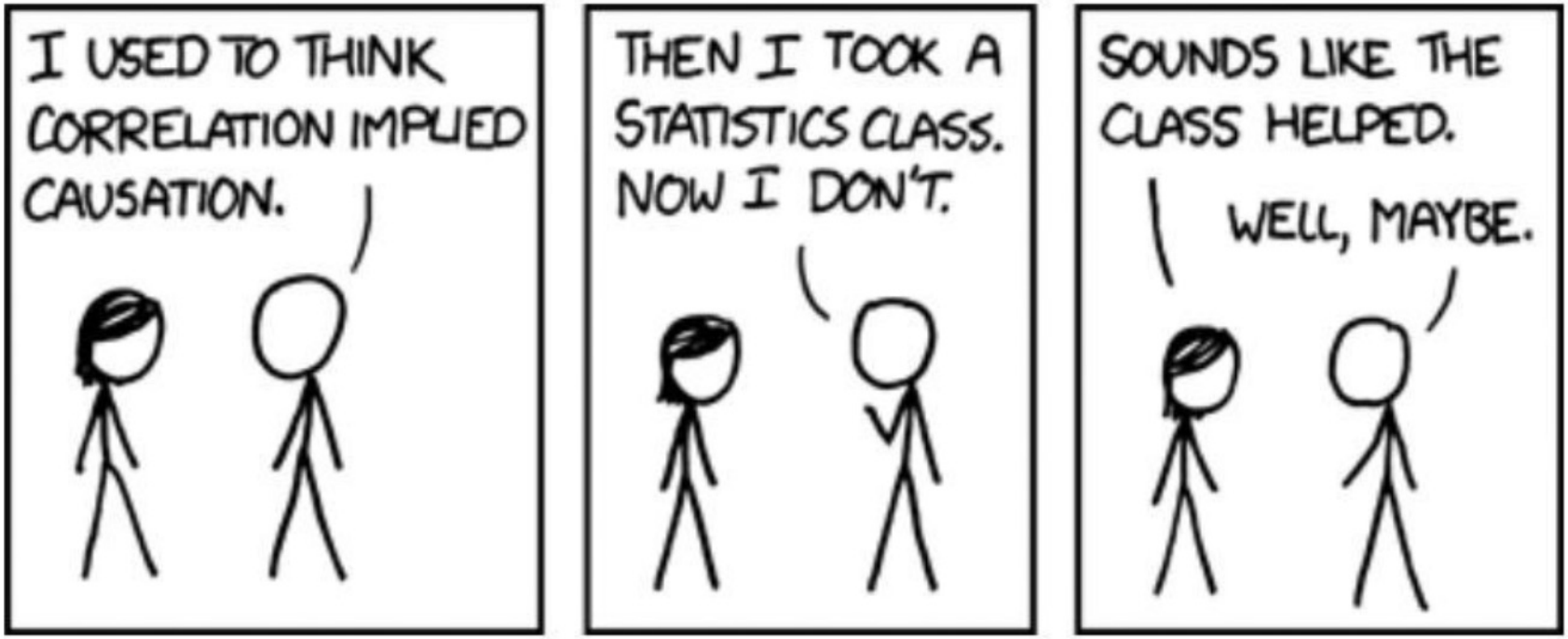
27th March 2023



HT framework:

1. H_0, H_A
2. Test statistic ✓
3. which tail → R, L, Two
4. p-value
5. p-value comparison

Marketting Campaign Impact:

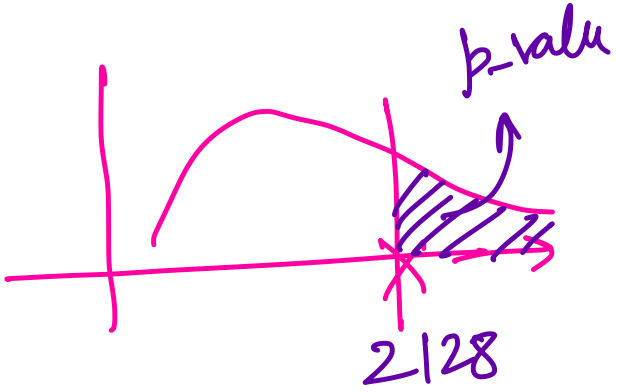
Avg. Rate/month

The average number of customers entering a store is 2000 per ' month A marketing company is hired to improve this number
The next month, number of customers was seen to be 2128
With 95% confidence, is this improvement statistically significant?

Soln:
- $H_0: \mu = 2000$
- $H_A: \mu > 2000$
- Test statistic ?
  - N = # of people entrij store.
  - $N_{obs} = 2128$
- Right Tailed.
- p-value : $P[N \geq 2128 \mid H_0 \text{ is True}]$

$= 1 - \text{poison.cdf}(k = 2127, mu = 2000)$

$= 0.002$



$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

$\Rightarrow \alpha = 0.05$ p-value $< \alpha$.

Reject H_0 ; Marketing worked

Recommendation System:

When a customer buys a T-shirt, a recommender algorithm also suggests a few related items The recommender system in production (legacy) that has a success rate of 10%

You and your team have developed a new deep learning algorithm for recommendation
It is tested before deploying.

Of the next 500 customers, 72 bought items recommended by the new model.

$\frac{72}{500} \Rightarrow 14\%$ — Observed

Is this improvement statistically significant or not?

Confidence of 95%.

Soln) $H_0: p = 0.1$

$H_A: p > 0.1$

- Test statistic → ? → Random Variable

X = # of people out of 500 buyer recom. product.

$X = 0, 1, 2, \ldots, 500$ ✓ $X \sim$ Binomial

Binomial ⟶ $n = 500$, $p = 0.1$

Under the assumption of H_0 being true

- $p\text{-value} = P[X \geq 72 \mid H_0 \text{ is True}]$

$= P[X \geq 72 \mid \text{Binom}(n = 500, p = 0.1)]$

$= 1 - \text{binom.cdf}[k = 71, n = 500, p = 0.1)$

p-value = 0.001

$\alpha = 0.05$

$p\text{-value} < \alpha$.

⇒ Reject H_0 → New model is Better.

Q.) Biased Coin.

$P(H) = 70\%$

$P(T) = 30\%$

} Simulate → A fair Coin.

⇒ Let's toss coin twice :

$SS = \{HH, HT, TH, TT\}$

100

$P(HH) = 0.7 \times 0.7 = 0.49$ → 49

$P(HT) = 0.7 \times 0.3 = 0.21$ → 21

$P(TH) = 0.3 \times 0.7 = 0.21$ → 21

$P(TT) = 0.3 \times 0.3 = 0.09$ → 09

HT, HT, ----

TH, TH, ----

21 H

21 T

Q) SQL.

SQL Queries are equally likely to:

a) Execute successfully in 1 minute,
b) Fail at 3 minutes
Upon failure, we run the query again till it is successful. Find the expected time to run this query

- S, FS, FFS, FFFS, - - - - - -
- 1, 4, 7, 10, 13, 16, 19, - - - - - -

$X \sim geom$

$E(X)$

start running

$E(X)$

$1/2$ → Success $E(X|S) = 1$

$1/2$ → fail. $E(X|F) = 3 + E(X)$

$$E(X) = \sum_i X_i \cdot p(X_i)$$ → weighted Avg.

$$E(X) = 1 \times \frac{1}{2} + (3 + E(X)) \times \frac{1}{2}$$

$$E(X) = 4 \quad *$$