

Introducción a la Inferencia Estadística vía Simulación



El problema: discriminación de género



- En 1972, como parte de un estudio más amplio de discriminación de género, se les brindó a 48 supervisores bancarios de sexo masculino el mismo legajo personal y se le pidió que evaluén si la persona del legajo debía ser promovida a un cargo superior o no, en un ascenso de rutina.
- Los legajos eran todos idénticos (hay un único legajo), excepto que a la mitad de los supervisores les entregaron legajos cuyos postulantes eran hombres, mientras que a la otra mitad legajos cuyos postulantes eran mujeres.
- La asignación de los legajos a los supervisores fue aleatoria.
- De los 48 legajos analizados, se aconsejó promocionar a 35 de ellos.
- El estudio intenta evaluar si existe discriminación de género.

Discriminación de género: los datos



		Promoción		Total
		Promovido	No promovido	
Sexo	masculino	21	3	24
	femenino	14	10	24
Total		35	13	48

Discriminación de género: planteo



- Nos enfrentamos a dos alternativas posibles
 - **Nada ocurre aquí (status quo):** la promoción a un cargo superior y el género del postulante **son** independientes. No hay discriminación de género. Las diferencias observadas entre las diferentes proporciones es sólo cuestión del azar. → **Hipótesis Nula**
 - **Algo está pasando aquí:** la promoción a un cargo superior y el género del postulante **no son** independientes. Hay discriminación de género. Las diferencias observadas entre las diferentes proporciones no se deben al azar. → **Hipótesis Alternativa**

Un juicio como un test de hipótesis...



- El test de hipótesis se parece mucho a un juicio



- La hipótesis nula H_0 : el acusado es inocente.
- La hipótesis alternativa H_A : el acusado es culpable.

- Se presenta la evidencia → *Se recogen los datos/muestra*
- Luego se evalúa la evidencia → *¿Podrían estos datos haber ocurrido sólo por cuestiones del azar si la hipótesis nula fuese verdadera?*
- Si la probabilidad de que la muestra obtenida sea muy improbable de ocurrir, la evidencia nos plantea una duda acerca de la validez de la H_0 .

Un juicio como un test de hipótesis...



- Si la evidencia no es lo suficientemente contundente para rechazar el supuesto de inocencia, el jurado declara que **el acusado no es culpable**.
 - El jurado no dice que el acusado es inocente, simplemente porque la evidencia recogida no es suficiente para condenarlo.
 - De hecho, el acusado puede ser inocente, pero el jurado no tiene manera de afirmarlo con certeza.
- En vocabulario estadístico se dice: **no se puede rechazar la hipótesis nula**
 - **Nunca se declara que la hipótesis nula es verdadera**, simplemente porque no hay forma de saber si es cierta o no
 - Por eso **nunca se acepta la hipótesis nula**

Un juicio como un test de hipótesis...



- En un juicio, la carga de la prueba recae sobre la acusación
- En un test de hipótesis, la carga de la prueba recae sobre la reclamo inusual
- La hipótesis nula se refiere al ***status quo***. Por lo tanto es la hipótesis alternativa la que uno considera inusual y para la cual debemos reunir pruebas.

Recapitulando: un juicio como un test de hipótesis.



- Se comienza con una **hipótesis nula** H_0 que representa el *status quo*
- Se dispone también de una **hipótesis alternativa** H_A que refiere a nuestra pregunta de investigación, i.e. lo que uno desea evaluar/ testear
- Finalmente uno realiza un test de hipótesis bajo el supuesto de que la **H_0 es verdadera**, ya sea **vía simulación** (lo que vamos a estudiar a continuación) o **vía métodos teóricos** (lo que estudiaremos luego, basándonos en el TCL) .
- Si los resultados del test sugieren que los datos no brindan evidencia convincente a favor de la hipótesis alternativa H_A , entonces nos quedamos con la hipótesis nula H_0 . Si por el contrario, la evidencia es contundente, rechazamos la H_0 a favor de la H_A .

La simulación del experimento



- ¿Cómo hacemos para simular el experimento de discriminación de género?
- Bajo el supuesto de independencia, i.e. es el azar quien “mueve los hilos” se tiene que
 - Si el resultado de **la simulación luce similar a los datos del experimento**, luego podemos determinar que las diferencias entre las proporciones de los legajos promovidos entre hombres y mujeres es producto del azar; i.e. la promoción a un puesto superior y el género del postulante **son** independientes.
 - Por el contrario, si **la simulación no luce parecido con los datos observados en el experimento**, luego podemos determinar que las diferencias entre las proporciones de los legajos promovidos entre hombres y mujeres no es producto del azar, sino que es consecuencia del género del postulante; i.e. la promoción a un puesto superior y el género del postulante **no son** independientes.



El esquema de simulación

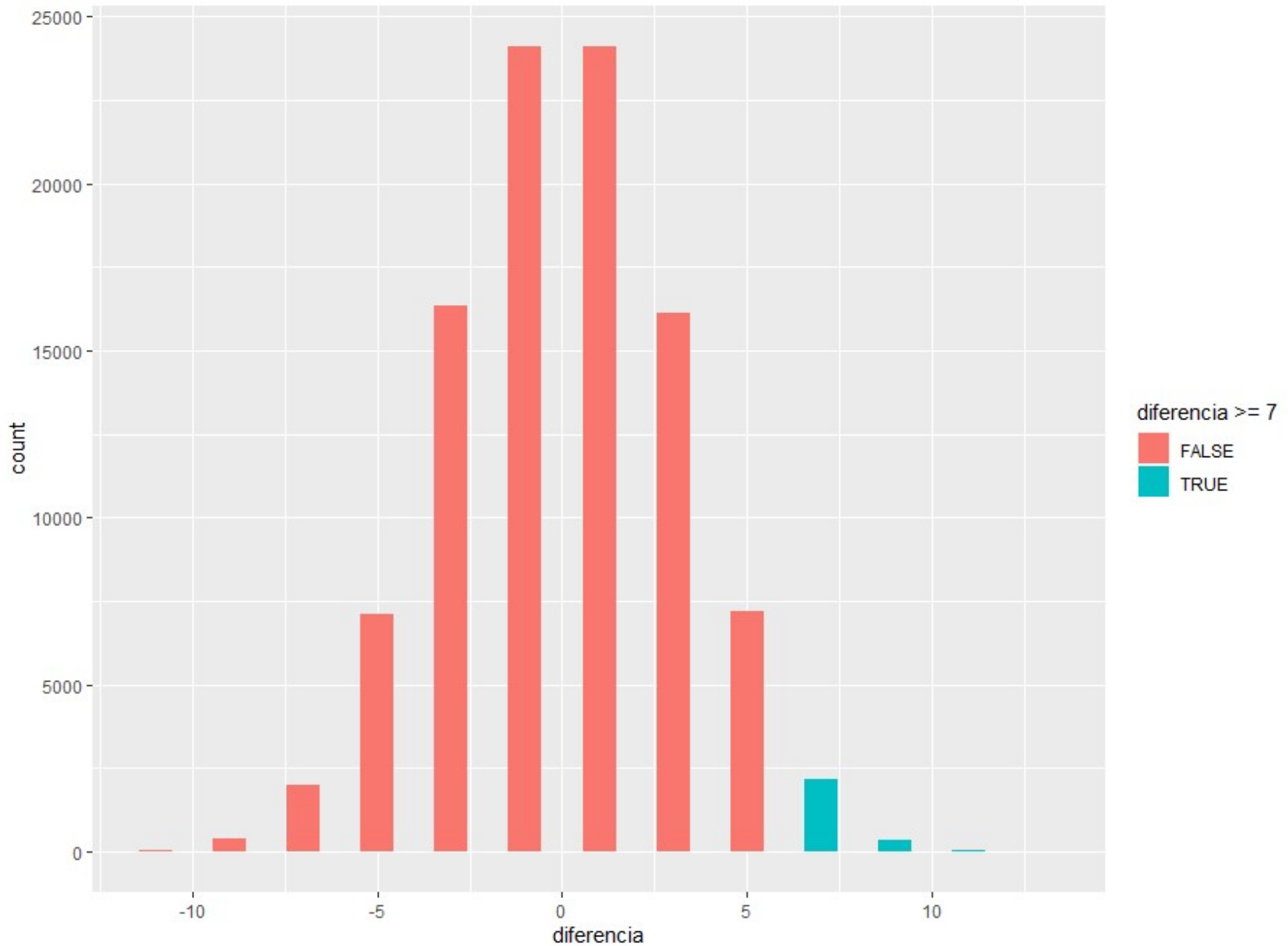
1. Conformar un grupo de 48 personas, donde 35 fueron promovidas y 13 no promovidas.
2. Separar luego de manera aleatoria dos grupos de 24 personas, donde uno corresponde a los legajos masculinos y el otro a los legajos femeninos.
3. Contabilizar dentro de cada grupo cuantos legajos fueron promovidos
4. Calcular la diferencia de entre los promovidos hombres y los promovidos mujeres
$$\text{Diferencia} = \text{Promovidos}_{\text{hombres}} - \text{Promovidos}_{\text{mujeres}}$$
5. Repetir pasos 1 a 5 muchas veces (100 o más)

Resultado de la simulación



- Representar gráficamente las diferencias registradas de las simulaciones, mediante un histograma
- Recordemos cuales eran nuestras hipótesis de trabajo
 - H_0 : la promoción a un cargo superior y el sexo del postulante **son** independientes. No hay discriminación de género.
 - H_A : la promoción a un cargo superior y el sexo del postulante **no son** independientes. Hay discriminación de género

Resultado de la simulación



Resumiendo el ejercicio de inferencia vía simulación



- Se especificaron una hipótesis nula y una hipótesis alternativa, que es la que se quiere evaluar.
- Se simuló el experimento asumiendo que la hipótesis nula es verdadera.
- Se evaluó la posibilidad de observar un resultado al menos tan extremo como el observado en los datos originales.
- Y si esto resulta muy poco probable, entonces rechazamos la hipótesis nula a favor de la hipótesis alternativa.