Hypothesis Testing

Chapters 8

Introduction

- Hypothesis testing is key to scientific inquiry
- We just have some hypothesis H
- To check it we collect data D
- We check if H is consistent with D
  - Consistency is probabilistic, and there is no 0/1 answer.
- Example:
  - Say you got a shipment of cables, and average breaking strength is claimed to be at least 7000 pounds per square inch (PSI).
  - D = You test 10 random cables and record PSI of each.
  - Hypothesis: PSI >= 7000.
  - Hypothesis testing: Is the hypothesis consistent with the observed data?

More real-world examples of hypothesis testing

- Number of clicks on the video is at least 100

- Average order value has increased since last financial year

- Investing in A brings a higher return than investing in B

- The new user interface converts more users into customers than the expected 30%

Types of hypothesis testing

- Parametric Vs Non-parametric
  - Parametric
    - Assume that the underlying data distribution has a known parametric form. E.g. Normal or exponential
  - Non-parametric
    - Unknown functional form of the distribution.

- One-sample Vs Two-sample
  - One-sample: D = sample from one distribution
  - Two-sample: D1, D2 = samples from two different distributions. Hypothesis is comparing them.
- Paired Vs Unpaired:

Parametric hypothesis testing for a single population

8.3.2 of Ross Text-book

Parametric Hypothesis testing

- Let $F_\theta(X)$ be a distribution on X with unknown parameters θ
- We want to test some property of θ E.g.

i) $\theta = \theta_0$ ii) $\theta \geq \theta_0$ iii) $\theta \leq \theta_0$

θ θ_0

  - Simple tests: all parameters fully specified

$\theta = \theta_0$ eg: $X \sim F_\theta(x) = Exp(\lambda)$; $\lambda = 2$

  - Complex tests

$\theta \geq \theta_0$ or $\theta \leq \theta_0$ or $\theta = [\theta_1, \theta_2]$; $\theta_1 = \theta_{1,0}$

θ_2 is unconstraints

$N(\mu, \sigma^2)$

$\mu = 2$

σ^2 is unspecified

- Step 1: Collect data (Evidence)

$D \equiv \{ X_1, X_2, X_3 \dots X_n \}$

Why not simple likelihood tests?

For simple hypothesis where all parameters are specified by the user: $\theta = \theta_0$

A default method

Measure log likelihood of D $\equiv \underbrace{\sum_{i=1}^{N} \log F_{\theta_0}(X_i)}_{LL(D|\theta_0)}$ ✓

If $LL(D|\theta_0)$ is high enough then accept hypothesis

Shortcomings:

1) Threshold is not specified.
2) Only applicable for simple hypothesis

Step 2: Formulate the question using a pair of hypothesis

- Null hypothesis H_0 that tests for equality (Reason will be clear later)

$\theta = \theta_0$ (H_1)

- Alternative hypothesis: Alternative values of the parameters

a) User is asserting that $\theta = \theta_0$, then

$H_1 : \theta \neq \theta_0$

b) User wants to test if $\theta \leq \theta_0$, then

$H_1 : \theta \leq \theta_0$

c) User wants to test if $\theta \geq \theta_0$ then

$H_1 : \theta \geq \theta_0$

Examples

- Is the average IQ of students in this class greater than 120?

$H_0: \mu = \mu_0 = 120$; $H_1: \mu > 120$

- Is the rise in temperature over the last ten years less than 2 degrees?

$H_0: \mu = 2$ deg ; $H_1: \mu \leq 2$

Step 3: Compute if D is extreme given hypothesis

- Step 3.1: Compute a test statistic T from the data --- some summary of the data D (Design step)

- Step 3.2: Identify the probability of T under the null hypothesis

  $P_{\theta_0}(T)$
  - For some F_{θ}, T it may be possible to show this in closed parametric form
  - For others, simulations may be required.

- Step 3.3: User specifies a significance level α, the error tolerance of rejecting the null hypothesis even if it is true

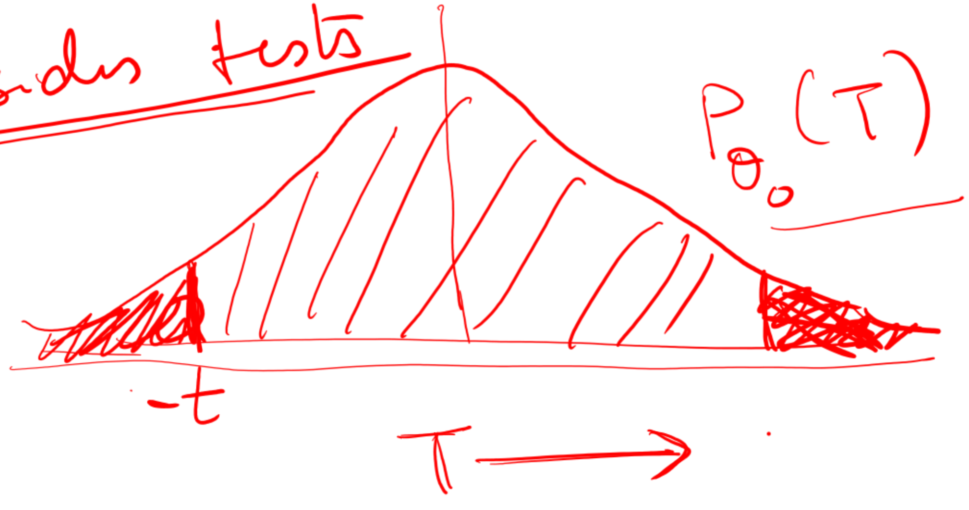
Step 3.4: Compute if T is extreme under $P_{\theta_0}(T)$

observed (t)

Two related ways:

- Define a critical region C such that D being in that region is unlikely if H_0 is true, and where H_1 is more likely, Or
- Compute a p-value: the probability — assuming the null hypothesis was true — of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed.

Two-sides tests



t = observed value of T from given data D.

$H_0: \theta = \theta_0$; $H_1: \theta \neq \theta_0$.

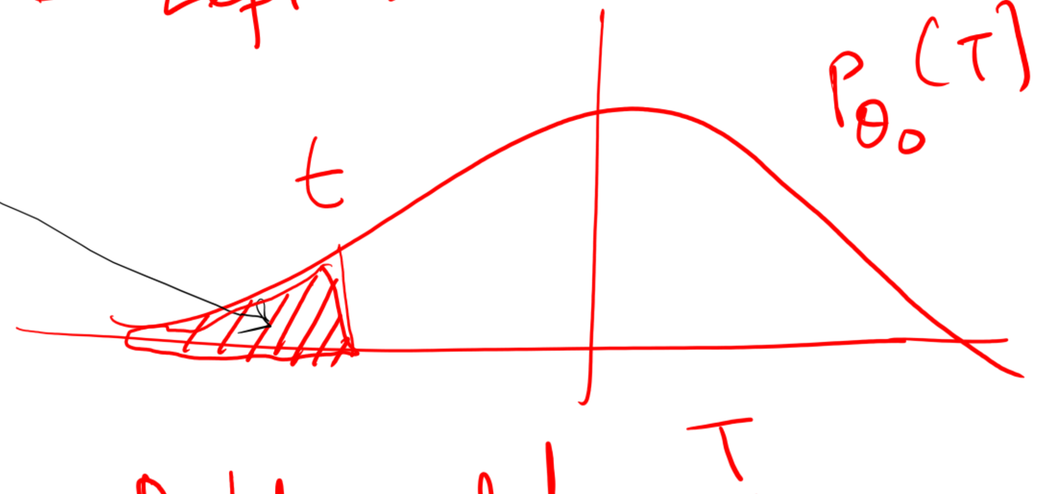
$\theta = \theta_0$

Assume $t > 0$

$P_{\theta_0}(T < -t \text{ or } T > t) \approx$ p-value.

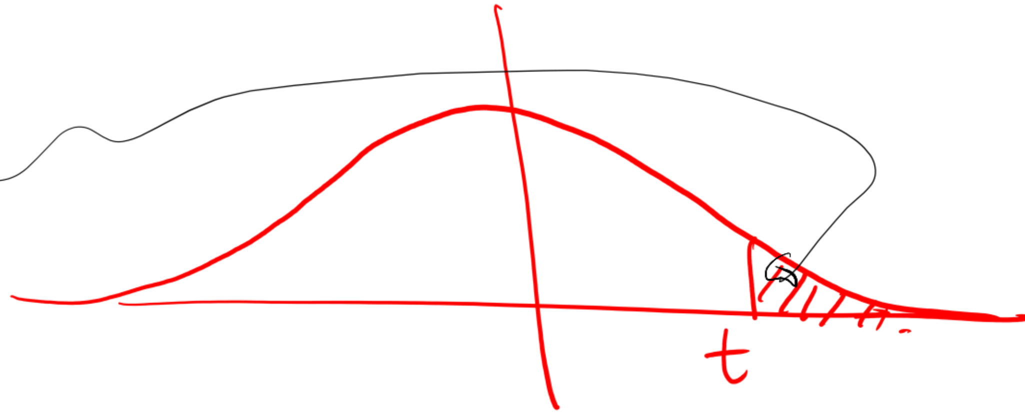
$H_0: \theta = \theta_0$; $H_1: \theta \le \theta_0$ — Left-sided test

Compute $P_{\theta_0}(T \le t)$: p-value.



$H_0: \theta = \theta_0$; $H_1: \theta \geq \theta_0$ Right-sided

$P(T \geq t)$: p-value.



Step 3.5: Accept/Reject decisions based on p-values

- Step 3.5: Compare p-value with given significance-level α

If p-value of test is $> \alpha$ then
then accept the null hypothesis H_0

else

accept the alternative H_1

Step 3.4 using critical regions method instead of p-values

- Define a Critical or reject region C such that the probability of T being in C is at most α, and if t lies in C then H_1 is more likely to be true than H_0
- Defining C for different hypothesis types

$P_{\theta_0}(T \in C) = \alpha$

Symmetric around 0

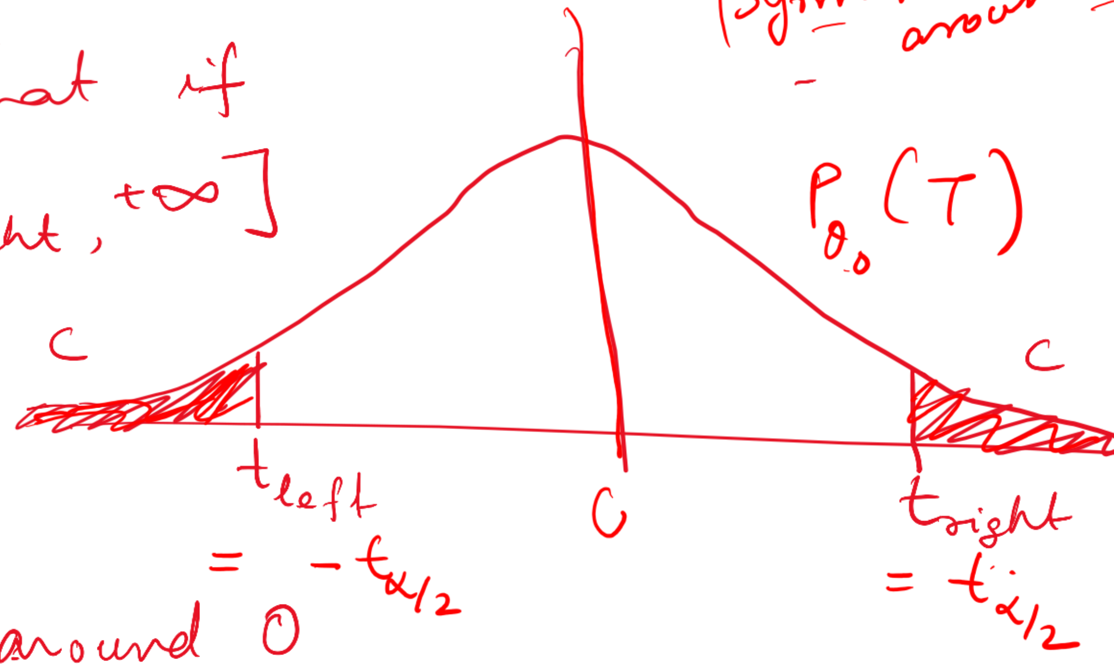
1) $H_0: \theta = \theta_0$; $H_1: \theta \neq \theta_0$

Define t_{left} & t_{right} such that if

$C = [-\infty, t_{left}] \cup [t_{right}, +\infty]$

$P(T \in C) = \alpha$

or $P(T \leq t_{left}$ or $T \geq t_{right}) = \alpha$.

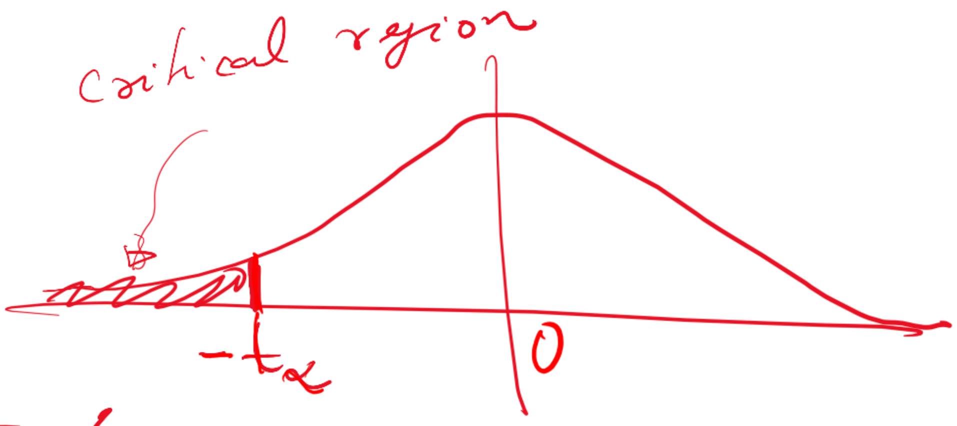


Often $P_{\theta_0}(T)$ will be symmetric around 0

So, find $t_{\alpha/2}$ s.t. $P(T \leq -t_{\alpha/2}) = \frac{\alpha}{2}$ & $P(T \geq t_{\alpha/2}) = \frac{\alpha}{2}$

Defining C for other hypothesis

2) $H_0 : \theta = \theta_0$; $H_1 : \theta \le \theta_0$



$C = [-\infty, t_\alpha]$ s.t

$P_{\theta_0}(T \in C) = P_{\theta_0}(T \le -t_\alpha) = \alpha$

$t_\alpha \equiv$ the upper percentile of $P_{\theta_0}(T)$

3) $H_0 : \theta = \theta_0$; $H_1 : \theta > \theta_0$

$C \equiv [t_\alpha, +\infty]$ s.t

$P_{\theta_0}(T \in C) = P_{\theta_0}(\theta \ge t_\alpha) = \alpha$

Step 3.5: Accept/Reject decisions based on critical regions

If $t \in C$ then reject null hypothesis
else accept " "

1) If $H_0: \theta = \theta_0$; $H_1: \theta \neq \theta_1$
$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

2)

3)

Example: Hypothesis test on mean of a normal distribution with unknown variance

Suppose that $X_1, \ldots, X_n$ is a sample of size n from a normal distribution having an unknown mean μ and a unknown variance $\cancel{\sigma^2}$ and suppose we are interested in testing the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis

$$H_1 : \mu \neq \mu_0$$

where μ_0 is some specified constant.

$$D = \{ x_1, x_2 - - x_n \}$$

Possible test statistic: $T = \left| \frac{\sum_{i=1}^{n} x_i}{n} - \mu_0 \right|$

variance is not known – so cannot define $P_{\theta_0}(T)$

A better choice of test statistic

$$\boxed{T = \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}}}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

$P_{\theta_0}(T)$??

$P(\bar{x}) \sim N\left(\mu_0 ; \frac{\sigma^2}{n}\right)$ unknown

- A good test-statistic is one where the distribution of T can be easily computed, and in that distribution the null hypothesis region is well separated from the alternative hypothesis.

Distribution of test statistic

- Property of sample mean and sample variance of a normal distribution

Theorem 6.5.1. If $X_1, \ldots, X_n$ is a sample from a normal population having mean μ and variance σ^2, then $\overline{X}$ and S^2 are independent random variables, with $\overline{X}$ being normal with mean μ and variance σ^2/n and $(n-1)S^2/\sigma^2$ being chi-square with $n-1$ degrees of freedom.

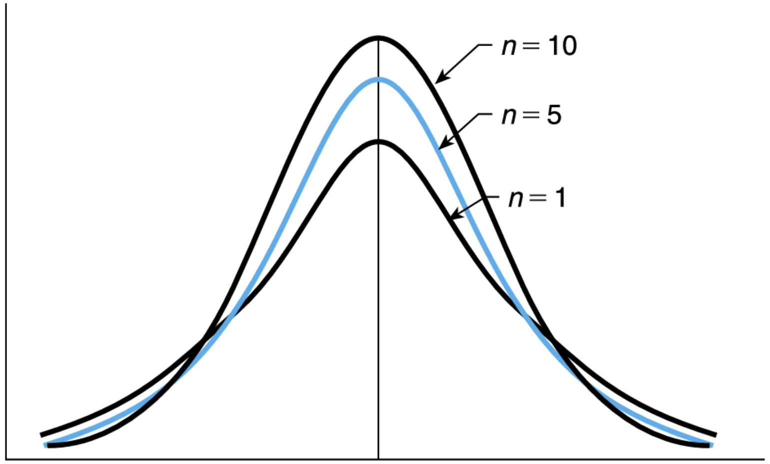
$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} + \left[\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right]^2$$

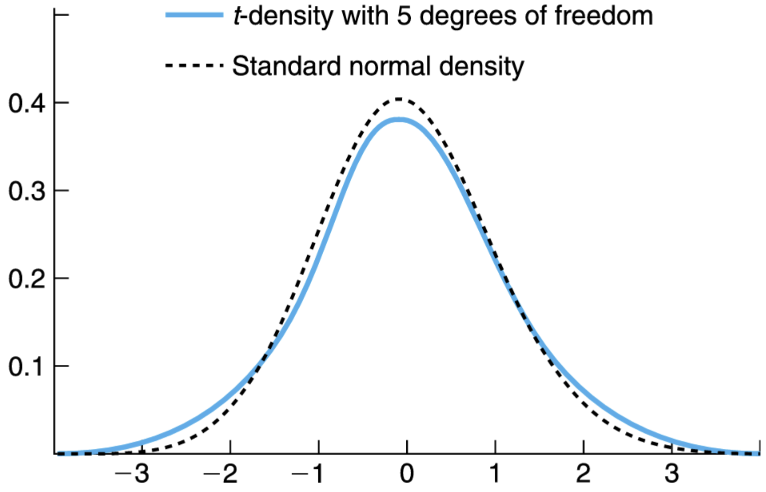
Distribution of test statistic (t-distribution)

If Z and χ_n^2 are independent random variables, with Z having a standard normal distribution and χ_n^2 having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

is said to have a *t-distribution with n degrees of freedom*. A graph of the density function of T_n is given in Figure 5.13 for $n = 1, 5,$ and 10.





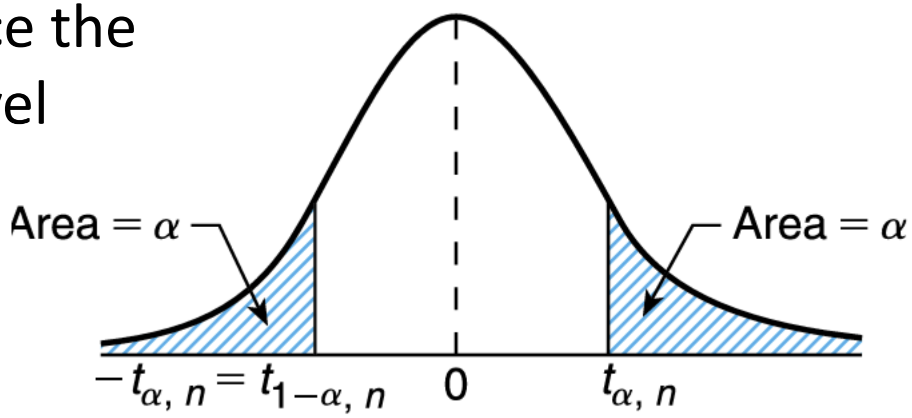
$$E[T_n] = 0, \qquad n > 1$$

$$\mathrm{Var}(T_n) = \frac{n}{n-2}, \qquad n > 2$$

Distribution of test statistic

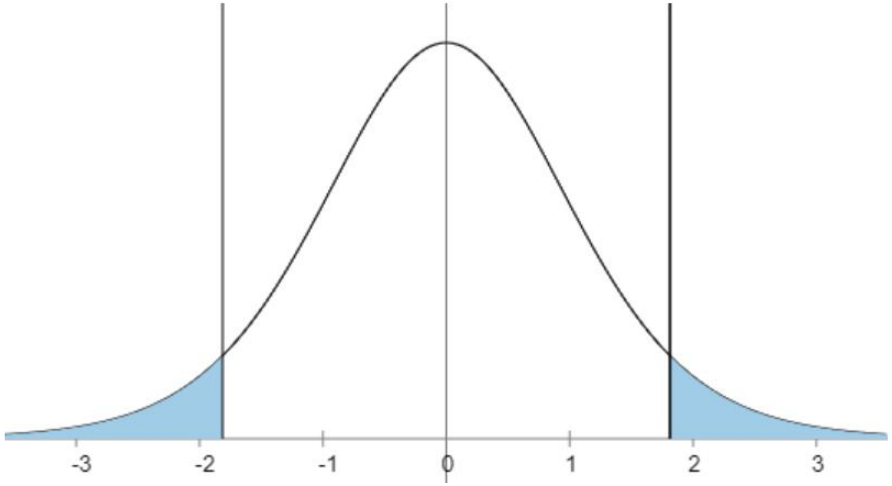
Since T depends only on n, we can compute in advance the cutoffs at which the area is less than a significance level

$$P\{T_n \geq t_{\alpha,n}\} = \alpha$$



$$P\{T_n \geq -t_{\alpha,n}\} = 1 - \alpha$$

$n = 10, \alpha = 0.05, t_\alpha = 1.812$



http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/pt.html

Defining critical region

$T = \dfrac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} = \dfrac{\sqrt{n}(\bar{X} - \mu_0)}{S}$ t-distribution with $n-1$ df.

Define

$$P_{\mu_0}\left\{-t_{\alpha/2,n-1} \le \underbrace{\boxed{\frac{\sqrt{n}(\overline{X}-\mu_0)}{S}}}_{} \le t_{\alpha/2,n-1}\right\} = 1-\alpha$$

T

$P_{\mu_0}(T) \sim t_{n-1}$

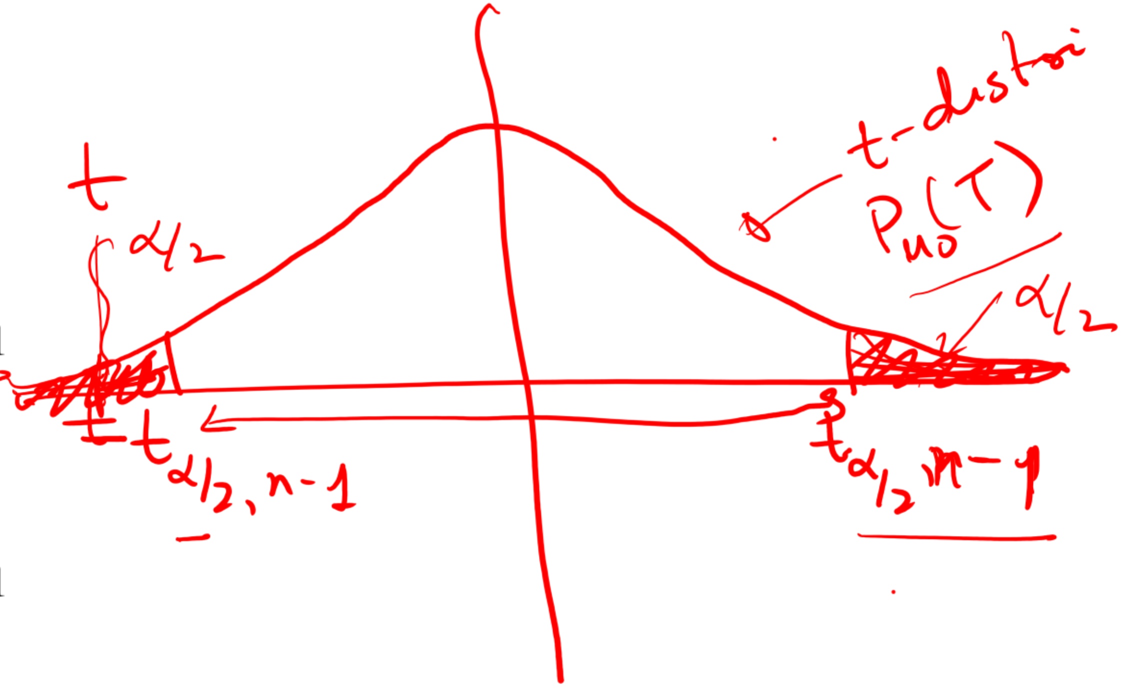
$t_{\alpha/2,n-1}$: upper $\alpha/2$ percentile of t_{n-1}

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0$$

is, when σ^2 is unknown, to

$$\text{accept} \quad H_0 \quad \text{if} \quad \left|\frac{\sqrt{n}(\overline{X}-\mu_0)}{S}\right| \le t_{\alpha/2,\,n-1}$$

$$\text{reject} \quad H_0 \quad \text{if} \quad \left|\frac{\sqrt{n}(\overline{X}-\mu_0)}{S}\right| > t_{\alpha/2,\,n-1}$$



Example 8.3.g. Among a clinic's patients having blood cholesterol levels ranging in the medium to high range (at least 220 milliliters per deciliter of serum), volunteers were recruited to test a new drug designed to reduce blood cholesterol. A group of 50 volunteers was given the drug for 1 month and the changes in their blood cholesterol levels were noted. If the average change was a reduction of 14.8 with a sample standard deviation of 6.4, what conclusions can be drawn?

Solution. Let us start by testing the hypothesis that the change could be due solely to chance — that is, that the 50 changes constitute a normal sample with mean 0. Because the value of the t-statistic used to test the hypothesis that a normal mean is equal to 0 is

$H_0 : \mu = 0 \; ; \quad H_1 : \mu > 0$

$n = 50$

$$T = \sqrt{n}\,\overline{X}/S = \sqrt{50}\,14.8/6.4 = 16.352$$

P-value $\quad P_{\theta_0}(T) =$ t-distribution with 49 degrees of freedom

$P_{\theta_0}(T > 16.352) = 5.6 \times 10^{-22} \leftarrow$ Very small

Reject null hypothesis. Drug was effective.

http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/pt.html

Example 8.3.h. A public health official claims that the mean home water use is 350 gallons a day. To verify this claim, a study of 20 randomly selected homes was instigated with the result that the average daily water uses of these 20 homes were as follows:

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Do the data contradict the official's claim?

Solution. To determine if the data contradict the official's claim, we need to test

$$H_0 : \mu = 350 \qquad \text{versus} \qquad H_1 : \mu \neq 350$$

This can be accomplished by noting first that the sample mean and sample standard deviation of the preceding data set are

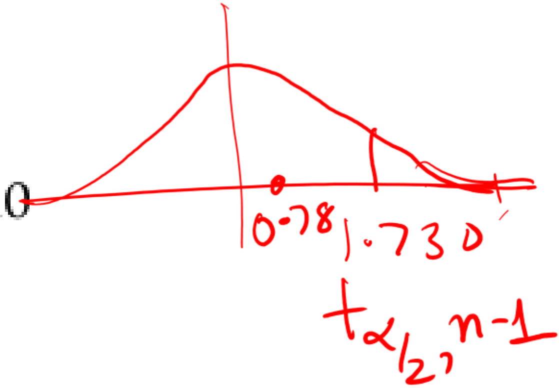
$$\overline{X} = 353.8, \qquad S = 21.8478$$

Thus, the value of the test statistic is

$$T = \frac{\sqrt{20}(3.8)}{21.8478} = .7778$$

$$T = \frac{\sqrt{20}(3.8)}{21.8478} = .7778$$

Because this is less than $t_{.05,19} = 1.730$, the null hypothesis is accepted at the 10 percent level of significance. Indeed, the p-value of the test data is



$$p\text{-value} = P\{|T_{19}| > .7778\} = 2P\{T_{19} > .7778\} = .4462$$

indicating that the null hypothesis would be accepted at any reasonable significance level, and thus that the data are not inconsistent with the claim of the health official. ■

Example 8.3.i. The manufacturer of a new fiberglass tire claims that its average life will be at least 40,000 miles. To verify this claim a sample of 12 tires is tested, with their lifetimes (in 1000s of miles) being as follows:

Practice question

Tire	1	2	3	4	5	6	7	8	9	10	11	12
Life	36.1	40.2	33.8	38.5	42	35.8	37	41	36.8	37.2	33	36

Test the manufacturer's claim at the 5 percent level of significance.

Solution. To determine whether the foregoing data are consistent with the hypothesis that the mean life is at least 40,000 miles, we will test

$$H_0 : \mu \geq 40{,}000 \qquad \text{versus} \qquad H_1 : \mu < 40{,}000$$

A computation gives that

$$\overline{X} = 37.2833, \qquad S = 2.7319$$

and so the value of the test statistic is

$$T = \frac{\sqrt{12}(37.2833 - 40)}{2.7319} = -3.4448$$

Since this is less than $-t_{.05,11} = -1.796$, the null hypothesis is rejected at the 5 percent level of significance. Indeed, the p-value of the test data is

$$p\text{-value} = P\{T_{11} < -3.4448\} = P\{T_{11} > 3.4448\} = .0027$$

indicating that the manufacturer's claim would be rejected at any significance level greater than .003. ■

Parametric hypothesis testing for two populations

Sections: 8.4.2 and 8.4.4 of Ross Text-book

Two-sample t-test

- We have two distributions $F_{\theta_x}(X)$ and $F_{\theta_y}(Y)$ with unknown parameters
- Hypothesis:
  - Whether the two parameters are equal.
  - Whether mean of one is greater than another.
- Data D: n samples of X, and m samples of Y

$x_1, x_2 \cdots x_n \quad y_1, y_2 \cdots y_m$

Special case: normal distributions with equal but unknown variance

$F_{\theta_x}(x) \sim N(\mu_x, \sigma^2)$ $\quad$ $F_{\theta_y}(Y) \sim N(\mu_y, \sigma^2)$

σ^2 is unknown.

- Hypothesis: are their means equal?

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x \neq \mu_y$$

$$\Leftrightarrow \quad H_0 : \mu_x - \mu_y = 0 \qquad H_1 : \mu_x - \mu_y \neq 0$$

- T-statistic:

Pooled or shared sample variance

$$S_x^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

$$S_y^2 = \frac{\sum_{i=1}^{m} (Y_i - \overline{Y})^2}{m-1}$$

$$\Rightarrow \quad S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

$$T \equiv \frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2(1/n + 1/m)}}$$

Distribution of test statistic T

- We can show that T follows a t-distribution with n+m-2 degrees of freedom.

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{S_p^2(1/n + 1/m)}} \sim t_{n+m-2}$$

- Proof: [Section 7.4 of text book]

Example:

Twenty-two volunteers at a cold research institute caught a cold after having been exposed to various cold viruses. A random selection of 10 of these volunteers was given tablets containing 1 gram of vitamin C. These tablets were taken four times a day. The control group consisting of the other 12volunteers was given placebo tablets that looked and tasted exactly the same as the vitamin C tablets. This was continued for each volunteer until a doctor, who did not know if the volunteer was receiving the vitamin C or the placebo tablets, decided that the volunteer was no longer suffering from the cold. The length of time the cold lasted was then recorded. At the end of this experiment, the following data resulted.

At the end of this experiment, the following data resulted.

10 12

Vit c placebo

Treated with Vitamin C	Treated with Placebo
5.5	6.5
6.0	6.0
7.0	8.5
6.0	7.0
7.5	6.5
6.0	8.0
7.5	7.5
5.5	6.5
7.0	7.5
6.5	6.0
	8.5
	7.0

Do the data listed prove that taking 4 grams daily of vitamin C reduces the mean length of time a cold lasts? At what level of significance?

Solution

- $H_0\colon \mu_C - \mu_P = 0, \quad H_1\colon \mu_C - \mu_P < 0$
- t =-1.8987, df = 20, p-value = 0.03606
- Accept the hypothesis that vitamin-C reduces duration of cold at 5 percent significance level

Paired t-test

Paired t-test

- For the same instance i, we have values before a treatment X_i, and after a treatment Y_i
- Example:
  - Suppose we are interested in determining whether the installation of a certain antipollution device will affect a car's mileage. To test this, a collection of n cars that do not have this device are gathered. Each car's mileage per gallon is then determined both before and after the device is installed.
  - How can we test the hypothesis that the antipollution control has no effect on gas consumption? The data can be described by the n pairs (Xi,Yi),i = 1,...,n, where Xi is the gas consumption of the ith car before installation of the pollution control device, and Yi of the same car after installation.

T-statistic for this test

- Assume that $\mathrm{W_i} = X_i - Y_i$ is Guassian with unknown mean and variance.
- Hypothesis to test:

$$H_0 : \mu_w = 0 \qquad \text{versus} \qquad H_1 : \mu_w \neq 0$$

- Test statistic for Gaussian with unknown variance as discussed earlier is

$$T = \frac{\sqrt{n}\,\overline{W}}{S_w} \qquad S_w^2 = \frac{1}{n}\left(\sum_{i=1}^{n}[X_i - Y_i] - \overline{w}\right)^2$$

- Accept/reject decision using critical regions

$$\text{accepting} \quad H_0 \quad \text{if} \quad -t_{\alpha/2,\,n-1} < \sqrt{n}\frac{\overline{W}}{S_w} < t_{\alpha/2,n-1}$$

$$\text{rejecting} \quad H_0 \quad \text{otherwise}$$

Example 8.4.c. An industrial safety program was recently instituted in the computer chip industry. The average weekly loss (averaged over 1 month) in labor-hours due to accidents in 10 similar plants both before and after the program are as follows:

Wi

Plant	Before	After	$A-B$
1	30.5	23	−7.5
2	18.5	21	2.5
3	24.5	22	−2.5
4	32	28.5	−3.5
5	16	14.5	−1.5
6	15	15.5	.5
7	23.5	24.5	1
8	25.5	21	−4.5
9	28	23.5	−4.5
10	18	16.5	−1.5

Test statistic with difference array:

```
>d = c(−7.5, 2.5, −2.5, −3.5, −1.5, .5, 1, −4.5, −4.5, −1.5)
>v = sqrt(10/var(d)) * mean(d)
>v
[1] − 2.265949
>pt(v, 9)
[1]0.02484552
```

$= T$

Thus, $v = -2.265949$, with resulting

$$p\text{-value} = P(T_9 \le -2.265949) = 0.02484552$$

Hypothesis test in Bernoulli population (Section 8.6 of Ross Book)

- The distribution changes to Bernoulli.

$p(x) = p^x (1-p)^{1-x} \quad x \in \{0, 1\}$

- Hypothesis:

$H_0: p = p_0$

$$H_0 : p \leq p_0 \quad \text{versus} \quad H_1 : p > p_0$$

- Data samples of size n will consist of 1s and 0s.

$x_1 \ldots x_n$

- T-statistic: number of 1s in D.

$T = \sum_{i=1}^{n} x_i$; Let t denote observed count.

- Distribution of T-statistic under a parameter p_0 will be Binomial(n, p_0)

$P_{p_0}(T) \sim \text{Binomial}(n, p_0)$

- Calculate p-value of this distribution

$P\{\text{Binomial}(n, p_0) \geq t\} = \sum_{i=t}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$

Example 8.6.a. A computer chip manufacturer claims that no more than 2 percent of the chips it sends out are defective. An electronics company, impressed with this claim, has purchased a large quantity of such chips. To determine if the manufacturer's claim can be taken literally, the company has decided to test a sample of 300 of these chips. If 10 of these 300 chips are found to be defective, should the manufacturer's claim be rejected?

Solution. Let us test the claim at the 5 percent level of significance. To see if rejection is called for, we need to compute the probability that the sample of size 300 would have resulted in 10 or more defectives when p is equal to .02. (That is, we compute the p-value.) If this probability is less than or equal to .05, then the manufacturer's claim should be rejected. Now

practice questions

$$
\begin{aligned}
P_{.02}\{X \geq 10\} &= 1 - P_{.02}\{X \leq 9\} \\
&= 1 - \text{pbinom}(9, 300, .02) \\
&= 0.08183807
\end{aligned}
$$

and so the manufacturer's claim cannot be rejected at the 5 percent level of significance. ■

Example 8.6.b. In an attempt to show that proofreader A is superior to proofreader B, both proofreaders were given the same manuscript to read. If proofreader A found 28 errors, and proofreader B found 18, with 10 of these errors being found by both, can we conclude that A is the superior proofreader?

practice questions

Solution. To begin note that A found 18 errors that B missed, and that B found 8 that A missed. Hence, a total of 26 errors were found by just a single proofreader. Now, if A and B were equally competent then they would be equally likely to be the sole finder of an error found by just one of them. Consequently, if A and B were equally competent then each of the 26 singly found errors would have been found by A with probability 1/2. Hence, to establish that A is the superior proofreader the result of 18 successes in 26 trials must be strong enough to reject the null hypothesis when testing

$$H_0: \ p \leq 1/2 \quad \text{versus} \quad H_1: p > 1/2$$

where p is a Bernoulli probability that a trial is a success. Because the resultant p-value for the data cited is

$$p\text{-value} = P\{\text{Bin}(26, .5) \geq 18\} = 0.03776$$

the null hypothesis would be rejected at the 5 percent level of significance, thus enabling one to conclude (at that level of significance) that A is the superior proofreader. ■