Week 8

Monday, 11 December 2023 9:39 AM

- For the sampling distributions for proportions and means, the standard deviations are based on population parameters
  - For proportions $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$
  - For means $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$

Qualitative data
one proportion week 6
Two proportions } week 7
Chi-square test }

Quantitative data
One mean
Paired Means
Two means

	Confidence interval	test statistic	
Wk 6 One proportion	proportion $\hat{p} \pm z^* \times \underbrace{\sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\text{Standard error}}$	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	NORMAL TABLE for p-value
Wk 8 One mean	Mean $\bar{y} \pm t^*_{n-1} \times \frac{s}{\sqrt{n}}$ standard error	$t = \frac{\bar{y}-\mu_0}{\frac{s}{\sqrt{n}}}$	T-tables for p-value
wk 7 Two proportions	$(\hat{p}_1 - \hat{p}_2) \pm z^* \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	$z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1}+\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}}$	NORMAL TABLE for P-value
Wk 9 Two means	$(\bar{y}_1 - \bar{y}_2) \pm t^*_{n_1+n_2-2} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$t = \frac{\bar{y}_1-\bar{y}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	T-Tables for p-value
Wk 8 Paired means	$\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}}$	$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$	T-Tables for P-value

HYPOTHESES

One proportion

One Mean

HYPOTHESES

One proportion

There are 20% people left handed

$H_0: p = 0.20$

$H_A: p \neq 0.20$

$p > 0.20$

$p < 0.20$

One Mean

The mean height of 14 year olds is 160cm

$H_0: \mu = 160$

$H_A: \mu \neq 160$

$\mu > 160$

$\mu < 160$

The t - distribution

- Student's *t*-models are unimodal, symmetric, and bell shaped, just like the Normal.

Assumptions and Conditions

Independence Assumption:

- Independence Assumption. The data values should be independent.
- Randomisation Condition: The data arise from a random sample or suitably randomised experiment.
- 10% Condition: When a sample is drawn without replacement, the sample should be no more than 10% of the population.

Normal Population Assumption:

- We can never be certain that the data are from a population that follows a Normal model, but we can check the
- Nearly Normal Condition: The data come from a distribution that is unimodal and symmetric.
  - Check this condition by making a histogram or Normal probability plot.

Critical value depends on confidence interval

<u>One mean</u> $\bar{y} \pm t_{n-1}^{*} \times \frac{s}{\sqrt{n}}$

Original Mean

$t = \frac{\bar{y}-\mu_0}{\frac{s}{\sqrt{n}}}$

$\bar{y} \rightarrow$ Sample mean

$s \rightarrow$ Sample S.D.

Sample statistics

$\bar{y}$ → Sample mean
s → Sample S.D.
n → Sample size

Sample Statistics

Question 2 : Confidence interval for one population mean

n

Consumer Reports tested 14 brands of vanilla yoghurt and found the following calories per 200g serving:

160 200 220 230 120 180 140 130 170 190 80 120 100 170

1. Create a 95% confidence interval for the average calorie content in 200g of vanilla yoghurt.
2. A diet guide claims that there are 120 calories in a 200g serving of vanilla yoghurt. What does this evidence indicate?

$\bar{x} = 157.9, \ s = 44.8$

$$\bar{x} = \frac{\sum x}{n}$$

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

1) $$\bar{y} \pm t^{*}_{n-1} \times \frac{s}{\sqrt{n}}$$

$$= 157.9 \pm 2.16 \times \frac{44.8}{\sqrt{14}}$$

$$= 157.9 \pm 25.86$$

$$= 132.04 \text{ and } 183.76$$

degrees of freedom, $df = n-1$
$= 14 - 1 = 13$

$t^{*}_{13}(95\%) = 2.16$

H_A						
≠ Two-sided test	~~0.20~~	~~0.10~~	~~0.05~~	~~0.02~~	~~0.01~~	P-value is LOW
> or < one-sided test	0.10	0.05	0.025	0.01	0.005	
df						df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	$\boxed{2.896}$	3.355	$\boxed{8}$ 4.27
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
$\boxed{13}$	1.350	1.771	$\boxed{2.160}$	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
confidence levels	80%	90%	$\boxed{95\%}$	$\boxed{98\%}$	99%	

P-value is HIGH

i.e., we are 95% confident that the average calorie content (μ) of all vanilla yoghurt is between 132.0 and 183.8 calories.

2) Diet guide is incorrect. There is more than 120 calories in 200g of Vanilla yoghurt as shown in the confidence interval.

Confidence Interval Calculator

The confidence interval calculator computes either the confidence interval of the **mean** or the confidence interval of the **standard deviation**, with calculation steps."

Confidence interval type: **Mean confidence interval**

Data is: Average, SD, n

Average ($\bar{x}$): 157.9

Sample size (n): 14

Do you know the population SD (σ)? No (use t-distribution)

Sample standard deviation (S): 44.8

Confidence Level (CL): 0.95

Rounding: 4

Mean confidence interval: [132.0333 , 183.7667].
Alternatively: **157.9 ± 25.8667**
Margin of Error (MOE): 25.8667.
Standard Error (S.E): 11.9733.

$$157.9 \pm 2.1604 * \frac{44.8}{\sqrt{14}}$$

Question 7 : Hypothesis test activity

Consumer Reports tested 14 brands of vanilla yoghurt and found the following calories per 200g serving:

n

160 200 220 230 120 180 140 130 170 190 80 120 100 170

Note: $\bar{x} = 157.9, \ s = 44.8$

μ

Does a 200g serving of vanilla yoghurt provide significantly more than 120 calories?
Perform a hypothesis test with α=0.025.

$H_0: \mu = 120$

$H_A: \mu > 120$ → one sided test

Hypotheses
Test - statistic
P - value
Compare P-value with α-level
Conclusion

$$t = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{157.9 - 120}{44.8/\sqrt{14}} = 3.1425$$

$t = 3.14$ $\quad$ $df = n - 1 = 14 - 1 = 13$

P-value is < 0.025

P-value is Low, Reject Null Hypothesis

∴ 200g of Vanilla yoghurt has significantly more than 120 calories.

One Sample T Test Calculator

Unknown standard deviation

Tails: Right (H_1: $\mu > \mu_0$)

Significance level (α): 0.025

Expected mean (μ_0): 120

Rounding: 4

Outliers: Included

Effect: Medium

Effect type: Standardized effect size

Effect Size: 0.5

Name: calorie content

Sample average ($\bar{x}$): 157.9

Sample size (n): 14

Sample SD (S): 44.8

The test statistic **T** equals **3.1654**,

$$t = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}}$$

The p-value equals **0.003724**,

Since the p-value < α, H_0 is rejected.
The population's average is considered to be greater than the expected average (120).
In other words, the sample average is greater than the expected average, and the difference is big enough to be statistically significant.

$$t(df) = \frac{\bar{x} - \mu_0}{S.E}$$

$$df = n - 1 = 14 - 1 = 13$$

$$S.E = \frac{S}{\sqrt{n}} = \frac{44.8}{\sqrt{14}} = 11.9733$$

$$t(13) = \frac{157.9 - 120}{11.9733} = 3.1654$$

PAIRED MEANS :

Question 8 : Paired data

Which of the following situations would result in paired data:

1. Comparing test results for one campus compared with another. TWO MEANS
2. Comparing individual students' Assignment 1 and 2 results. PAIRED MEANS
3. Investigating blood pressure change from a new drug. PAIRED MEANS

Mean difference

Paired means $\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}}$

$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$

Assumptions and Conditions

Paired Data Assumption:

- The data must be paired.

Independence Assumption:

- The differences must be independent of each other.

Normal Population Assumption: We need to assume that the population of *differences* follows a Normal model.

- Nearly Normal Condition: Check this with a histogram or Normal probability plot of the differences.

Question 10 : Paired data example

A group of 9 randomly selected adults were given self defence lessons. Prior to the course, they were tested to determine their self-confidence. After the course they were given the same test. A high score on the test indicates a high degree of self-confidence. The scores and their differences (after – before) are given in the table below.

$n=9$

$df = n-1$
$=9-1$
$=8$

$d = (\text{After} - \text{Before})$

Adult	Before	After	Differences, d
1	6	8	2
2	10	12	2
3	8	9	1
4	6	6	0
5	5	7	2
6	4	5	1
7	3	4	1
8	8	9	1
9	5	5	0

$\bar{d} = \frac{\Sigma d}{n} = 1.11$

$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = 0.78$

Note $\bar{x}_d = 1.11$ and $s_d = 0.78$

a) Do the results indicate that the course significantly increases the self-confidence of adults? Use α=0.01.

b) i) Calculate a 98% confidence for the true mean difference between the scores.

ii) Comment on your result in relation to your conclusion from a).

(a) $H_0: \mu_d = 0$

$H_A: \mu_d > 0$

HYPOTHESES H_0, H_A

TEST STATISTIC

P-VALUE

P-VALUE & α-level

P-VALUE
✓ P-VALUE & α-level
✓ CONCLUSION

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{1.11}{0.78/\sqrt{9}} = 4.27$$

$t = 4.27$ and $df = 8$

T-table for P-value → P-value < 0.01 ✓
P-value is LOW ✓
P-value < α ✓

P-value is LOW, Reject Null Hypothesis.
∴ Self defence lessons has increased self-confidence ✓

Paired T Test Calculator (Dependent T test)

Tails: Right (H_1: After - Before > d)

Significance level (α): 0.01

Outliers: Included

Effect: Medium

Effect type: Cohen's d (Standardized)

Effect Size: 0.5

Expected difference (d): 0

Digits: 4

Before	After
6, 10, 8, 6, 5, 4, 3, 8, 5	8, 12, 9, 6, 7, 5, 4, 9, 5

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

The test statistic **T** equals **4.264**,

The p-value equals **0.001373**,

Since the p-value < α, H_0 is rejected. ✓
The **After** population's average is considered to be greater than the **Before** population's average. ✓ OR
In other words, the sample average of After is greater than Before, and the difference is big enough to be statistically significant. ✓

(b) $t^*_8(98\%) = 2.896$

$$\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}} = 1.11 \pm 2.896 \times \frac{0.78}{\sqrt{9}}$$

$= 1.11 \pm 0.75$

$= 0.36$ AND 1.86

We are 98% Confident that true mean difference between the Scores is between 0.36 and 1.86

Means Difference Confidence Interval Calculator

Calculates the confidence interval for the difference between two population means: **equal variances, unequal variances**, and **paired groups**.

Data is: Raw data

Type: Matched samples (Paired)

Group-1: 8, 12, 9, 6, 7, 5, 4, 9, 5

Group-2: 6, 10, 8, 6, 5, 4, 3, 8, 5

Outliers: Included

Confidence Level (CL): 0.98

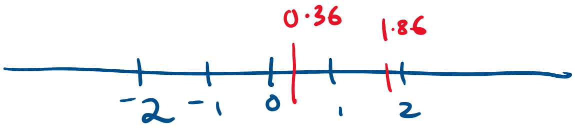
Rounding: 4

Means' difference confidence interval: [0.3564 , 1.8659].
Alternatively: **1.1111 ± 0.7548**
Margin of error (MOE): 0.7548.
Means difference standard error (S.E): 0.2606.

$$\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}}$$

CI = 1.1111 ± 2.8965 * 0.260579.
CI = 1.1111 ± 0.7548

(ii) There is no "zero" in the confidence interval, Reject Null Hypotheses.



If there is a "zero" in the Confidence interval, FAIL to Reject Null Hypotheses.