Abdelhafid Boussouf University Center – Mila

Institute of Natural and Life Sciences

Department of Environment and Ecology

Academic Year: 2025/ 2026

Module: Biostatistics

Instructor: R. Boustila

Practical Sessions: Series 2

Exercise 1: Small Sample (Student t Test)

Problem:
Measure nitrate concentration (mg/kg) in 10 soil samples:

12.5, 13.2, 11.8, 12.9, 13.0, 12.4, 11.9, 12.6, 12.7, 12.3

Tasks:

1. Compute sample mean and standard deviation.
2. Compute the standard error (SE).
3. Determine the 95% confidence interval for the mean.
4. Interpret the result.

Step 0: Enter the data

In **Column A**, enter your 10 soil sample values:

A
12.5
13.2
11.8
12.9
13.0
12.4
11.9
12.6
12.7
12.3

Step 1: Sample mean

In any empty cell, e.g., **B1**, type:

```
=AVERAGE(A1:A10)
```

- This will give **12.53**

Step 2: Sample variance and standard deviation

- **Sample variance (using your formula $S^2 = \Sigma(xi - \bar{x})^2 / n$):**

In **B2**, type:

```
=VAR.P(A1:A10)
```

- This gives **0.1841** (population variance formula, since n in denominator).

Note: Excel's `VAR.P` divides by n; `VAR.S` divides by n-1.

- **Sample standard deviation:**

In **B3**, type:

```
=SQRT(B2)
```

- This gives **0.429**

Step 3: Standard error (SE)

Using formula SE = S / \sqrt{n-1}

In **B4**, type:

```
=B3/SQRT(COUNT(A1:A10)-1)
```

- This gives **0.143**

Step 4: 95% Confidence Interval

- Degrees of freedom: df=n−1=9
- t-value for 95% CI (two-tailed) can be calculated in Excel using:

```
=T.INV.2T(0.05, COUNT(A1:A10)-1)
```

- In **B5**, type:

```
=T.INV.2T(0.05, COUNT(A1:A10)-1)
```

- This gives **2.262**
- **Margin of error:**

```
=B4*B5
```

- In **B6**, type:

```
=B4*B5
```

- Gives **0.323**
- **Lower CI:**

```
=B1-B6
```

- **Upper CI:**

```
=B1+B6
```

- Results:

CI	Value
Lower	12.21
Upper	12.85

✅ Step 5: Summary in Excel Table

Statistic	Formula	Value
Mean ( )	`=AVERAGE(A1:A10)`	12.53
Variance (S^2)	`=VAR.P(A1:A10)`	0.1841
Std. dev (S)	`=SQRT(B2)`	0.429
SE	`=B3/SQRT(COUNT(A1:A10)-1)`	0.143
t (95% CI)	`=T.INV.2T(0.05, COUNT(A1:A10)-1)`	2.262
Margin of error	`=B4*B5`	0.323
95% CI	Lower: `=B1-B6` Upper: `=B1+B6`	[12.21, 12.85]

📘 Exercise 2 : Large Sample (Grouped Data)

Problem:
A quality control test for 150 tablets; sodium bicarbonate content (mg):

Class	[1610,1615)	[1615,1620)	[1620,1625)	[1625,1630)	[1630,1635)
Frequency	7	8	42	75	18

Tasks:

1. Compute sample mean and standard deviation.
2. Compute the standard error (SE).
3. Determine the 95% confidence interval for the mean.
4. Interpret the result.

Step 0: Organize data

Class	Frequency (f)	Midpoint (x)
1610-1615	7	1612.5
1615-1620	8	1617.5
1620-1625	42	1622.5
1625-1630	75	1627.5
1630-1635	18	1632.5

Step 1: Sample mean

In Excel:

1. Column D: `f * x` → `=B2*C2`
2. Sum of D: `=SUM(D2:D6)`
3. Sum of frequencies: `=SUM(B2:B6)`
4. Mean: `=SUM(D2:D6)/SUM(B2:B6)`

- **Result:**
  x̄=1625.25 mg

Step 2: Sample standard deviation

Formula for **grouped data**:

1. Column E: `(x - mean)^2` → `=(C2 - mean)^2`
2. Column F: `f * (x - mean)^2` → `=B2*E2`
3. Sum of F: `=SUM(F2:F6)`
4. Variance: `=SUM(F2:F6)/SUM(B2:B6)`
5. Standard deviation: `=SQRT(variance)`

- **Result:** S≈4.79 mg

Step 3: Standard error using n−1n-1n−1

- Total sample: n=∑fi=150n = \sum f_i = 150n=∑fi=150
- Formula in Excel:

```
= S / SQRT(SUM(B2:B6)-1)
```

- **Result:**

SE≈4.79/149≈0.392

Step 4: 95% Confidence Interval

Use **Z = 1.96** (large sample):

$CI=\bar{x}\pm Z*SE$

- Margin of error: `=1.96*SE` → ≈ 0.769
- Lower limit: `= mean - margin` → 1625.25 - 0.769 ≈ **1624.48**
- Upper limit: `= mean + margin` → 1625.25 + 0.769 ≈ **1626.02**

Step 5: Summary Table

Statistic	Value
Mean ($\bar{x}$)	1625.25
Std. dev (S)	4.79
SE	0.392
95% CI	[1624.48, 1626.02]

✅ **Interpretation:**
With 95% confidence, the **true mean sodium bicarbonate content** is **between 1624.48 mg and 1626.02 mg**.

📗 Exercise 3 : One-Sample t-Test (Test of Conformity)

Objective: Test whether the mean soil organic matter content differs from the national reference value m0=4.5%m_0 = 4.5\%m0=4.5%.

Data:

Sample No.	1	2	3	4	5	6	7	8	9	10	11	12
Organic Matter (%)	4.2	4.8	5.0	4.1	4.4	4.7	4.3	4.9	4.5	4.6	4.0	4.8

Step 1: Enter Data

A
4.2
4.8
5.0
4.1
4.4
4.7
4.3
4.9
4.5

A

4.6

4.0

4.8

(Paste these in cells A1:A12)

Step 2: Compute Sample Statistics

Cell	Formula	Explanation
B1	`=COUNT(A1:A12)`	Sample size n = 12
B2	`=SUM(A1:A12)`	Sum of observations = 54.3
B3	`=AVERAGE(A1:A12)`	Sample mean = 4.525
B4	`=VAR.S(A1:A12)`	Sample variance = 0.09854
B5	`=STDEV.S(A1:A12)`	Sample standard deviation = 0.3140
B6	`=B5/SQRT(B1-1)`	Standard error = 0.0946

Step 3: Compute t-Statistic

Cell	Formula	Explanation
B7	`=B3-4.5`	Difference from reference = 0.025
B8	`=B7/B6`	t statistic = 0.2641
B9	`=B1-1`	Degrees of freedom = 11

Step 4: Critical Value (Two-Tailed Test, α = 0.05)

Cell	Formula	Explanation
B10	`=T.INV.2T(0.05,B9)`	Two-tailed critical t ≈ 2.201

Step 5: Decision

Cell	Formula	Explanation
B11	`=IF(ABS(B8)>B10,"Reject H0","Do not reject H0")`	Result: “Do not reject H0”

📒 Exercise 4:

Problem: We want to compare the average exam scores of two independent classes to determine

whether there is a significant difference between their mean scores.

Data

• Class A (n1 = 8): 78, 82, 85, 90, 76, 88, 84, 79

• Class B (n2 = 7): 72, 75, 80, 77, 74, 79, 70

Significance level: α = 0.05 (two-tailed)

Tasks:

1. Compute the sample means of sample 1 and 2.

2. Compute the sample variances of sample 1 and 2.

3. Perform a two-sample t-test to compare the two population means.

Step 1: Enter the Data

Class A	Class B
78	72
82	75
85	80
90	77
76	74
88	79
84	70
79	

(Class A in A1:A8, Class B in B1:B7)

Step 2: Compute Sample Means

- **Class A mean** → `=AVERAGE(A1:A8)` → **82.75**
- **Class B mean** → `=AVERAGE(B1:B7)` → **75.29**

Step 3: Compute Sample Variances

- **Class A variance** → `=VAR.S(A1:A8)` → **20.935**
- **Class B variance** → `=VAR.S(B1:B7)` → **11.515**

Use `VAR.S` because these are **sample variances**.

Step 4: Compute Pooled Variance (Equal Variance Assumption)

1. Weighted sum of squares:
   `=(COUNT(A1:A8)-1)*VAR.S(A1:A8)+(COUNT(B1:B7)-1)*VAR.S(B1:B7)` → 248.083
2. Pooled variance:
   `=F1/(COUNT(A1:A8)+COUNT(B1:B7)-2)` → **19.083**
3. Standard error:
   `=SQRT(F2*(1/COUNT(A1:A8)+1/COUNT(B1:B7)))` → **2.260**

Step 5: Compute t-Statistic

`t_obs = (Mean_A - Mean_B)/SE` → `=(D1-D2)/F3` → **3.29**

Step 6: Degrees of Freedom and Critical Value

- Degrees of freedom: `=COUNT(A1:A8)+COUNT(B1:B7)-2` → **13**
- Critical t (two-tailed, α=0.05): `=T.INV.2T(0.05,H1)` → **2.160**

Step 7: Decision

- Compare t_obs with t_critical:
  `t_obs > t_critical → Reject H0` → **There is a significant difference between the classes.**

Step 8: 95% Confidence Interval for Difference

- Difference of means: `=D1-D2` → 7.46
- Lower bound: `=J1-H2*F3` → 2.56
- Upper bound: `=J1+H2*F3` → 12.36

✅ CI: `[2.56, 12.36]`