ADS2 Week10-The t-test

2019-11-19

Wanlu Liu

wanluliu@intl.zju.edu.cn

Lab website: labw.org

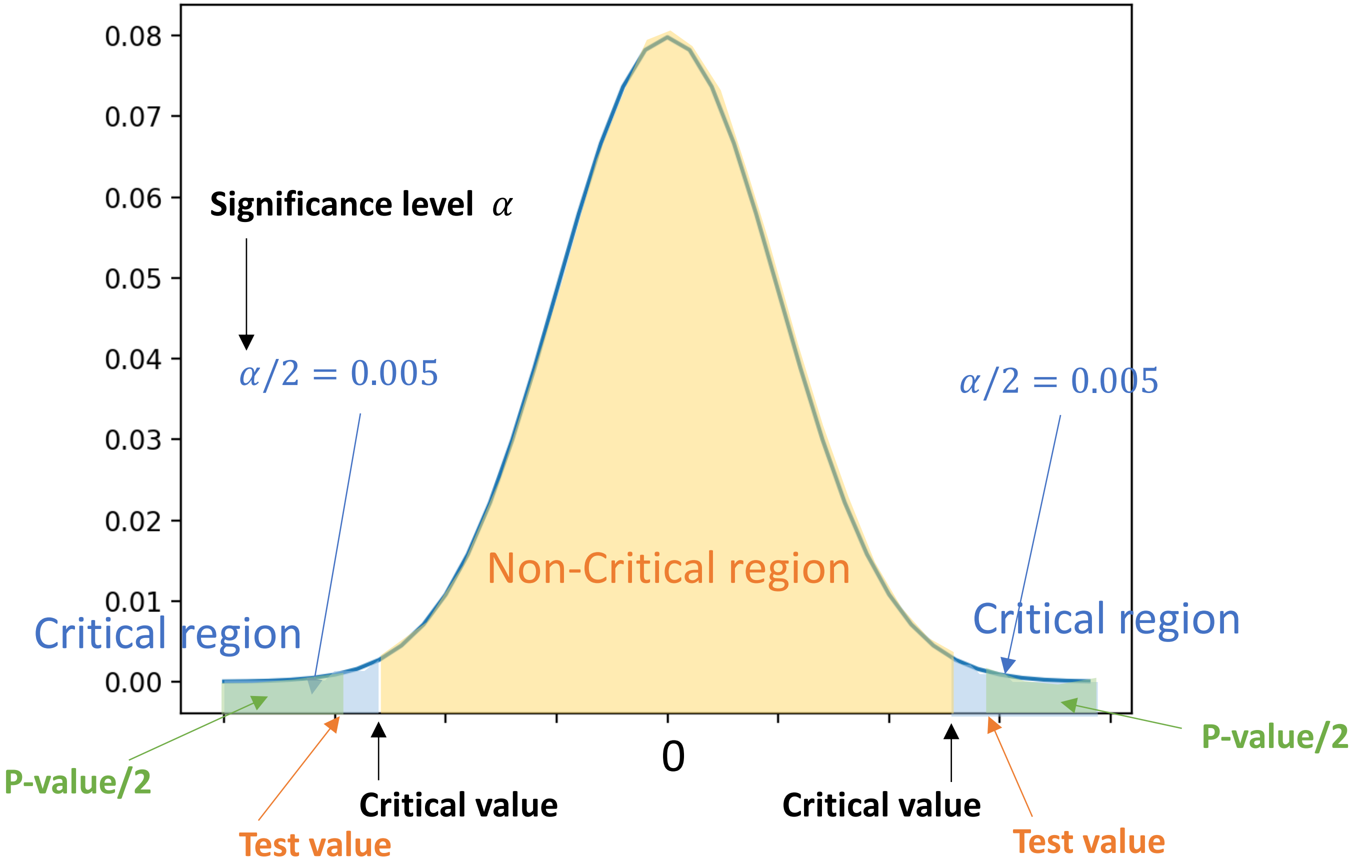
Learning Objectives

- **The T test**
  - Reveal the steps of hypothesis testing
  - Understand the basics on z-test and t-test
  - Explain what is One-sample t-test
  - Distinguish difference between Two-sample paired and unpaired t-test
- **R session**
  - Doing one sample/two sample, one tailed/two tailed, paired/unpaired t-test in R

Some REVIEW

- Hypothesis Testing? Null vs alternative hypothesis?
- What is significance level?
- What is P-value?
- What is critical value? What is critical region and what is non-critical regions?
- What is test value?

Two tailed test, $\alpha = 0.01$



Steps for Hypothesis Testing

1. State the hypotheses and identify the claim.
2. Characteristics of the comparison distribution
3. Find the critical value(s).
4. Compute the test value.
5. Make the decision to reject or not reject the null hypothesis.
6. Summarize the results.

Normality

- Say that $x_1, x_2, \ldots, x_n$ are i.i.d. observations from a Gaussian distribution
  - "i.i.d." stands for independent and identically distributed
- Their sample mean is:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- Sample variance is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (xi - \bar{x})^2$$

Large sample Hypothesis Test **z-statistics**

- Assumption based on **Normal distribution.**

$$\mathcal{N} \sim (\mu, \sigma^2)$$

- The **z test** is a statistical test for the **mean of a population**. It can be used when **n$\geq$ 30**, or when the population is **normally distributed** and **σ is known**.

- The formula for z test is:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

, where $\bar{x}$ = sample mean ($\bar{x} = \frac{x_1 + x_2 + \cdots + xn}{n}$)
μ = population mean
σ = standard deviation of population
n = sample size (number of observations)

My real life problem – z test example

- I got this Mi band which can track my daily sleeping time, and also telling me the mean/SD of sleeping time for other users in the population with the same gender, age groups and city. I want to know whether I sleep more than average or not.

My real life problem – z test example

I think my sleep duration is longer than the population. My average sleep time in August per day is 7.77h. The population with 100 people tested has an average of 7.6h with standard deviation of 1.2h. At α = 0.05, is there enough evidence to support my claim?

Questions:

- This is a **one tailed** test or **two tailed** test? If one tailed, this is a **left** or **right** tailed test?

Step 1: State the hypotheses and identify the claim.

$$H_0 : \mu \leq 7.6 \; and \; H_a: \mu > 7.6 \; (claim)$$

Step 2: what distribution to use?

Normal distribution (n>=30), z test.

Step 3: Find the critical value.

Since α=0.05 ad the test is a right-tailed test, the critical value is z = +1.65.

Step 4: Compute the test value.

$$z = \frac{7.77-7.6}{1.2/\sqrt{31}}=0.789$$

Step 5: Make the decision.

Since the test value, 0.789 < 1.65, falls in the noncritical region, the decision is to not reject the null hypothesis.

Step 6: Summarize the results. There is no enough evidence to support the claim that I slept more than the general public.

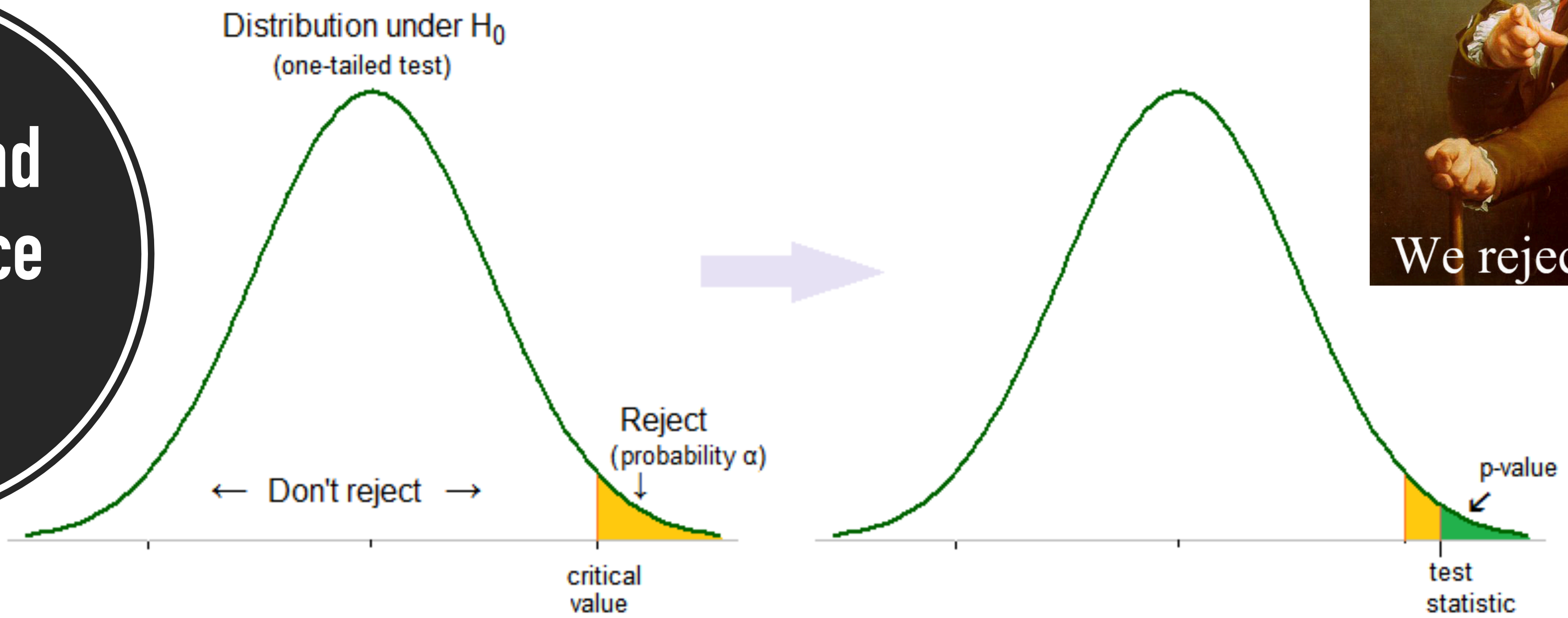
P-value and Significance level

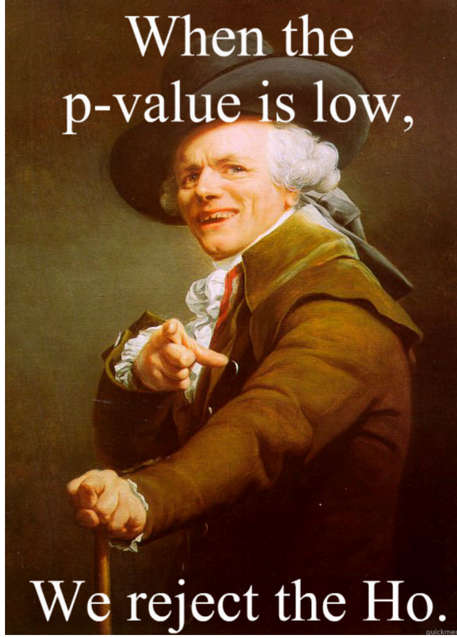
- Besides listing an α value, many computer statistical packages give a P-value for hypothesis tests.

- The P-value is the actual probability of getting the sample mean value or a more extreme sample mean value in the direction of the alternative hypothesis (> or <) if the null hypothesis is true.

- The P-value is the actual area under the standard normal distribution curve (or other curve, depending on what statistical test is being used) representing the probability of a particular sample mean or a more extreme sample mean occurring if the null hypothesis is true.

P-value and Significance level





- If P-value ≤ α, reject the null hypothesis.
- If P-value > α, do not reject the null hypothesis.

How about test for small size sample?

Student's t-Distribution

W.S. Gosset (1876-1937) was a chemist who was a brewer and agricultural statistician for the famous Guinness brewing company in Dublin. It insisted that its employees keep their work secret, so he published under the pen name 'Student' the distribution in 1908. This was one of the first results in modern small-sample statistics.

http://brainstrom.org/why-is-the-the-t-test-called-as-such-why-is-it-named-so-guinness-beer/

Small sample Hypothesis Test t-statistics

- Test for sample mean
- When the population standard deviation is **unknown** and $\boldsymbol{n < 30}$, the z test is inappropriate for testing hypotheses involving means.
- The **t test** is used in this case.
- The formula for t test is:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

, where $\bar{x}$ = sample mean ($\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$)

μ = population mean

s = sample variance ($s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$)

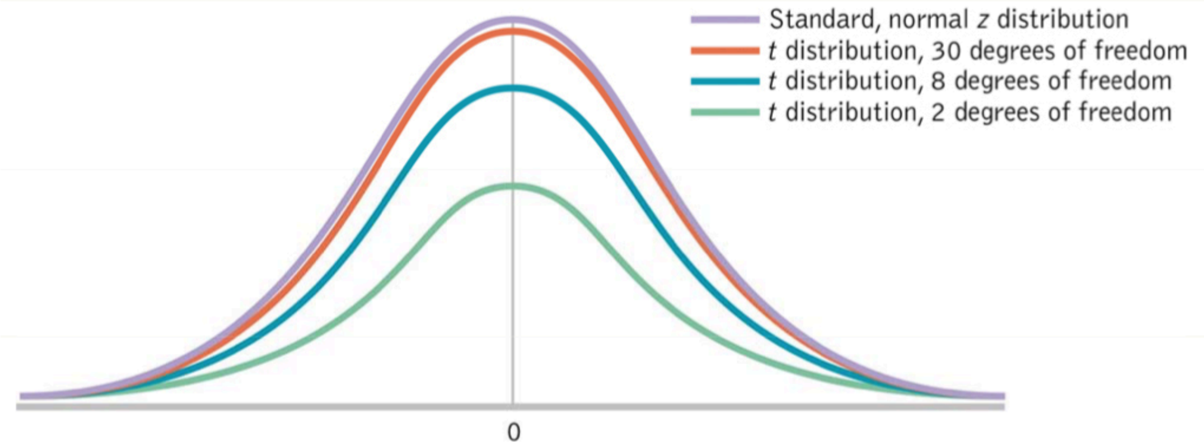
n = sample size

Small sample Hypothesis Test t-statistics

- **Interpretation:** The number of (estimated) standard deviations of the sample mean from its expected value μ

- **The t-value follows a Normal distribution?**
  - No
  - We are using sample variance s^2 , not true variance
  - Closer to normal distribution as the bigger n is.

- The quantity (n-1) is called the **degrees of freedom** of the t value



Types of t-test

- One sample t test (compare one sample to the population)
  - One sample Two tailed t test
  - One sample One tailed t test
- Two sample t test (compare between two samples)
  - Two sample Paired t test (paired condition)
  - Two sample unpaired t test (unpaired condition)

One tailed vs two tailed

My real life problem – t test example

Mi band can track my daily steps. In the last week, the steps I walked for each day is (9486,6875,2721,1778,9338,6867,3826), and my average steps per day for last month is 6040. Does my workout in last week matches my average working out amount?

One sample (last week) compared to the population.

My real life problem – t test example

Mi band can track my daily steps. In the last week, the steps I walked for each day is (9486,6875,2721,1778,9338,6867,3826), and my average steps per day for last month is 6040. Does my workout amount in last week significantly different from my average working out amount?

Questions:

- This is a **one tailed** test or **two tailed** test? If one tailed, this is a **left** or **right** tailed test?

Step 1: State the hypotheses and identify the claim.

$$H_0 : \mu = 6040 \; and \; H_a : \mu \neq 6040 \; (claim)$$

Step 2: what distribution to use?

One sample two tailed t test

Step 3: Find the critical value.

Since α=0.05 ad the test is a two tailed test, and the df=7-1=6, the critical value is t = $\pm$2.447

My real life problem – t test example

Step 4: Compute the test value.

$$\bar{x} = \frac{x_1+x_2+\cdots+xn}{n} = \frac{9486+6875+2721+1778+9338+6867+3826}{7} = 5841.571$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(xi-\bar{x})^2}{n-1}}$$

$$\sqrt{\frac{(9486-5841.6)^2+(6875-5841.6)^2+(2721-5841.6)^2+(1778-5841.6)^2+(9338-5841.6)^2+(6867-5841.6)^2+(3826-5841.6)^2}{6}} = 3107.503$$

So $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{5841.571-6040}{3107.503/\sqrt{7}} = -0.1689439$

Step 5: Make the decision.

Since the test value, -0.168 > -2.447, falls in the noncritical region, the decision is to not reject the null hypothesis.

Step 6: Summarize the results.

There is no enough evidence to support the claim that the workout amount is different than my average workout amount.

R code

```
t.test(x, y = NULL,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, …)
```

$> \ b = c(9486{,}6875{,}2721{,}1778{,}9338{,}6867{,}3826)$
$> \ t.test(b, mu = 6040, aliternative = c(\text{“}two.sided\text{”}))$

```
        One Sample t-test

data:  b
t = -0.16894, df = 6, p-value = 0.8714
alternative hypothesis: true mean is not equal to 6040
95 percent confidence interval:
 2967.610 8715.533
sample estimates:
mean of x
 5841.571
```

My real life problem – t test example

I was playing a Mobile RPG game called Yin yang shi, and the official chance for getting a Specially Super Rare (SSR) card is 1.4%. However, I drew 600 times and never got a SSR card. Am I really that unlucky and should I quit the game?

Suggestions: In order to answer this question scientifically, we need hypothesis testing.

My real life problem – t test example

Back to the Yin yang shi example :), the official data suggest that the average probability for getting a SSR is 1.4%. My eight other friends and I are playing this game, and our probability for get a SSR is (0.1%, 2%, 0.03%, 3%, 0.05%, 0.2%, 0.7%, 0.2%, 0.01%) . Since our average probability (0.70%) of getting a SSR card is smaller than the official claims, we want to know whether we should quit this game because we are unlucky in this game. Is there any evidence to support our claim at $\alpha = 0.05$?

Try to solve out yourself (5min), you can use your computer/calculator!

My real life problem – t test example

Back to the Yin yang shi example :), the official data suggest that the average probability for getting a SSR is 1.4%. My eight other friends and I are playing this game, and our probability for get a SSR is (0.1%, 2%, 0.03%, 3%, 0.05%, 0.2%, 0.7%, 0.2%, 0.01%) . Since our average probability (0.70%) of getting a SSR card is smaller than the official claims, we want to know whether we should quit this game because we are unlucky in this game. Is there any evidence to support our claim at $\alpha = 0.05$?

Step 1: State the hypotheses and identify the claim.

$$H_0 : \mu \geq 1.4\% \; and \; H_a : \mu < 1.4\% \; (claim)$$

Step 2: what distribution to use? One sample left tailed t test

Step 3: Find the critical value.

Since α=0.05 ad the test is a left tailed test, and the df=9-1=8, the critical value is t = -1.860

My real life problem – t test example

Step 4: Compute the test value.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + xn}{n} = \frac{0.1+\ 2+\ 0.03+\ 3+\ 0.05+\ 0.2+\ 0.7+\ 0.2+\ 0.01}{9} = 0.70$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n-1}}$$

$$\sqrt{\frac{(0.1-0.70)^2+(2-0.70)^2+(0.03-0.70)^2+(3-0.70)^2+(0.05-0.70)^2+(0.2-0.70)^2+(0.7-0.70)^2+(0.2-0.70)^2+(0.01-0.70)^2}{8}} = 1.07$$

So $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{0.70 - 1.4}{1.07/\sqrt{9}} = -1.96$

Step 5: Make the decision. Since the test value, -1.96 < -1.86, falls in the critical region, the decision is to reject the null hypothesis.

Step 6: Summarize the results. There is enough evidence to support the claim that the me and my friends are super unlucky in this game, thus we all decide to quit it!

R code

t.test(x, y = NULL,
alternative = c("two.sided", "less", "greater"),
mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, …)

$> a = c(0.1, 2, 0.03, 3, 0.05, 0.2, 0.7, 0.2, 0.01)$
$> t.test(a, mu = 1.4, var.equal = TURE, alternative = c("less"))$

```
        One Sample t-test

data:  a
t = -1.9627, df = 8, p-value = 0.04265
alternative hypothesis: true mean is less than 1.4
95 percent confidence interval:
     -Inf 1.363151
sample estimates:
mean of x
0.6988889
```

Paired (dependent) Two-sample t test

- Given the following data (expression levels of n genes):

$x_1, x_2, \dots, x_n$ Before drug treatment

$y_1, y_2, \dots, y_n$ After drug treatment

- Measure whether the "after" member of the pair is different from the "before" member

$d_1, d_2, \dots, d_n$ After drug treatment Difference $d_i = x_i - y_i$

- Hypothesis testing:
  - Null hypothesis H_0 : The mean of this sample of differences is 0
  - Alternative hypothesis H_a : The mean is not 0

- It is just a one-sample t-test of sort we used above

Un-Paired(independent) Two-sample t test

- Supposed that two samples are drawn independently

$x_1, x_2, \dots, x_n$ Assumed to have a normal distribution with mean μ_x

$y_1, y_2, \dots, y_n$ Assumed to have a normal distribution with mean μ_y

- Is the difference in means that we observe between two groups significant more than we'd expect to see based on chance alone?

- Hypothesis testing:
  - Null hypothesis H_0 : The means of the two samples are equal $\mu_x = \mu_y$
  - Alternative hypothesis H_a : Not equal $\mu_x \neq \mu_y$

Un-Paired(independent) Two-sample t test

$$t = \frac{[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)]}{s_{Difference}}$$

$$s^2{}_{Difference} = \left(\frac{s^2{}_{pooled}}{\mathrm{n}_x}\right) + \left(\frac{s^2{}_{pooled}}{\mathrm{n}_y}\right)$$

$$s^2{}_x = \frac{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2}{n_x - 1}$$

$$s^2{}_y = \frac{\sum(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^2}{n_y - 1}$$

$$s^2{}_{pooled} = \left(\frac{df_x}{df_{total}}\right) s^2{}_x + \left(\frac{df_y}{df_{total}}\right) s^2{}_y$$

$$df_{total} = df_x + df_y$$

Problem– OCT4 gene expression in women vs men

- Is there any difference in OCT4 gene expression in women vs men? Significance α = 0.05

<u>OCT4 gene expression level observed</u>

Women: 84, 97, 58,90

Men: 88,94,52,97,86

https://www.weizmann.ac.il/WeizmannCompass/sections/briefs/in-sickness-and-in-health

Problem– OCT4 gene expression in women vs men

- **Identify the questions:**
- Population
  - Population 1: expression level of OCT4 in 4 women
  - Population 2: expression level of OCT4 in 5 men
- Distribution: differences between mean
  - Not mean differences
- Assumptions are the same

Step 1: State the hypotheses and identify the claim.

$$H_0 : \mu_x = \mu_y \; and \; H_a : \mu_x \neq \mu_y (claim)$$

Step 2: what distribution to use?

Two sample unpaired two tailed t test

Step 3: Find the critical value.

Since α=0.05 ad the test is a two tailed test, and the df_{total}=(4-1)+(5-1)=7, the critical value is t = $\pm$2.365

Problem- OCT4 gene expression in women vs men

Step 4: Compute the test value.

$\bar{x}$=82.25

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
84	1.75	3.063
97	14.75	217.563
58	-24.25	588.063
90	7.75	60.063

$$s^2{}_x = \frac{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2}{n_x - 1} = \frac{868.725}{4-1} = 289.584$$

$\bar{y}$=83.4

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
88	5.4	29.16
94	11.4	129.96
52	-30.6	936.36
97	14.4	207.36
86	3.4	11.56

$$s^2{}_y = \frac{\sum(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^2}{n_y - 1} = \frac{1314.44}{5-1} = 328.6$$

Problem– OCT4 gene expression in women vs men

Step 4: Compute the test value.

Pooled Variance

Weighted average of the two estimates of variance – one from each sample – that are calculated when conducting an independent samples t test.

$$s^2{}_{pooled} = \left(\frac{df_x}{df_{total}}\right) s^2{}_x + \left(\frac{df_y}{df_{total}}\right) s^2{}_y$$

$$s^2{}_{pooled} = \left(\frac{3}{7}\right) 289.584 + \left(\frac{4}{7}\right) 328.6 = 311.878$$

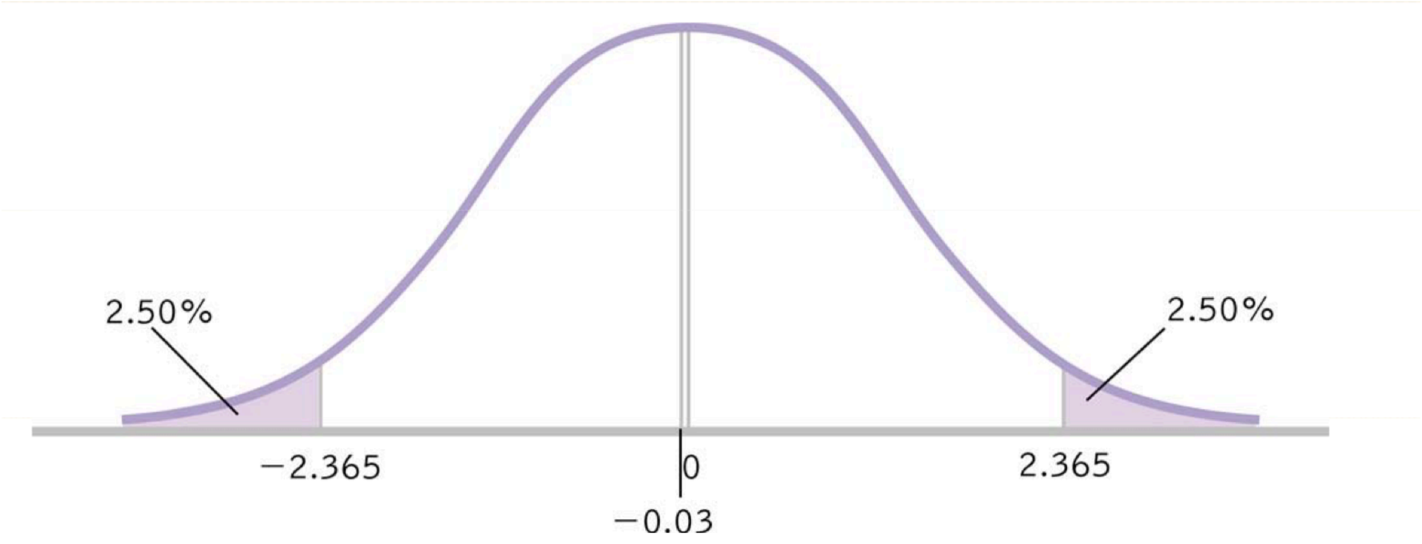
$$s^2{}_{Difference} = \left(\frac{s^2{}_{pooled}}{\mathrm{n}_x}\right) + \left(\frac{s^2{}_{pooled}}{\mathrm{n}_y}\right) = \left(\frac{311.878}{4}\right) + \left(\frac{311.878}{5}\right) = 140.346$$

$$s_{Difference} = \sqrt{140.346} = 11.847$$

Problem- OCT4 gene expression in women vs men

Step 4: Compute the test value.

$$t = \frac{[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)]}{s_{Difference}} = \frac{(\bar{x} - \bar{y})}{s_{Difference}} = \frac{(82.25 - 83.4)}{11.847} = -0.0971$$



Step 5: Make the decision.

Since the test value, -0.097 > -2.365, falls in the noncritical region, fail to reject the null hypothesis.

Step 6: Summarize the results.

There is no evidence to support the claim that OCT4 expression level is different in men vs women.

R code

```
t.test(x, y = NULL,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)
```

$> \ women = c(84{,}97{,}58{,}90)$
$>$ *men= c(88,94,52,97,86)*
$> \ test = t.test(women, men, alternative = \text{“}two.sided\text{”}, paired = FALSE, var.equal = TRUE)$

```
        Two Sample t-test

data:  women and men
t = -0.097144, df = 7, p-value = 0.9253
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -29.14256  26.84256
sample estimates:
mean of x mean of y
    82.25     83.40
```

Summary

- **The T test**
  - steps of hypothesis testing
    - 6 steps
  - Understand the basics on z test and t-test
    - formula
  - One-sample t-test
    - When and how to use one sample t test
  - Two-sample paired and unpaired t-test
    - What’s the difference between paired and unpaired t-test
- **R session**
  - Know how to perform T test in R