Stat 88: Probability & Mathematical Statistics in Data Science

Lecture 28 PART 2: 4/2/2021

Section 9.1

Introducing Hypothesis tests

Hypothesis tests

- Hypothesis tests or *tests of significance* are tests in which we use data to draw conclusions, or *make inferences* about the process that generated the data, or the population from which we drew the sample.
- Underlying idea: Observed values can't be too far from the expected value, *if our assumptions are correct*.
- What if they are not correct? ***How far is too far***?

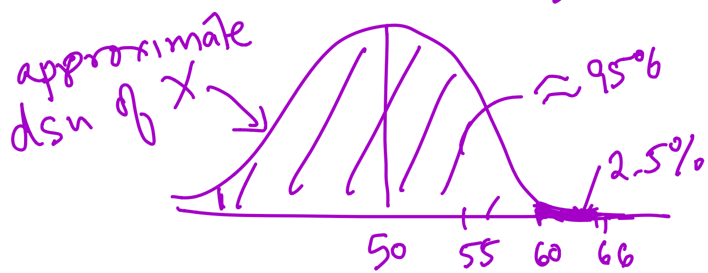
$\boxed{X \sim Bin(100, \frac{1}{2})}$

- Toss a coin 100 times. See 54 heads. Do you have reason to believe that the coin is not fair?

$E(X) = 50$

$SD(X) = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 5$

- What about 60 heads? 66 heads?
- How would we decide?



- Let's look at some scenarios.

Example: Gender bias

A large supermarket chain in Florida occasionally selects employees to receive management training. A group of women there claimed that female employees were passed over for this training in favor of their male colleagues. The company denied this claim.

Suppose that the large employee pool of the Florida chain that can be tapped for management training is half male and half female. Let's assume they have 1000 employees, 500 females

Since this program began, *none* of the 10 employees chosen have been female. What would be the probability of 0 out of 10 selections being female, if there truly was no gender bias?

Example: Woburn

(from Intro to Stats : De Veaux, Velleman et. al)
A Civil Action

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?

8 cases in a population of 35,000

Example: The pill

A pharmaceutical company advertises for their birth control pill has an efficacy of 99.5% in preventing pregnancy. However, under typical use, the real efficacy is only about 95%. That is, 5% of the women taking that pill for a year will experience an unplanned pregnancy that year. A gynecologist looks back at a random sample of 200 medical records from patients who have been prescribed this pill one year before.

1. She finds that 14 women had become pregnant within 1 year while taking the pill. Is this surprising?
2. What about 20 pregnant women in the sample? Is this surprising?

What would you conclude in (1)? What about (2)?

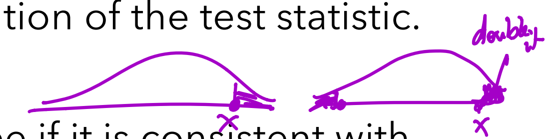
Hypotheses tests: Review of steps

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

2. State an appropriate ***alternative*** hypothesis. Note that this should not overlap with the null hypothesis. You should state both the hypotheses in informal terms and in terms of random variables.

random variable

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. The null hypothesis will specify the distribution of 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 " "

5. If this probability is too small, then something is wrong, perhaps with your assumption (null hypothesis).

Vocabulary review

- null hypothesis (H_0) H-zero or H-naught
- alternative hypothesis (H_1) (H_A)
- *P-value or observed significance level*
  - 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. P("data observed" | null is true)
  - The smaller the P-value, the stronger the evidence ***against*** the null and ***towards*** the alternative (in the direction of the alternative)
  - "Small" is for you to decide. 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.

Fix your tolerance before looking at the data.
This level is called SIGNIFICANCE LEVEL or α
usually set at 5% or 1%

Example: Gender bias

A large supermarket chain in Florida occasionally selects employees to receive management training. A group of women there claimed that female employees were passed over for this training in favor of their male colleagues. The company denied this claim.

Suppose that the large employee pool of the Florida chain that can be tapped for management training is half male and half female.

Since this program began, *none* of the 10 employees chosen have been female. What would be the probability of 0 out of 10 selections being female, if there truly was no gender bias?

X = # of females selected in 10 years

H_0: There is no gender bias in selection

$H_0 \Longrightarrow X \sim HG(\underset{N}{1000}, \underset{G}{500}, 10) \leftarrow$

H_1: There is a gender bias $\leftarrow$ no mathematical statement

Test statistic $= X$ observed value of $X = 0$

$\text{P-value} = P(X \leq 0 \mid \underbrace{X \sim HG(1000, 500, 10)}_{\text{specified by } H_0})$

$P(X \overset{\text{"}}{=} 0)$

P-value $= P(X=0 \mid H_0 \text{ is true})$

$$P(X=0) = \frac{\binom{500}{0}\binom{500}{10}}{\binom{1000}{10}}$$

≈ 0.00093

Approximate using $X \sim Bin(10, \frac{1}{2})$

Data contradict Ho, so reject Ho.

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? 8 cases

$n = 35{,}000$ (popn of Woburn in early '90s)

Using #s from that time, $p = \frac{30{,}800}{280{,}000{,}000} \approx 0.00011$

30,800 ← Leukemia cases

280,000,000 ← US popn (BACK THEN)

p is small, n is large. Let X = # of cases in Woburn

assuming that whether a resident has leukemia or not is independent of all other residents.

$X \sim Bin(35000, 0.00011)$

$E[X] \approx 3.85$

Ho: No problem, just chance that we would have 8 cases.

H_1: # of cases is not due to chance (some other reason)

Test statistic is X

p-value - use Poisson dsn to compute this, $\lambda = 3.85$

$P(X \geq 8) = 1 - \text{stats.poisson.cdf}(7, 3.85) \approx 0.0427$

Result: is significant but not highly significant

Do the data contradict H_0? If so, reject H_0.

Example: The pill

A pharmaceutical company advertises for their birth control pill has an efficacy of 99.5% in preventing pregnancy. However, under typical use, the real efficacy is only about 95%. That is, 5% of the women taking that pill for a year will experience an unplanned pregnancy that year. A gynecologist looks back at a random sample of 200 medical records from patients who have been prescribed this pill one year before.

1. She finds that 14 women had become pregnant within 1 year while taking the pill. Is this surprising?

2. What about 20 pregnant women in the sample? Is this surprising?

What would you conclude in (1)? What about (2)?

X = # of women on the pill who got pregnant (among sample of 200)

H_0: The efficacy is 95% $\longrightarrow$ $\underbrace{X \sim Bin(200, 0.05)}_{\text{specified by } H_0}$ $E(X) = 1$

H_1: The efficacy is NOT 95%

Test statistic $= X$

P-value for observed value of $X = 14$: $P(X \geq 14 \mid H_0) = 1 - \sum_{k=0}^{13} \binom{200}{k} (0.05)^k (0.95)^{200-k}$

1-stats.binom.cdf(13, 200, 0.05)

≈ 0.1299

P-value for observed value of $X = 20$: $P(X \geq 20 \mid H_0) \approx 0.001$

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?

$H_0: \mu = 98.6$

$H_1: \mu < 98.6$

Test statistic : average temp recorded for 100 patients

Observed value : 98.2 deg F

P-value: Note that A_{100} = avg. temp = $\dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$

Use CLT, $A_{100} \sim$ normal,

$\mu = 98.6$ (by H_0)

$SD(A_{100}) \approx \dfrac{1.5}{\sqrt{100}} = 0.15$

Observed value = 98.2

p-value $= P(A_{100} < 98.2)$ $\qquad Z = \dfrac{A_{100} - 98.6}{0.15}$

$= P\left(Z < \dfrac{98.2 - 98.6}{0.15}\right)$

$= P(Z < -2.6667) \approx 0.0038$

p-value very small, reject the null.

<u>Conclusion</u>: Data support the alternative

(mean temp < 98.6).