Chapter 7: One sample tests

(Ott & Longnecker Sections: 5.4-5.7, 5.9)

Duzhe Wang

Part 2

https://dzwang91.github.io/stat371/

UNIVERSITY OF WISCONSIN–MADISON

What do we study

Key concepts: Power calculation, sample size calculation, sign test

Power of a test

Definition

The power of a test is the probability of rejecting H_0 given that a specific alternative hypothesis is true. That is

$$\text{Power} = 1 - \beta$$

where β is the Type II error.

Power of a test

Definition

The power of a test is the probability of rejecting H_0 given that a specific alternative hypothesis is true. That is

$$\text{Power} = 1 - \beta$$

where β is the Type II error.

- How do we calculate the power of a test given a specific alternative hypothesis?

An example

Story

A paint shop uses an automatic device to apply paint to engine blocks. It is important that the amount applied is of a minimum thickness.

Primary Research Question

Its customer, a manufacturer wants to know the average thickness of paint in a warehouse. It is supposed to be 1.50mm.

$\mu = 1.50\text{mm}?$

Sampling

16 blocks are selected randomly and then measured from thousands of blocks in the warehouse.

$n = 16$

Data

1.29, 1.12, 0.88, 1.65, 1.48, 1.59, 1.04, 0.83, 1.76, 1.31, 0.88, 1.71, 1.83, 1.09, 1.62, 1.49

(in mm)

$\bar{x} = 1.358$

$s = 0.3385$

An example

- Step 1:

$$H_0\colon \mu = 1.5,\ H_A\colon \mu \neq 1.5$$

- Step 2: Choose $\alpha = 0.05$
- Step 3: Use the T test statistic: $T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$. In this example, $t_{obs} = \frac{1.348 - 1.50}{\frac{0.3385}{\sqrt{16}}} = -1.796$. The rejection region is $T < -t_{n-1,\alpha/2}$, $T > t_{n-1,\alpha/2}$. In this example, $t_{15,0.025} = 2.13$, so the rejection region is $T < -2.13$ or $T > 2.13$.
- Step 4: Make a conclusion: since $t_{obs} = -1.796$ does not fall in the rejection region, so we do not reject the null.

Power computation

- Suppose $H_A : \mu = 1.4$,

$$\begin{aligned}\text{Power} = P(\text{Reject } H_0 | H_0 \text{ false}) &= P\left(|T| > 2.13 \middle| \mu = 1.4\right) \\ &= P\left(\left|\frac{\bar{X} - 1.5}{S/\sqrt{n}}\right| > 2.13 \middle| \mu = 1.4\right) = ?\end{aligned}$$

Power computation

- Suppose $H_A : \mu = 1.4$,

$$\begin{aligned} \text{Power} = P(\text{Reject } H_0 | H_0 \text{ false}) &= P\left(|T| > 2.13 \middle| \mu = 1.4\right) \\ &= P\left(\left|\frac{\bar{X} - 1.5}{S/\sqrt{n}}\right| > 2.13 \middle| \mu = 1.4\right) = ? \end{aligned}$$

- Under $\mu = 1.4$, $\frac{\bar{X}-1.5}{S/\sqrt{n}}$ is not a t distribution! But $\frac{\bar{X}-1.4}{S/\sqrt{n}}$ is a t distribution.

Power computation

- Suppose $H_A : \mu = 1.4$,

$$\text{Power} = P(\text{Reject } H_0 | H_0 \text{ false}) = P\left(|T| > 2.13 \middle| \mu = 1.4\right)$$
$$= P\left(\left|\frac{\bar{X} - 1.5}{S/\sqrt{n}}\right| > 2.13 \middle| \mu = 1.4\right) = ?$$

- Under $\mu = 1.4$, $\frac{\bar{X}-1.5}{S/\sqrt{n}}$ is not a t distribution! But $\frac{\bar{X}-1.4}{S/\sqrt{n}}$ is a t distribution.
- We rewrite $\frac{\bar{X}-1.5}{S/\sqrt{n}} = \frac{\bar{X}-1.4}{S/\sqrt{n}} + \frac{1.4-1.5}{S/\sqrt{n}}$

Power computation continued

- Now under $\mu = 1.4$,

$$Power = P(|\frac{\bar{X} - 1.5}{S/\sqrt{n}}| > 2.13|\mu = 1.4) =$$

$$P(|\frac{\bar{X} - 1.4}{S/\sqrt{n}} + \frac{1.4 - 1.5}{S/\sqrt{n}}| > 2.13|\mu = 1.4)$$

- Since $S = 0.3385, n = 16$, so

$$Power = P(|\frac{\bar{X} - 1.4}{S/\sqrt{n}} - 1.18| > 2.13|\mu = 1.4)$$

$$= P(\frac{\bar{X} - 1.4}{S/\sqrt{n}} > 3.31|\mu = 1.4) + P(\frac{\bar{X} - 1.4}{S/\sqrt{n}} < -0.95|\mu = 1.4)$$

$$= P\left(T_{15} < -0.95\right) + P\left(T_{15} > 3.31\right) = 0.179 + 0.002 = 0.181$$

Power computation continued

- Now under $\mu = 1.4$,

$$Power = P(|\frac{\bar{X} - 1.5}{S/\sqrt{n}}| > 2.13|\mu = 1.4) =$$

$$P(|\frac{\bar{X} - 1.4}{S/\sqrt{n}} + \frac{1.4 - 1.5}{S/\sqrt{n}}| > 2.13|\mu = 1.4)$$

- Since $S = 0.3385, n = 16$, so

$$Power = P(|\frac{\bar{X} - 1.4}{S/\sqrt{n}} - 1.18| > 2.13|\mu = 1.4)$$

$$= P(\frac{\bar{X} - 1.4}{S/\sqrt{n}} > 3.31|\mu = 1.4) + P(\frac{\bar{X} - 1.4}{S/\sqrt{n}} < -0.95|\mu = 1.4)$$

$$= P(T_{15} < -0.95) + P(T_{15} > 3.31) = 0.179 + 0.002 = 0.181$$

- This power is low. This means that if the true population mean really was $\mu = 1.4$, we would be very unlikely to reject the null based on a sample of size 16. How can we increase the power?

Sample size computation

- We can control the probability of Type I and Type II error simultaneously provided we can collect the requisite number of samples. What is the required sample size to achieve some power?
- It can be shown that if σ is known, the sample size n required to achieve power $1-\beta$ for a test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ at a given alternative $\mu = \mu_A$ at level α is approximately:

  $$n = \left(\frac{\sigma(z_{\alpha/2}+z_\beta)}{\mu_0-\mu_A}\right)^2$$

  where $z_{\alpha/2}$ and z_β are the $\alpha/2$ and β right-tailed critical values of the standard normal distribution.

Sample size computation continued

- Now suppose we wanted to determine the sample size to have power 0.8 to reject when the true mean was $\mu_A = 1.4$.
- Thus β is 1-0.8 $= 0.2$. Use $S = 0.3385$ as an estimator of σ. Using the standard normal table we have $z_{0.05/2} = 1.96$ and $z_{0.2} = 0.84$. By formula we would need:

$$n = \left(\frac{0.3385(1.96+0.84)}{1.5-1.4}\right)^2 = 89.8, \text{ round up to } 90.$$

Sign tests

- We'll introduce a nonparametric(distribution free) statistical test for median.
- The advantage of sign tests compared with t tests is they do not require strong distributional assumptions.

Sign tests

- We'll introduce a nonparametric(distribution free) statistical test for median.
- The advantage of sign tests compared with t tests is they do not require strong distributional assumptions.
- **Example:** The sanitation department in a large city is considering separating recyclable material out of the trash to save on landfill space and make money selling the recyclables. Based on data from other cities, it is determined that if at least half of the households in the city have 4.6 lbs or more of reclaimable recyclable material, then the separation will be profitable for the city. A random sample of 11 households yields the following data on pounds of recyclable material found in the trash:

  $$14.2, 5.3, 2.9, 4.2, 1.2, 4.3, 1.1, 2.6, 6.7, 7.8, 25.9$$

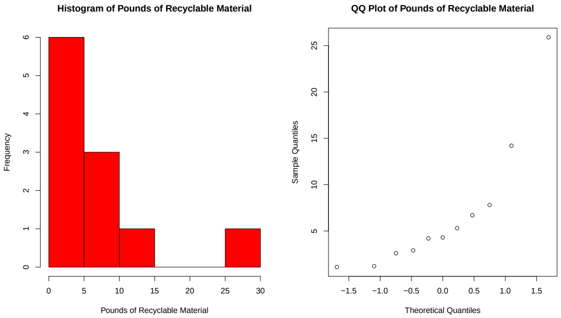
  Note that the median of the sample is 4.3 lbs. We are interested in whether the median of the whole city is 4.6lbs.

Example continued

- Let m be the population test, we want to test: $H_0 : m = 4.6$ vs. $H_A : m > 4.6$.

Example continued

- Let m be the population test, we want to test: $H_0 : m = 4.6$ vs. $H_A : m > 4.6$.
- A QQ plot and histogram:



The histogram looks neither normal nor symmetric, and the QQ plot also does not support normality.

- Since we only have $n = 11$, the CLT is risky.

Example continued

- If the null hypothesis is correct, then by the definition of median, the sample should have about half of the observations greater than 4.6 and half less than 4.6.
- Thus a natural choice of test statistic is the number of observations in the sample that are greater than 4.6, which we might call B.
- If the null hypothesis is correct, $B \sim Bin(11, 0.5)$.
- Our observed test statistic is $B_{obs} = 5$. Hence, the p-value is $P(B \geq 5) = 0.726$. This was found by adding Binomial$(11, 0.5)$ probabilities for $B = 5$, $B = 6$, ..., $B = 11$ when $n = 11$ and $p = 0.5$.
- Set $\alpha = 0.05$, then there is not sufficient evidence to reject the null. There is not strong evidence that separation of the recyclables would be profitable in this city.

Sign test

- Assumption. Let $X_1, \ldots, X_n$ be an i.i.d. sample from some population with population median m.

Sign test

- Assumption. Let $X_1, \ldots, X_n$ be an i.i.d. sample from some population with population median m.
- Hypothesis. We wish to test $H_0 : m = m_0$ vs. H_A. Possible alternative hypotheses: $H_A : m > m_0$(one-tailed), $H_A : m < m_0$(one-tailed), or $H_A : m \neq m_0$(two-tailed).

Sign test

- Assumption. Let $X_1, \ldots, X_n$ be an i.i.d. sample from some population with population median m.
- Hypothesis. We wish to test $H_0 : m = m_0$ vs. H_A. Possible alternative hypotheses: $H_A : m > m_0$(one-tailed), $H_A : m < m_0$(one-tailed), or $H_A : m \neq m_0$(two-tailed).
- Test statistic. $B =$ number of data values greater than m_0. (Ignore values tied with m_0.) Note that if H_0 is true, $B \sim \text{Binomial}(n^*, 0.5)$, where n^* is the number of data points not equal to m_0.

Sign test

- Assumption. Let $X_1, \ldots, X_n$ be an i.i.d. sample from some population with population median m.
- Hypothesis. We wish to test $H_0 : m = m_0$ vs. H_A. Possible alternative hypotheses: $H_A : m > m_0$(one-tailed), $H_A : m < m_0$(one-tailed), or $H_A : m \neq m_0$(two-tailed).
- Test statistic. $B =$ number of data values greater than m_0. (Ignore values tied with m_0.) Note that if H_0 is true, $B \sim \text{Binomial}(n^*, 0.5)$, where n^* is the number of data points not equal to m_0.
- p-value. Let b be the observation of B. If:
  - $H_A : m > m_0$: p-value is $P(B \geq b) = P(B = b) + P(B = b + 1) + \ldots + P(B = n^*)$.
  - $H_A : m < m_0$: p-value is $P(B \leq b) = P(B = b) + P(B = b - 1) + \ldots + P(B = 1) + P(B = 0)$.
  - $H_A : m \neq m_0$: p-value is $2 \min\{P(B \geq b), P(B \leq b)\}$.

What's the next?

We'll discuss how to test population proportions in the next lecture.