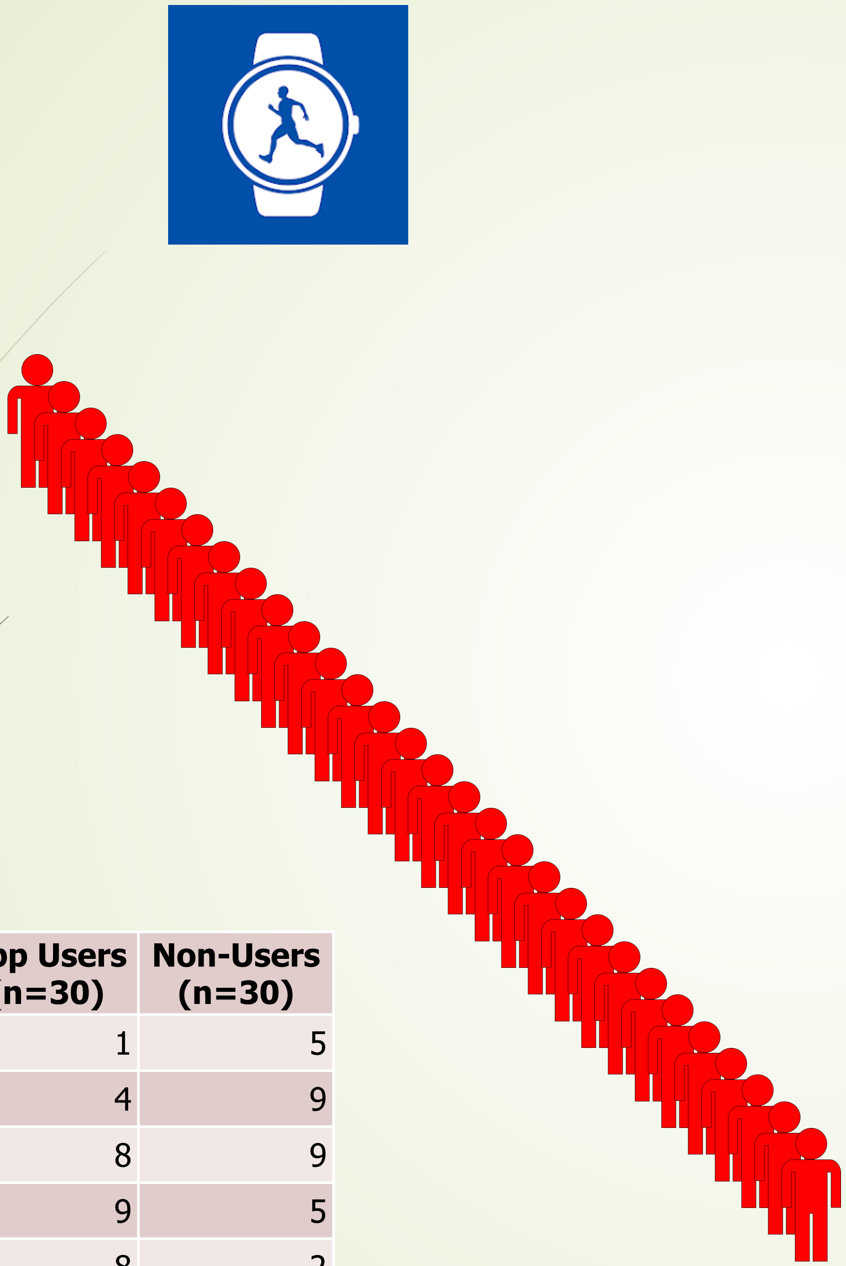
Stat C131A: Statistical Methods for Data Science

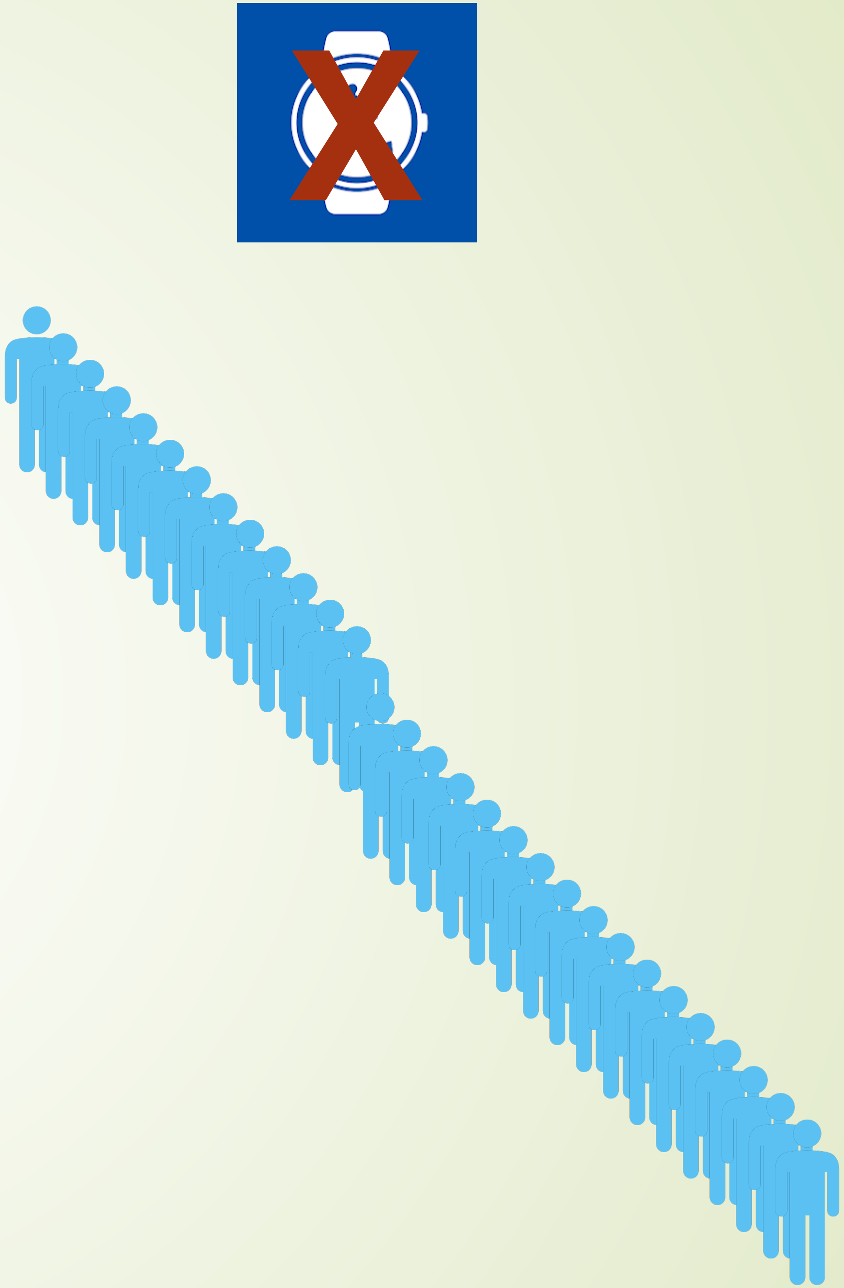
Lecture 9: Hypothesis Testing II

Sep 25 2025

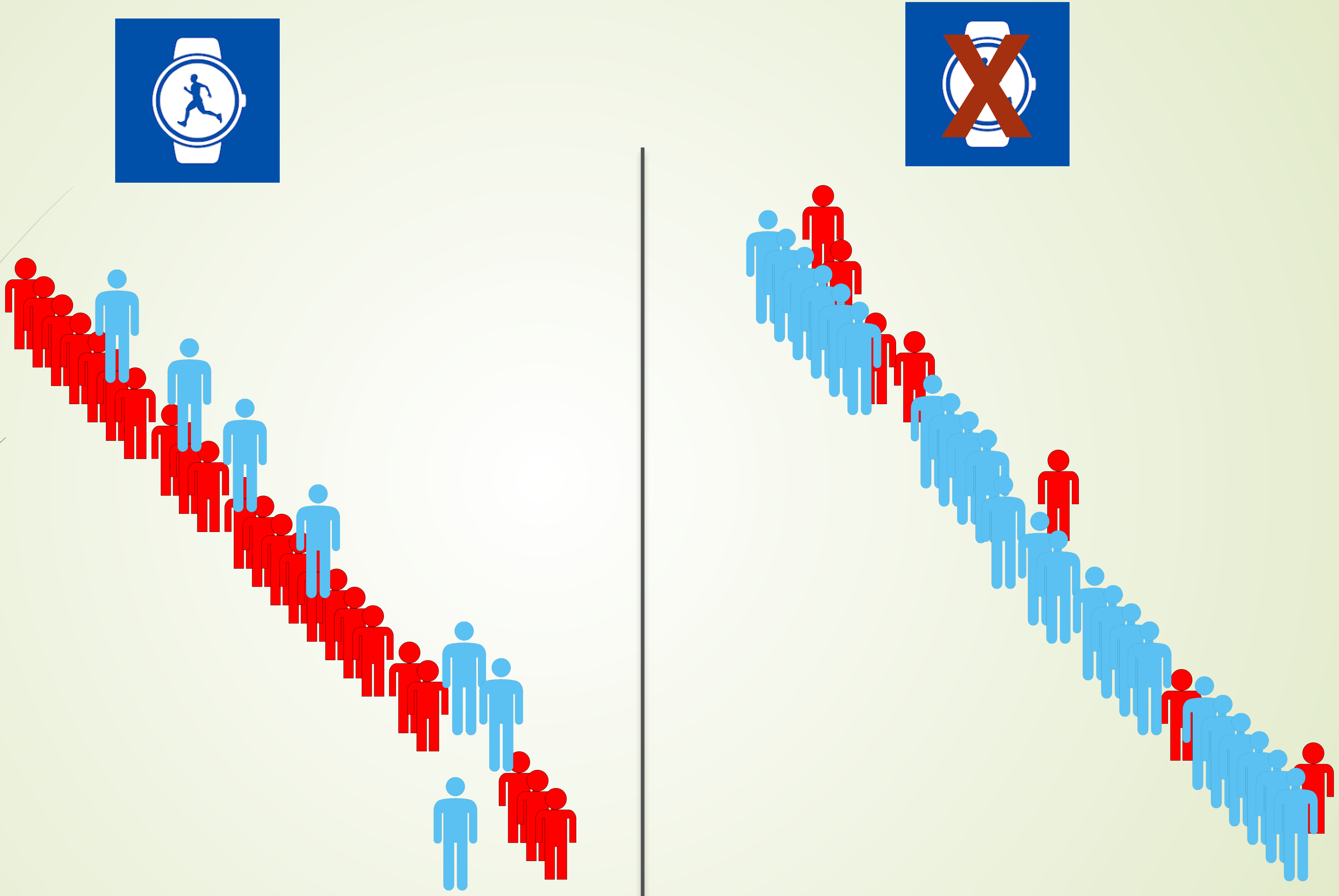
Clarifications on Permutation Test

- It does **NOT assume Normality**, or any underlying specific distribution for that matter, for the Population
- Only assumes **exchangeability of the labels**, and **under the null hypothesis**
- What is **"exchangeability of labels"** ?
  - And which labels exactly ?





Daily Steps Bin	App Users (n=30)	Non-Users (n=30)
< 4,000	1	5
4,000–5,999	4	9
6,000–7,999	8	9
8,000–9,999	9	5
≥ 10,000	8	2



Daily Steps Bin	App Users (n=30)	Non-Users (n=30)
< 4,000	1	5
4,000–5,999	4	9
6,000–7,999	8	9
8,000–9,999	9	5
≥ 10,000	8	2

Daily Steps Bin	App Users (n=30)	Non-Users (n=30)
< 4,000	4	4
4,000–5,999	4	7
6,000–7,999	5	11
8,000–9,999	7	4
≥ 10,000	10	4

What did we compute ?

- We determined that the **average difference in steps person,** between the (true) app users versus the non app users was 1933 steps
- We observed that this difference was met/exceeded only 40 out 10,000 times in our shuffling experiment
- This is our observed statistic, of 40
  - that occurs with a probability of 40/10000= 0.004
- The probability of seeing a difference this extreme (only 40/10000) assuming the null hypothesis to be true (aka there is no effect) is 0.004
  - Which is, our p-value = 0.004
- So we can reject the null hypothesis at the 5% significance level (or even at 1% or 0.5% significance level)

Recap

- Two ways to test
  - Permutation
    - Does not assume underyling population distribution
  - Parametric tests
    - Does assume normal
    - Statistical parameters, of the Population
- The p-value, is under the null hypothesis

Let's work out an example

Kraft to pay fines for underweight meat packages

01.31.2013 *By Staff*

MADISON, Wis. – Northfield, Ill.-based Kraft Foods Group agreed to pay a fine of $13,911.50 and take system-wide corrective actions after the Wisconsin Department of Agriculture, Trade and Consumer Protection discovered underweight packages of Oscar Mayer-brand lunch meat in several stores.

The fine is the company's third in two years for violations of Wisconsin's weights and measures statutes. As part of the agreement, Kraft admitted no wrongdoing.

The underweight packages included eight packages of Oscar Mayer brand cooked ham and one package of Oscar Mayer brand honey ham, according to the agency. The underweight packages were found in Beloit, Dodgeville, Plover, Racine, Watertown, Waukesha and Wisconsin Rapids. The problem was discovered by the state's Consumer Protection Bureau during inspections conducted in June and July 2012.

Problem

Scenario: *There is a concern that a certain brand of lunch packets may be* ***underweight*** *compared to the advertised* ***300 g*** *label claim. To investigate, an agency takes a* ***random sample*** *of* ***36 packets****. The sample has a* ***mean weight of 298.2 g*** *and a sample* ***standard deviation of 3.9 g***

- Question: Is the **company selling underweight packets** ?

Scenario: *There is a concern that a certain brand of lunch packets may be* ***underweight*** *compared to the advertised* ***300 g*** *label claim. To investigate, an agency takes a* ***random sample*** *of* ***36 packets****. The sample has a* ***mean weight of 298.2 g*** *and a sample* ***standard deviation of 3.9 g***

- Question: Is the **company selling underweight packets** ?

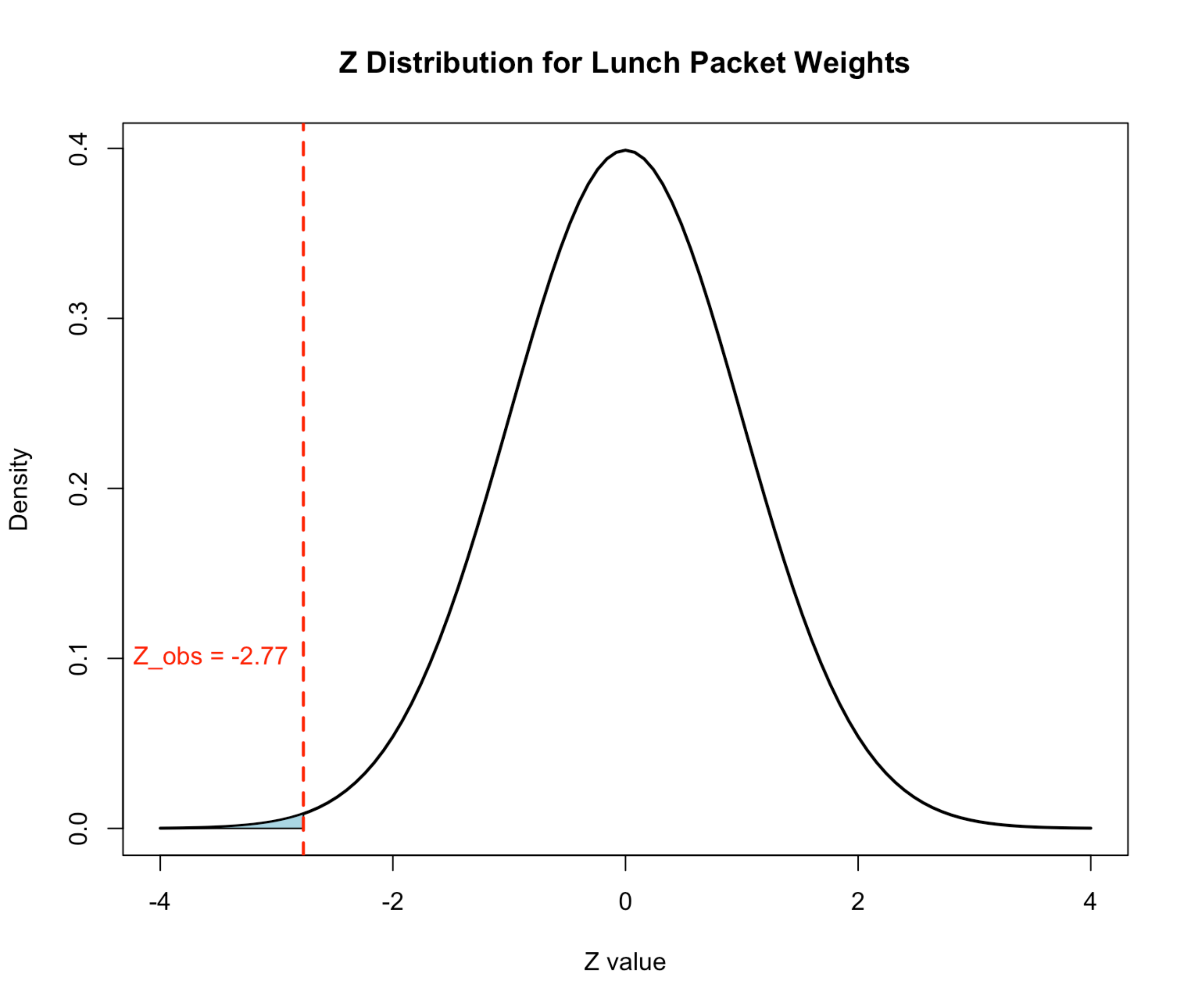
  - μ = Population mean
  - σ = Population standard deviation
  - Null Hypothesis: H_0 : μ = 300
  - Alternate Hypothesis: H_1 : μ = 298.2

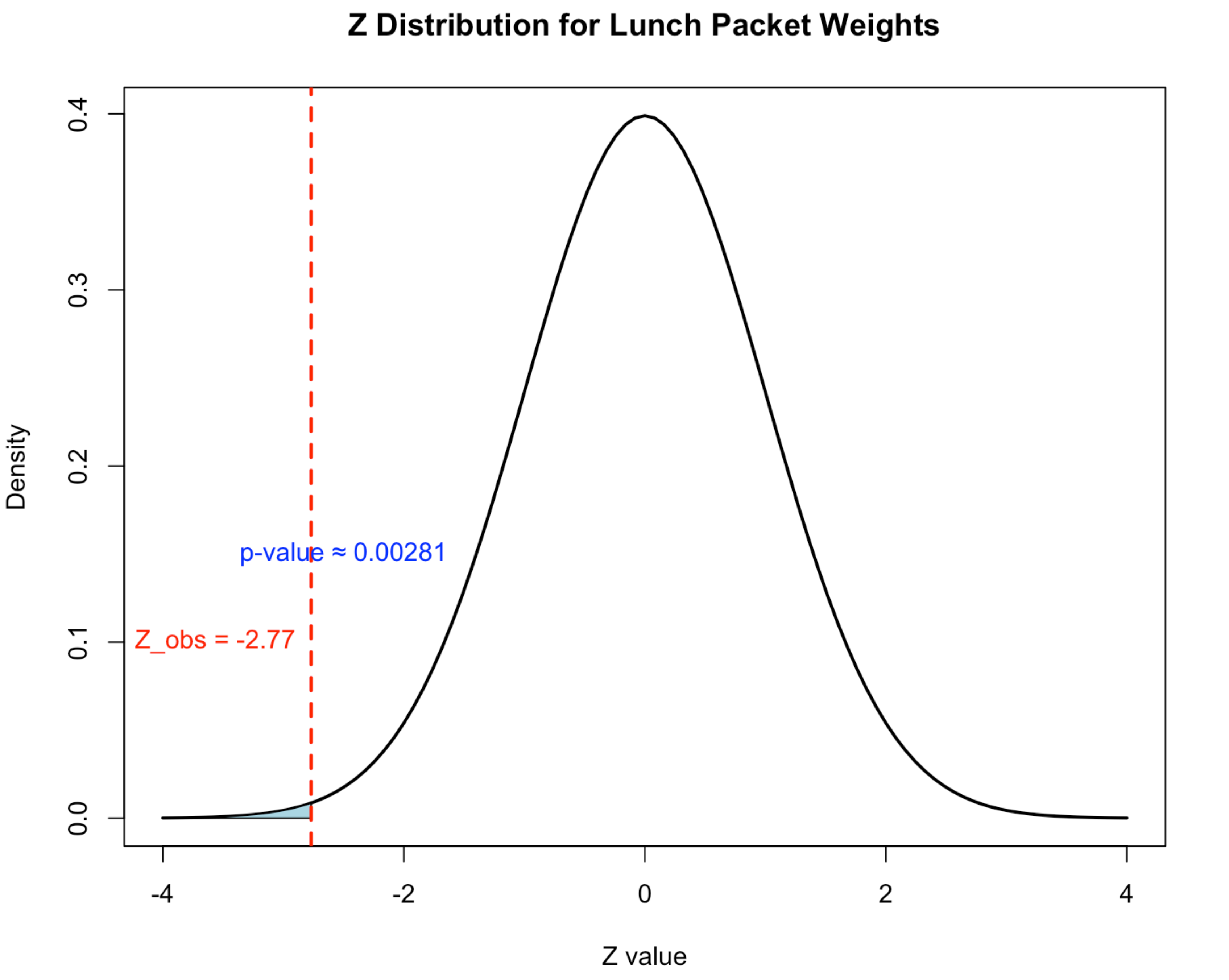
The Z-Stat: Observation

- Idea: compare the **observed sample mean** (or proportion) to the **hypothesized population mean**
- $$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$
- **Standard Error (SE)** = typical variation of a sample statistic due to random sampling.
  - For a mean: $\mathrm{SE} = \dfrac{\sigma}{\sqrt{n}}$
  - For a proportion: $\mathrm{SE} = \sqrt{\dfrac{p(1-p)}{n}}$
- The **bigger** $|Z|$ is, the **more evidence** we have **against** H_0 !
- The Z distribution is a Standard Normal Distribution
  - mean = 0 and standard deviation = 1

Determine Z

- $$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{298.2 - 300}{3.9/\sqrt{36}} = \frac{-1.8}{0.65} \approx -2.77$$


Z Distribution for Lunch Packet Weights
Density
0.0
0.1
0.2
0.3
0.4
Z_obs = -2.77
-4
-2
0
2
4
Z value


Z Distribution for Lunch Packet Weights
Density
0.0
0.1
0.2
0.3
0.4
p-value ≈ 0.00281
Z_obs = -2.77
-4
-2
0
2
4
Z value

Decision

- We have obtained a p-value of 0.00281
- Default α = 0.05
  - The p-value of 0.00281 < α
- We can reject the null hypothesis that says that the lunch packets are not underweight
- We can conclude, with a significance level of 5%, that these lunch packets are indeed underweight

Now thus far

- We had sufficiently large samples
- We knew (could estimate) the population variance

What if …

- We had only very few samples ?
- And thus could not estimate the population variance ?

The **Student's** **t-distribution**

- Used when the **population standard deviation** σ is **unknown**
- Replace σ with the **sample standard deviation** s
- The test statistic becomes:
  - $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
- s is the **sample** standard deviation

Degrees of Freedom

- Distribution depends on degrees of freedom (df = n − 1).
- Compared to Z, the t-distribution is:
  - Also
    - Centered at 0
    - Symmetric

Degrees of freedom (df)

- Idea: number of independent values that can vary after applying constraints.
- Always linked to the denominator of the test statistic (variance estimate).
- One sample t-test
  - $\mathrm{df} = n - 1$
- Two sample t-test assuming **equal** variances
  - $\mathrm{df} = n_1 + n_2 - 2$
- Two sample t-test assuming **unequal** variances (Welch's formula)
  - $$\mathrm{df} = \frac{\frac{s^2}{n_1} + \frac{s^2}{n_2}^{2}}{\frac{\frac{s^2}{n_1}^{2}}{n_1 - 1} + \frac{\frac{s^2}{n_2}^{2}}{n_2 - 1}}$$

Student's t-test

- William Sealy Gosset (1876–1937) : statistician & chemist at Guinness in Dublin.
- Problem: quality control with **small sample sizes** (beer ingredients, yields, taste panels).
- Standard Normal (Z) methods required large n, but Gosset often had $n < 10$.
- Developed the t-distribution to handle extra uncertainty from estimating σ with small samples.
- Published in Biometrika (1908) under the pseudonym "Student"
  - Company policy forbade staff publishing
  - Worked later with Pearson
- The "Student's t-distribution" is central to inference with small samples

t-distribution: Example

*A winery uses **reverse osmosis** to dealcoholize its wine and must confirm that **polyphenols remain at or above 450 mg/L**. The quality team samples **8** batches and finds a **mean polyphenol concentration** of **456.5** mg/L with a **sample standard deviation** of **8** mg/L.*

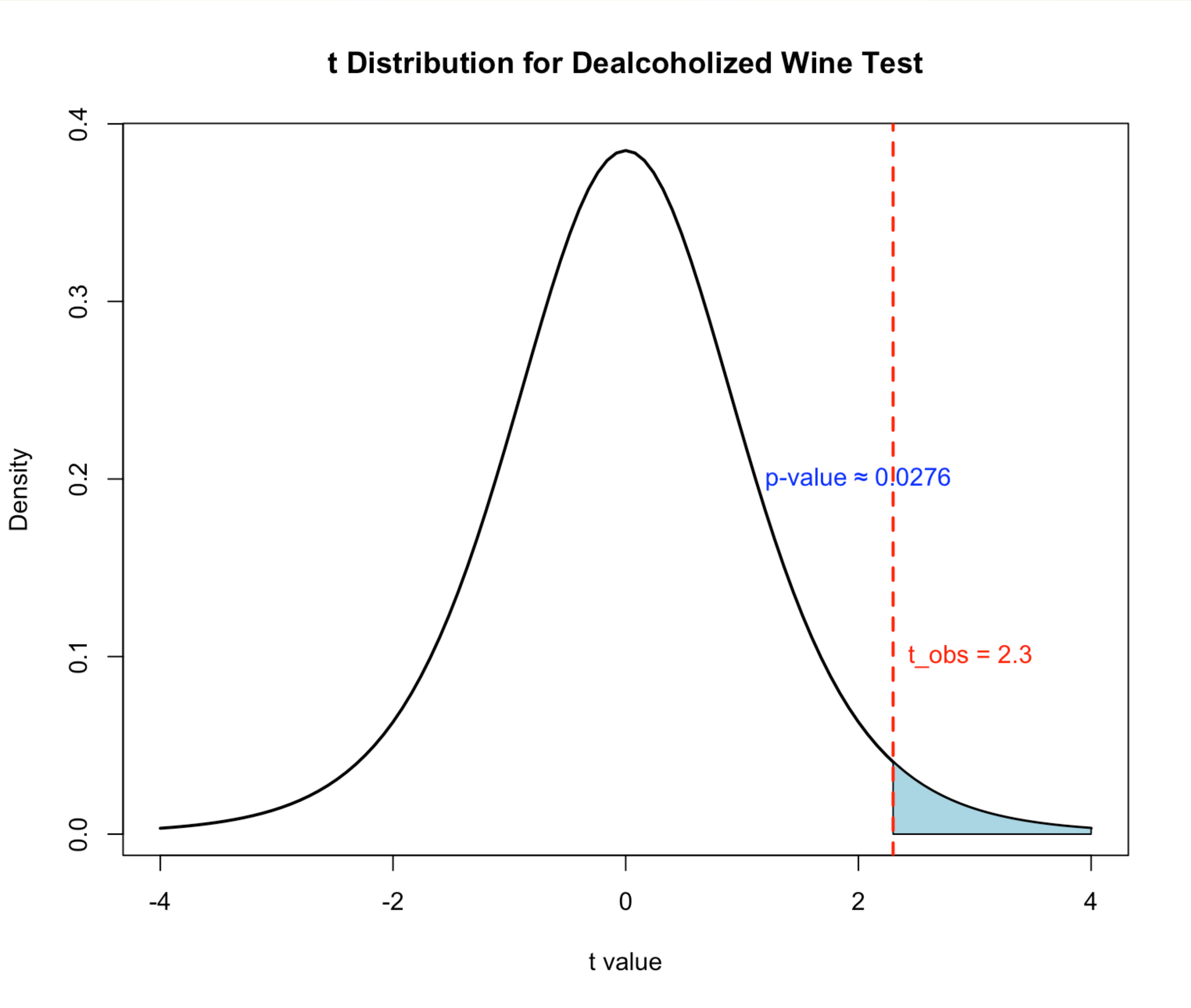
- Question: Is the dealcoholization process (reverse osmosis) not detrimental to the polyphenol concentrations we want to maintain ?

t-distribution: Example

A winery uses ***reverse osmosis*** *to dealcoholize its wine and must confirm that* ***polyphenols remain at or above 450 mg/L****. The quality team samples* ***8*** *batches and finds a* ***mean polyphenol concentration*** *of* ***456.5*** *mg/L with a* ***sample standard deviation*** *of* ***8*** *mg/L.*

- Question: Is the dealcoholization process (reverse osmosis) not detrimental to the polyphenol concentrations we want to maintain ?
  - μ = Population mean
  - σ = Sample standard deviation
  - Null Hypothesis: H_0 : μ < 450
  - Alternate Hypothesis: H_1 : μ > = 450

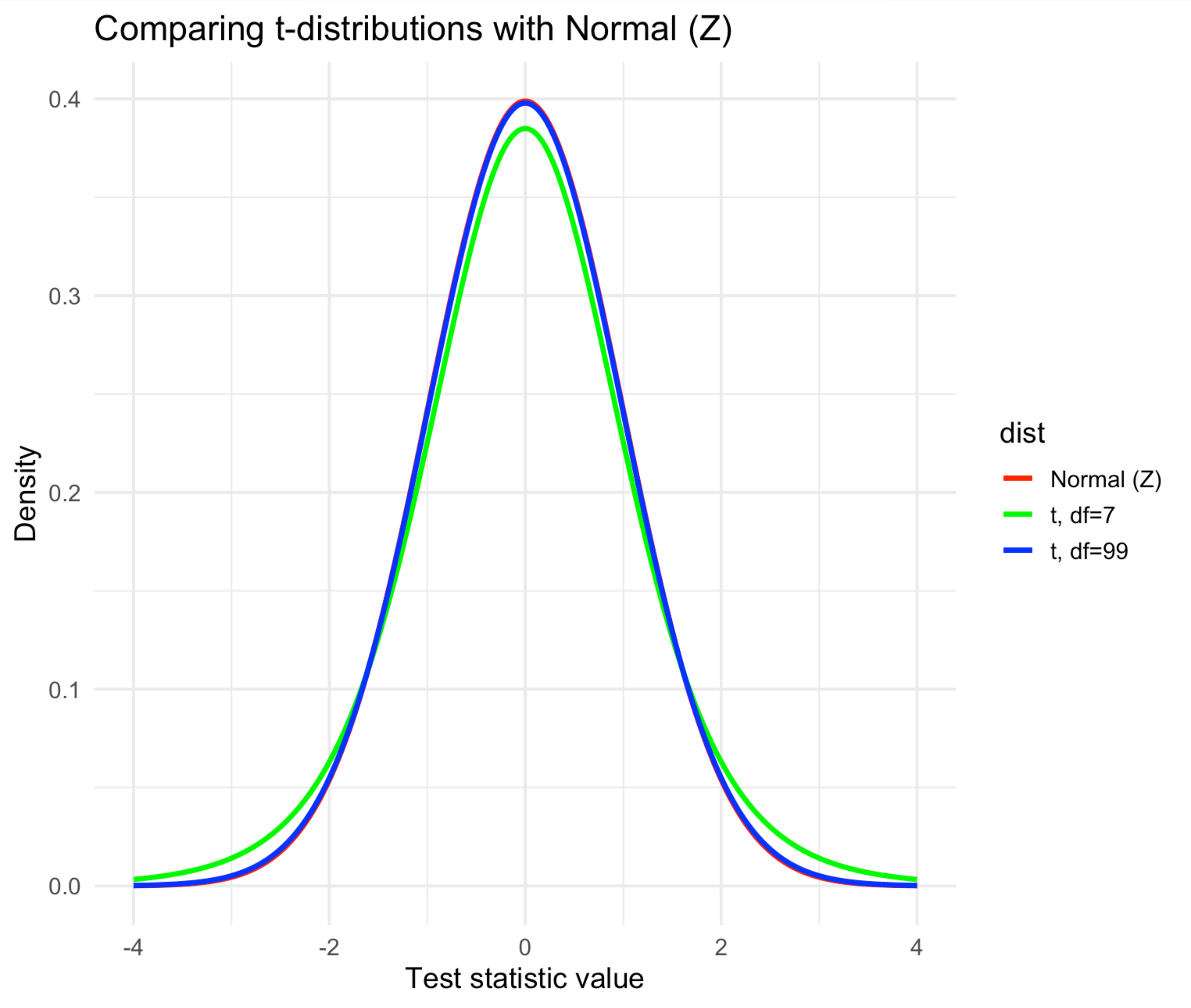
- $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{456.5 - 450}{8/\sqrt{8}} = \dfrac{6.5}{2.83} \approx 2.30$



t Distribution for Dealcoholized Wine Test
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
t value
p-value ≈ 0.0276
t_obs = 2.3

Decision

- We have obtained a p-value of 0.0276
- Default α = 0.05
  - The p-value of 0.0276 < α
- We can reject the null hypothesis that says that the dealcoholization process (of reverse osmosis) IS detrimental to maintaining desired polyphenol levels in the wine
- We can conclude, with a significance level of 5%, that our dealcoholozation process is NOT detrimental to maintaing desired polyphenol levels in the wine



Comparing t-distributions with Normal (Z)
Density
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
Test statistic value
dist
Normal (Z)
t, df=7
t, df=99