Practical exercise

Maria Beatriz Walter Costa
Viral Ecology and Omics (VEO)

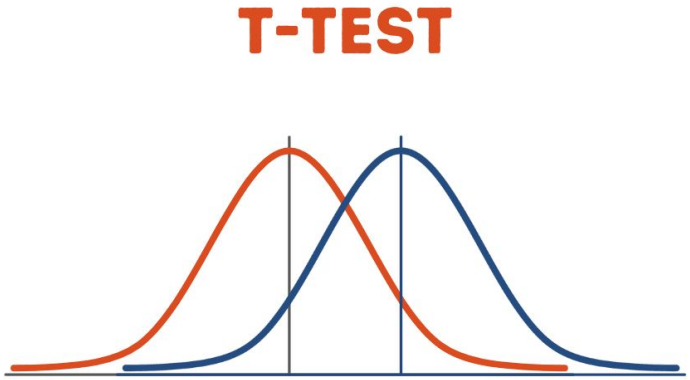
Foto: Anne Günther

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

INSTITUTO FEDERAL
Goiano
Campus Rio Verde

Question: are the cultivation temperatures of two Bacterial groups different?

- **Null hypothesis:** cultivation temperatures of the two groups are not different
- **Samples: independent and unpaired**



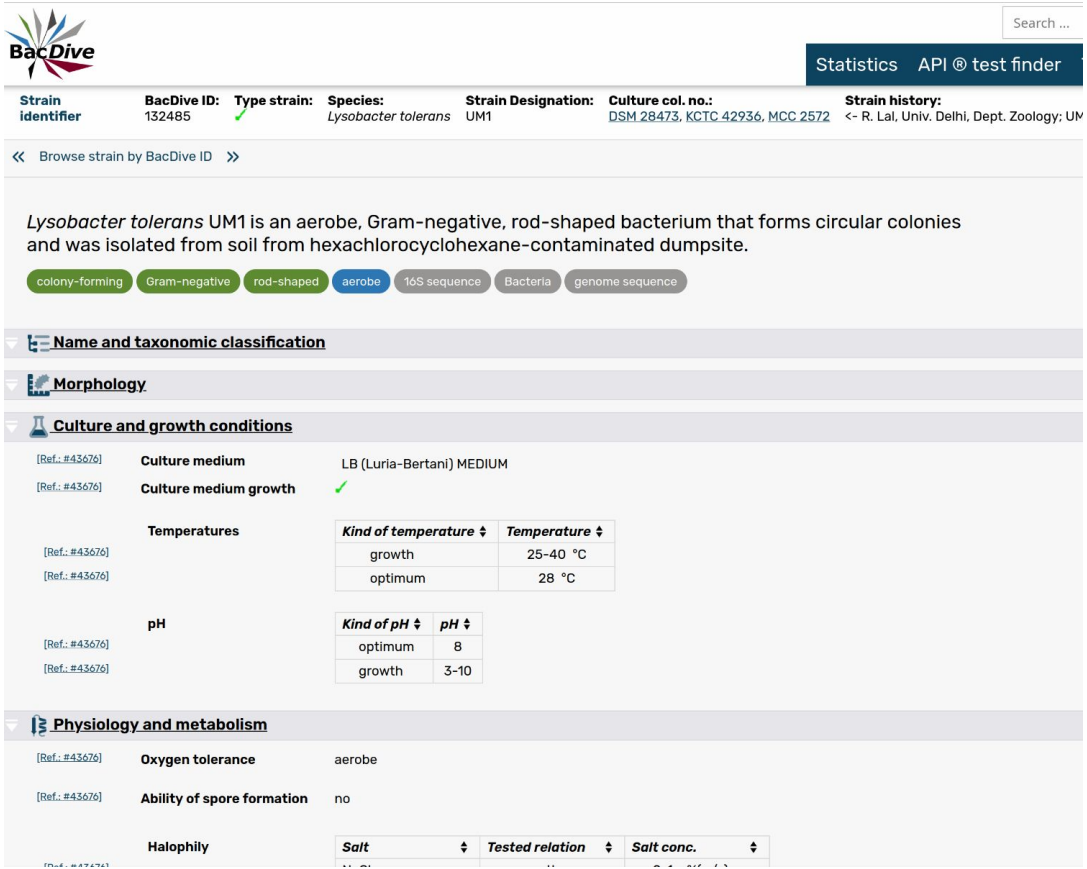
Images: https://olhardigital.com.br and https://www.kdnuggets.com

Outline of the practical exercise

- **Collect data**
  - BacDive: https://bacdive.dsmz.de/
- **Prepare data for t-test**
  - calculate: mean, standard deviation, sample size, degrees of freedom
- **Calculate t-value**
- **t-test: compare t-value to critical value**
- **Interpret results**

- **Slides at:**
  https://github.com/waltercostamb/statistics

BacDive: cultivation metadata of temperature



Collect data

- **Go to BacDive**: https://bacdive.dsmz.de/
- **Collect the species name and the first (top) reported temperature for the isolate IDs:**

- **Group 1**
  - OUC007
  - DSM 9849
  - EBR-02E-0045
  - DSM 1237

- **Group 2**
  - 318
  - DSM 23807
  - DSM 12644

Collect data: solution

- **Group 1**
  - *Abyssibacter profundi* OUC007: **37 C**
  - *Abiotrophia defectiva* DSM 9849: **37 C**
  - *Acetivibrio clariflavus* EBR-02E-0045: **55 C**
  - *Acetivibrio thermocellus* DSM 1237: **55 C**

- **Group 2**
  - *Achromobacter denitrificans* 12 (318): **30 C**
  - *Achromobacter insolitus* DSM 23807: **28 C**
  - *Acidovorax defluvii* DSM 12644: **30 C**

Prepare data for t-test

- **Calculate sample size for group 1 (n1)**
- **Calculate sample size for group 2 (n2)**

- **Calculate mean for group 1**
- **Calculate mean for group 2**

- **Calculate standard deviation for group 1**
- **Calculate standard deviation for group 2**

- **Calculate the degrees of freedom (df) for the t-test**
  - **df = n1 + n2 -2**

$$\text{Mean} = \frac{\text{Sum of All Data Points}}{\text{Number of Data Points}}$$

Fórmulas

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ = desvio padrão da população

N = o tamanho da população

x_i = cada valor da população

μ = a média da população

Steps to calculate standard deviation (in portuguese):
https://www.calculadora.app/matematica/desvio-padrao

Prepare data for t-test: solution

- Sample size for group 1 (n1): **4**
- Sample size for group 2 (n2): **3**

- Mean for group 1: **46**
- Mean for group 2: **29.33**

- Standard deviation for group 1: **9**
- Standard deviation for group 2: **0.94**

- Degrees of freedom (df) for the t-test: **5**

- **Note that, in this exercise, we used the standard deviation formula for population**
- **In a scientific experiment, we should use the standard deviation formula for sample**

Formulas for Standard Deviation

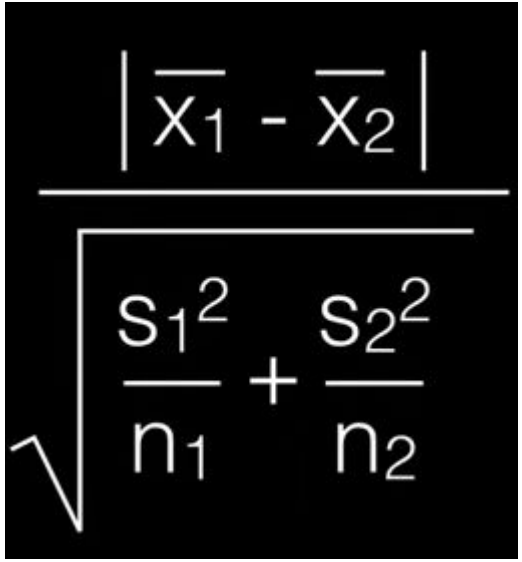
Population Standard Deviation Formula	$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$
Sample Standard Deviation Formula	$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

Notations for Standard Deviation

- σ = Standard Deviation
- x_i = Terms Given in the Data
- $\bar{x}$ = Mean
- n = Total number of Terms

Formulas for population and sample: https://byjus.com/standard-deviation-formula/;

Calculate t-value



- **x1 = mean of group 1**
- **x2 = mean of group 2**

- **s1 = standard deviation of group 1**
- **s2 = standard deviation of group 2**

- **n1 = sample size of group 1**
- **n2 = sample size of group 2**

- **We will see the balance between signal and noise**

Calculate t-value: solution: **3.7**

- x1, mean of group 1 = **46**
- x2, mean of group 2 = **29.33**

- s1, standard deviation of group 1 = **9**
- s2, standard deviation of group 2 = **0.94**

- n1, sample size of group 1 = **4**
- n2, sample size of group 2 = **3**

- **Conclusion: signal is larger than noise (t-value > 1)**

$$\frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t-test: compare t-value to critical value on t-table

- **We know, our signal is larger than noise**
- **Now we compare t-value to critical value to know if this is due to chance**
- **t-value < critical value**: we do not reject the null hypothesis
- **t-value > critical value**: we reject the null hypothesis

Critical value for df=5 and p-value = 0.05 on t-table: 2.571

df/α	0.9	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	1	2	3.078	6.314	12.706	31.821	64	637
2	0.142	0.816	1.386	1.886	2.92	4.303	6.965	10	31.598
3	0.137	0.765	1.25	1.638	2.353	3.182	4.541	5.841	12.929
4	0.134	0.741	1.19	1.533	2.132	2.776	3.747	4.604	8.61
5	0.132	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.718	1.134	1.44	1.943	2.447	3.143	3.707	5.959
7	0.13	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.13	0.706	1.108	1.397	1.86	2.306	2.896	3.355	5.041
9	0.129	0.703	1.1	1.383	1.833	2.263	2.821	3.25	4.781
10	0.129	0.7	1.093	1.372	1.812	2.228	2.764	3.169	4.587

t-test: compare t-value to critical value on t-table

- **We compared t-value = 3.7 to critical value = 2.571**
- **t-value > critical value**: we reject the null hypothesis
- **Conclusion: there is a statistical difference between temperatures of groups 1 and 2!**
- **Remember that: t-tests assume normal distributions of samples and similar variance**
  - We do not meet these criteria, so for a scientific experiment, we should do a non-parametric test instead!
  - We used the t-test today for learning purposes

df/α	0.9	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	1	2	3.078	6.314	12.706	31.821	64	637
2	0.142	0.816	1.386	1.886	2.92	4.303	6.965	10	31.598
3	0.137	0.765	1.25	1.638	2.353	3.182	4.541	5.841	12.929
4	0.134	0.741	1.19	1.533	2.132	2.776	3.747	4.604	8.61
5	0.132	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.718	1.134	1.44	1.943	2.447	3.143	3.707	5.959
7	0.13	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.13	0.706	1.108	1.397	1.86	2.306	2.896	3.355	5.041
9	0.129	0.703	1.1	1.383	1.833	2.263	2.821	3.25	4.781
10	0.129	0.7	1.093	1.372	1.812	2.228	2.764	3.169	4.587

YouTube Channel “Bozeman Science”, Student’s t-test,
t-table from: https://www.sthda.com/english/wiki/t-distribution-table

Thank you!