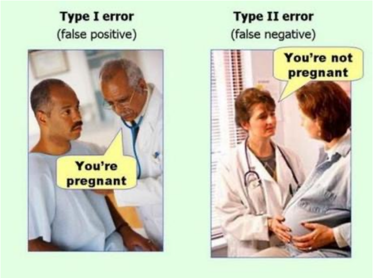
Hypothesis Testing with 2 Samples

부산대학교 정보·의생명공학대학
정보컴퓨터공학부

Hypothesis Testing revisited



❖ Hypothesis Testing

- Null Hypothesis (귀무가설) : H_0
- Alternative Hypothesis (대립가설) : H_1
- p-value and Significance Level : α
  - p-value is the probability that, if the null hypothesis is true, the results from another randomly selected sample will be as extreme or more extreme as the results obtained from the given sample.
- Decision : To reject or not reject H_0
- If $\alpha > p$-value, reject H_0.
  - The results of the sample data are significant. There is sufficient evidence to conclude that H_0 is an incorrect belief and that the alternative hypothesis, H_1 may be correct.
- If $\alpha \leq p$-value, do not reject H_0.
  - The results of the sample data are not significant. There is not sufficient evidence to conclude that the alternative hypothesis, H_1, may be correct.
  - When you "do not reject H_0", it does not mean that you should believe that H_0 is true. It simply means that the sample data have failed to provide sufficient evidence to cast serious doubt about the truthfulness of H_0.
- Conclusion: After you make your decision, write a thoughtful conclusion about the hypotheses in terms of the given problem

Table of error types		Null hypothesis (H_0) is	
		True	False
Decision about null hypothesis (H_0)	Don't reject	Correct inference (true negative) (probability = $1 - \alpha$)	**Type II error (false negative)**
	Reject	**Type I error (false positive)** (probability = α)	Correct inference (true positive)

❖ Example

- It's a Boy Genetics Labs claim their procedures improve the chances of a boy being born. The results for a test of a single population proportion are as follows:
- $H_0 : p = 0.5,\ H_1 : p > 0.5$
- $\alpha = 0.01, p$-Value = 0.025
- Interpret the results and state a conclusion in simple, non-technical terms

Hypothesis Testing with 1 sample

- Test of a single population mean μ
  - 단일 평균의 검정

- Distribution Needed for Hypothesis Testing
  - a Student's t-distribution (often called a t-test),
    - Simple random sample from a population that is approximately normally distributed.
    - Use the sample standard deviation to approximate the population standard deviation.
    - Note that if the sample size is sufficiently large, a t-test will work even if the population is not approximately normally distributed.
  - a Normal distribution (often called a z-test),
    - Simple random sample from the population.
    - The population you are testing is normally distributed or your sample size is sufficiently large.
    - You know the value of the population standard deviation which, in reality, is rarely known

- t-score/ t-statistic

sample.mean()

$$t = \frac{\bar{X} - \mu_0}{\sqrt{\frac{s^2}{n}}} \sim t(\nu)$$

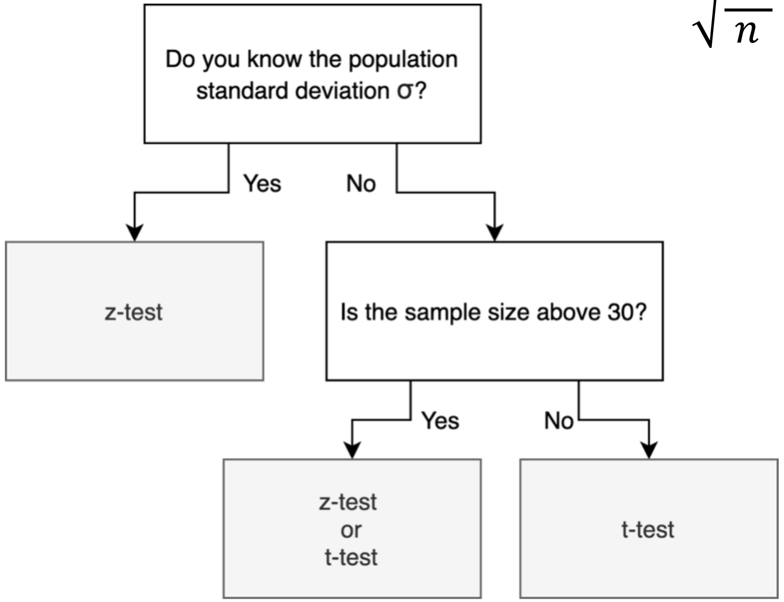
ν : degree of freedom = $n-1$

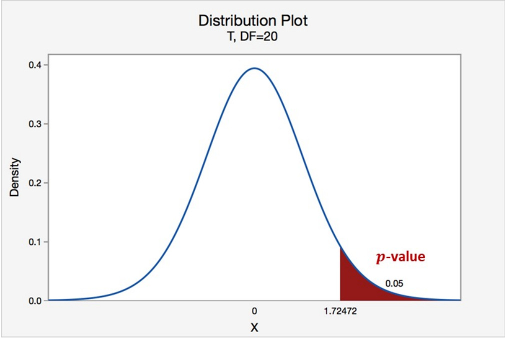
sample.var()

sample.count()

- z-score/z-statistic/Standard score

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$





One sample t-test example

❖ Energy Bar Test

- Imagine we have collected a random sample of 31 energy bars from a number of different stores to represent the population of energy bars available to the general consumer. The labels on the bars claim that each bar contains 20 grams of protein.
- You can find the actual protein contained in the collected energy bars in the variable -"ebp_df" in the code below (ebp stands for energy_bar_protein). It shows that some bars have less than 20 grams of protein. Other bars have more.
- You might think that the data support the idea that the labels are correct. Others might disagree.
- We decide to perform one-sample t-test to test the following hypotheses
- $H_0: \mu = 20$ vs. $H_1: \mu \neq 20$

```
ebp_df = pd.DataFrame(
data=[20.7, 27.46, 22.15, 19.85, 21.29, 24.75, 20.75,
      22.91, 25.34, 20.33, 21.54, 21.08, 22.14, 19.56,
      21.1, 18.04, 24.12, 19.95, 19.72, 18.28, 16.26,
      17.46, 20.53, 22.12, 25.06, 22.44, 19.08, 19.88,
      21.39, 22.33, 25.79], columns=['protein'])
```

❖ Questions

- What is the degrees of freedom of the sample ?

  # of samples – 1 = 30

- Calculate the t-statistic and the p-value of the sample
- For 2 significance levels $\alpha_1 = 0.05, \alpha_2 = 0.01$ for the test, what should you conclude regarding the hypothesis?

```
df = sample.count() - 1
t_stat = (sample.mean()-
mu)/math.sqrt(sample.var()/sample.count())
print(f'\n t-statistic = {t_stat}')
rv_t = stat.t(df)
print(f'\n p-value (!= {mu})= {(1-rv_t.cdf(t_stat))*2}')

# using ttest_1samp()
result1 = stat.ttest_1samp(ebp_df, mu)
print(f'\n {result1}')
```

```
t-statistic = 3.066831635284074
p-value (!= 20)= 0.0045526210606354756
Ttest_1sampResult(statistic=array([3.06683164]),
pvalue=array([0.00455262]))
```

부산대학교 PUSAN NATIONAL UNIVERSITY

Hypothesis Testing with 2 samples

❖ A/B Test

- An A/B test is an experiment with two groups to establish which of two treatments, products, procedures, or the like is superior.
- Often one of the two treatments is the standard existing treatment, or no treatment. If a standard (or no) treatment is used, it is called the control. A typical hypothesis is that a new treatment is better than the control.

❖ Key terms for A/B Testing

- Treatment : Something (drug, price, web headline) to which a subject is exposed.
- Treatment group (실험군, 처리군): A group of subjects exposed to a specific treatment.
- Control group (대조군): A group of subjects exposed to no (or standard) treatment.
- Randomization : The process of randomly assigning subjects to treatments.
- Subjects : The items (web visitors, patients, etc.) that are exposed to treatments.
- Test statistic (검정 통계량): The metric used to measure the effect of the treatment.

❖ Some Examples

- Testing two web headlines to determine which produces more clicks
- Testing two web ads to determine which generates more conversions
- Testing two therapies to determine which suppresses cancer more effectively
- Testing two prices to determine which yields more net profit

❖ Data Types & related tests

- Numerical : t-test
- Categorical : χ^2 test

t-Test for 2 population means

❖ The comparison of two population means

- With Unknown standard deviations → t-test
- With Known standard deviations → z-test

❖ Assumptions

- The **two independent samples** are
  simple random samples from two distinct populations.
- For the two distinct populations:
  - if the sample sizes are small, the distributions are important (should be normal)
  - if the sample sizes are large, the distributions are not important (need not be normal)

❖ Independent two-sample t-test

- Equal sample sizes and variances
- Equal or unequal sample sizes, unequal variances →
  Welch's t-test

❖ Python/Colab – scipy.stats.ttest_ind()

- ttest_ind(a,b, **equal_var=True**, alternative='two-sided')
- a,b : Samples
- equal_var = True
  - Equal sample sizes and variances
- eqaul_var = False
  - Unequal variances, Welch's t-test
- alternative = 'two-sided', 'less', 'greater'

Independent two-sample t-test

❖ Equal sample sizes and variances

- t-statistic

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{n}}}$$

  - $\bar{X}_1, \bar{X}_2$: Sample Mean
  - $s_{X_1}^2, s_{X_2}^2$: **Unbiased Estimator** of the variances
    - Not the exactly same to the sample variance
    - $\hat{s}^2 = \begin{cases} \frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \mu_X)^2 & \text{If } \mu_X \text{ is known} \\ \frac{1}{n-1} \cdot \sum_{i=1}^{n}(x_i - \hat{\mu})^2 & \text{If } \mu_X \text{ is unknown} \end{cases}$
    - Sample.var() $\times n/(n-1)$
  - n : Number of samples of each group ($n = n_1 = n_2$)
- Degree of freedom for student's t-distribution = $2n - 2$

❖ Welch's t-test – Unequal Variance

- t-statistic

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_{X_1}^2}{n_1} + \frac{s_{X_2}^2}{n_2}}}$$

- Degree of freedom for student's t-distribution

$$= \frac{\left(\frac{s_{X_1}^2}{n_1} + \frac{s_{X_2}^2}{n_2}\right)^2}{\frac{\left(\frac{s_{X_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_{X_2}^2}{n_2}\right)^2}{n_2 - 1}}$$

Independent two-sample t-test Example

❖ Online vs Face-to-face Class

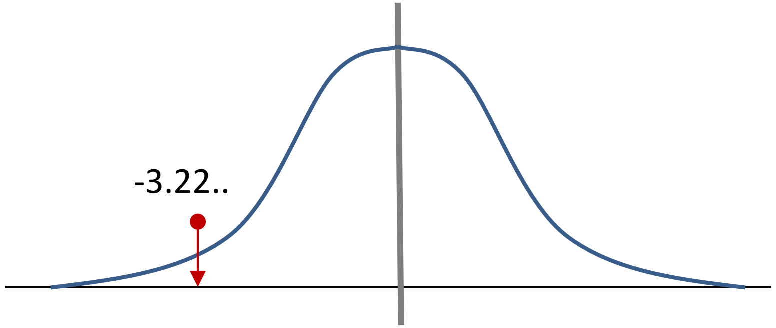
- A professor at a large community college wanted to determine whether there is a difference in the means of final exam scores between students who took his statistics course online and the students who took his face-to-face statistics class.
- He believed that the mean of the final exam scores for 2 class would be same. Was the professor correct? The randomly selected 30 final exam scores from each group are listed in the following tables

X_1 - Online Class									
67.6	41.2	85.3	55.9	82.4	91.2	73.5	94.1	64.7	64.7
70.6	38.2	61.8	88.2	70.6	58.8	91.2	73.5	82.4	35.5
94.1	88.2	64.7	55.9	88.2	97.1	85.3	61.8	79.4	79.4

X_2 - Face-to-face Class									
77.9	95.3	81.2	74.1	98.8	88.2	85.9	92.9	87.1	88.2
69.4	57.6	69.4	67.1	97.6	85.9	88.2	91.8	78.8	71.8
98.8	61.2	92.9	90.6	97.6	100	95.3	83.5	92.9	89.4

❖ Statistics

- $n_1 = n_2 = 30$
- $\bar{X}_1 = 72.98, \bar{X}_2 = 84.98$
- $s_{X_1}^2 = 286.24.., s_{X_2}^2 = 137.22..$
  - $Var(X_1) = 276.70 \ldots,\ Var(X_2) = 132.64 \ldots$
  - $s_{X_1}^2 = Var(X_1) \cdot \frac{n}{n-1}$
- t-statistic = -3.2285832980227287
  - The same for both equal or unequal variance as $n_1 = n_2$
- Df = 58 in equal variance test
  - $\Pr(X < t$-stat$)$= 0.0010246…
  - p-Value for 2-sided test = 2×0.0010246=0.002…



Independent two-sample t-test Example

❖ $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2, \alpha = 0.05$

- Are the population standard deviations known or unknown?
  - Yes/No
- Which distribution do you use to perform the test?
  - Student's t distribution
- Is this test greater, less, or two sided?
  - Two sided
- What is the p-value ?
  - 0.002049… for equal variance
  - 0.002164… for non-equal variance
- Do you reject or not reject the null hypothesis?
  - Reject H_0

```
import numpy as np
import scipy.stats as stats

val_online = np.array([67.6,41.2,85.3,55.9,…
val_f2f = np.array([77.9, 95.3, 81.2,…

print(stats.ttest_ind(val_online,val_f2f))
print(stats.ttest_ind(val_online,val_f2f, equal_var=False))

---------------------------------------

Ttest_indResult(statistic=-3.2285832980227287,
pvalue=0.0020492135340385376)
Ttest_indResult(statistic=-3.228583298022783,
pvalue=0.0021648992238072)
```

❖ $H_0: \mu_1 = \mu_2, H_1: \mu_1 < \mu_2, \alpha = 0.05$

- Is this test greater, less, or two sided?
- What is the p-value ?
- Do you reject or not reject the null hypothesis?