Stat 88: Prob. & Mathematical Statistics in Data Science

https://xkcd.com/892/

Lecture 32 : 4/12/2024

Section 9.3, 9.4

A/B testing & confidence intervals

Hypotheses tests: Review of steps

Example is a coin fair?

1. State the ***null*** hypothesis (H_0) – that is, what is the assumption we are going to make. This will determine how we compute probabilities

H_0: coin is fair ($P(H)=\frac{1}{2}$)

H_0: coin is unfair ($P(H)\neq\frac{1}{2}$)

2. State an ***alternative*** hypothesis (H_A) Note that this should not overlap with the null hypothesis, and it may or may not define probabilities (example: there was gender bias etc)

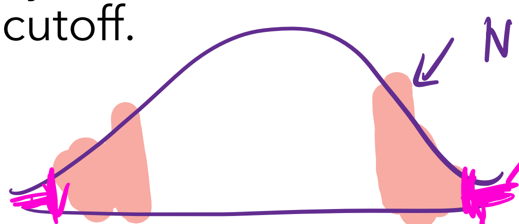
3. Decide on a test statistic to use that will help you decide which of the two hypotheses is supported by the evidence (data). Usually there is a natural choice. Use the null hypothesis to specify probabilities for the test statistic.

H_0: specifies a distribution for the test statistic

4. Find the observed value of the test statistic, and see if it is ***consistent*** with the null hypothesis. That is, compute the chance that we would see such an observed value, or more extreme values of the statistic (p-*value*).

P-value: P(obs value or more extreme | H_0 is true)

5. State your conclusion: whether you reject the null hypothesis or not. This is based on your chosen cutoff ("level of the test"). Reject if the p-value is less than the cutoff.



Example: Woburn

In the early 1990s, a leukemia cluster was identified in the Massachusetts town of Woburn. Many more cases of leukemia, a malignant cancer that originates in a bone marrow cell, appeared in this small town than would be predicted. Was this evidence of a problem in the town or just chance?

At the time of this case, US popn ≈ 280 mill.

# of leukemia cases 30,800 (in US)

$p = \frac{30{,}800}{280 \times 10^6} \approx 0.00011$

Population of Woburn ≈ 35000 = n

Expected # of leukemia cases = np = 3.85

Observed # of leukemia cases = 8

Generic null : "obs. diff is due to chance"

Step 1 define H_0, H_A

H_0: $p = 0.00011$ or $np = \mu = 3.85$ (excess cases random event)

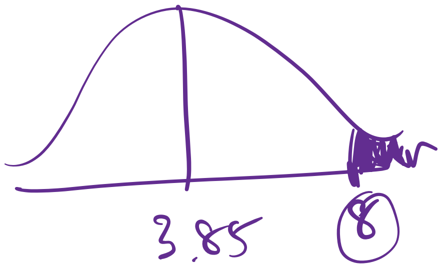
H_A: $\underline{p > 0.00011}$ ($np = \mu > 3.85$)

Step 2 : define test stat X = # of cases in Woburn

$X \sim Bin(n = 35000, p = 0.00011)$ ← determined by H_0.

Step 3: Compute prob of observed data using dsn specified by H_0

$P(X \geq 8)$ (& then we will double)



Going to use the Poisson dsn to approximate the prob. since n is large & p is tiny.

poisson ($\lambda = 3.85$)

$\text{Prob}(X \geq 8) = 4.27\%$

$= 1 - \text{stats.poisson.cdf}(7, 3.85)$ = p-value

NOTE

The p-value shape depends on H_A

Ex $H_A: p \neq p_0$

2-sided alternative, 2 sided P-value

If $H_A: p > p_0$

If $H_A: p < p_0$

Exercise: complete this.

If value is $< 5\%$ Reject H_0

" " " $\geq 5\%$ Do not reject H_0.

Could compare to 5% → Reject H_0

" " " 1% → Do not reject H_0

The threshold we compare to is called the significance level.

Observed significance levels (a.k.a p-values)

- The p-*value* decides if observed values are *consistent* with the null hypothesis. It is a *tail* probability (also called *observed significance level)*, and is the chance, **assuming that the null hypothesis is true**, of getting a test statistic equal to the one that was observed or even more in the direction of the alternative.
- If this probability is too small, then something is wrong, perhaps with your assumption (null hypothesis). That is, the data are unlikely if the null is true and therefore, your data are ***inconsistent*** with the null hypothesis.
- p-value is **not** the chance of null being true. The null is either true or not.
- The p-value is a *conditional probability* since it is computed *assuming* that the null hypothesis is true.
- The smaller the p-value, the stronger the evidence ***against*** the null and ***towards*** the alternative (in the direction of the alternative).
- Traditionally, below 5% ("result is statistically significant") and 1% ("result is highly significant") are what have been used. Significant means the p-value is small, not that the result is important.

Ex. 9.5.1

- All the patients at a doctor's office come in annually for a check-up when they are not ill. The temperatures of the patients at these check-ups are independent and identically distributed with unknown mean μ.
- The temperatures recorded in 100 check-ups have an average of 98.2 degrees and an SD of 1.5 degrees. Do these data support the hypothesis that the unknown mean μ is 98.6 degrees, commonly known as "normal" body temperature? Or do they indicate that μ is less than 98.6 degrees?

obs. valued sample mean

(1) $H_0: \mu = 98.6$

$H_A: \mu < 98.6$

$\sigma \approx 1.5$ (sample SD)

$n = 100$ $\quad \frac{\sigma}{\sqrt{n}} \approx \frac{1.5}{\sqrt{100}} = 0.15$

(2) Test statistic is A_n ← normalized, $E(A_n) = \mu$

Observed value of A_n is 98.2 deg F

$$T = \frac{A_{100} - \mu}{\sigma/\sqrt{100}} \approx N(0,1)$$

P value $= P\left(T < \underbrace{\frac{98.2-98.6}{0.15}}_{-2.66}\right) = 0.0038$

Reject the Null.