UNIVERSIDAD DEL SALVADOR
FACULTAD DE CIENCIAS SOCIALES
Estadística y Técnicas de Computación Aplicadas a las Ciencias Sociales

LA PRUEBA DE T DE STUDENT PARA DIFERENCIA DE MEDIAS Y PROPORCIONES

Horacio Chitarroni*

1. El caso de las muestras independientes

¿Ayuda a conseguir trabajo haber terminado el secundario?

Los comienzos laborales de los jóvenes suelen ser difíciles. Cuando se busca el primer empleo se carece de experiencia previa y eso es una dificultad. El capital con que se afrontan las primeras búsquedas de trabajo, se supone que es el que se ha adquirido en la educación formal. Hoy día, se supone que contar con, al menos, el secundario completo, es casi imprescindible. Pero también se dice que esa es una condición necesaria pero no suficiente para una búsqueda exitosa.

¿Tardan efectivamente menos en encontrar trabajo los jóvenes que han completado la educación media que aquellos que no lo han hecho?

El cuadro que sigue forma parte de una investigación más amplia sobre las trayectorias laborales juveniles realizada en un momento de elevada incidencia del desempleo y sugiere una respuesta al interrogante arriba planteado.

Y curiosamente –como sucede tantas veces– la evidencia empírica contraría nuestras presunciones. En vez de conseguir empleo más rápido, los jóvenes que completaron el secundario tardan, en promedio, un poco más en hacerlo. ¿Es que no conviene estudiar?

Cuando las cosas no son como esperábamos, hay que buscarles explicaciones: plantear hipótesis. Una posible es que los jóvenes que han completado el secundario provengan de hogares de ingresos más altos, donde su trabajo no es tan imprescindible y pueden permitirse búsquedas más prolongadas, a la espera de que surjan oportunidades más acordes con sus pretensiones. Otra, sugiere que cuando han completado estudios medios, los jóvenes ponen mayores condiciones a sus búsquedas y no aceptan cualquier empleo. Hay formas de indagar en estas hipótesis...

Pero bien mirado, la diferencia no es grande: apenas dos meses. Y los datos provienen de una muestra: bien podría ser que todo fuera efecto de la casualidad. Antes de seguir adelante, sería bueno comprobar si no es así...

* Profesor asociado.

Jóvenes desocupados de 18 a 24 años: duración media del desempleo (en meses)
Según nivel educativo
Total de aglomerados urbanos

Jóvenes desocupados de 18 a 24 años	Hasta secundario incompleto	Secundario completo y más
tiempo medio de búsqueda (meses)	7,3	9,1
Desvío estándar	7,867	8,665
Tamaño muestral	910	810

Fuente: EPH-INDEC (octubre de 2002)

Los diseños experimentales y la hipótesis de nulidad

La situación se asemeja no poco a la de los llamados *diseños experimentales*, en donde un grupo (el grupo *experimental*) es sometido a un cierto *tratamiento*, del cual se espera que produzca un cierto efecto. Y se lo compara con otro grupo (grupo de *control*) que difiere del anterior en que no ha recibido el tratamiento. Se compara en ambos grupos el valor que asume una cierta variable dependiente, sobre la que se supone que ejerce algún efecto el *tratamiento*. Si fuera así, esta variable debiera mostrar diferencias entre ambos grupos: el *experimental* y el de *control*. Claro que en estos diseños hay que precaverse de que los grupos no difieran en otros aspectos, ajenos a haber sido objeto o no del *tratamiento*, y que pudieran influir sobre la variable dependiente bajo observación. Normalmente, esto resulta garantizado por el hecho de que estos grupos hayan sido obtenidos como muestras independientes de la misma población, o bien que resulten de la partición por métodos de azar de una sola muestra. Recién luego de conformados estos grupos en forma aleatoria, debiera introducirse el *tratamiento* –también denominado usualmente *estímulo*– en uno de ambos y no en el otro.

En el caso de los jóvenes que buscan trabajo tenemos dos grupos: los que han completado el secundario y los que no lo han hecho. Y pretendemos observar si la titulación ejerce algún efecto sobre el tiempo que dura el desempleo. Pero aquí los grupos no fueron conformados al azar sino que difieren en haber completado o no el secundario, que es una condición previa de los jóvenes. Y ello permite suponer que también puedan diferir en otros aspectos influyentes sobre el tiempo que duran sus búsquedas laborales, tal como se lo ha hipotetizado más arriba.

Pero en cualquier caso es importante descartar la posibilidad de que ni siquiera se trate de una verdadera diferencia, sino de un azar del muestreo: es decir que si pudiéramos indagar a toda la población, los tiempos de búsqueda de empleo no difirieran. Esto es, precisamente, lo que afirmará en este caso la **hipótesis nula** (H_0) que debiéramos poder descartar. Es decir, la hipótesis nula dirá que en las respectivas poblaciones de jóvenes desocupados con y sin secundario completo, el tiempo medio que duran las búsquedas de empleo no difiere. Que la diferencia observada en la muestra es una jugarreta del azar...

La distribución de muestreo de la diferencia de medias

La estructura de esta prueba resulta de una extensión del teorema del límite central y la ley de los grandes números, expuestos en un punto anterior. Habíamos visto allí que si se obtuvieran de una población repetidas muestras al azar de un tamaño suficiente, las medias de las muestras acabarían por distribuirse normalmente con una media equivalente a la de la población y con un desvío estándar que también resultaría igual al poblacional, pero dividido por la raíz cuadrada del tamaño muestral.

Supongamos ahora que tuviésemos no una sino dos poblaciones. Y que fuésemos obteniendo sendas muestras, tomadas de a pares, una de cada población. Y así como lo hacemos calculáramos las medias de estas muestras en una variable cualquiera (en nuestro ejemplo, el tiempo de búsqueda de empleo), para luego obtener la diferencia entre las medias de una y otra muestra. Cuando termináramos de hacer esto con cada par de muestras, tendríamos una larga colección de diferencias.

Si hubiéramos respetado el orden, siempre obteniendo primero la media de la muestra proveniente de la población 1 (la de los jóvenes sin secundario, por caso) y luego la media de la muestra de la población 2 (los jóvenes con secundario, en nuestro ejemplo), para luego restar la segunda de la primera, estas diferencias serían algunas veces positivas (cuando la primera media fuera mayor que la segunda) y a veces negativas (en la situación inversa). Pero estas diferencias se distribuirían a su vez en torno a una media que sería equivalente a la diferencia existente entre las medias de las respectivas poblaciones.

Y si fuera cierta la hipótesis nula (H_0), que afirma que las medias poblacionales no difieren entre sí (es decir que la diferencia $\mu_1 - \mu_2 = 0$), entonces las diferencias de medias muestrales se distribuirían en torno a una media equivalente a cero.

Esa distribución, además, se convierte en una distribución normal toda vez que el tamaño sumado de las dos muestras que se comparan alcanza o supera 120 casos. Pero cuando los tamaños son menores, se trata de una distribución diferente que se denomina, precisamente, distribución de t de Student[1].

La distribución de muestreo de la t de Student nos permite saber con precisión qué proporción de las diferencias de medias de muestras pueden alejarse de la diferencia entre las medias poblacionales (que sería igual a cero bajo la hipótesis nula) en cierta cantidad de desvíos estándar de la propia distribución. Y si las muestras son lo suficientemente grandes como para normalizar la distribución, entonces, será cierto que el 95% de las diferencias de medias no podrían superar el valor de +- 1,96 desvíos estándar, si efectivamente la diferencia entre las medias poblacionales ($\mu_1 - \mu_2$) fuera cero. En otros términos, no se alejarían de cero en más de 1,96 veces el valor del desvío.

Cuando las muestras no alcanzan el tamaño necesario par normalizar la distribución, entonces la cantidad de desvíos a un lado y otro de la media entre los que quedan encerradas el 95% de las diferencias son algo mayores, pues se trata de una distribución más aplanada. Puede vérsela en la tabla III del anexo I.

<u>Estimación del desvío estándar de la distribución de medias muestrales y cálculo de t</u>

El desvío estándar de la distribución de las diferencias de medias de muestras independientes puede ser estimado a partir de las varianzas de las mismas muestras, toda vez que lo habitual es desconocer las varianzas poblacionales. La expresión de ese desvío estándar es:

[1] Quien la estudió y desarrolló matemáticamente utilizó ese seudónimo para publicar sus trabajos: *Student.*

$$\sqrt{\left(\frac{s2(1)}{n(1)-1}+\frac{s2(2)}{n(2)-1}\right)}$$

Donde:

$s^2_{(1)}$ y $s^2_{(2)}$ son las varianzas de las muestras 1 y 2

$n_{(1)}$ y $n_{(2)}$ son los tamaños de las muestras 1 y 2 (se les resta 1 para convertirlos en los grados de libertad)

Si esta es la estimación del desvío estándar de la distribución muestral, entonces bastará dividir la diferencia hallada entre ambas medias por esa expresión, para hallar el valor de t de Student:

$$t = \frac{\overline{X(1)} - \overline{X(2)}}{\sqrt{\left(\frac{s2(1)}{n(1)-1}+\frac{s2(2)}{n(2)-1}\right)}}$$

Donde:

$\overline{X(1)}$ y $\overline{X(2)}$ son las respectivas medias de la muestra 1 y la muestra 2

Ahora podremos valernos de esta fórmula para dar respuesta al interrogante referido a la duración del desempleo en los jóvenes con y sin secundaria completa. No tendremos más que reemplazar en ella:

$$t = \frac{7{,}3-9{,}1}{\sqrt{\left(\frac{61{,}9}{910}+\frac{75{,}1}{810}\right)}} = -4{,}449$$

Al aplicar la fórmula, en este caso hemos omitido descontar un grado de libertad de cada muestra, tal como aparecía en la fórmula original. La razón para esa omisión consiste en que con tamaños muestrales superiores a 150 grados de libertad (calculados como $n_1 + n_2 - 2$) la distribución de muestreo se hace equivalente a la normal y no hace falta ya hacerlo así.

La tabla de t: cómo leerla. Pruebas unilaterales y bilaterales

El valor calculado de t debe compararse con un valor crítico, que es preciso igualar o superar para rechazar la hipótesis nula. Ese valor podemos hallarlo en la tabla de la distribución de t, disponible usualmente en los textos de estadística[2]. La estructura de esta tabla se asemeja a la ya vista de la distribución de chi cuadrado. En la primera columna de la izquierda tenemos los grados de libertad (n_1+n_2-2). Veremos que cuando sobrepasamos los 120 grados de libertad los

[2] Por ejemplo: GARCIA FERRANDO, Manuel (1985) *Socioestadística*. Madrid: Alianza Universidad.

valores de la tabla se igualan a la distribución normal: el valor crítico es 1,96 para una significación –que se busca en los cabezales de las columnas– de 0,05 (es decir, para una confianza de 95%).

Sin embargo, al mirar esto último advertiremos que la tabla de t tiene una doble fila de cabezales: la primera para una prueba "de dos colas" (o bilateral) y la segunda para una prueba "de una sola cola" (o unilateral). Si examinamos con más cuidado, veremos que hay un corrimiento: el valor crítico de t que corresponde a 0,05 de significación en la prueba unilateral se corresponde con 0,10 en la bilateral. Por ejemplo, un valor obtenido de t = 1,64 alcanzaría a rechazar la hipótesis nula si usáramos la primera alternativa pero no con la segunda. ¿De qué depende el uso de una u otra?

Pues depende de que al plantearnos la prueba, hubiéramos podido o no anticipar el sentido o signo de la diferencia. ¿Qué se quiere decir con esto? Pues bien, al comparar las duraciones medias del desempleo entre jóvenes con y sin secundario, no sabíamos decir con certeza cuál sería mayor. Inclusive, hasta nos sorprendió que tardaran más en conseguir trabajo los más calificados, aunque luego formuláramos una hipótesis para explicarlo. Es decir que al restar una media de la otra, no estábamos en condiciones de anticipar el signo de la diferencia. Cuando esto es así, tenemos que usar una prueba bilateral, porque no sabemos de qué lado de la media cero que plantea la hipótesis nula saldrá la qué nosotros obtengamos: tanto puede ser negativa (hacia el lado izquierdo de la curva) como positiva (hacia el lado derecho). Y en estas condiciones, el 5% de diferencias de medias muestrales (si fuera ese el nivel de riesgo de error de tipo I asumido) que nos inducirían a rechazar equivocadamente la hipótesis nula se repartiría equitativamente a los dos lados de la curva.

Pero supongamos que hubiéramos tenido una presunción firme al respecto y hubiéramos esperado encontrar más tiempo desempleados a los jóvenes con mayor educación. En ese caso también esperaríamos que la segunda media fuese mayor que la primera y que, en consecuencia, el signo de la diferencia resultase negativo, tal como ocurrió. De ser así, ¡nos habríamos ganado media curva! Pues predijimos que la diferencia se situaría del cero hacia la derecha y así fue. Por debajo del cero se ubican el 50% de las diferencias posibles. Luego, para llegar a 95% necesitamos adicionar 45% más. Y para el 45% de área necesitaremos alejarnos menos de la media hacia la derecha que si tuviéramos que llegar al 47,5%, dejando el margen de error repartido en ambas colas. Por ejemplo, cuando la curva se normaliza necesitaríamos 1,96 desvíos para llegar a 47,5%, pero solamente 1,64 para llegar a 45%. Es decir que si empleamos una prueba unilateral podremos rechazar la hipótesis nula con un valor de t de Student algo menor. ¡Es que la osadía de plantear una hipótesis más precisa, que además de esperar una diferencia anticipe su sentido, tiene un premio...! Claro que tal hipótesis hubiera resultado refutada sin más si el signo de t hubiera resultado opuesto al esperado, independientemente de su valor absoluto.

Es que, ya lo decía Popper, cuanto mayor contenido informativo tenga una proposición, tanto más expuesta estará a ser refutada. Pero a la inversa, si no es potencialmente refutable es porque no dice nada: no forma parte de las proposiciones científicas.

<u>Los supuestos de normalidad y homocedasticidad</u>

Pero la aplicación del test de t de Student exige el cumplimiento de supuestos de *normalidad* y *homocedasticidad*: lo primero significa que en las poblaciones las distribuciones de la variable que se compara debieran ser aproximadamente normales. En cuanto a lo segundo, significa que, también en la población, las varianzas de las duraciones del desempleo entre jóvenes con y sin secundario debieran ser similares.

Por cierto que, en general, es poco lo que podemos saber acerca de las poblaciones de las que provienen las muestras que comparamos. Siempre nos manejamos con datos muestrales. Pero de todas formas, mirando las muestras uno puede inferir la distribución poblacional: ¡en eso consiste el juego!

Si la variable considerada (en el ejemplo, la duración del desempleo) mostrara una distribución muy alejada de la normalidad en las muestras –por caso, muy asimétrica– debiéramos temer que no se cumpla el supuesto de normalidad. Esto podríamos saberlo graficando la distribución o examinando la media y la mediana (que tienden a coincidir en caso de normalidad) y el coeficiente de sesgo (que debiera tender a cero en distribuciones normales).

Y si la varianza de esa misma variable difiriera mucho de una muestra a otra, será probable que esté ausente la homocedasticidad.

Aunque la violación de cualquiera de estos supuestos quita robustez al procedimiento, tornándolo menos confiable, se ha señalado que estas consecuencias se agravan con tamaños muestrales reducidos, mientras que generan menores distorsiones con muestras grandes.

La salida del SPSS

Cuando se dispone de un programa de procesamiento y análisis estadístico de información como el SPSS, entonces algunos problemas aparecen fácilmente resueltos. La pantalla de resultados nos mostraría lo siguiente:

NIVREC	N	Mean	Std. Deviation	Std. Error Mean
Sin secund	910	7,34	7,87	0,26
Con secund	810	9,07	8,67	0,30

Esta primera tabla de resultados es meramente descriptiva en su contenido. Tenemos los valores de las medias de ambas muestras, sus tamaños, así como los respectivos desvíos estándar. Así éstos últimos fuesen muy diferentes entre sí, ello sería un primer indicio de no cumplimiento del requisito de la homogeneidad de las varianzas (homocedasticidad): no olvidemos que la varianza es el cuadrado del desvío estándar. Por fortuna, aquí no parece suceder así.

	Levene test		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	F	Sig.							
								Lower	Upper
Equal variances assumed	1,63	0,202	-4,33	1718	0,0000	-1,73	0,398	-2,51	-0,95
Equal variances not assumed			-4,31	1643,99	0,0000	-1,73	0,400	-2,514	-0,94

Para saber si el supuesto de homocedasticidad se cumple, el SPSS aplica otro test de significación auxiliar, que es el Test de Levene. Lo que hace esta prueba es comparar las varianzas de dos muestras, bajo la hipótesis nula de que las poblaciones de las que provienen tienen varianzas iguales o, al menos, no significativamente diferentes. Si la significación de Levene es

0,05 o más pequeña, entonces se rechaza esta hipótesis nula y se concluye que las varianzas poblacionales no son iguales. Allí el supuesto no se cumpliría. Sin embargo, en nuestro ejemplo, la significación de Levene es 0,202, de manera que si rechazáramos la hipótesis nula que afirma la igualdad de las varianzas, correríamos un riesgo de cometer error de tipo I de 20%. La aceptamos, entonces, y afirmamos que las varianzas poblacionales no difieren significativamente: es decir, que se cumple el requisito de homocedasticidad.

¡Es una suerte que así sea! Pero aunque no hubiera ocurrido así, todo tendría solución. Pues la salida de SPSS nos ofrece dos alternativas para la t de Student: en la primera fila, un valor de t (-4,33) para cuando –como sucedió aquí– podemos asumir la igualdad de varianzas. Y con ese valor de t, ya vemos que es posible rechazar la hipótesis nula (ahora la de la prueba t, que afirma que las medias poblacionales no difieren) con una significación de 0,0000: esto no significa seguridad total sino que el riesgo de cometer error de tipo I es tan pequeño, que serían necesarios más dígitos decimales para poder leerlo.

¿Qué hubiera pasado si no se cumplía la igualdad de varianzas: es decir si nos veíamos forzados a rechazar la hipótesis de nulidad de Levene. Pues en ese caso tendríamos que haber leído el valor y la significación de t en la segunda fila de la tabla. Hay allí una corrección consistente en calcular t con unas distribuciones recortadas de sus puntuaciones extremas, con la finalidad de atenuar la dispersión[3]: por esta razón se pierden grados de libertad y el valor de t resulta ligeramente menor (-4,31). Alcanza, sin embargo, para rechazar la hipótesis nula con gran confianza.

¿Qué más nos muestra la tabla? Pues tenemos el error estándar de la estimación de la verdadera diferencia existente entre las medias poblacionales, así como un intervalo de confianza dentro del cual, con 95% de probabilidad, se encuentra situada esa diferencia. Dicho intervalo resulta de sumar y restar 1,96 veces el error estándar a la estimación muestral de esta diferencia:

$-1{,}73 \pm 1{,}96 * 0{,}398 = -2{,}51 \leftrightarrow -0{,}95$

¿Por qué hacemos esto? Pues porque ya se ha explicado más arriba que todas las diferencias de medias muestrales posibles han de distribuirse normalmente en torno a la diferencia existente entre las medias de las poblaciones. Y en una distribución normal, el 95% de los casos no se alejan de la media poblacional en más de 1,96 veces el desvío estándar de dicha distribución. De manera que la diferencia entre las medias de *nuestras* muestras –con un 95% de probabilidad– no debería guardar una distancia superior a esa con respecto a la verdadera diferencia existente entre las medias poblacionales.

2. El caso de las proporciones

¿Ayuda el secundario...?

Toda vez que el tamaño de la muestra determinó una elevada potencia de la prueba, es decir incrementó –acaso hasta en exceso- su sensibilidad para rechazar la hipótesis de nulidad, bien vale la pena reflexionar un poco más sobre la cuestión (¡siempre vale la pena reflexionar!). Si la mayor celeridad en la obtención de un empleo se debiera a que los jóvenes que no tienen el secundario completo provienen de hogares con mayores urgencias económicas, donde sus ingresos resultan imprescindibles por exiguos que sean, entonces sería de esperar que estos jóvenes se vuelquen a las búsquedas en mayor medida que los que no padecen tales presio-

[3] Pues si una de las varianzas fuese significativamente superior a la otra, ello podría estar motivado por la presencia de puntuaciones muy extremas.

nes. Si halláramos que no es así, tenderíamos a pensar que la razón tal vez sea otra: al tener estudios medios aumentan sus expectativas y son más selectivos.

En las ciencias empíricas suela hablarse de un *experimento crucial* para designar la situación en que dos hipótesis rivales concurren a explicar los mismos hechos (como sucede aquí con la urgencia de trabajar de los "sin secundario" vs. las mayores pretensiones de los "con secundario" para dar cuenta de la diferente duración del desempleo). ¿Cómo elegir?: si pudieran desprenderse de ambas hipótesis consecuencias empíricas de signo opuesto, entonces estaríamos ante un *experimento crucial* que nos permitiría decidir. En este caso no llegamos a tanto, pues no hallar diferencias en las tasas de actividad permitirá descartar la primera hipótesis pero no necesariamente reforzar la segunda.

Pero con eso ya habríamos avanzado bastante, desbrozando la maleza del conocimiento empírico. Ya decía Popper que se avanzaba descartando falsedades y no afirmando certezas. Entonces, acaso convenga comparar las tasas de actividad de los jóvenes con y sin secundario. Es lo que muestra el próximo cuadro:

Jóvenes de 18 a 24 años	Hasta secundario incompleto	Secundario completo y más
Tasa de actividad	54,1	44,2
Tamaño muestral	2907	2492

Fuente: EPH-INDEC (octubre de 2002)

En este caso, tenemos nuevamente dos muestras: la de jóvenes con secundario completo y la de los que no lograron completar el nivel. Desde ya, son ambas muy grandes, porque no incluyen solo a los desocupados, sino también a los ocupados (que junto con los anteriores conforman los económicamente activos) y a los inactivos (que no trabajan ni buscan hacerlo). Precisamente, lo que comparamos entre ambas submuestras, son las respectivas proporciones de activos: 54% y 44%. Y estas proporciones sugieren que, efectivamente, en estas edades tratan de trabajar en mayor medida los que no han completado estudios medios, presuntamente pertenecientes a hogares más pobres. Pero una vez más, podemos preguntarnos: ¿será esta una diferencia real o nos engañarán las muestras? Habrá una hipótesis nula que afirmará esto último: es decir, que en las poblaciones respectivas estas proporciones no difieren.

Afortunadamente, la prueba de t puede adaptarse sin problemas al caso de una comparación entre proporciones. Y la razón es muy simple: ya se ha explicado en un capítulo anterior, al ocuparnos de las estimaciones muestrales, que si en una variable dicotómica (activos/no activos) puntuamos con uno (1) a los casos que asumen una de ambas categorías (por ejemplo, a los activos) y con cero (0) a los restantes, entonces, la proporción de unos será exactamente la media aritmética de esa variable. Si diez jóvenes tuviéramos seis activos, la proporción de valores 1 sería 0,6 (60%) y la media de la variable sería, también, 0,6. En otras palabras, una proporción (que multiplicada por 100 se convierte en un porcentaje) no es sino un caso particular de la media. Veremos cómo queda la prueba:

t = $$\frac{p(1)-p(2)}{\sqrt{\left(\frac{p(1)*q(1)}{n(1)-1}+\frac{p(2)*q(2)}{n(2)-1}\right)}}$$

Donde:

Reemplazando:

t = $$\frac{54{,}1-44{,}2}{\sqrt{\left(\frac{54{,}1*45{,}9}{2907}+\frac{44{,}2*55{,}8}{2492}\right)}}$$ = 7,29

Una vez más podemos rechazar sin inconvenientes la hipótesis nula, aunque con una contribución muy importante del tamaño muestral, claro está. ¿Qué hubiera sucedido si tuviésemos una muestra mucho menor, por ejemplo, diez veces menor? Por curiosidad, podemos calcular la prueba con unos n divididos por diez: nos tranquilizará comprobar que, aún así, el valor de t sería suficiente para rechazar la hipótesis de nulidad con un nivel de significación de 0,05 (o una confianza de 95% en que nuestra decisión es acertada).

En definitiva, parece ser cierto que los jóvenes que no completaron el secundario permanecen menos tiempo desocupados porque se ven más urgidos a trabajar, y esto debe relacionarse con características de sus hogares. Pero para estar más seguros, sería preciso indagar si las diferencias subsisten cuando se mantienen bajo control los ingresos de los hogares. Por supuesto que ello es posible y tal sería el camino a seguir por parte del investigador…

3. El caso de las muestras apareadas

¿Ganan más los jóvenes que han recibido entrenamiento laboral?

¿Dan resultado los programas de capacitación laboral para jóvenes?

Cuando se aplican políticas públicas, es cada vez más usual que se pretenda evaluar si las mismas han cumplido con las metas esperadas de ellas. Esto sucede con los programas sociales, que son sometidos a evaluaciones. Generalmente se procede a medir una serie de indicadores sobre los que se espera que impacte el programa antes de su iniciación –lo que se suele denominar *línea de base*– y más tarde se los vuelve a medir a ver si se han producido los cambios esperados.

Por ejemplo, si se implementa un programa de entrenamiento a jóvenes trabajadores, destinado a lograr que sus ingresos mejoren a mediano plazo, los efectos del programa podrán medirse comparando los ingresos medios que obtienen antes de recibir ese entrenamiento y los que logran tiempo después, ya con sus nuevas calificaciones.

Cuando comparamos a jóvenes que han terminado o no sus estudios secundarios podemos hacer de cuenta que tenemos dos muestras independientes[4]. Pero, a veces, se trata de la misma muestra de sujetos que han sido medidos en dos oportunidades diferentes. En estos casos, si hay dos mediciones sobre la misma muestra, nuevamente podríamos abrigar dudas respecto de los efectos apreciados. ¿Puede tratarse de una diferencia meramente azarosa? Es decir, ¿podríamos atribuirla a la casualidad?

Otra formulación de la misma prueba t de Student puede brindarnos la respuesta.

En este caso, la fórmula de t se obtiene de una manera diferente:

$$t = \frac{\overline{Xd}}{\frac{Sd}{\sqrt{n-1}}}$$

Donde:
$\overline{Xd}$: media aritmética de las diferencias (entre la primera y la segunda medición)
S*d:* desvío estándar de la diferencias
n: tamaño de la muestra

[4] Aunque en rigor, la EPH obtiene una sola muestra y luego , el investigador la subdivide en distintas submuestras: los jóvenes y los que no lo son, los que terminaron sus estudios o los que no lo hicieron, etc.

En el ejemplo:

Salario medio antes de la capacitación	953
Salario medio luego de la capacitación	1152
Tamaño de la muestra	1720
Media aritmética de las diferencias	-198,7
Desvío estándar de las diferencias	520,5

t = $\dfrac{-198{,}7}{\dfrac{520{,}5}{\sqrt{1720}}}$ = -15,83

Una vez más es innecesario restar un grado de libertad, pues el tamaño de la muestra determina que la distribución resulte normal. Así, en este caso, el valor de t equivale a un valor de z. Y ni siquiera es necesario que recurramos a la comparación con la tabla, pues el valor de t debe compararse con un valor crítico para saber si permite o no rechazar la hipótesis nula. En este caso, puesto que además hemos previsto el signo de la diferencia (esperábamos que los salarios medios se incrementaran y así resultó) podemos trabajar con una prueba de una sola cola. Y ya sabemos que el valor de t a superar es 1,64 para una significación de 0,05. De manera que –una vez más favorecidos por el tamaño de la muestra– podemos desdeñar la hipótesis nula con una elevada confianza a la vista del alto valor de t obtenido: será posible hacerlo aun con una significación de 0,01 o de 0,001.

Puesto que hemos hallado elementos suficientes para desdeñar la posibilidad de que la diferencia obedezca al azar –que es precisamente lo que afirma la hipótesis de nulidad– entonces se fortalecería la presunción de que *algo* que sucedió entre ambas mediciones provocó esta diferencia. No podemos, sin embargo, asegurar que ese algo fue la capacitación brindada, porque en el lapso transcurrido habrán ocurrido, sin duda, muchas otras cosas, sobre las que la encuesta no ha recogido información alguna. Pero al menos podremos fortalecer un poco nuestra conjetura.

Siempre sin seguridad total, claro está, porque persistirá el riesgo de incurrir en el error de tipo I descartando equivocadamente una hipótesis nula cierta: pero en este caso ese riesgo resulta mínimo.