

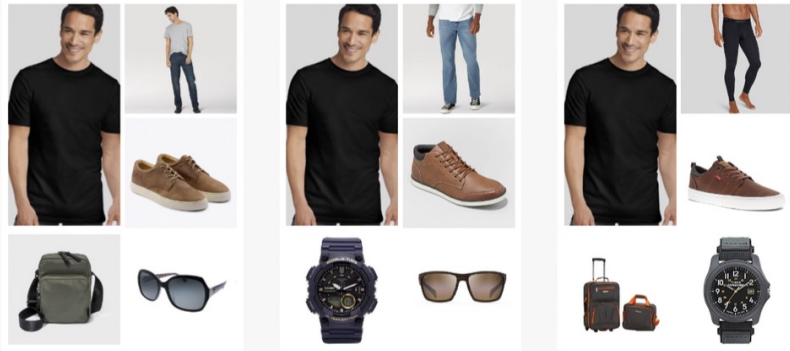
Hypothesis Testing

3

— Interview Question

Recommender System

Find the perfect way to style it
Build a complete look with our inspiration boards



The image displays three distinct styling boards for a male model wearing a black t-shirt. Each board presents a different set of accessories and footwear to complete the look. Board 1 features tan sneakers, a green messenger bag, and sunglasses. Board 2 features brown sneakers, a blue watch, and sunglasses. Board 3 features brown sneakers, a black watch, and a black bag. Each board is labeled at the bottom.

View Board 1 View Board 2 View Board 3

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.

$$\frac{72}{500} \Rightarrow 14.4\%$$

Is the improvement brought by the new model statistically significant at 95% confidence?

$$\alpha = 0.05$$

Solⁿ. $H_0: p = 0.1$
 $H_A: p > 0.1$

• Test statistic: $X \rightarrow$ # of people who bought the recommended items.

what distribution?

$0, 1, 2, \dots, 500$ \rightarrow Binomial
 $(n=500)$

$$X \sim \text{Binomial}(n=500, p=0.1)$$

\Rightarrow Right tailed Test ✓

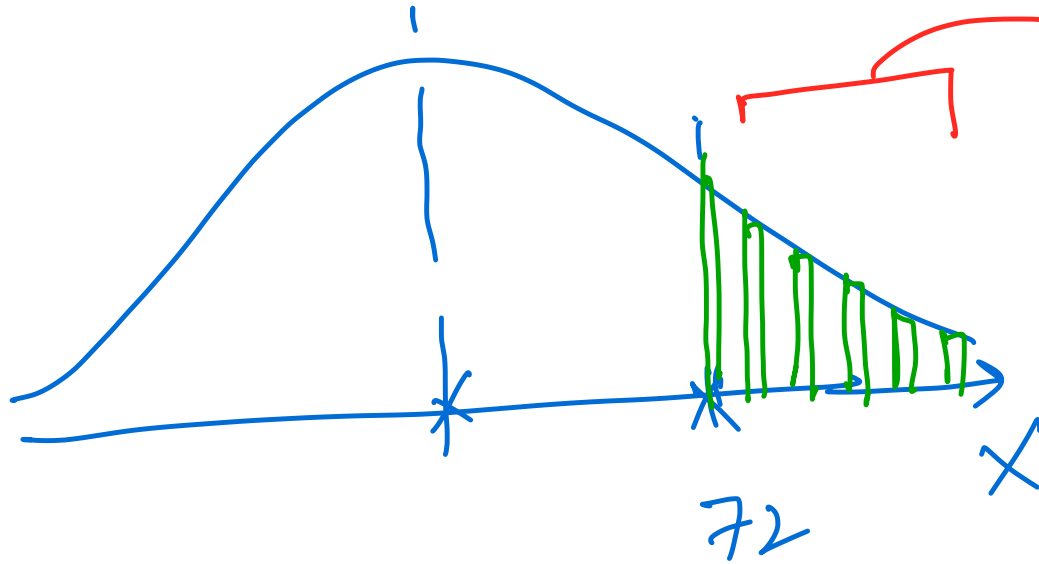
p-value:

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

$$= 1 - P[X \leq 71 | H_0 \text{ is True}]$$

$$= 1 - \text{Binom.cdf.}$$

$$\left(\begin{array}{l} n=500, \\ R=71, \\ p=0.1 \end{array} \right)$$



$$= \underline{\underline{0.001}}$$

$$\text{p-value} = \underline{\underline{0.001}}$$

$$\bullet \alpha = 0.05$$

$\text{p-value} < \alpha$
Reject H_0 ✓