide la variable que pretende medir. La validez se define en el contexto del uso de la prueba. Según Salkind (1997), la validez tiene tres aspectos relacionados: a) la validez se refiere a los resultados de una prueba y no a la prueba en sí misma; b) la validez no es una cuestión de "Todo o nada"; esta puede variar desde poco válida hasta muy válida; y c) la validez de los resultados se deben interpretar dentro del contexto en el que ocurre la prueba. Por lo tanto, la forma de examinar la validez de un instrumento o de una prueba es determinar si la prueba se concentra en los resultados

GRamónS

de un estudio y si los resultados se entienden dentro del contexto del propósito de la investigación.

Debido a que existen diferentes propósitos al usar cualquier procedimiento de medición, existen diferentes clases de validez⁴:

1. Validez lógica: aunque parece una redundancia, la validez lógica se refiere a que la medición implique la actividad o aspecto que está siendo medido(a). Significa que la prueba es válida por definición. Una prueba de balance estático que utiliza el balance sobre un solo pie tiene una validez lógica.
2. Validez de contenido : grado en que la medición representa al concepto medido. Es por tanto, un tipo de validez que deben cumplir los cuestionarios o pruebas de conocimiento de una determinada materia o área. Un instrumento de medición debe contener todos los ítems de dominio de contenido de las variables a medir.
3. Validez de criterio: la validez de un instrumento se debe comparar con algún criterio externo, considerado como un estándar. Según Thomas y Nelson (1998), la validez de criterio puede ser de dos clases:
 - 3.1. *Validez de concurrencia*. Cuando, simultáneamente a la medición realizada, se realiza otro procedimiento que lo valida. En el caso de la medición del consumo de oxígeno, un investigador puede usar la Bolsa de Douglas y simultáneamente el procedimiento de medición mediante ergoespirómetro. Si las mediciones obtenidas son similares o dicho en términos estadísticos, no existen diferencias significativas entre las mediciones, se considera que la bolsa de Douglas es válida para medir consumo de oxígeno. En los casos de experimentos sobre ejecución de movimientos como en la gimnasia artística, el juzgamiento (medición) de los que se inician es comparado con los jueces más veteranos. Cuando el juzgamiento de los novatos coincide con el de los veteranos, se considera que existe validez. Dicho de otra manera, en todo experimento se debe destacar y hacer explícita la manera como se determinó la validez de(l) instrumento(s).
 - 3.2. *Validez de predicción*. Una prueba se puede usar para predecir lo que sucederá con una determinada variable en el transcurso del tiempo. Un ejemplo puede ser la talla: algunos investigadores han tomado la talla a los dos años para determinar la talla que una persona tendrá a los 18 años.
4. Validez de constructo : se refiere al grado en que una medición se relaciona consistentemente con otras mediciones de acuerdo con hipótesis derivadas teóricamente y que conciernen a los conceptos que están siendo medidos. Un **constructo** es una variable medida que tiene lugar dentro de una teoría o esquema teórico. En el campo de la actividad física, podría ser el caso de las

⁴ American Psychological Association (APA). 1994. Publication manual of the American Psychological Association. 4 ed. Washington, DC: Author.

pruebas que miden una misma variables: test para la potencia anaeróbica. El test de Margaria puede ser validado con pruebas de velocidad en 50 o 60 metros, con el salto alto o con el salto largo.

Se puede realizar validez de constructo cuando, al realizar mediciones de una determinada variable, luego de una intervención sobre dicha variable, el instrumento también capta las nuevas variaciones producidas en la variable en estudio. Igualmente, si el instrumento es capaz de diferenciar muy bien los sujetos, se dice que tiene validez de constructo.

¿Qué factores que pueden afectar la confiabilidad y la validez?

- a) La improvisación. Dada la condición de confiabilidad y validez, ningún instrumento de medición se debe diseñar a la ligera o improvisadamente. Los investigadores experimentados tardan algún tiempo en diseñar o seleccionar el instrumento a emplear en sus mediciones.
- b) Uso de instrumentos no validados. Todo instrumento de medición debe ser reconocido y aprobado por la comunidad científica respectiva. Así, cuando se determinar la masa de un sujeto en una balanza, dicho instrumento debe tener algún grado de reconocimiento el cual viene junto con las instrucciones del aparato. De no ser así, el instrumento debe ser comparado, o validado que es lo mismo, con un instrumento previamente reconocido.
- c) Instrumento no adecuado para quien lo utiliza. En el caso de una medición en bicicleta ergométrica, cuando se quiere evaluar niños, dicha bicicleta debe tener la posibilidad de adecuarse a las medidas antropométricas de ellos.
- d) Condiciones o normas con las cuales se debe realizar la medición. La mayoría de los protocolos de medición en las ciencias aplicadas a la actividad física y el deporte están estandarizados para ser realizados con unas determinadas condiciones. En el test de Abalakov (salto con ambas piernas sin impulso de los brazos y sin contramovimiento), el deportista no debe mover el tronco y la articulación de la rodilla debe presentar un ángulo de 90°. Si no se cumplen dichas condiciones, la medición no tiene validez.

2.1.3 Control o validez interna de la situación experimental.

El término control deriva etimológicamente del inglés *counter-roll*, que significa lista principal con la que toda lista subsiguiente debe ser contrastada y si fuera preciso, corregirla⁵. De aquí, se asume el término como el intento de eliminar el efecto de variables extrínsecas a la investigación, de tal modo que pueda existir la

⁵ Sierra Bravo, R. 1996. Tesis doctorales y trabajos de investigación. 4 ed. Madrid: Paraninfo

GRamónS

seguridad de que los efectos hallados son debidos únicamente a la variable independiente por probar o probada.

Se puede deducir de lo anterior que el control tiene una estrecha relación con las variables externas. De hecho, cualquier variable externa que influya en la investigación y no sea controlada (neutralizada o eliminada), producirá efectos no deseados y contaminantes del verdadero papel que cumple la variable independiente. Spector (1982)⁶ afirmaba que el control es la piedra angular del diseño experimental, pero se puede extender a todo tipo de diseño.

Algunas expresiones al respecto del control son:

- ⇒ Cuando hay control se puede conocer o estimar la relación causal; cuando no se logra el control, no se puede conocer dicha relación.
- ⇒ Lograr control en un experimento es controlar la influencia de otras variables extrañas en las variables dependientes, para que se pueda de esta manera saber si las variables independientes que interesan tienen o no efecto sobre las dependientes.

¿Existen procedimientos para lograr control en un experimento?

Según Sierra Bravo⁷, existen los siguientes procedimientos para establecer control:

- a) La eliminación. Si una variable externa incide de manera significativa en un experimento, se debe eliminar. Si el hecho de ser mujer influye de manera significativa, se elimina dicha característica y luego se estudia en otra situación experimental.
- b) Mantener constantes. Una manera de eliminar el efecto de una variable no estudiada (interviniente) es manteniendo constante su efecto. Para el caso de variables físicas, es mantener la Temperatura del aire constante así como el grado de humedad, la hidratación, la nutrición, el mismo grado de actividad por fuera del experimento, entre otros ejemplos.
- c) Aleatorización. Consiste en elegir por sorteo o por medios aleatorios (azar) los sujetos para la conformación de los grupos. Si se tienen varios grupos, se debe asignar un proceso aleatorio para la pertenencia a cada uno de ellos. La razón de la aleatoriedad radica en el hecho de que de acuerdo a las leyes del azar, si un proceso es asignado de dicha manera, con un nivel aceptable de probabilidad no existirán diferencias significativas en las distintas distribuciones y asignaciones resultantes; dicho de otra manera, las variables extrañas no controladas serán distribuidas de manera similar, no afectando a algún grupo en particular.
- d) Igualación. La presencia de variables extrañas en todos los grupos o situaciones de la investigación debe ser similar o igual. Una manera de igualar el ordenando los sujetos de mayor a menor (o viceversa) y asignarlos en orden a los grupos. Si se

⁶ Spector, P.E. 1982. Research design. 2 ed. Beverly Hills: Sage Pub.

⁷ Sierra Bravo, opus cit

GRamónS

tienen dos grupos, el primero pertenecerá al grupo 1, el segundo al grupo 2; el tercero al grupo 1 y así sucesivamente.

- e) Repetición. La repetición controla las variables de los sujetos puesto que si son los mismos sujetos y las mismas situaciones experimentales, los resultados deben ser los mismos. Se pueden hacer repeticiones de diversos tratamientos/mismos sujetos o grupos, mismo tratamiento/diversos grupos, o mismo tratamiento/mismos grupos o sujetos.
- f) Grupo de control. Se busca un grupo en el cual no será sometido a la influencia de la variable independiente, grupo denominado de control. Este grupo deberá poseer las mismas características del grupo experimental. Se presupone que las variables extrañas actuarán similarmente en ambos grupos, por lo que las diferencias entre los resultados observados en ambos grupos se pueden atribuir a la variable independiente, que ha actuado solo en el grupo experimental.

Fuentes de invalidación interna⁸:

1. Maduración de los participantes. En el ámbito de la actividad física y el deporte, y sobre todo cuando se estudian niños en fase de crecimiento o desarrollo, estos factores por si solos explican fenómenos como el incremento de la fuerza, de la resistencia o de otras capacidades condicionales. Las modificaciones que se sucedan en el grupo experimental no pueden generalizarse porque una variable interviniente como es la maduración de los participantes no se controló.
2. Inestabilidad de mediciones, participantes y/o componentes. Si las personas que realizan las mediciones no tienen una preparación previa, es decir, no se han *estandarizado*, las mediciones no tendrán validez. Lo mismo puede suceder si no se garantizan la calidad de los implementos o instrumentos de medición. No está por demás destacar la influencia que pueden tener las variaciones del día en las mediciones fisiológicas: la secreción de hormonas (tipo cortisol) no es la misma durante el día (ciclo circadiano), por lo tanto si en una investigación acerca de fuerza, unos sujetos se miden por la mañana y otros por la tarde, significa que hay inestabilidad en las mediciones.
3. Administración de pruebas. Se deben seguir los mismos protocolos de administración de las pruebas a todos los sujetos. Resaltar aquí la influencia que puede tener el calentamiento sobre la medición de las capacidades condicionales. Para el caso de medición de la resistencia física, no se obtiene el mismo resultado cuando se realiza un calentamiento que sin calentamiento; cuando se realiza un calentamiento, no debe realizarse al azar (todos lo hacen al gusto) que seguir un protocolo de calentamiento. No es lo mismo medir en un laboratorio que medir en una pista de atletismo. En general, se deben seguir los

⁸ Cambell y Stanley (1966), Cambell (1975) y Babbie (1979).

procedimientos establecidos para la medición de las variables en todos los sujetos, con alto grado de rigurosidad.

4. Cambios en los instrumentos o métodos de medición. Si a unos sujetos se mide con un instrumento y otros sujetos se miden con otro, se debe probar que ambos instrumentos miden lo mismo con un alto grado de confiabilidad. De lo contrario, los resultados que se obtengan no son válidos.
5. Métodos de selección. En los casos experimentales, la selección de los sujetos a dos grupos experimentales debe hacerse por el método del azar. Este método de selección garantiza que los dos grupos son equivalentes, es decir, que tienen una composición semejante. En la selección de la muestra de un diseño descriptivo, la manera como se seleccione, pues cuando la muestra es intencionada, se puede incurrir en un sesgo, o lo que es lo mismo, que la muestra examinada tenga valores mayores o menores de la población.
6. Mortalidad experimental. Se debe tener presente, sobre todo en los estudios experimentales que los sujetos de los grupos no abandonen el experimento, evento denominado muerte experimental, o en el caso de que la tengan, que sea similar.

El experimentador como fuente de invalidación interna:

- No debe interactuar con los sujetos de investigación
- No debe manipular los resultados de la investigación.
- Los sujetos no deben conocer los objetivos de la investigación, las hipótesis, las condiciones experimentales, porque pueden alterar los resultados.

¿Cómo se logra el control y la validez interna?

- * Seleccionando varios grupos de comparación (dos como mínimo). Al momento de iniciarse (equivalencia inicial) un experimento se debe garantizar una equivalencia que no es individual sino grupal. Se logra fundamentalmente realizando una asignación aleatoria o al azar.
- * Durante el experimento deben observarse y tenerse en cuenta las mismas instrucciones para todos los grupos. Se deben estandarizar las personas que toman parte en la investigación.

¿Qué es la validez externa ?

Se refiere a qué tan generalizables son los resultados de un experimento a situaciones no experimentales y a otros sujetos y poblaciones. La pregunta fundamental es: ¿Lo encontrado en la investigación, a qué grupos, contextos o situaciones se aplica o se puede aplicar?

GRamónS

Las siguientes son fuentes de invalidación externa :

1. Efecto reactivo o de interacción de las pruebas (preprueba). Cuando se intenta medir varias capacidades condicionales, no se pueden medir todas el mismo día. Así, si se desea medir tanto la fuerza máxima como el máximo consumo de oxígeno el mismo día o una después de otra, en la medición de fuerza (suponiendo que se realiza primero) puede lograr mediciones confiables pero en la medición de consumo de oxígeno, dado que la prueba máxima de fuerza agotó reservas energéticas, la medición obtenida no es confiable por el efecto de cansancio o agotamiento producido por la otra prueba.
2. Interacción entre los errores de selección y el tratamiento experimental. Si un criterio de selección, para una investigación que intenta ver el efecto de un plan de entrenamiento de fuerza en personas mayores, fue el no haber realizado entrenamiento previos o no se atleta de rendimiento y dicho criterio no se cumple a cabalidad, puede suceder que ingresen al grupo experimental sujetos con un alto grado de desarrollo de la fuerza de manera que el plan de entrenamiento propuesto no surta el debido efecto, motivo por el cual se pierde el objetivo del experimento.
3. Efectos reactivos de los tratamientos experimentales (condiciones experimentales). El experimentador debe hacer que los sujetos se olviden de que están en un experimento y no deben sentirse observados.
4. Interferencia de tratamientos múltiples (Tratamiento con efecto irreversible). Se debe garantizar, en procesos de investigación relacionados con actividad física y deporte, que los sujetos sólo realicen el plan de entrenamiento propuesto y no realicen mas actividad en lugares como gimnasios, clubes, etc.
5. Imposibilidad de replicar los tratamientos. Si el tratamiento o el procedimiento de medición es tan complejo y difícil de realizar, no se puede cumplir el postulado de reproducibilidad de todo proceso de investigación por lo cual se invalida el experimento o la investigación.

Los **EXPERIMENTOS VERDADEROS** tienen las siguientes características:

- Tienen grupo control
- Tienen grupos de comparación
- Tienen equivalencia de grupos

Tipos de experimentos verdaderos:

1. Únicamente con postprueba y grupo control:

$$\begin{array}{l} RG_1 \quad X \quad O_1 \\ RG_2 \quad - \quad O_2 \end{array}$$

No se hace medición previa puesto que la asignación por azar garantiza la igualdad de los grupos. Algunos autores afirman que no es necesario la medición inicial puesto que la medición previa puede ser un sesgo.

Se pueden establecer más de dos grupos

Entre las desventajas se enumeran dos: 1) si la aleatoriedad no fue apropiada, los grupos no son equivalentes, 2) no se puede usar la previa para asignar las personas a los grupos.

Este tipo de diseño logra controlar todas las fuentes de invalidación interna.

2. Con preprueba-postprueba y grupo control

$$\begin{array}{l} RG_1 \quad O_1 \quad X_1 \quad O_2 \\ RG_2 \quad O_3 \quad X_2 \quad O_4 \\ RG_3 \quad O_5 \quad - \quad O_6 \end{array}$$

Dado que tanto el grupo control como los experimentales son asignados al azar, todos los grupos son equivalentes. Por lo tanto, las variaciones en la variable dependiente se pueden atribuir solo a la variable dependiente.

La preprueba puede servir para confirmar la aleatoriedad empleada, sobre todo en grupos menores de 15 sujetos, en los cuales la aleatoriedad no funciona muy bien.

Este tipo de diseño también logra controlar todas las fuentes de invalidación interna.

3. Cuatro grupos de Solomon

$$\begin{array}{l} RG_1 \quad O_1 \quad X \quad O_2 \\ RG_2 \quad O_3 \quad - \quad O_4 \\ RG_3 \quad - \quad X \quad O_6 \\ RG_4 \quad - \quad - \quad O_6 \end{array}$$

Este tipo de diseño, creado por Salomón (1949) es una suma de los dos tipos de diseños citados anteriormente:

Este tipo de diseño tiene dos grupos experimentales (G_1 y G_3) y dos grupos de control (G_2 y G_4). El objetivo de este tipo de diseño es analizar la influencia que puede tener la medición previa sobre los resultados, al comparar los grupos 1 y 2, con los grupos 3 y 4. Se puede también analizar la efectividad del tratamiento al comparar los grupos 1 y 2. Si se desea analizar la influencia de la prueba previa,

GRamónS

se comparan los grupos 1 y 3. La gran desventaja de este tipo de diseño es su gran requerimiento de grupos y por tanto de dinero, tiempo y recursos.

4. De series cronológicas múltiples (antes o después de X)

RG_1	X_1	O_1	O_2	O_3
RG_2	X_2	O_4	O_5	O_6
RG_3	-	O_7	O_8	O_9

RG_1	O_1	O_2	O_3	X_1	O_4	O_5	O_6
RG_2	O_1	O_2	O_3	X_2	O_4	O_5	O_6
RG_3	O_1	O_2	O_3	-	O_4	O_5	O_6

Cuando se está interesado en medir efectos a largo plazo, aparecen los diseños cronológicos. Debido a que no se conoce cuando se producen o desaparecen los efectos, se realizan mediciones periódicas luego de realizado el experimento. Las mediciones se pueden llevar a cabo antes del tratamiento. Estas mediciones aseguran que los valores previos al experimento no variaron.

Se pueden presentar los siguientes casos:

RG_1	10	10.5	10.2	X_1	14	16	18
RG_2	11	11.2	11.1	X_2	14	12	10
RG_3	10	11	10.5	-	10.8	10	10.5

En el grupo 1, la mediciones previas fueron muy similares. Luego del tratamiento, el efecto de produce de manera progresiva en el transcurso del tiempo. En el grupo 2, igualmente las mediciones previas son muy similares. Luego del experimento, las mediciones incrementaron pero rápidamente tendieron a disminuir. En el grupo 3 las mediciones, tanto previas como posteriores al experimento no variaron.

5. Diseños factoriales

En este tipo de diseños, el investigador manipula dos o mas variables independientes e incluyen dos o mas niveles de presencia en cada una de ellas. La denominación de este tipo de diseños se hace por el número de variables independientes y el número de niveles de cada una de ellas. Como ejemplos tenemos: a) un diseño 2×2 , quiere decir que se emplean dos variables independientes con dos niveles, mas específicamente, podría ser una investigación de fuerza con dos tipos de entrenamiento en los cuales se emplea una diferencia de género (hombres y mujeres). b) un diseño 3×3 quiere decir que se emplearán tres tipos de variables independientes con tres categorías o

GRamónS

niveles., c) 3 X 2, d) 4 x 2. Como se puede ver, los números no necesariamente tienen que ser iguales.

En este tipo de diseño, el número de grupo se calcula multiplicando los números que lo componen: para un diseño 2 x 2, el total de grupos experimentales es de 4; para un diseño 3 x 3 x 2, el total de grupos será 18. La denominación de cada uno de estos grupos se puede hacer mediante la metodología propuesta por Christensen (2000)⁹. La asignación de los sujetos a los diferentes grupos debe hacerse al azar. En el cuadro 2 se muestran varios ejemplos.

Cuadro 2. Diseños factoriales 2 x 2 y 3 x 2, con la denominación de los grupos.

	A ₁	A ₂		A ₁	A ₂	A ₃
B ₁	A ₁ B ₁	A ₂ B ₁	B ₁	A ₁ B ₁	A ₂ B ₁	A ₃ B ₁
B ₂	A ₁ B ₂	A ₂ B ₂	B ₂	A ₁ B ₁	A ₂ B ₂	A ₃ B ₂

La gran ventaja de este tipo de diseños radica en el hecho de que le permite al investigador evaluar los efectos de cada variable independiente sobre la dependiente por separado, además de evaluar los efectos de las variables independientes conjuntamente. Hay dos tipos de efectos a evaluar: los efectos de cada variable independiente, también llamados efectos principales y los efectos de interacción entre dos o más variables independientes.

En el cuadro 3 se muestra un ejemplo de posibles resultados de un diseño 3x2 en el cual se pretende estudiar el efecto que produce un plan de entrenamiento de fuerza a intensidades de 20%, 50% y 80%, sobre la variable género (hombres-mujeres). Los resultados se dan en unidades relativas (peso levantado / peso corporal). En los casos A y B se dan solo efectos principales sin efectos interactivos; en los casos C y E, solo efectos interactivos; en el caso F, no existen efectos.

⁹ Christensen, L.B. (2000). *Experimental methodology*. 8 ed. Boston, Mass.: Allyn and Bacon, Inc.

Cuadro 3. Ejemplos de resultados en un diseño factorial 3x2, en el cual se pretende analizar el efecto de un plan de entrenamiento de fuerza con tres niveles de intensidad sobre la variable género. Los resultados se presentan en forma relativa (Peso levantado / peso corporal). Y se analizan los efectos principales e interacciones.

A				B			
Intensidad				Intensidad			
	20%	50%	80%		20%	50%	80%
Hombres	2.00	1.50	1.00	Hombres	2.00	2.00	2.00
Mujeres	2.00	1.50	1.00	Mujeres	1.00	1.00	1.00
La variable intensidad produce efectos, mientras que la variable género no produce efectos. Existen efectos principales sin efectos interactivos.				La variable intensidad no produce efectos, mientras que la variable género produce efectos. Existen efectos principales sin efectos interactivos.			
C				D			
Intensidad				Intensidad			
	20%	50%	80%		20%	50%	80%
Hombres	2.00	1.50	1.00	Hombres	2.00	1.75	1.50
Mujeres	1.00	1.50	2.00	Mujeres	1.7	1.50	1.25
Ninguna de las variables tiene efectos por si misma. No existen efectos principales pero si efectos interactivos.				Las dos variables tienen efecto pero no hay interacción			
E				F			
Intensidad				Intensidad			
	20%	50%	80%		20%	50%	80%
Hombres	2.00	1.50	1.00	Hombres	0.5	0.6	0.55
Mujeres	1.75	1.25	1.00	Mujeres	0.58	0.59	0.54
Las dos variables producen efectos y existen efectos interactivos.				No existe algún tipo de efecto			

2.2. Cuasiexperimentos.

En los diseños denominados cuasiexperimentales, existe una manipulación intencionada de una o varias variables independientes con el objeto de observar y analizar el efecto que produce sobre una variable dependiente, pero la asignación de los grupos y de los sujetos a los grupos no se realiza al azar, siendo esta última característica la que los diferencia de los diseños verdaderos. Los cuasiexperimentos difieren de los experimentos verdaderos en la equivalencia inicial de los grupos.

La asignación de los grupos se hace por la situación real, es decir, son grupos intactos. En términos reales, se puede hacer un estudio con escolares de manera que los grupos se conforman de acuerdo a la distribución que posee la escuela, es decir, se toman tal cual están en la realidad; en el campo del deporte, sería un estudio comparativo entre las selecciones de fútbol que posee una liga.

Estos diseños se utilizan cuando no es posible asignar los sujetos en forma aleatoria a los grupos que recibirán los tratamientos experimentales. Esta falta de Aleatorización produce problemas de validez interna y externa. Según Weiss (1990)¹⁰, estos diseños deben luchar con la selección como fuente posible de equivocación o de error. Campbell y Stanley (1966) afirman que "precisamente por la falta de control experimental total, es imprescindible que el investigador conozca a fondo cuáles son las variables particulares que su diseño específico no controla. Así, estará mas pendiente de su posible influencia y tendrá mejores elementos para evaluarla". Cuando el investigador no encuentre razones para equiparar los grupos, debe limitarse a describir y/o correlacionar los resultados, pero no podrá hacerlo con fines explicativos.

A pesar de las teóricas desventajas de aleatorización, el método cuasiexperimental es indispensable por la siguiente razón: permite explorar temas que de otra manera no podrían explorarse debido a cuestiones éticas, morales y/o prácticas. Veamos algunos ejemplos:

- Diferencias entre el desarrollo del equilibrio entre niños normales y niños con problemas del oído interno,
- Diferencias entre un plan de actividad aeróbica en sujetos que con sobrepeso y obesos.
- Diferencias en el desarrollo motor en niños y niñas con síndrome Down.

En todos estos proyectos, la asignación a los grupos debe hacerse de forma real puesto que no podríamos someter a niños sanos a daños en el oído para estudiar el efecto de la actividad física, o tomar unos niños y sobrealimentarlos para que se vuelvan obesos y luego estudiarlos bajo esta condición. Estas condiciones naturales de "anormalidad" permiten estudiarlos por lo que los individuos ya han sufrido el padecimiento, motivo por el cual también se conocen como estudio "expost facto" (después del hecho).

DISEÑOS CUASIEXPERIMENTALES

- En este tipo de diseño, el suceso que según la hipótesis, causa las diferencias entre los grupos, ya ocurrió.
- Los grupos son "intactos", es decir, ya en la realidad pertenecen a una determinada categoría o variable.
- Tienen grupo control.

¹⁰ Weiss, C.H. (1990). Investigación evaluativa: Métodos para determinar la eficiencia de los programas de acción. México: Trillas.

¿Qué tipo de diseños cuasiexperimentales existen?

Los diseños cuasiexperimentales pueden ser del mismo tipo que los enumerados para los diseños experimentales. Como ya se estableció, la gran problemática es la asignación de los sujetos al azar, en lo demás, los tipos de diseño son los mismos:

- 1). Con posprueba y grupos intactos:
 $G_1 \quad X \quad O_1$
 $G_2 \quad - \quad O_2$
- 2). Con preprueba-postprueba y grupo control
 $G_1 \quad O_1 \quad X_1 \quad O_2$
 $G_2 \quad O_3 \quad X_2 \quad O_4$
 $G_3 \quad O_5 \quad - \quad O_6$
- 3). De series cronológicas múltiples (antes o después de X)
 $G_1 \quad X_1 \quad O_1 \quad O_2 \quad O_3$
 $G_2 \quad X_2 \quad O_4 \quad O_5 \quad O_6$
 $G_3 \quad - \quad O_7 \quad O_8 \quad O_9$

2.3. Preexperimentos.

Los preexperimentos son la clase de diseños experimentales que se caracterizan por el grado mínimo de control. Es decir, no tienen grupo control y la asignación de los grupos y de los sujetos no se hace al azar. Por estas características, no son adecuados para el establecimiento de relaciones explicativas entre una variable independiente y otra dependiente. Se usan en estadios exploratorios de experimentos verdaderos y su interpretación es muy cautelosa debido a la invalidez tanto interna como externa que presentan.

Este tipo de experimentos se ven en la práctica diaria. Un profesor que trata de validar un método de enseñanza, toma unos sujetos intencionadamente (no al azar) y no compara sus resultados con otro grupo control. Posteriormente defiende los resultados ante la comunidad académica. ¿Este procedimiento cumple con los requisitos de la ciencia y de los protocolos experimentales?, ¿Puede tener algún grado de confiabilidad este estudio?

3. ¿Cómo se selecciona la muestra en un diseño experimental?

Para seleccionar una muestra, en primera instancia se debe delimitar la población. Para delimitar la población se debe tener claro el planteamiento del problema. La población debe precisarse en cuanto a sus características, el lugar y/o el tiempo. Un ejemplo de delimitación de una posible población sería: la población comprende todos aquellos sujetos de género masculino, que pertenezcan a los clubes deportivos de fútbol, que estén reconocidos como tales ante la respectiva liga, en edades

GRamónS

comprendidas entre los 15 y 20 años, que hayan competido en al menos un campeonato municipal. Una vez definida la población, se realiza un censo o se cuentan los sujetos que pertenecen a dicha población. Para nuestro ejemplo, se contaron 1243 niños, que cumplen dichos requisitos.

Dado que en algunos casos no se puede someter a toda la población a la variable independiente, se debe seleccionar una muestra, definida ésta como un subconjunto de la población. La problemática generada ahora con la muestra es que, los resultados obtenidos con dicha muestra se puedan generalizar a la población, concepto definido como **generalizabilidad**.

En el campo de la ciencia, aparecen dos maneras de obtener una muestra: muestreo probabilístico y no probabilístico. Estos muestreos hacen relación a la probabilidad que tiene cada sujetos elegido en la muestra. En el muestreo probabilístico, todos los sujetos tienen la misma probabilidad de ser elegidos, mientras que en el no probabilístico no se conoce la probabilidad de ser elegido o todos no tienen la misma probabilidad de serlo. Como hemos visto en las características de los diseños experimentales, una de las grandes características de los experimentos verdaderos es que la selección sea al azar, es decir, una selección probabilística en la cual todos los sujetos tengan la misma probabilidad de ser elegidos.

¿Qué maneras existen para realizar un muestreo probabilístico?

Existen las siguientes maneras de realizar un muestreo probabilístico:

- a) Muestreo aleatorio simple
- b) Muestreo sistemático
- c) Muestreo estratificado

Muestreo aleatorio simple

Aquí, cada miembro de la población tiene una probabilidad igual e independiente de ser seleccionado como parte de la muestra. Con esta premisa se está buscando que la población y la muestra tengan las mismas características, basado en el teorema de la estadística que dice la media de las medias de una población, elegidas al azar, es idéntica a la media de la población.

El proceso consta de cuatro pasos:

- 1). Definir la población
- 2). Listar todos los miembros de la población
- 3). Asignar números a cada miembro de la población
- 4). Aplicar un criterio para seleccionar la muestra.

El mejor criterio del paso 4 es una tabla de números aleatorios, en la cual se cumple con el criterio de igualdad e independencia de la selección. Otra opción es

GRamónS

colocar los rótulos de toda la población en una bolsa e ir sacando los sujetos al azar.

☑ Muestreo sistemático.

En este muestreo se escoge al azar un número (k-ésimo) que relaciona el tamaño de la muestra y la población. Si la población son 50 y la muestra son 10, pues el k-ésimo es 5. Se escoge al azar un sujeto de la lista y se cuentan los siguientes 5 sujetos y dicho sujeto formará parte de la muestra. Se repite este procedimiento hasta que sean seleccionados los 10 sujetos.

Este tipo de muestreo es más fácil de llevar a cabo pero viola el precepto de que cada miembro debe tener la misma probabilidad e independencia de ser elegido. Cuando se elige cada quinto, la independencia puesto que se rige por los k-ésimos.

☑ Muestreo estratificado

En algunas oportunidades, la conformación de la población tiene características especiales que merecen ser preservadas en la conformación de la muestra. Por ejemplo, una población tiene el 40% de hombres y el 60% de mujeres, por lo que la muestra también debe conservar dicha proporción. A este tipo de situaciones, se les llama muestreo estratificado.

¿Qué tan grande debe ser la muestra?

La muestra debe ser lo suficientemente grande como para que represente la población. La selección del tamaño se hace por procedimientos estadístico-matemáticos. Hernández y col (2003) proponen el siguiente procedimiento:

- a) Se debe conocer el tamaño de la población (N)
- b) Se debe establecer el error estándar (se), es decir, el error que aceptamos cometer al realizar el estudio. Dicho valor se da como una fracción de 1 y se interpreta como se hace con la probabilidad. Un error estándar de 0.1 quiere decir que aceptamos cometer un error en diez casos o 10 en cien. Dicho error estándar lo fija el investigador.
- c) Se determina la varianza de la población, la cual es igual a se^2
- d) Se determina la varianza de la muestra, la cual es igual a $p(1-p)$. Para el caso, p es la probabilidad de ocurrencia del evento. También el investigador escoge el nivel con el cual desea trabajar.
- e) Se calcula el tamaño de la muestra sin ajustar (n') mediante la siguiente ecuación:

$$n' = \frac{s^2}{V^2}; \quad s^2 = p(1-p); \quad V^2 = se^2$$

GRamónS

f) Se calcula el tamaño de la muestra (n) mediante la siguiente ecuación:

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

En un caso hipotético, se tienen 945 deportistas y se quiere seleccionar una muestra con un error estandar (se) de 0.015. ¿Cuál es el tamaño de la muestra.

- a) N = 945
- b) error estandar (se) = 0.015
- c) V = 0.015²
- d)

$$s^2 = 0.9(1 - 0.9) = 0.09; \quad V^2 = 0.015^2 = 0.000225$$

$$n' = \frac{0.09}{0.00025}$$

$$n' = 400$$

e) Se calcula el tamaño de la muestra (n) mediante la siguiente ecuación:

$$n = \frac{400}{1 + \frac{400}{945}} = \frac{400}{1 + 0.423} = \frac{400}{1.423} = 281$$

Aunque este tipo de procedimientos se realizan en la selección de una muestra, no es concebible que en un diseño experimental se pueda trabajar con una muestra de 280 personas o sujetos. En los diseños experimentales en el campo de la actividad física y el deporte, en los cuales se somete a un grupo de personas a un tratamiento, el cual debe tener un alto grado de control, el tamaño de la muestra se hace de manera que la muestra no sobrepase un determinado número o que sea un mínimo. Así, una muestra no sobrepasa los 20 sujetos pero no se menor de 10 sujetos.

4. ***¿Cómo se puede concluir que un determinado tratamiento produjo efectos positivos en los sujetos experimentales?***

El este aparatado se analizan las técnicas estadísticas para comparar el efecto del tratamiento sobre dos grupos, tal como sería el caso del efecto de diferentes tipos de entrenamiento sobre dos grupos. Se presenta la prueba más simple de comparación entre dos grupos como lo es el *t Test*. También se describe el análisis de la varianza como una manera de probar la diferencia significativa entre dos o más

GRamónS

grupos. Finalmente, también se describe el uso del análisis factorial de la varianza, en los cuales dos o más variables independientes pueden ser comparadas.

En el proceso de investigación experimental se plantean las hipótesis del mismo. Estas hipótesis generalmente son de dos tipos: a) la hipótesis nula (H_0) y b) la hipótesis alterna (H_A).

La H_0 afirma que no existieron efectos significativos de la variable independiente sobre la dependiente. Es decir, que los resultados no variaron. Una manera de ver esta hipótesis desde la óptica matemática-estadística es:

$$H_0 \rightarrow \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$$

Donde μ_1 y μ_2 son las medias de los valores al inicio y al final del experimento.

La H_A afirma que si existieron diferencias entre los resultados al inicio del experimento y al final del mismo. La simbología matemática-estadística es

$$H_A \rightarrow \mu_1 \neq \mu_2 \rightarrow \mu_1 - \mu_2 \neq 0$$

Donde μ_1 y μ_2 son las medias de los valores al inicio y al final del experimento.

El propósito de la estadística es evaluar la hipótesis nula para un nivel dado de probabilidad (por ejemplo, $p < 0.05$). En otras palabras, hacer la siguiente pregunta: ¿dados dos niveles de diferente tratamiento, cuál es la probabilidad de que dichas diferencias ocurran por azar en un 5 por ciento de las veces?. Todo lo que la estadística puede hacer es aceptar o rechazar la hipótesis nula. La estadística no puede aceptar la hipótesis de investigación. Solamente un razonamiento lógico, un buen diseño experimental y una teorización apropiada pueden hacerlo. La estadística sólo puede determinar si dos grupos son diferentes, pero no porqué son diferentes.

Siempre que se estima un parámetro poblacional mediante una muestra, se hacen preguntas como. «¿Qué calidad tiene la estimación obtenida? ¿Cabe llegar a la conclusión de que el parámetro poblacional es idéntico al estadígrafo muestral, o es posible que exista alguna discrepancia? Si es así, ¿Cuál es la magnitud de ésta discrepancia?». Para responder a cada una de estas preguntas, se deben comparar los resultados muestrales con los resultados «esperados». Los resultados esperados están dados por la distribución muestral apropiada. La siguiente pregunta es: ¿Cómo es la distribución de un estadígrafo particular? ¿Cómo se puede conocer la forma de la distribución, y por tanto, cuáles son los resultados esperados? Puesto que las inferencias que se hacen implican el conocimiento de la forma de la distribución muestral, es necesario establecer ciertos modelos idealizados. La curva normal y la distribución binomial son dos modelos cuyas propiedades matemáticas se conocen. En consecuencia, estas dos distribuciones se emplean frecuentemente como modelos para describir determinadas distribuciones muestrales. Así por ejemplo, si se sabe que la distribución muestral de determinado estadígrafo toma la forma de una distribución normal, se pueden usar las propiedades conocidas de esta distribución para hacer inferencias y predicciones acerca del estadígrafo.

Puesto que rara vez se puede estudiar exhaustivamente una población, se recurre a tomar muestras de dicha población para llegar a una hipótesis referente a las características o parámetros de la población. Cuando se realiza una encuesta con cien personas, el deseo no es conocer lo que piensan esas cien personas sino realizar una estimación acerca de toda la población. Para comprender cómo se pueden hacer inferencias acerca de una población a partir de una muestra es necesario analizar el concepto de *distribución muestral*.

Una distribución muestral es aquella distribución teórica de probabilidades de los valores posibles de una prueba estadística (estadígrafo) muestral que ocurriría si se obtuvieran todas las muestras posibles, de un tamaño fijo, de una población dada.

En la práctica, las inferencias acerca de los parámetros de una población se hacen basándose en estadígrafos que se calculan mediante una muestra de N observaciones tomadas al azar de esta población. Si se tomaran, de manera indefinida muestras de tamaño N de esta población, no sería de extrañar encontrar algunas diferencias entre los valores de los estadígrafos muestrales obtenidos. Sería muy raro encontrar siempre el mismo conjunto de observaciones.

Hay una distribución muestral para cada estadígrafo, tal como la media, la desviación estándar, etc. Este es uno de los conceptos más importantes de la estadística inferencial. En la tabla 1 se muestra un ejemplo de distribuciones tomado de Runyon y Haber. Se trata de distribuciones muestrales basadas en la extracción, con reposición, de todas las posibles muestras de N=2, N=3 y N=4 de una población de cuatro números (0,3,6,9).

Tabla 1. Distribuciones muestrales correspondientes a medias de muestras de N=2, N=3, N=4 tomadas de una población de números (0,3,6,9; $\mu=4.5$, $\sigma=3.94$).

N = 2		N = 3		N = 4	
Media	p(Media)	Media	p(Media)	Media	p(Media)
9.0	0.0625	9.0	0.0156	9.00	0.0039
7.5	0.1250	8.0	0.0469	8.25	0.0156
6.0	0.1875	7.0	0.0938	7.50	0.0391
4.5	0.2500	6.0	0.1562	6.75	0.0781
3.0	0.1875	5.0	0.1875	6.00	0.1211
1.5	0.1250	4.0	0.1875	5.25	0.1562
0.0	0.0625	3.0	0.1562	4.50	0.1719
		2.0	0.0938	3.75	0.1562
		1.0	0.04969	3.00	0.1211
		0.0	0.0156	2.25	0.0781
				1.50	0.0391
				0.75	0.0156
				0.00	0.0039
Media = 4.5		Media = 4.5		Media = 4.5	
DS = 2.37		DS = 1.94		DS = 1.68	

GRamónS

Se extrae al azar una muestra de cuatro números de una cierta población y se obtiene una media de 6,0. Ahora la pregunta es. ¿Constituye esta media un suceso ordinario o un suceso raro? Si no se dispone de un marco de referencia, la pregunta carece de sentido. Si por el contrario, se dispone de una distribución muestral y sabiendo que esta distribución es la presentada en la tabla 1, cuando $N=4$. La respuesta es que se obtendrá un 12% ($p = 0.1211$) de las veces. De la misma manera, una media mayor o igual a 6.00 se obtendrá aproximadamente un 26% de las veces ($p = 0.1211 + 0.0871 + 0.0391 + 0.0156 + 0.0039 = 0.2578$).

En este mismo ejemplo, si una media de 7,50 aparece un 5% de las veces, ¿éste resultado es confiable? o por contrario, ¿resulta de una manipulación externa?. La probabilidad según la distribución es de 3.91% y la diferencia con 5% es muy pequeña, por lo que se podría responder afirmativa o negativamente. Entonces, aparece otra pregunta, ¿dónde se traza la línea que determina la inferencia que hagamos del experimento?

La respuesta a esta pregunta revela la naturaleza básica de la ciencia: su orientación es más bien probabilística que absoluta. En las ciencias sociales, la mayoría de los investigadores han adoptado como base para inferir la intervención de factores no aleatorios, uno de los dos puntos de separación:

1. **El nivel de significación de 0.05 o nivel del 5%:** Cuando el suceso, u otro que representa una mayor desviación, ocurre por azar el 5% o menos de las veces, algunos investigadores están dispuestos a afirmar que los resultados se deben a factores ajenos al azar.
2. **Nivel de significación de 0.01 o nivel del 1%:** Cuando el suceso, u otro que representa una mayor desviación, ocurre por azar el 1% de las veces o menos, otros investigadores están dispuestos a afirmar que los resultados se deben a factores ajenos al azar.

Si retornamos al ejemplo anterior de una media de 7.50 ocurra un 5% de las veces, usando el nivel de significación del 0.05 se debe deducir que el experimento está manipulado ($p < 0.05$) o que está fuera de los límites posibles. Pero si se escoge un nivel de 0.01, el resultado es posible, puesto que $p > 0.01$.

De este ejemplo se establece una gran premisa del diseño de experimentos: no es posible que se realice un estudio, se analicen los resultados, se llegue a un valor de probabilidad y luego se decida el nivel de significación. El nivel de significación debe especificarse previamente al estudio como una parte general del diseño del experimento.

† Test o t de Student:

Cuando se conoce la media de una población pero no su desviación estándar (σ_x), se puede calcular este estadígrafo a partir de datos muestrales. En consecuencia, la σ_x estimada (s_x) debe basarse en la σ estimada que es \hat{S} , de manera que:

$$s_{\bar{X}} = \frac{\hat{S}}{\sqrt{N}}$$

Si se sustituye \hat{S} por σ , se obtiene una aproximación razonablemente buena de la distribución de medias muestrales. A comienzos del siglo XX, William Gosset, un estadístico que se firmaba con el seudónimo de Student, observó que la aproximación de \hat{S} por σ era deficiente, sobre todo para muestras pequeñas, debido a que \hat{S} tiende a subestimar a σ más de la mitad de las veces. La mayor contribución de Gosset consistió en la determinación del cociente t:

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$$

donde μ_0 es el valor de la media poblacional bajo la hipótesis nula.

El cociente t de Student es un examen fuerte, lo que significa que las inferencias estadísticas son posiblemente válidas aun cuando la distribución poblacional se aparte mucho de la normalidad. Por lo tanto, si se tienen dudas acerca de la normalidad de la población, es aconsejable aumentar el tamaño N de la muestra.

Características de la distribuciones t:

Si se comparan las distribuciones t con la curva normal estándar, se tiene que:

1. Ambas distribuciones son simétricas respecto a una media de cero. Por tanto, la proporción del área situada a la derecha de un valor positivo particular de t es igual a la proporción a la izquierda del correspondiente valor negativo de t.
2. Las distribuciones t son más dispersas que la curva normal. En consecuencia, la proporción de área más allá de un valor específico de t es mayor que la proporción de área mas allá del valor correspondiente de z. Sin embargo, cuanto mayores sean los grados de libertad (N-1), mayor será la semejanza entre las distribuciones t y la curva normal estándar.

En la gráfica 1 se muestra las distribuciones muestrales de una t de Student con 3 y 10 grados de libertad (gl), comparadas con una distribución Z.

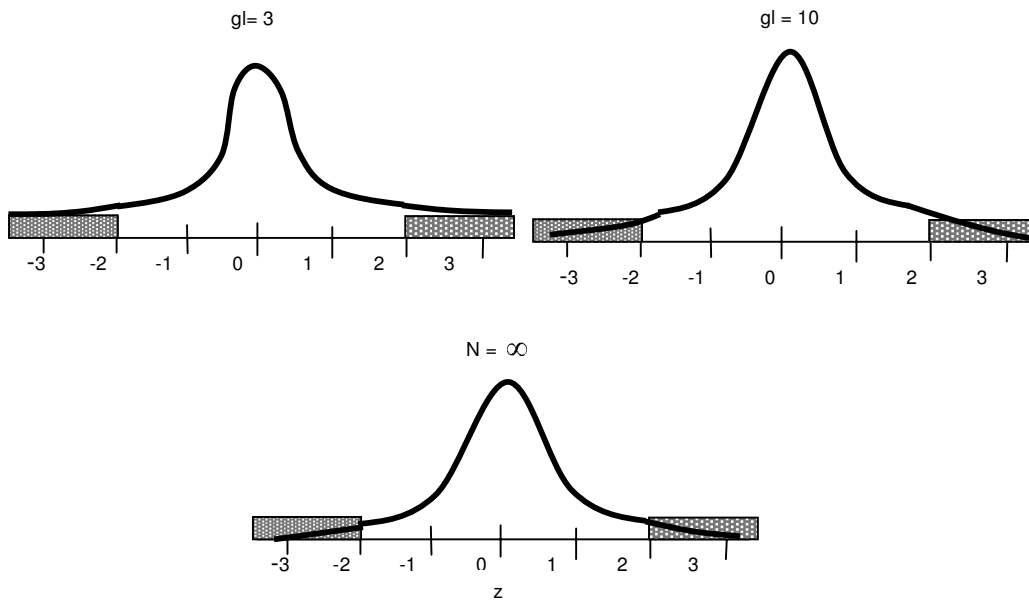


Gráfico 1. En la parte superior se muestran distribuciones muestras de t Student con 3 y 10 grados de libertad (gl). Abajo, una distribución de Z.

Al examinar la gráfica 1, se puede apreciar que, en la curva normal estándar, $|z| \geq 1.96$ define la región de rechazo al nivel de significación de 0.05. Sin embargo, cuando $gl = 3$, una región descrita por $|t| \geq 1.96$ incluye aproximadamente el 15 por ciento del área total. Por lo tanto, si se emplea la curva normal para examinar estadísticamente hipótesis cuando N es pequeño y σ desconocida, se cometería un error de tipo I, es decir, rechazar H_0 siendo verdadera. Por otra parte, se requiere un valor mucho mayor de t para determinar los límites de la región crítica o de rechazo, así, para $gl = 3$, el valor absoluto de la t obtenida debe ser igual o mayor que 3.18 para rechazar H_0 al nivel de significación 0.05 (examen bilateral o de dos colas). Sin embargo, cuando gl crece, las diferencias en proporciones de área bajo la curva normal y bajo la distribución t se hacen insignificantes.

Los valores de t encontrados en las tablas son valores críticos, es decir, son aquellos valores que limitan las regiones de rechazo correspondientes a varios niveles de significación.

Pruebas t de Student o t Test.

1. t Test para muestras independientes.
2. t Test para muestras dependientes.

1. † Test para muestras independientes.

En este tipo de casuística, se tienen inicialmente dos grupos independientes sometidos a diferentes tipos de tratamiento.

Uno de los requisitos para iniciar un experimento es que los dos grupos sean equivalentes. Para cumplir este requisito, existe el siguiente procedimiento estadístico (Rogers, Howard y Vessey, 1993; Westlake, 1983):

1º) Calcular la desviación estándar de la diferencia entre las medias de los grupos:

$$s_{n_1-n_2} = \sqrt{\left[\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \right] \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

donde n es el número de sujetos, s es la desviación estándar de cada uno de los grupos.

2º) Calcular el valor de z para la hipótesis nula, que para el caso, debe ser que los dos grupos no son equivalentes.

$$z = \frac{(M_1 - M_2) - (X_{dif})}{s_{m_1-m_2}}$$

donde M es la media de cada uno de los grupos,

X es la diferencia establecida como aceptable (generalmente el 10% del valor de la media del primer grupo).

Realicemos un ejemplo ilustrativo.

Se tienen dos grupos de jóvenes en edades entre 12 y 14 años. Se desea saber si los dos grupos son equivalentes en su masa corporal. El primer grupo tiene una media de 36.7 kg (3.67) y el segundo tiene 39.7 kg (3.08). Ambos grupos tienen 24 sujetos. La diferencia permitida en masa es de 3.7 kg (10% de 36.7 kg).

Al aplicar las fórmulas obtenemos:

$$s_{n_1-n_2} = \sqrt{\left[\frac{(24-1)3.67^2 + (24-1)3.08^2}{24+24-2} \right] \left[\frac{1}{24} + \frac{1}{24} \right]} = 0.977$$

$$z = \frac{(36.7 - 39.7) - (3.7)}{0.977} = -0.749$$

Este valor de z, tiene una p de 0.45. Por lo tanto, se acepta la hipótesis nula. En otras palabras, los dos grupos no son equivalentes.

GRamónS

Retornando al caso de comparar dos grupos independientes, de las dos muestras se calcula la media y la desviación estándar de cada grupo. Para calcular la *t* de Student se emplea la siguiente fórmula:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1} - s_{\bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{SC_1 + SC_2}{N_1 + N_2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Veamos un caso hipotético. Se tienen dos grupos de 20 jóvenes, entre los 10 y 12 años, a los cuales se les aplicaron dos métodos de entrenamiento de la movilidad articular (estático y dinámico) sobre la articulación de la cadera, al realizar el movimiento de extensión. Los resultados (ver cuadro 7.1) para el grupo 1 fueron los siguientes: una media de ganancia en movilidad articular fue 10 cm con una desviación estándar de 1.08. Para el grupo 2, la media de ganancia en movilidad articular fue de 9 cm con una desviación estándar de 1.69.

La hipótesis nula para esta situación establece que la diferencia entre las medias de los grupos 1 y 2 es igual a cero. La hipótesis alterna es que la diferencia no es igual a cero, es decir, que las medias son estadísticamente diferentes.

Para establecer la diferencia, establecemos un nivel de significancia de 0.05. La distribución muestral será la *t* de Student con $gl = N_1 + N_2 - 2 = 20 + 20 - 2 = 38$. La región crítica será $|t| \geq 2.021$ y puesto que es un estudio no direccional de dos colas, la región consta de todos los valores de *t* tales que $|t| \geq 2.021$ y $|t| \leq -2.021$. El valor hallado para la *t* fue de 2.24. Este valor está por fuera de la región crítica, por cuanto se rechaza la hipótesis nula y se acepta la hipótesis alterna. Es decir, el método empleado con el grupo 1 es mejor para incrementar el rango de movilidad articular de la cadera, de jóvenes entre 10 y 12 años.

• Homogeneidad de la varianza:

No podemos olvidar que para tener seguridad de los resultados, se deben realizar pruebas acerca de la homogeneidad de la varianza, es decir, que la varianza de los dos grupos sea equivalente. Aunque pequeñas violaciones de este aspecto no representan un problema mayor, las violaciones más grandes son más comunes si los grupos no son aproximadamente iguales. Para el ejemplo citado, vemos que sus varianzas son diferentes, aunque el número de sujetos en cada muestra es igual. Los programas actuales de estadísticas poseen pruebas como la de Levene para determinar si la varianza es o no homogénea. Para el caso, el estadístico de Levene tiene un valor de 1,76 con $p = 0.192$, por lo que se acepta que la varianza de los dos grupos es semejante.

GRAMÓN S

Cuadro 7.1. Valores hipotéticos encontrados en dos grupos experimentales (X₁, X₂). En la parte inferior se muestran los cálculos de la media, la desviación estándar y de la t de Student.

N	X ₁	X ₁ ²	X ₂	X ₂ ²
1	12	144	10	100
2	11	121	9	81
3	10	100	8	64
4	9	81	9	81
5	11	121	10	100
6	10	100	5	25
7	9	81	8	64
8	9	81	12	144
9	9	81	12	144
10	10	100	11	121
11	11	121	10	100
12	10	100	10	100
13	12	144	9	81
14	9	81	8	64
15	8	64	7	49
16	10	100	9	81
17	11	121	9	81
18	9	81	7	49
19	10	100	8	64
20	10	100	9	81
Suma	200	2022	180	1674
SC	22.00		54.00	
Varianza =	1.2		2.8	
Media =	10.00		9.00	
DS =	1.08		1.69	
$t = ((10 - 9) / ((22 + 54) / 38) (1/20) (1/20))^{0.5}$ $t = 2.24 \quad t^2 = 5.01$ $w^2 = (t^2 + 1) / (t^2 + N1 + N2 - 1) = 5.01 + 1 / 5 + 38 = 0.13$				

Frecuentemente una diferencia significativa en las varianzas (en particular cuando la varianza del grupo experimental es mayor que la del grupo control) indica un efecto dual de las condiciones experimentales, es decir, que a unos sujetos les produce un efecto positivo y a otras, uno negativo. Una mayor varianza indica más valores extremos en las dos colas de una distribución. El investigador atento tomará estos hechos como base para explorar la posibilidad de que existan efectos duales.

- Estimación del grado de asociación entre la variable experimental y la variable dependiente:

La obtención de un valor estadísticamente significativo implica que existe algún grado de asociación entre la variable experimental y la variable dependiente. Sin embargo, el hecho de que un resultado sea estadísticamente significativo no concede

GRamónS

importancia de manera automática a lo hallado. Sin N es suficientemente grande, incluso una diferencia trivial puede resultar estadísticamente significativa.

Una manera de establecer la significancia de una diferencia estadísticamente significativa, es mediante la aplicación de la medida de asociación ω^2 (omega cuadrado). Esta medida se calcula mediante la siguiente fórmula:

$$\omega^2 = \frac{t^2 - 1}{t^2 + N_1 + N_2 - 1}$$

donde t = t Student, N_1 = número de sujetos de la muestra 1 y N_2 = número de sujetos de la muestra 2.

El resultado puede interpretarse de la misma forma que el coeficiente de determinación (r^2), en el cual un valor de 0.50 equivale a decir que el 50% de la varianza de la variable dependiente ha sido tomada en cuenta por las variaciones en X. De la misma forma, ω^2 se puede considerar como el porcentaje en el cual la variable dependiente explica los valores hallados o el porcentaje de la varianza en la variable dependiente que ha sido tomada en cuenta en función de las variaciones en la variable del tratamiento.

Para el presente ejemplo, el resultado se ha presentado en la parte inferior del cuadro 7.1. El valor de w^2 fue de 0.13, es decir, tan sólo un 13% de la variable dependiente es tomada en cuenta en la variable experimental.

Aunque este ejemplo no ilustra muy bien el hecho, se debe tener presente que un valor significativo estadísticamente en la t no siempre tiene una significancia real.

2. t Test para muestras dependientes.

En los apartados 1 y 2 se han examinado los casos en los cuales se compara un grupo con la media poblacional y el caso en que se comparan dos grupos independientes. Un tercer caso es aquel en el cual se compara un grupo al cual se le ha realizado una medida inicial (pretest) y otra medida final (postest), luego de aplicársele un tratamiento. Estamos entonces ante el caso de t Test para muestras dependientes.

Para calcular la t de Student de este caso, uno de los métodos empleados es el denominado *Diferencia Directa* (Runyon, Harber, 1992; Tomas y Nelson, 1996).

El método de la diferencia directa consiste en hallar las diferencias de las calificaciones obtenidas antes y después para cada pareja de sujetos y tratar estas diferencias como si fuesen calificaciones originales. Mediante este procedimiento, se transforma un caso de dos muestras en un caso de una muestra. Se halla la media muestral y un error estándar basado en una desviación estándar. La hipótesis nula es que la media obtenida de la diferencia entre las calificaciones ($\Sigma D/N = MDD$) proviene

GRamónS

de una distribución muestral de diferencia de medias (la población = s_{DD}) en que la diferencia media tiene algún valor específico.

Veamos un ejemplo real, en el cual se sometieron 12 sujetos en edades comprendidas entre los 18 y 22 años, a un entrenamiento pliométrico con el ánimo de observar su efecto sobre la fuerza de miembros inferiores. Para medir este efecto, se empleó una alfombrilla de salto, que permite calcular la altura del salto realizado desde una flexión de 90 de rodillas (SJ) o en forma de contramovimiento (CMJ). Los resultados se muestran en la tabla 7.2.

Tabla 7.2. Altura del salto (cm) en SJ y SQJ, de un grupo de 20 sujetos de sujetos sometidos a entrenamiento pliométrico.

Sujeto	SJ				Sujeto	CMJ			
	PRETEST	POSTEST	DD	DD ²		PRETEST	POSTEST	DD	DD ²
1	37.6	35.7	-2.0	3.8	1	39.2	40.1	1.0	0.9
2	28.1	40.0	11.9	141.6	2	40.3	40.6	0.3	0.1
3	36.9	36.8	-0.1	0.0	3	40.9	38.6	-2.4	5.6
4	37.0	35.6	-1.4	2.0	4	38.7	37.7	-1.0	1.0
5	38.0	37.3	-0.7	0.5	5	39.4	38.4	-1.0	1.0
6	36.0	38.5	2.5	6.0	6	38.4	43.3	4.8	23.1
7	34.0	34.2	0.2	0.0	7	37.1	38.2	1.2	1.3
8	36.5	39.8	3.3	10.9	8	40.0	44.5	4.5	20.4
9	41.9	46.2	4.3	18.3	9	42.9	47.2	4.3	18.7
10	39.1	40.1	1.0	0.9	10	41.9	41.2	-0.7	0.5
11	31.5	32.1	0.6	0.3	11	33.9	33.2	-0.7	0.5
12	33.9	34.9	1.0	0.9	12	36.2	40.3	4.1	16.8
		MDD	1.70				MDD	1.20	
		Suma	20.42	185.45			Suma	14.36	89.96
		Suma ²	416.98				Suma ²	206.21	
		SC	150.70				SC	72.77	
		s_{DD}	1.07				s_{DD}	0.74	
		t	1.59				t	1.61	

El planteamiento del problema, en términos estadísticos formales es el siguiente:

1. La hipótesis nula es que la diferencia entre medias (pretest y postest) es igual a cero.
2. La hipótesis alterna es que la diferencia es diferente de cero.
3. La prueba estadística empleada es la t de Student para muestras correlacionadas.
4. Nivel de significancia es de 0.05
5. La distribución muestral es la t de Student con $gl = N - 1 = 11$.
6. La región crítica, dado que no es una hipótesis direccional en un determinado sentido, el estudio será de dos colas donde el valor de t es: $t_{(0,05)} = t \leq 2.201$ y $t \geq -2.201$

GRamónS

$$t = \frac{MDD}{s_{DD}} = \frac{MDD}{\sqrt{\frac{\sum DD^2 - \frac{(\sum DD)^2}{N}}{N(N-1)}}$$

La t de Student para el SJ será:

$$t = MDD / s_{DD}$$

$$MDD = (\sum DD/N) = 20.42 / 12 = 1.70$$

$$s_{DD} = (SC_{DD} / (N-1))^{0.5} = ((\sum DD^2 - \sum (DD)^2/N) / (N(N-1)))^{0.5} = ((185.45 - (416.98 / 12 * 11)) / 11)^{0.5}$$

$$s_{DD} = 1.07$$

$$t = 1.70 / 1.07 = 1.59$$

La t de Student para el CMJ, calculada mediante el mismo procedimiento fue de 1.61.

Como conclusión, se acepta la hipótesis nula. Es decir, bajo las condiciones del experimento, el entrenamiento pliométrico no aumenta de manera significativa la fuerza de miembros inferiores.

t Test y Potencia en Investigación.

La potencia puede ser definida como la probabilidad de rechazar la hipótesis nula cuando la hipótesis es falsa. El obtener potencia en investigación es muy deseable.

La siguiente es la fórmula para el cálculo de la t de Student para muestras independientes:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Se puede apreciar que es un quebrado donde el denominador es otro quebrado. En este caso, para incrementar el valor de t se debe aumentar el numerador y/o disminuir el denominador.

El numerador es la diferencia entre las medias de los grupos analizados. Para incrementar la diferencia entre estas medias, la respuesta es de sentido común: se deben emplear tratamientos de una manera acertada, de manera intensiva y concentrada, no dejando al azar ninguna de las variables intervinientes. Así, si se aplica dos semanas un tratamiento, los resultados no serán tan buenos como si le aplicamos tres semanas o cuatro. De esta manera, las diferencias entre el inicio y el final del tratamiento serán mayores.

El denominador son las varianzas de cada grupo, así como el número de sujetos. Si las variaciones llegan a ser muy pequeñas (cerca de la media), la varianza también es

GRamónS

muy pequeña. La manera de obtener menos variaciones en los efectos del tratamiento es aplicándolos de una manera consistente y racional. Entonces, si el numerador permanece constante y la varianza se hace pequeña, la t aumenta su valor.

El tercer factor es el número de sujetos, que es el denominador del denominador. Entonces, aumentando el número de sujetos, el quebrado del denominador disminuye y el resultado final es que la t aumenta de valor. Por lo tanto, una manera de incrementar la potencia aumentando el número de los sujetos de una investigación.

Un último factor para incrementar la potencia es disminuyendo el valor de alfa, pero como ya se estudió, el disminuir el valor de alfa aumenta la probabilidad de errores tipo II, es decir, aceptar hipótesis falsas.

La relación de la t de Student tiene un numerador y un denominador. Desde un punto de vista teórico, el numerador puede ser considerado como la varianza verdadera y el denominador, el error de la varianza. Si no existiesen diferencias entre los dos grupos, las dos varianzas serían iguales y la relación sería igual a 1. Cuando se encuentra una t cuyo valor es considerado como significativo, lo que realmente se está diciendo es que la varianza verdadera excede al error de la varianza hasta un cierto grado. La cantidad que debe exceder de uno para ser significativo depende del número de sujetos y del nivel de significancia establecido.

La estimación de la fortaleza de la relación entre la variable dependiente y la independiente se representa por la relación entre la varianza verdadera y la varianza total. Es decir, $w^2 = \text{Varianza verdadera} / \text{Varianza total}$.

¿Se pueden realizar estos cálculos en un programa de computador como Excell?

Afortunadamente los computadores traen incorporados algunos programas para manejar datos, las llamadas hojas de cálculo para manejar números o para bases de datos, que permiten realizar todos los anteriores cálculos.

Una hoja de cálculo en el programa Excell está estructurada en forma de cuadrícula en la cual, en las columnas aparece en su parte superior una letra (desde A hasta IV, mientras que las en filas aparecen numeradas en el lado izquierdo, iniciando en 1 y terminando en 65.536. De esta manera, cada cuadro o celda tiene una denominación que combina la letra de la columna y el número de la fila.

Entre los íconos que tiene esta hoja en su parte superior (barras de estado o barra estandar) aparece una f_x que quiere decir fórmulas o funciones. Al hacer clic en este ícono, aparece una ventana con el título de ***Pegar función*** y dos ventanas: categoría de la función y nombre de la función. En la categoría de la función aparece una categoría que es la de las funciones estadísticas. Al hacer clic en esta categoría aparecen en el lado derecho todas los estadísticos que calcula. Para el caso de la t de Student, se busca ***PRUEBA. T.*** Al hacer clic sobre esta función, aparece otra ventana que tiene cuadro íconos: Matriz1, Matriz2, Colas y Tipo. Cada uno de estos íconos tiene un espacio en blanco (para escribir los rangos de celdas solicitados) y al final de este espacio un icono con colores: rojo en el centro, azul y blanco en los extremos.

GRamónS

Para el caso de la Matriz1, cuando se hace doble clic en este icono, aparece una ventana sola que permite seleccionar de la hoja de cálculo el rango de celdas en los que se encuentra los resultados del pretest o del primer grupo. Para hacer la selección, solo tiene que hacer clic en el primer dato y mantener el botón derecho del mouse hasta el último dato. Una vez seleccionado el rango de datos, se vuelve a hacer clic en la ventana de colores. Se realiza el mismo proceso para la Matriz2.

En la ventana de Colas, el programa la suministra una ayuda: una cola de distribución = 1; dos colas de distribución = 2. En este caso, como hemos visto en las hipótesis, la hipótesis alterna puede ser limitada a dos casos: 1) cuando las medias son mayores o menores, denominadas unidireccionales o también 1 cola, o 2) cuando las medias son diferentes, es decir que puede ser mayor o menor simultáneamente, bidireccionales, o estudios de 2 colas. Entonces, basado en las hipótesis que se hayan planteado, usted escoge la opción 1 o 2, y la escribe en la ventana.

En la ventana de Tipo, la ayuda le dice: pareado = 1; dos muestras con igual varianza = 2; dos muestras varianza distinta = 3. Como hemos visto, un estudio con igual varianza es una prueba pretest-postest con el mismo grupo. Un estudio con dos grupos distintos es un estudio con dos grupos de diferente varianza. De acuerdo a la situación específica del experimento, usted deberá introducir la opción respectiva.

Una vez haya introducido todos los datos, debe hacer clic en aceptar. El resultado de este procedimiento es el cálculo de la probabilidad de que las medias de los dos conjuntos de datos puedan ser iguales o no. Para la interpretación de esta probabilidad se debe tener en cuenta el nivel de significancia que se haya planteado. Si el nivel de significancia es ≤ 0.05 , entonces cualquier valor menor que este nivel de significancia se interpretará como que fue "estadísticamente significativo", es decir, que existe diferencia estadística entre las series de datos, por lo que se acepta la hipótesis alterna y se rechaza la nula. En el caso de la probabilidad se mayor o igual que 0.05, se acepta la hipótesis nula (no existen diferencias entre las medias) y se rechaza la alterna. Por otra parte, si el nivel de significancia es 0.01, si la probabilidad es menor de dicho valor, entonces se acepta al hipótesis alterna; si el valor es mayor de 0.01, se acepta la hipótesis nula.

GRamónS

Ejemplo práctico.

Tabla 1. Datos ficticios de un grupo experimental en el cual se estudia el efecto que un plan de entrenamiento pliométrico produce sobre la saltabilidad de un grupo de jugadores de voleibol.

	A	B	C	D
1				
2		Salto alto (cm)		
3	Sujeto	Preprueba	Postprueba	
4				
5	1	280	282	
6	2	288	290	
7	3	295	300	
8	4	300	298	
8	5	297	300	
10	6	301	300	
11	7	300	399	
12	8	287	297	
13	9	298	300	
14	10	297	300	
15				
16	Media	294,3	306,6	
17				
18	t Student	0,26		

En una hoja de Excell, introduzca los datos como se muestran en la tabla 1. Como se puede apreciar, la media del grupo en la preprueba fue de 294.3 cm y en la posprueba fue de 306.6.

Para calcular la t de Student, coloque el ratón o el indicador en la celda B18. Una vez situado en dicha celda, haga clic en el ícono f_x y busque la ventana de categoría estadísticas. Luego busque PRUEBA.T y haga doble clic en ella. En la Matriz1 deberá introducir B5:B14 o seleccionar estos datos en la hoja. En la Matriz2, selecciona el rango C5:C14. En el espacio para Colas, escriba 2. En el espacio para tipo, introduzca 2, que es el caso para dos grupos con la misma varianza.

El resultado del cálculo es 0.26. Con base en este dato y asumiendo un nivel de significancia de 0.05, se acepta la Hipótesis nula de igualdad de las medias y se rechaza la alterna de diferencia entre ellas.