

Estudio sobre hábitos y rutinas de la población a la hora de hacer deporte



Universitat d'Alacant
Universidad de Alicante

GRADO EN MATEMÁTICAS
INFERENCIA ESTADÍSTICA

Alejandro Sellés Pérez
Aitana Soto García

Índice

| | |
|--|-----------|
| 1. Introducción | 2 |
| 2. Contrastes planteados | 3 |
| 2.1. ¿La gente que está apuntada al gimnasio entrena menos días a la semana que la gente que realiza otros deportes? | 3 |
| 2.2. ¿Los chicos entrenan semanalmente más horas que las chicas? . . | 4 |
| 2.3. ¿Los adultos acuden más a las clases que los jóvenes? | 5 |
| 2.4. ¿Los adultos están dispuestos a pagar más por el gimnasio? . . . | 6 |
| 2.5. ¿Más del 30 % de los apuntados al gimnasio toman proteínas u otros suplementos para mejorar su rendimiento? | 8 |
| 2.6. ¿Un cuarto de los jóvenes toman bebidas energéticas o elevadas dosis de cafeína para mejorar su rendimiento en el gimnasio? . . | 8 |
| 3. Conclusiones | 10 |

1. Introducción

El ejercicio físico y el deporte han adquirido una relevancia creciente en la búsqueda de una vida saludable y equilibrada. La práctica regular de actividad física no solo beneficia nuestra condición física, sino que también tiene un impacto positivo en nuestra salud mental, nuestro bienestar general y nuestra calidad de vida. Por ello, en estos últimos años hemos sido partícipes de la creación de muchos centros deportivos en nuestras ciudades, en los que el número de integrantes no para de crecer. En este trabajo, queremos hacer un estudio sobre los diferentes hábitos y rutinas que tiene la gente a la hora de ir a entrenar.

Para obtener las respuestas de nuestra población, hemos elaborado una encuesta mediante Google Forms. Estas cuestiones han sido respondidas por 431 personas, y gracias a ellas hemos ido clasificando cada individuo según su rango de edad, el tipo de deporte que practican y la frecuencia con la que lo realizan. Además, tenemos 227 respuestas de personas apuntadas al gimnasio, de las cuales hemos obtenido más información sobre sus hábitos. Para este estudio realizaremos un total de 6 contrastes de hipótesis.

Por un lado, vamos a realizar dos contrastes en los cuales utilizaremos la muestra completa. Es decir, tanto las personas que acuden al gimnasio como las que realizan otros deportes. En esta parte del trabajo nos centraremos en cuestiones más generales; estudiaremos si los hombres entrenan más que las mujeres y si la gente apuntada al gimnasio entrena menos días que el resto.

Por otro lado, en los otros cuatro contrastes nos centraremos exclusivamente en la gente que acude al gimnasio, y estudiaremos algunos de sus hábitos y rutinas. Analizaremos si toman proteínas o bebidas energéticas para mejorar su rendimiento, si van más a las clases o a la sala de máquinas, y cuánto dinero están dispuestos a pagar mensualmente según la edad.

Finalmente, para resolver estos contrastes utilizaremos el programa R, y durante todo el trabajo vamos a usar un nivel de significación fijo, el cual será $\alpha = 0,05$.

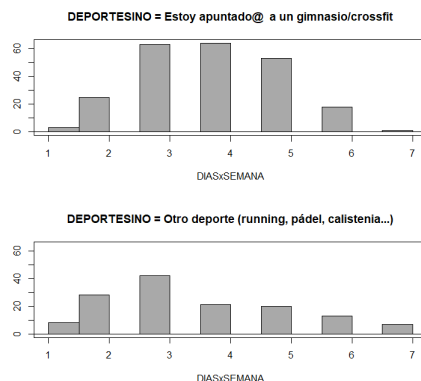
2. Contrastes planteados

2.1. ¿La gente que está apuntada al gimnasio entrena menos días a la semana que la gente que realiza otros deportes?

Al realizar esta pregunta, tuvimos el pensamiento de que el entrenamiento relacionado con la fuerza y hacer pesas, es un tipo de entreno que produce mucha fatiga muscular; por tanto, requiere de un tiempo de descanso mayor a cualquier otro deporte. Consideramos 2 grupos de personas: por un lado, la gente que va al gimnasio; y, por otro lado, la gente que realiza otro deporte. Dentro de estos grupos, estudiamos el número de días que entrenan semanalmente. Definimos por tanto las variables:

$$\begin{cases} X & \equiv \text{Número de días por semana que entrena una persona apuntada al gimnasio} \\ Y & \equiv \text{Número de días por semana que entrena una persona que realiza otro deporte} \end{cases}$$

Estos son los datos obtenidos en nuestra encuesta:



Gráficamente podemos observar que la distribución de los que no van al gimnasio no se asemeja a una normal. Sin embargo, la otra gráfica podríamos pensar que sí. De todas maneras, vamos a comprobarlo en R mediante el test de Shapiro-Wilk:

```
> normalityTest(DIASxSEMANA ~ DEPORTESINO, test="shapiro.test", data=vagos)

-----
DEPORTESINO = Estoy apuntado@ a un gimnasio/crossfit

Shapiro-Wilk normality test

data: DIASxSEMANA
W = 0.93449, p-value = 0.0000000156

-----
DEPORTESINO = Otro deporte (running, pádel, calistenia...)

Shapiro-Wilk normality test

data: DIASxSEMANA
W = 0.92772, p-value = 0.000001573
```

En ambos casos, obtenemos un p-valor mucho menor a nuestro nivel de significación. Por lo tanto, no podemos asumir que sigan distribuciones normales. Así, para poder obtener una solución vamos a trabajar con las medianas de ambas variables y aplicar el test de Wilcoxon para el siguiente contraste:

$$\begin{cases} H_0 : M_X \geq M_Y \\ H_1 : M_X < M_Y \end{cases}$$

Después de introducir los datos en R, obtenemos:

```
> wilcox.test(DIASxSEMANA ~ DEPORTESINO, alternative='less', exact=FALSE, correct=TRUE, data=vagos)
Wilcoxon rank sum test with continuity correction

data: DIASxSEMANA by DEPORTESINO
W = 17989, p-value = 0.9895
alternative hypothesis: true location shift is less than 0
```

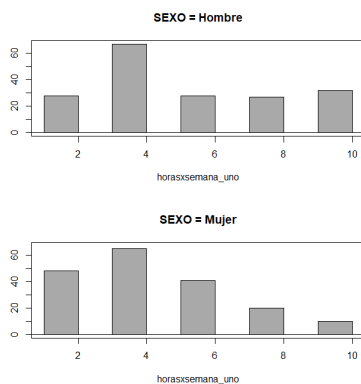
Por tanto, como hemos obtenido un p-valor de 0.9895, y éste es mayor a nuestro nivel de significación (0.05), entonces no tenemos evidencia suficiente para rechazar la hipótesis nula, por tanto, no podemos concluir que la gente apuntada al gimnasio entrena menos días que la gente que realiza otros deportes.

2.2. ¿Los chicos entrenan semanalmente más horas que las chicas?

A lo largo de los últimos años hemos observado tanto en Instagram y en nuestros círculos cercanos que cada vez hay más mujeres que hacen deporte. Aún así, pensamos que los hombres suelen hacerlo con más frecuencia que ellas. Para este estudio, hemos dividido nuestra población entre hombres y mujeres, y hemos analizado las horas que entrenan unos y otros a la semana. Tenemos entonces las variables:

$$\begin{cases} X \equiv \text{Horas por semana que entrena un hombre al azar} \\ Y \equiv \text{Horas por semana que entrena una mujer al azar} \end{cases}$$

Estos han sido los resultados obtenidos en la encuesta:



A simple vista, podemos ver que ninguna de las dos variables sigue una distribución normal. Vamos a comprobarlo mediante el test de Shapiro-Wilk en R:

```
> normalityTest(horasxsemana_uno ~ SEXO, test="shapiro.test", data=vagos)

-----
SEXO = Hombre

      Shapiro-Wilk normality test

data:  horasxsemana_uno
W = 0.88871, p-value = 2.106e-10

-----
SEXO = Mujer

      Shapiro-Wilk normality test

data:  horasxsemana_uno
W = 0.89033, p-value = 2.223e-10
```

Como el p-valor obtenido en ambos casos es inferior a nuestro nivel de significación (0.05). No podemos asumir normalidad en ninguna de las dos variables. Por tanto, para obtener una solución, vamos a trabajar con las medianas y plantear el siguiente contraste:

$$\begin{cases} H_0 : M_X \leq M_Y \\ H_1 : M_X > M_Y \end{cases}$$

Ahora, vamos a resolverlo mediante el test de Wilcoxon en R. Obtenemos los siguientes resultados:

```
> wilcox.test(horasxsemana_uno ~ SEXO, alternative='greater', exact=FALSE, correct=TRUE, data=vagos)

      Wilcoxon rank sum test with continuity correction

data:  horasxsemana_uno by SEXO
W = 19960, p-value = 0.0005054
alternative hypothesis: true location shift is greater than 0
```

Como el p-valor obtenido es 0.0005054, y es menor que el nivel de significación que estamos utilizando (0.05), entonces tenemos evidencias suficientes para rechazar la hipótesis nula. Y, por tanto, podemos afirmar que los chicos entrenan más horas semanalmente que las chicas.

2.3. ¿Los adultos acuden más a las clases que los jóvenes?

En los últimos años hemos podido apreciar como los jóvenes se preocupan más por su cuerpo y a la hora de entrenar lo hacen más en la sala de pesas. Por otro lado, creemos que la gente más mayor hace ejercicio atendiendo a las clases dirigidas del gimnasio. Veamos si podemos afirmar que la proporción de adultos que acuden a las clases dirigidas es mayor que la de los jóvenes menores de 25 mediante un test de proporciones para dos variables.

Definimos X e Y dos variables que siguen una distribución Bernoulli, $X \sim b(\theta_1)$ e $Y \sim b(\theta_2)$.

$$X = \begin{cases} 0 & \text{si un encuestado al azar menor de 25 años va clases dirigidas} \\ 1 & \text{si un encuestado al azar menor de 25 años no va a clases dirigidas} \end{cases}$$

$$Y = \begin{cases} 0 & \text{si un encuestado al azar mayor de 25 años va a clases dirigidas} \\ 1 & \text{si un encuestado al azar mayor de 25 años no va a clases dirigidas} \end{cases}$$

$$\begin{cases} H_0 : \theta_1 \leq \theta_2 \\ H_1 : \theta_1 > \theta_2 \end{cases}$$

Y obtenemos el siguiente resultado:

```

          FUERZA_O_CLASES
EDAD_BIN_BUENO  0      1 Total Count
0      7.5  92.5    100    161
1     34.8  65.2    100     66

2-sample test for equality of proportions with continuity correction

data: .Table
X-squared = 24.879, df = 1, p-value = 0.0000003053
alternative hypothesis: less
95 percent confidence interval:
-1.0000000 -0.1609641
sample estimates:
prop 1      prop 2 
0.07453416 0.34848485

```

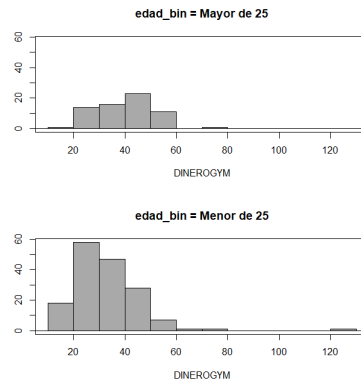
Hemos obtenido un p-valor de 0.0000003053, menor que nuestro nivel de significación (0.05) por lo tanto podemos rechazar la hipótesis nula y afirmar que la proporción de personas mayores de 25 que asiste a las clases dirigidas es mayor que la de los jóvenes menores de 25.

2.4. ¿Los adultos están dispuestos a pagar más por el gimnasio?

A la hora de plantear este contraste, pensamos que esto podía ser algo evidente, ya que los adultos tienen un mayor poder adquisitivo. Sin embargo, también tenemos que tener en cuenta que a lo largo de los últimos años, para muchos jóvenes el gimnasio se está convirtiendo en una de sus prioridades a la hora de gastar su dinero. Vamos a trabajar solamente con la población de la gente que acude al gimnasio. Dividiremos ésta entre los jóvenes (menores de 25) y los adultos (mayores de 25), y estudiaremos cuántos euros están dispuestos a pagar mensualmente. Definimos entonces las siguientes variables:

$$\begin{cases} X & \equiv \text{Dinero que un adulto al azar está dispuesto a pagar por el gimnasio} \\ Y & \equiv \text{Dinero que un joven al azar está dispuesto a pagar por el gimnasio} \end{cases}$$

Los datos obtenidos en la encuesta son:



A simple vista, podemos observar que ninguna parece seguir una distribución normal. Vamos a comprobarlo mediante el test de normalidad de Saphiro-Wilk en R:

```
> normalityTest(DINEROGYM ~ edad_bin, test="shapiro.test", data=sologymbros)
-----
edad_bin = Mayor de 25
      Shapiro-Wilk normality test

data:  DINEROGYM
W = 0.93899, p-value = 0.002852
-----
edad_bin = Menor de 25
      Shapiro-Wilk normality test

data:  DINEROGYM
W = 0.85665, p-value = 3.041e-11
```

De la misma manera que los contrastes anteriores, como el p-valor obtenido es menor que el p-valor con el que estamos trabajando (0.05), entonces no podemos asumir normalidad en ninguna de las variables. Por lo tanto, para obtener la solución, trabajamos con las medianas y planteamos el siguiente contraste:

$$\begin{cases} H_0 : M_X \leq M_Y \\ H_1 : M_X > M_Y \end{cases}$$

Aplicamos el test de Wilcoxon en R y obtenemos:

```
> Tapply(DINEROGYM ~ edad_bin, median, na.action=na.omit, data=sologymbros) # medians by group
Mayor de 25 Menor de 25
      50         35

> wilcox.test(DINEROGYM ~ edad_bin, alternative='greater', exact=FALSE, correct=TRUE, data=sologymbros)

Wilcoxon rank sum test with continuity correction

data:  DINEROGYM by edad_bin
W = 7344, p-value = 0.000002287
alternative hypothesis: true location shift is greater than 0
```

Ahora, como el p-valor obtenido es 0.000002287, y éste es menor que el nivel de significación con el que trabajamos (0.05). Entonces, tenemos evidencia suficiente para rechazar la hipótesis nula, y por tanto, podemos afirmar que los adultos están dispuestos a pagar más dinero mensualmente por el gimnasio.

2.5. ¿Más del 30 % de los apuntados al gimnasio toman proteínas u otros suplementos para mejorar su rendimiento?

Hemos observado que en los últimos años el mundo del fitness se ha puesto de moda, por ello creemos que la proporción de personas que consumen suplementos ha aumentado. Según una noticia del periódico 'La Razón' el 30 % de las personas apuntadas a un gimnasio consumían este tipo de sustancias en 2015. Veamos si podemos afirmar que este porcentaje ha aumentado.

Para llevar a cabo el contraste definimos la variable discreta X que sigue una distribución Bernoulli:

$$X = \begin{cases} 1 & \text{si un encuestado al azar toma proteínas o suplementos} \\ 0 & \text{si un encuestado al azar no toma proteínas o suplementos} \end{cases}$$

Planteamos el contraste:

$$\begin{cases} H_0 : \theta \leq 0,3 \\ H_1 : \theta > 0,3 \end{cases}$$

Y obtenemos el siguiente resultado:

```
PROTEINA_BIN
  0  1
143 84

1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.7
X-squared = 5.3033, df = 1, p-value = 0.01064
alternative hypothesis: true p is less than 0.7
95 percent confidence interval:
 0.0000000 0.6808467
sample estimates:
p
0.6299559
```

Hemos obtenido un p-valor de 0.01064 menor que nuestro nivel de significación (0.05) por lo tanto podemos rechazar la hipótesis nula y afirmar que hoy en día más del 30 % consumen algún suplemento para mejorar su rendimiento en el gimnasio.

2.6. ¿Un cuarto de los jóvenes toman bebidas energéticas o elevadas dosis de cafeína para mejorar su rendimiento en el gimnasio?

Hemos observado en nuestros gimnasios y en nuestro círculo cercano, que el consumo de elevadas dosis de cafeína se está haciendo algo habitual para ir a entrenar al gimnasio. Por ello, queremos ver si podemos afirmar que más del 25 % de los jóvenes menores de 25 años consumen, al menos una vez a la semana, una elevada cantidad de cafeína.

Para llevar a cabo el contraste definimos la variable discreta X que sigue una distribución Bernoulli:

$$X = \begin{cases} 1 & \text{si un encuestado menor de 25 años al azar toma productos con exceso de cafeína} \\ 0 & \text{si un encuestado menor de 25 años al azar no toma productos con exceso de cafeína} \end{cases}$$

Planteamos el contraste:

$$\begin{cases} H_0 : & \theta \leq 0,25 \\ H_1 : & \theta > 0,25 \end{cases}$$

Y obtenemos el siguiente resultado:

```
Frequency counts (test is for first level):
ENERGETICA
No  Si
111  50

1-sample proportions test with continuity correction

data: rbind(.Table), null probability 0.75
X-squared = 2.8344, df = 1, p-value = 0.04613
alternative hypothesis: true p is less than 0.75
95 percent confidence interval:
 0.0000000 0.7487709
sample estimates:
p
0.689441
```

Hemos obtenido un p-valor de 0.04613, menor que nuestro nivel de significación (0.05) por lo tanto podemos rechazar la hipótesis nula y afirmar que hoy en día más del 25 % de los jóvenes consumen bebidas energéticas o elevadas dosis de cafeína para mejorar su rendimiento en el gimnasio.

3. Conclusiones

Tras analizar los resultados obtenidos en nuestros contrastes hemos podido observar que estábamos muy equivocados en la idea de que la gente que acude al gimnasio necesita más días de recuperación. Después de pensar en ello, hemos llegado a la conclusión de que la gente hace del hecho de ir al gimnasio una rutina; sin embargo, cuando realizan otros deportes como pádel o fútbol, suelen necesitar de más gente para practicarlos, y por tanto, es más difícil ponerse de acuerdo y entrenar más días.

Por otro lado, podemos concluir que hay grandes diferencias entre los jóvenes y adultos cuando hablamos del mundo del gimnasio. Nos ha parecido interesante este análisis, ya que a la hora de elegir un lugar para entrenar, las cuestiones del coste mensual y la oferta de clases dirigidas son las que marcan la diferencia entre unos y otros. Por lo que, pensamos que estos gimnasios antes de abrir sus puertas, deben preguntarse hacia qué público quieren dirigirse y trabajar a partir de ello.

Finalmente, pese a que sabíamos que los datos iban a ser muy altos, ya que día a día es lo que vemos en nuestros gimnasios, nos han sorprendido mucho los resultados de la gente que toma proteína y los jóvenes que consumen cafeína. Nos parece que los porcentajes son muy elevados y que quizás la gente no está muy bien informada a la hora de tomar estos suplementos, ya que no siempre van de la mano con una buena salud y un mejor rendimiento.

$$\ell = (k2) \frac{k!}{2!(k-2)!} \frac{k(k-1)}{2}$$