Recap

06 April 2023 19:43

Session 2 → $\boxed{HT}$ → Why

→ What

→ Null and alternate

→ Z-test

→ 1 tailed vs 2 tailed

→ Type 1 vs type 2 error

significance level (α)

Session 2

P-values

t-test

P-value

06 April 2023 06:48

of 53

→ 53 head

0.07

$P(H>53)$ $\boxed{p, n}$

p_value

experiment

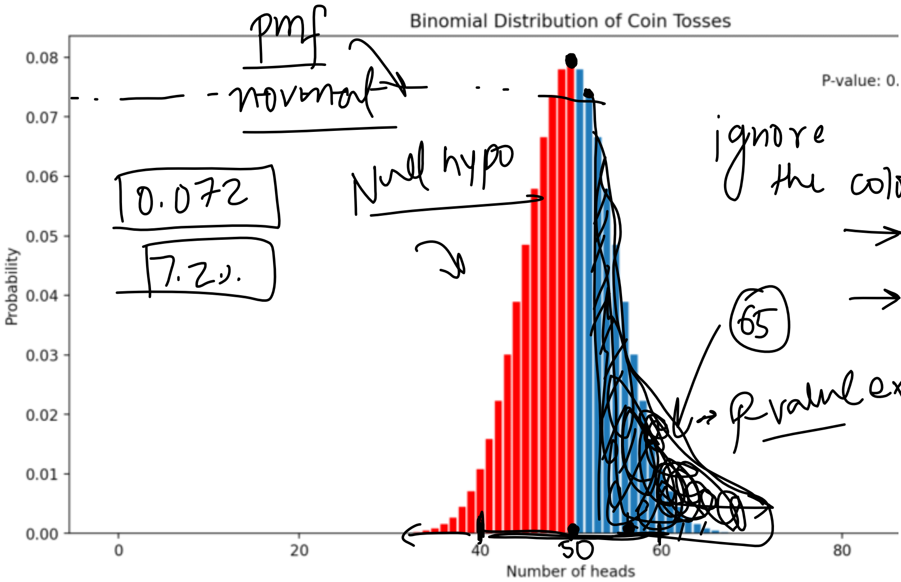
↓

1 coin → $\boxed{\text{100 times toss}}$

binomial distribution

#heads

P-value is the $\boxed{\text{probability}}$ of getting a sample as or more $\boxed{\text{extreme}}$(having more evidence against H0) than our own sample given the Null Hypothesis(H0) is true.



ignore the colors

→ H_0: $P(H) = P(T)$

→ H_a: $P(H) > P(T)$

65

p-value exp → 100 times → 53 heads

exp → $\boxed{\text{100 times}}$

↑

30 times

$\boxed{P = 0.3}$

53 H

2 times

In simple words p-value is a measure of the strength of the evidence against the Null Hypothesis that is provided by our sample data.

$\boxed{\text{Null hyp}}$

100 exp → 80 times → $\boxed{0}$

Interpreting p-value

06 April 2023 08:25

$\rightarrow$ $\boxed{\text{reject your } H_0}$

With significance value

$\alpha = 0.05 / 0.01 \rightarrow$ p-value $\leq \alpha$

$\boxed{0.05}$

Without significance value $\boxed{\alpha} \rightarrow$

1. Very small p-values (e.g., $p < 0.01$) indicate strong evidence against the null hypothesis, suggesting that the observed effect or difference is unlikely to have occurred by chance alone.
2. Small p-values (e.g., $0.01 \leq p < 0.05$) indicate moderate evidence against the null hypothesis, suggesting that the observed effect or difference is less likely to have occurred by chance alone.
3. Large p-values (e.g., $0.05 \leq p < 0.1$) indicate weak evidence against the null hypothesis, suggesting that the observed effect or difference might have occurred by chance alone, but there is still some level of uncertainty.
4. Very large p-values (e.g., $p \geq 0.1$) indicate weak or no evidence against the null hypothesis, suggesting that the observed effect or difference is likely to have occurred by chance alone.

α

P-value in context of Z-test

06 April 2023 07:08

{rejection region approach} p-value approach

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day and the pop std is 4. The company wants to know if the new training program has significantly increased productivity.

$\mu = 50 \quad n = 30 \quad \bar{x} = 53$

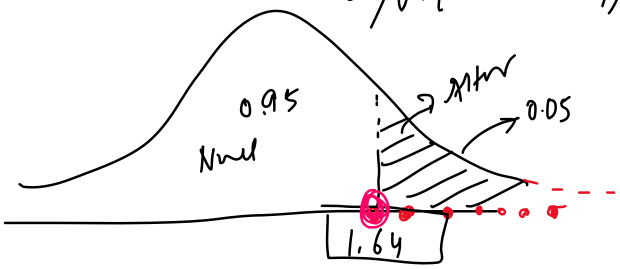
$\sigma = 4 \quad \alpha = 0.05$

$H_0: \mu = 50$

$H_a: \mu > 50$

$$\boxed{Z\text{-stat}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{53-50}{4/\sqrt{30}} = \frac{3 \times \sqrt{30}}{4} = \boxed{4.10} / \boxed{1.7}$$

reject



$\boxed{Z = 4.10}$

$1 - \boxed{0.95} \rightarrow \boxed{0.05}$

p-value

p-value → critical point

$\boxed{0.999} = \boxed{0.0001}$

p-value < 0.05

reject Null hypo

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a pop standard deviation of 5 grams.

$\mu = 50 \quad n = 40 \quad \bar{x} = 49$

$\sigma = 5 \quad \alpha = 0.05$

$H_0: \boxed{\mu = 50}$

$H_a: \mu \neq 50$

p-value

2 tailed

$\boxed{0.206} > 0.05$

$$Z = \frac{49 - 50}{5/\sqrt{40}} = \frac{-\sqrt{40}}{5} = \boxed{-1.26}$$

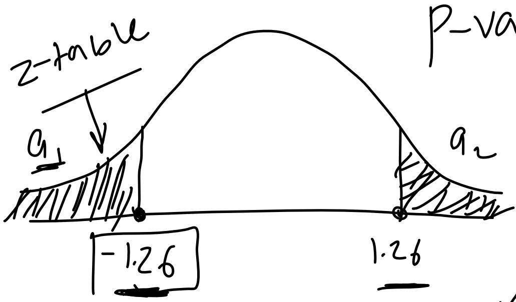
$\boxed{\text{p-val}} \; \boxed{0.206}$

$\text{p-value} = a_1 + a_2$

$5/\sqrt{40}$ 5

$p\text{-value} = a_1 + a_2$ $\boxed{p\text{-val}}\ \boxed{0.206}$

$a_1 = 0.103 \times 2$

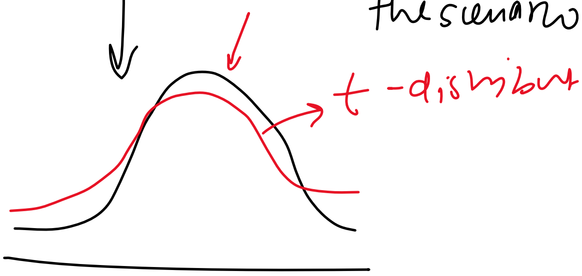


$\boxed{\text{Z test}}$ → $\boxed{\sigma}$ ✓

$\boxed{\text{t-test}}$ → $\boxed{\sigma}$ ✗ S ✓

smaller sample sizes

most of the scenario



for rejection region approach
P- value approach

T-tests

06 April 2023 14:14

A t-test is a statistical test used in hypothesis testing to compare the means of two samples or to compare a sample mean to a known population mean. The t-test is based on the t-distribution, which is used when the population standard deviation is unknown and the sample size is small.

There are three main types of t-tests:

1 sample $\rightarrow \bar{X} \rightarrow \mu$

With the help of sample mean we try perform hypothesis testing on population mean

σ $\bar{X}$

One-sample t-test: The one-sample t-test is used to compare the mean of a single sample to a known population mean. The null hypothesis states that there is no significant difference between the sample mean and the population mean, while the alternative hypothesis states that there is a significant difference.

1 class $\rightarrow$ test A $\rightarrow$ popl

1 class $\rightarrow$ test B $\rightarrow$ pop

Independent two-sample t-test: The independent two-sample t-test is used to compare the means of two independent samples. The null hypothesis states that there is no significant difference between the means of the two samples, while the alternative hypothesis states that there is a significant difference.

Paired t-test (dependent two-sample t-test): The paired t-test is used to compare the means of two samples that are dependent or paired, such as pre-test and post-test scores for the same group of subjects or measurements taken on the same subjects under two different conditions. The null hypothesis states that there is no significant difference between the means of the paired differences, while the alternative hypothesis states that there is a significant difference.

Single Sample t-test

06 April 2023 14:14

t-test

σ ✗ s ✓

days → 50gm

40 days

A one-sample t-test checks whether a sample mean differs from the population mean.

Assumptions for a single sample t-test

→ sample normally distri

1. Normality - Population from which the sample is drawn is normally distributed
2. Independence - The observations in the sample must be independent, which means that the value of one observation should not influence the value of another observation.
3. Random Sampling - The sample must be a random and representative subset of the population.
4. Unknown population std - The population std is not known.

Suppose a manufacturer claims that the average weight of their new chocolate bars is 50 grams, we highly doubt that and want to check this so we drew out a sample of 25 chocolate bars and measured their weight, the sample mean came out to be 49.7 grams and the sample std deviation was 1.2 grams. Consider the significance level to be 0.05

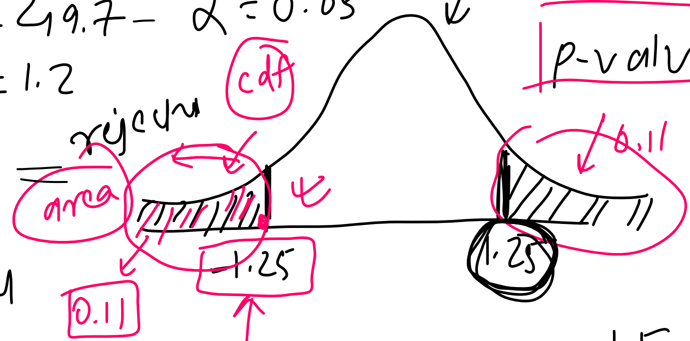
$\mu = 50$ $n = 25$ $\bar{x} = 49.7$ $\alpha = 0.05$

$s = 1.2$

$H_0: \mu = 50$

$H_a: \mu \neq 50$

assuming it is normal



$p\text{-value} = 0.22$

> 0.05

Null reject ✗

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{49.7 - 50}{1.2/\sqrt{25}} = \frac{-0.3 \times 5}{1.2} = \frac{-1.5}{1.2} = -1.25$$

$df = n - 1 = 24$

$H_0: \boxed{\mu = 35}$ →

$H_a: \mu < 35$

25 sample age → normal

t-test

Shapiro-Wilk test

$\boxed{\mu = 35}$ σ ✗

$\bar{x}$, s, $\alpha = 0.05$

$p\text{-value} < 0.05$ not normal

$p\text{-value} > 0.05$ normal

Python Case Study 1

06 April 2023 17:27

Independent 2 sample t-test

06 April 2023 14:15

titanic

males avg age (25) > females avg age (25)

An independent two-sample t-test, also known as an unpaired t-test, is a statistical method used to compare the means of two independent groups to determine if there is a significant difference between them.

Assumptions for the test:

1. Independence of observations: The two samples must be independent, meaning there is no relationship between the observations in one group and the observations in the other group. The subjects in the two groups should be selected randomly and independently.
2. Normality: The data in each of the two groups should be approximately normally distributed. The t-test is considered robust to mild violations of normality, especially when the sample sizes are large (typically n ≥ 30) and the sample sizes of the two groups are similar. If the data is highly skewed or has substantial outliers, consider using a non-parametric test, such as the Mann-Whitney U test.
3. Equal variances (Homoscedasticity): The variances of the two populations should be approximately equal. This assumption can be checked using F-test for equality of variances. If this assumption is not met, you can use Welch's t-test, which does not require equal variances.
4. Random sampling: The data should be collected using a random sampling method from the respective populations. This ensures that the sample is representative of the population and reduces the risk of selection bias.

QQ, hist → shapiro-wilk

A B

$$\sigma_A^2 = \sigma_B^2$$

levene

p-value < 0.05

$$\sigma_A^2 \neq \sigma_B^2$$

p value > 0.05

$$\sigma_A^2 = \sigma_B^2$$

reject my H_0

Suppose a website owner claims that there is no difference in the average time spent on their website between desktop and mobile users. To test this claim, we collect data from 30 desktop users and 30 mobile users regarding the time spent on the website in minutes. The sample statistics are as follows:

→ avg time

desktop users = [12, 15, 18, 16, 20, 17, 14, 22, 19, 21, 23, 18, 25, 17, 16, 24, 20, 19, 22, 18, 15, 14, 23, 16, 12, 21, 19, 17, 20, 14]

mobile_users = [10, 12, 14, 13, 16, 15, 11, 17, 14, 16, 18, 14, 20, 15, 14, 19, 16, 15, 17, 14, 12, 11, 18, 15, 10, 16, 15, 13, 16, 11]

avg time

$$H_0 = \mu_d = \mu_m$$

$$H_a = \mu_d \neq \mu_m$$

Desktop users:

- Sample size (n1): 30
- Sample mean (mean1): 18.5 minutes
- Sample standard deviation (std_dev1): 3.5 minutes

Mobile users:

- Sample size (n2): 30 → 25
- Sample mean (mean2): 14.3 minutes
- Sample standard deviation (std_dev2): 2.7 minutes

check assumptions

We will use a significance level (α) of 0.05 for the hypothesis test.

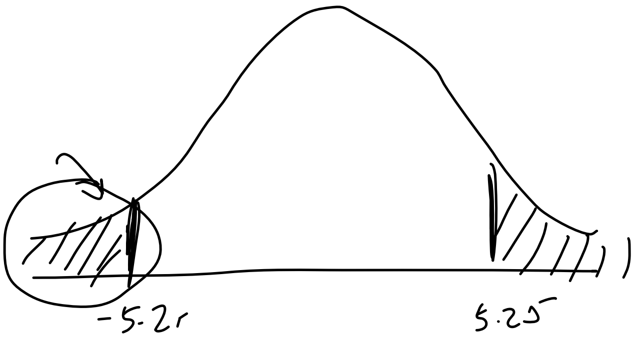
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t-statistic

$$t = \frac{18.5 - 14.3}{\quad} = \frac{4.2}{\quad} = 5.25$$

$$t = \frac{[illegible]}{\sqrt{\frac{(3.5)^2}{30} + \frac{(2.7)^2}{30}}} = \frac{4.2}{\sqrt{\frac{19.54}{30}}} = \boxed{5.25}$$



$\boxed{n-1}$ t

$n_1 + n_2 - 2$

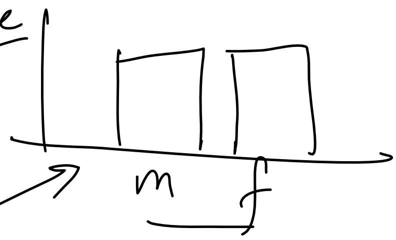
$30+30-2 = 58$

$\boxed{\mu_m > \mu_f}$ → ~~avg~~ $\boxed{\mu_m = \mu_f}$

age

age and gender

pop



$H_0: \mu_m = \mu_f$

$H_1: \mu_m > \mu_f$

$\alpha = 0.05$

Python Case Study 2

06 April 2023 17:27

Paired 2 sample t-test

06 April 2023 14:21

A paired two-sample t-test, also known as a dependent or paired-samples t-test, is a statistical test used to compare the means of two related or dependent groups.

Common scenarios where a paired two-sample t-test is used include:

1. Before-and-after studies: Comparing the performance of a group before and after an intervention or treatment.
2. Matched or correlated groups: Comparing the performance of two groups that are matched or correlated in some way, such as siblings or pairs of individuals with similar characteristics.

	I	II	d
A	50	55	-5
B	60	60	0
C	76	66	10
D	40	60	-20
E	25	100	-75

normal

Assumptions

1. Paired observations: The two sets of observations must be related or paired in some way, such as before-and-after measurements on the same subjects or observations from matched or correlated groups.
2. Normality: The differences between the paired observations should be approximately normally distributed. This assumption can be checked using graphical methods (e.g., histograms, Q-Q plots) or statistical tests for normality (e.g., Shapiro-Wilk test). Note that the t-test is generally robust to moderate violations of this assumption when the sample size is large.
3. Independence of pairs: Each pair of observations should be independent of other pairs. In other words, the outcome of one pair should not affect the outcome of another pair. This assumption is generally satisfied by appropriate study design and random sampling.

Let's assume that a fitness center is evaluating the effectiveness of a new 8-week weight loss program. They enroll 15 participants in the program and measure their weights before and after the program. The goal is to test whether the new weight loss program leads to a significant reduction in the participants' weight.

$\mu_{dif} = \mu_{bef} - \mu_{af} = 0$

$H_0: \boxed{\mu_{before} = \mu_{after}} =$

$H_1: \mu_{before} > \mu_{after}$

> 0.05

Before the program:
```
[80, 92, 75, 68, 85, 78, 73, 90, 70, 88, 76, 84, 82, 77, 91]
```
After the program:
```
[78, 93, 81, 67, 88, 76, 74, 91, 69, 88, 77, 81, 80, 79, 88]
```

Significance level (α) = 0.05

name	wt before	wt after	diff
A	80	78	x_1
B	92	93	x_2
C	75	.	x_3
⋮	⋮	⋮	⋮
K	91	88	x_{15}

diff

normal dist

$\bar{X}_{diff}$ S_{diff}

$t = \bar{X}_{diff}$ 15

$t = \frac{\bar{x}_{diff}}{s_{diff}/\sqrt{n}}$ 1.5

0.1

$\boxed{0.54 > 0.05} \rightarrow \boxed{X}$

ANOVA → Analysis of variance

One Way ANOVA → 1 variable

Two Way ANOVA → 2 Variable

Use to compare more than one sample means.

↓ uses

F - distribution

Doing 2 things → Variation in sample (within)
→ Variation in b/w the sample

Null hypothesis: samples in all groups are drawn from the same populations

: $\mu1 = \mu2 = \mu3$

Alternate hypothesis: : $\mu1 \neq \mu2 \neq \mu3$

example of ANOVA with the F distribution:

Suppose you want to compare the average weight of three different types of apples (A, B, and C). You randomly select 10 apples from each type, weigh them, and record the weights. You want to determine if there is a significant difference in the average weight of the three types of apples.

To perform an ANOVA on this data, you would calculate the mean weight of each type of apple, as well as the overall mean weight. Then, you would calculate the sum of squares between groups (SSB) and the sum of squares within groups (SSW).

The SSB measures the variation in the sample means between the three groups, while the SSW measures the variation within each group. The ratio of SSB to SSW follows an F distribution, and we can use this distribution to calculate a p-value to test the null hypothesis that there is no significant difference between the groups.

Suppose we obtain the following results:

Mean weight of apples in group A = 150 grams
Mean weight of apples in group B = 140 grams
Mean weight of apples in group C = 160 grams
Overall mean weight = 150 grams
SSB = 200
SSW = 1200
Degrees of freedom (df) between groups = 2,= k - 1, where k is the number of groups being compared.
Degrees of freedom within groups = 27 = N - k ,where N is the total sample size and k is the number of groups being compared.

Using an F distribution table or statistical software, we can find the F statistic associated with these results. Let's say we obtain an F statistic of 3.6 with a p-value of 0.042.

Since the p-value is less than 0.05, we can reject the null hypothesis and conclude that there is a significant difference in the mean weight of the three types of apples. We cannot, however, determine which specific groups are significantly different from each other using ANOVA alone. A post-hoc test, such as Tukey's HSD or Bonferroni, would be needed for that.

To calculate the sum of squares between groups (SSB) and the sum of squares within groups (SSW), we use the following formulas:

SSB = ni(xi - x)^2 - (T)^2/N
where ni is the sample size of the ith group, xi is the mean of the ith group, x is the overall mean, T is the total number of observations, and N is the total number of groups.

SSW = (xi - xi)^2
where xi is the jth observation in the ith group, xi is the mean of the ith group, and j ranges from 1 to ni.

In these formulas, SSB represents the variation in the sample means between the groups, while SSW represents the variation within each group.