Análisis y Visualización de Datos

Diplomatura CDAAyA 2020

Prueba o Test de Hipótesis

Test de Hipótesis

En muchos problemas se requiere tomar una decisión entre aceptar o rechazar una proposición sobre algún parámetro o alguna afirmación de una población.

Esta proposición o afirmación recibe el nombre de **hipótesis estadística**, y el procedimiento de toma de decisión sobre la hipótesis se conoce como **prueba o test de hipótesis**.

El problema

En la empresa ComaMenosVaca se usan algoritmos de aprendizaje automático para encontrar las combinaciones de ingredientes que más se asemejen a la carne vacuna, con el objetivo de reducir el (costoso) consumo de este tipo de alimentos.

Un equipo de desarrollo ha implementado un nuevo modelo con una mejor arquitectura, y es hora de decidir si reemplazar el modelo viejo con el modelo nuevo.

Experimentos A/B

Se divide el conjunto de clientes en dos grupos: el grupo A (o control) al que le se le vende productos generados con el modelo original, y el grupo B (o tratamiento) al que se le vende productos generados con el modelo nuevo.

Para saber qué modelo es mejor, necesitamos una característica o **métrica** que cuantifique la posible mejora. Se elige el volumen promedio de gasto de los clientes.

El **estadístico** que se elige es la diferencia entre los promedios de ventas de ambos grupos.

Experimentos A/B

Ponemos en marcha el experimento, y 3 meses después analizamos los resultados:

Observamos que el grupo A gasta en promedio $500 por compra, y el grupo B $510 ($10 de diferencia, 2% de aumento)

Asumiendo que el costo de usar ambos modelos es el mismo, esta evidencia, ¿es suficiente para cambiar de modelo?

Test de hipótesis para experimentos A/B

Recordemos que nuestro estadístico (la diferencia en el promedio de ventas) es un **estimador puntual**, calculado en base a nuestra muestra. Por lo tanto, tiene un error.

Para tener una estimación de qué tan confiable es nuestro estadístico, decidimos realizar un test de hipótesis. Como lo que nos interesa es saber si existe o no diferencia entre los modelos, formulamos nuestra hipótesis nula como H_0: *el modelo nuevo no tiene efecto, son los dos iguales.*

H_0: el modelo nuevo no tiene efecto, son los dos iguales.

Entonces, siguiendo la metodología de test de hipótesis, nos preguntamos:

Asumiendo que no hay diferencia entre los modelos (H_0 es verdadera), ¿cuál es la probabilidad de que hayamos visto una muestra con una diferencia de promedios de venta (estadístico) de $10?

Asumiendo que no hay diferencia entre los modelos (H_0 es verdadera), ¿cuál es la probabilidad de que hayamos visto una muestra con una diferencia de promedios de venta (estadístico) de $10?

Necesitamos saber la distribución de nuestro estadístico.

Muy convenientemente, por el teorema central del límite, el promedio de las ventas se distribuye siguiendo una normal $N(\mu_A, \sigma_A^2)$ y $N(\mu_B, \sigma_B^2)$.

La suma (o resta) de dos variables con distribución normal también se distribuye normalmente $N(\mu_A - \mu_B, \sigma_A^2 + \sigma_B^2)$.

Asumiendo que no hay diferencia entre los modelos (H_0 es verdadera), ¿cuál es la probabilidad de que hayamos visto una muestra con una diferencia de promedios de venta (estadístico) de $10?

El estadístico (diferencia entre los promedios de ventas) sigue una distribución normal.

Si H_0 es cierta, entonces la media de la distribución de nuestro estadístico debería ser 0. Podemos usar un test de hipótesis de doble cola para calcular esta probabilidad.

Asumiendo que no hay diferencia entre los modelos (H_0 es verdadera), ¿cuál es la probabilidad de que hayamos visto una muestra con una diferencia de promedios de venta (estadístico) de $10?

El p-valor nos dice que tan extremo es el valor del estadístico obtenido de la muestra.

Si es menor que un nivel de significancia alpha, entonces podemos decir que asumiendo H_0, la muestra es muy poco probable. Luego, dado la muestra fue observada, **indicaría** que H_0 es falsa.

Lo que queríamos saber es: ¿qué tan probable es que, dados los datos que obtuve a partir de mi muestra, H_0 sea verdadera?

Lo que el test de hipótesis nos da: asumiendo que H_0 es verdadera, ¿cuál es la probabilidad de los datos observados?

Si el p-valor **no** es menor que el nivel de significancia, la muestra era probable.

No se puede concluir que H_0 sea verdadera, (que no hay diferencia entre el promedio de los grupos). Sólo podemos decir que **no encontramos evidencia** de que hay una diferencia.

Lo que queríamos saber es: ¿qué tan probable es que, dados los datos que obtuve a partir de mi muestra, H_0 sea verdadera?

Lo que el test de hipótesis nos da: asumiendo que H_0 es verdadera, ¿cuál es la probabilidad de los datos observados?

Si el p-valor es menor que el nivel de significancia alpha, no se puede concluir **nada** sobre la distribución del estadístico (la diferencia en promedio de ventas).

Lo que queríamos saber es: ¿qué tan probable es que, dados los datos que obtuve a partir de mi muestra, H_0 sea verdadera?

Lo que el test de hipótesis nos da: asumiendo que H_0 es verdadera, ¿cuál es la probabilidad de los datos observados?

No importa cuánto valga el p-valor, el tests de hipótesis no permiten asegurar **nada** sobre la distribución del estadístico bajo la hipótesis alternativa (que existe una diferencia entre los promedios de compra de ambos grupos).

¡Más cosas!

Lo que no tuvimos en cuenta antes de empezar:

- ¿Qué nivel de significancia elegir?
- ¿Qué poder estadístico tiene el test?
    - Si i) existe una diferencia entre ambos modelos y ii) pudiéramos repetir este mismo experimento en una copia exacta de nuestro universo muchas veces, ¿cuántas veces el experimento detectaría que existe una diferencia? ¿Vale la pena todo el trabajo de realizar el experimento si en realidad tenemos pocas chances de tener una conclusión válida?
- ¿Qué efecto tiene el test?
- ¿Cuántos clientes en la muestra son necesarios?

Potencia estadística [Opcional]

Probabilidad de que la hipótesis alternativa sea aceptada **cuando** la hipótesis alternativa es verdadera (1 - beta).

P(rechazar H0 | H0 es falsa)

P(p_value < alpha | H0 es falsa)

Depende de:

- Tamaño del efecto (magnitud del estadístico)
- Tamaño de la muestra
- Nivel de significancia

Volvemos 12:05!