Université
de Technologie
Tarbes
Occitanie Pyrénées

THE **COMPLETE GUIDE** TO

INFERENTIAL STATISTICS

< By Raymond HOUE NGOUNA />

Université de Technologie Tarbes Occitanie Pyrénées

</SECTION 4>

TWO-SAMPLE T-TEST

SECTION OUTLINE

- ❑ Case study
- ❑ Key concepts
  - ✓ Dependent variable and Independent variable
  - ✓ Test for normality
  - ✓ Test for equality of variance
  - ✓ Paired two-sample t-test
- ❑ Summary
  - ✓ Two-sample t-test implementation

CASE STUDY

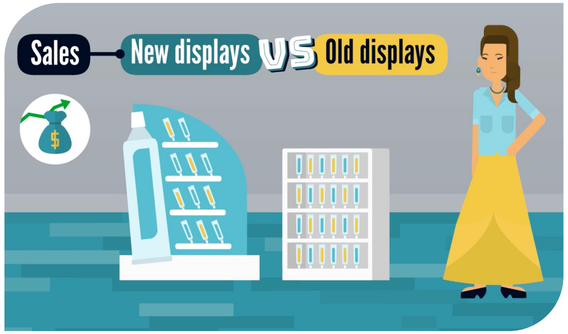
CASE STUDY

- ✓ Carmen works in the trade marketing department of a company that sells toothpaste.

  - The products are distributed through hypermarkets, supermarkets, and convenience stores.
  - She wants to introduce a new type of in-store display but is unsure if the concept will be effective.

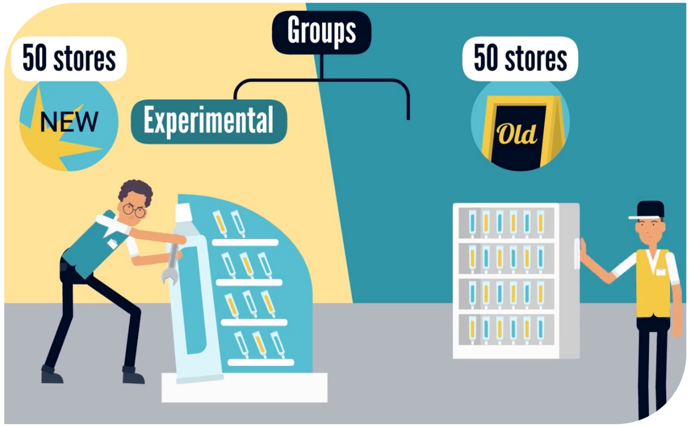
  - Instead of rolling out the visual merchandising solution across the entire retail network, she decides to first conduct an experiment.

CARMEN'S EXPERIMENT

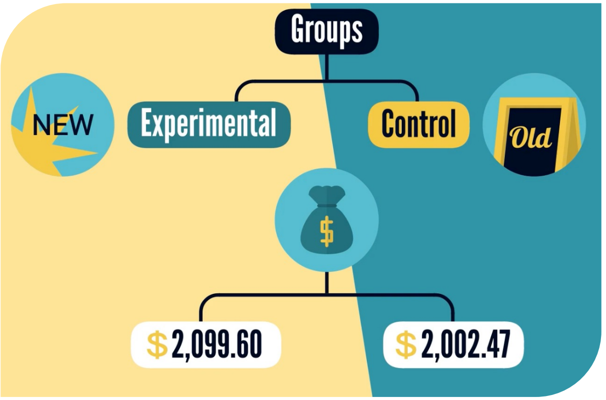


✓ Carmen aims to determine whether there's a difference in sales volume between stores using new display setups and those with traditional displays.

- Her team installed the new displays in 50 stores across the country, forming the experimental group.
- Meanwhile, another 50 stores within the company's distribution network kept the regular displays, serving as the control group.
- Over several days, Carmen and her manager monitored sales volume as the dependent variable.



OUTCOME OF THE EXPERIMENT



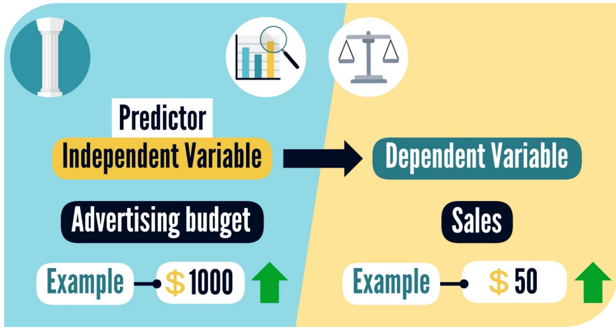
- ✓ The data shows that the new display type averages \$2,099.60 in sales:
  - outperforming the old display type's average of \$2,002.47,
  - suggesting that the newer displays lead to higher sales volumes.

- ✓ To ensure the difference is statistically significant, Carmen conducts a two-sample t-test with a significance level of 0.05:
  - t = 1.7693 with 98 degrees of freedom.
  - p-value = 0.0400 ➔ Carmen is happy (why?).

DEPENDENT / INDEP. VARIABLES

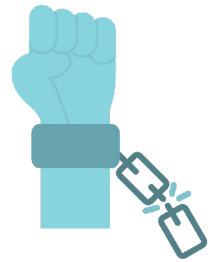
DEPENDENT & INDEPENDENT VARIABLES

- ✓ In statistics, the dependent and independent variables are the two primary variables studied to understand their relationship.



- ✓ The connection between these variables is one of the focus of statistical analysis.

INDEPENDENT VARIABLES

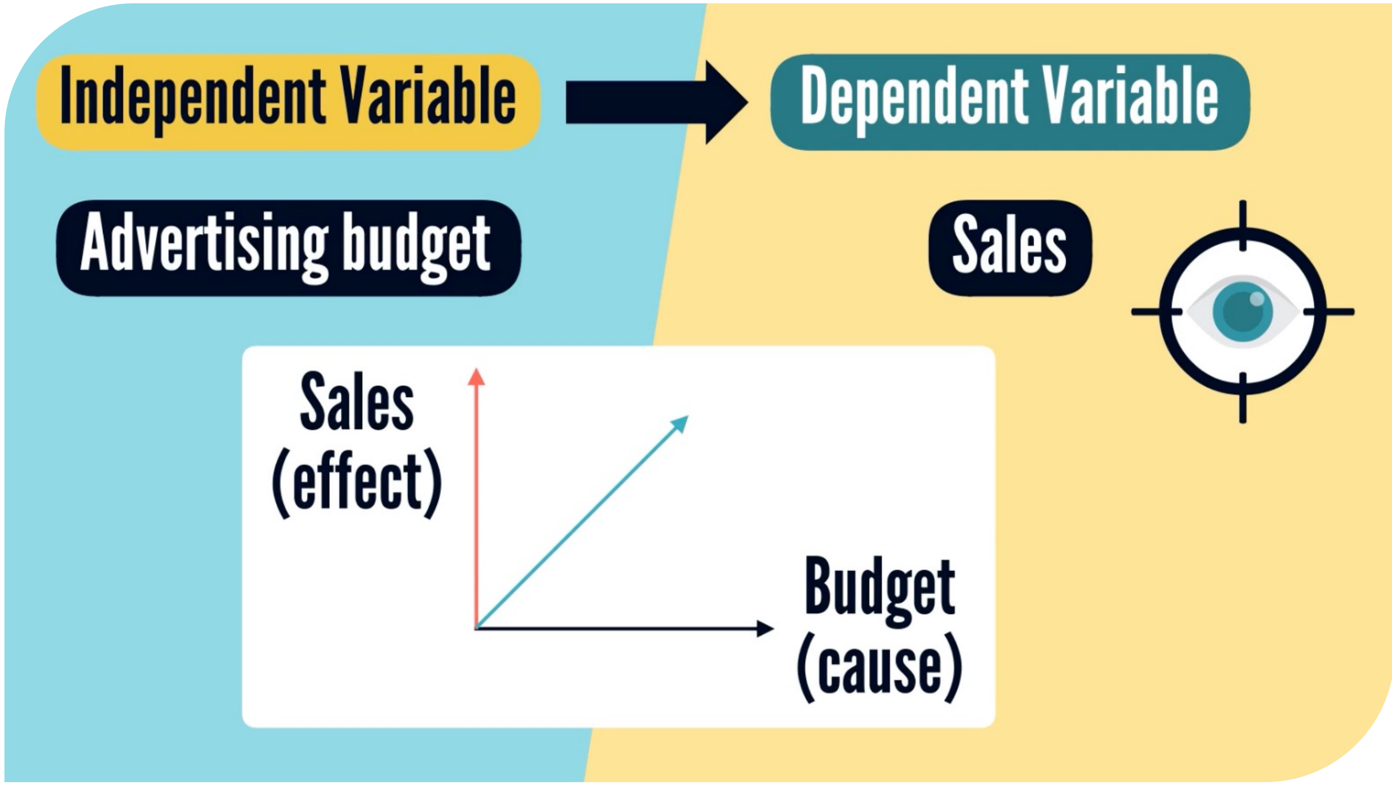


- ✓ The independent variable (IV), also known as the predictor or explanatory variable, is manipulated or controlled in an experiment or study.
  - Its values are deliberately changed or varied to observe their effect on the dependent variable.
  - It is called "independent" because it stands alone and is not influenced or altered by other measured variables.
  - In a cause-and-effect relationship, independent variables are expected to cause changes in the dependent variable.
- ✓ In visual analyses, IVs are represented on a chart’s x-axis.

DEPENDENT VARIABLES

- ✓ The dependent variable (DV), also known as the response or outcome variable, is the primary focus of observation in an experiment or study.
  - It is what the researcher measures and is influenced by the experimental conditions, with its variation depending on changes in the independent variable in a cause-and-effect relationship.
- ✓ The dependent variable responds to changes in the independent variable and is **plotted on the** y-axis.

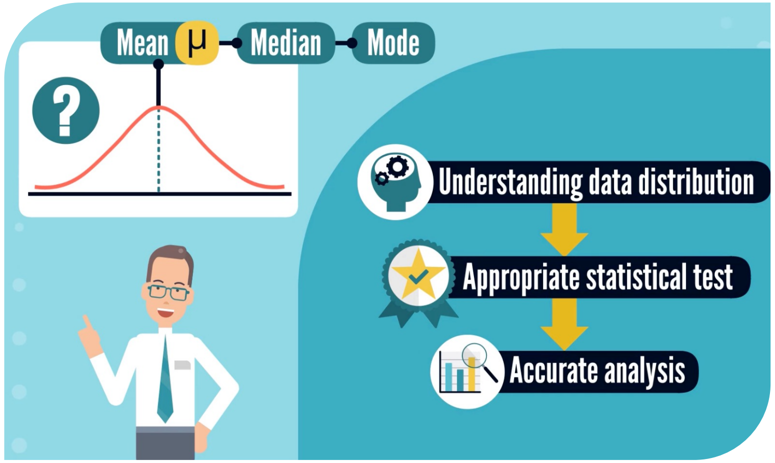
VISUALIZATION OF DV AND IV



NORMALITY TEST

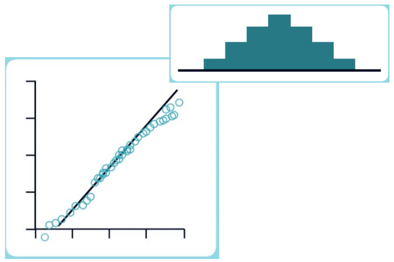
Université de Technologie Tarbes Occitanie Pyrénées

NORMALITY TEST



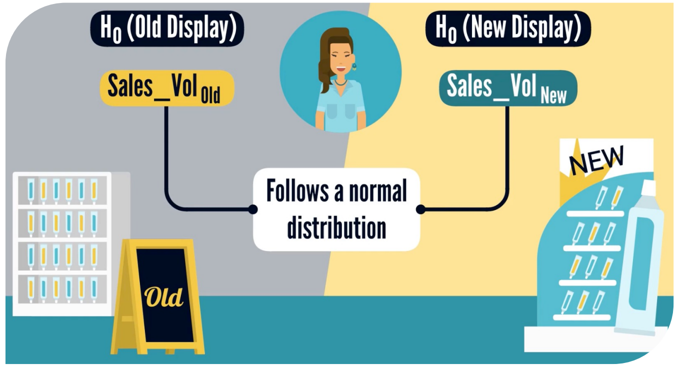
- ✓ This section presents another technique to determine whether a dataset follows a normal distribution.
  - This knowledge is crucial because understanding the distribution of data can significantly influence the choice of statistical methods and the accuracy of your analyses.

- As described earlier, normality can be visually assessed using a histogram or a QQ-plot.
- Additionally, it can be evaluated mathematically using the Shapiro-Wilk test, a specific statistical test for normality.



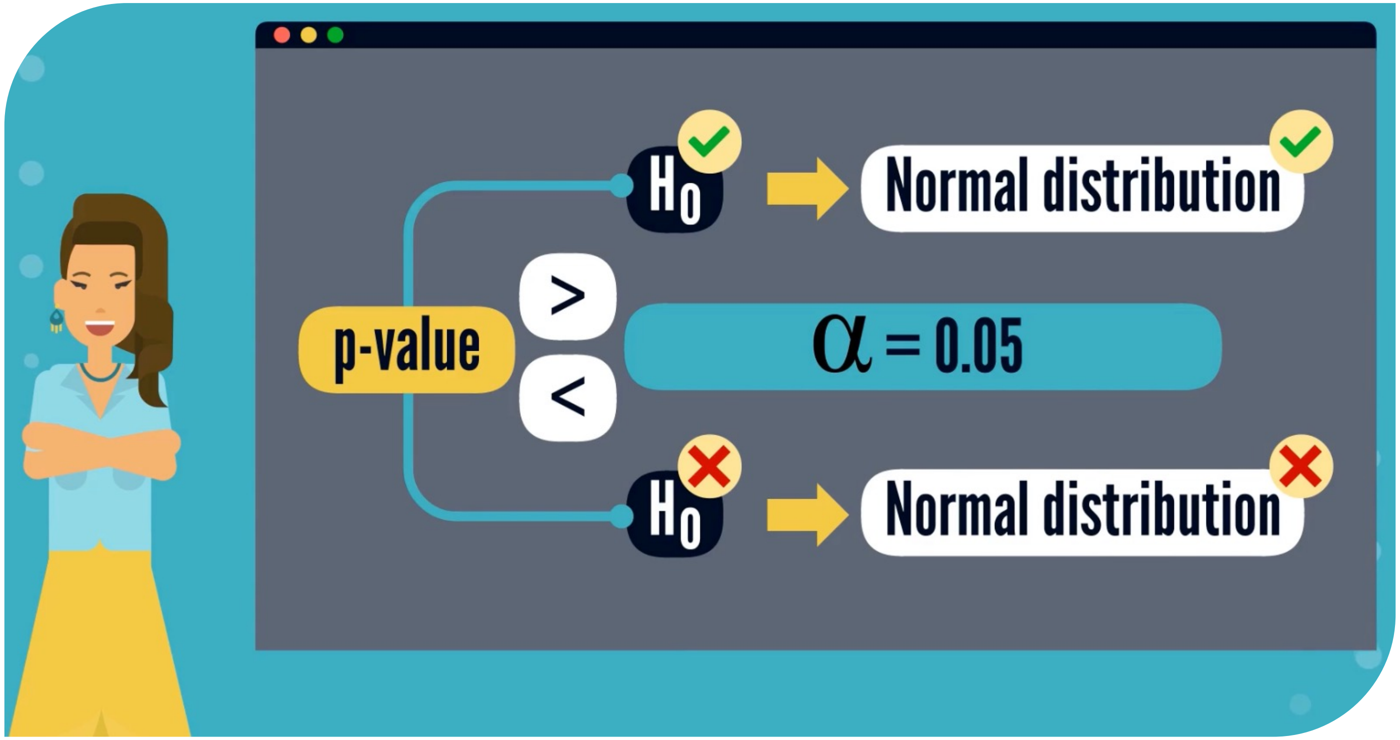
SHAPIRO-WILK TEST

- ✓ The Shapiro-Wilk test is a dedicated statistical test for assessing normality.
- ✓ Its primary objective is to quantitatively evaluate the likelihood that the data in a sample were drawn from a normal distribution.
  - The null hypothesis of the test suggests that the data follow a normal distribution.



Carmen's case study

CARMEN’S STUDY TEST FOR NORMALITY



EXPECTATION IN APPLYING SHAPIRO-WILK TEST

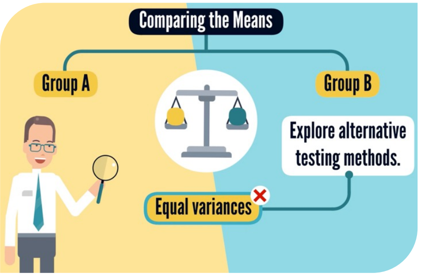
TEST FOR EQUALITY OF VARIANCE

Université de Technologie Tarbes Occitanie Pyrénées

RELIABILITY OF A MEAN COMPARISON TEST

✓ The reliability of certain statistical tests, including mean comparisons, depends on a key assumption:

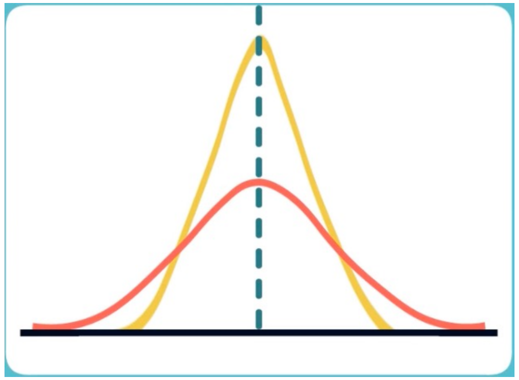
- The 2 samples must have equal variances.
- This means that the variability within each group from which the samples are drawn should be similar.



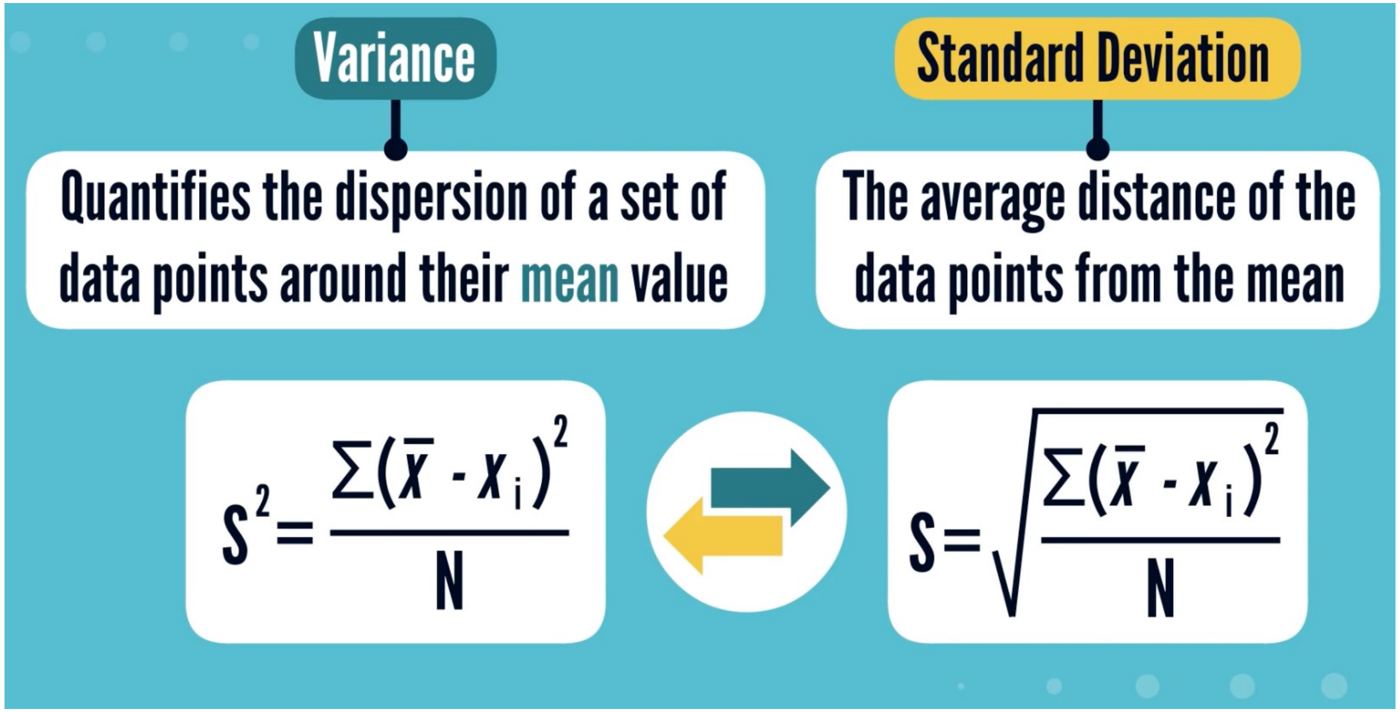
- If this condition is not met, the test's p-value may be unreliable.
- In situations where variances are unequal, alternative testing methods may be necessary to address this discrepancy and yield more robust results.

LEVENE'S TEST FOR EQUALITY OF VARIANCES

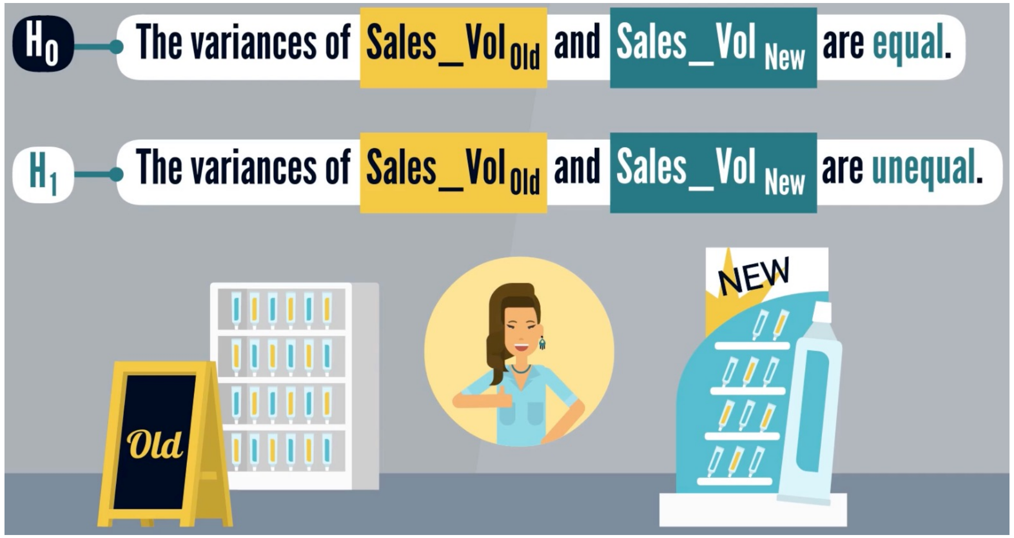
- ✓ One commonly used technique to assess whether the variances of two or more groups differ significantly is Levene's test for equality of variances.
  - Variance is a measure of spread that quantifies how data points are dispersed around their mean value.
  - It captures the variability within a data set, highlighting how much the data points differ from the average.

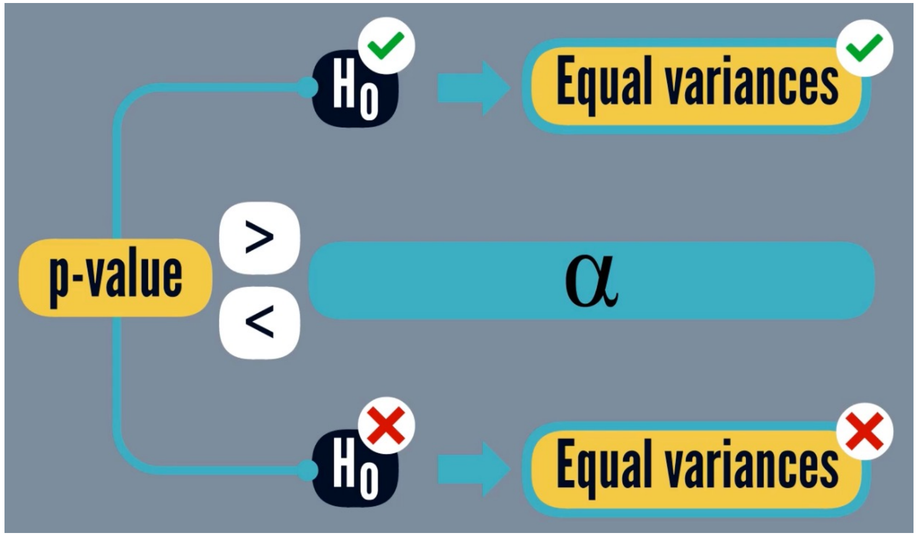


LINK WITH STANDARD DEVIATION



LEVENE’S TEST APPLIED TO CARMEN’S STUDY

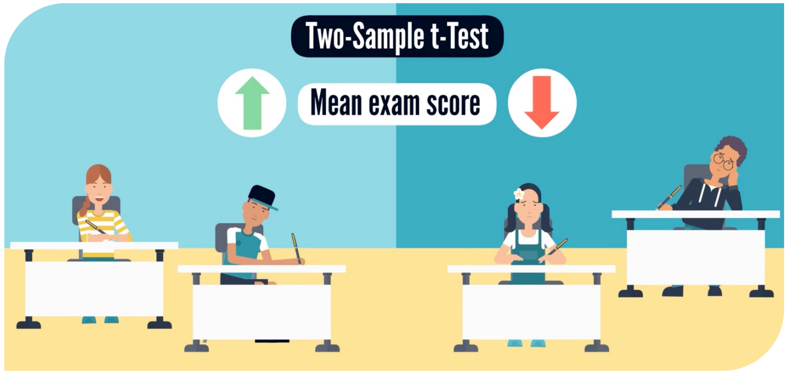




TWO-SAMPLE T-TEST

Université de Technologie Tarbes Occitanie Pyrénées

TWO-SAMPLE T-TEST



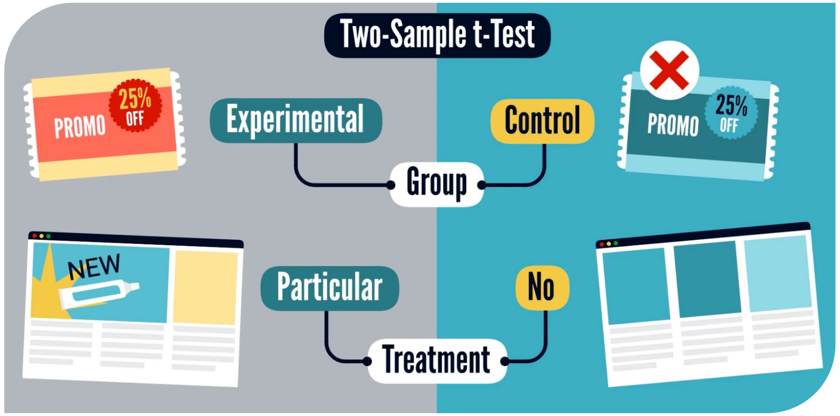
- ✓ The two-sample t-test is a statistical hypothesis test used to assess whether there is a significant difference between the means of 2 independent groups or samples.
  - This test is commonly applied when comparing the means of two populations or treatment groups, to determine whether the observed difference reflects a true effect or is simply due to chance.
- ✓ In essence, it helps evaluate whether the differences in mean values are statistically significant.

BRIEF HISTORY

- ✓ There are two main types of t-tests:
  - The two-sample (unpaired) t-test, also known as Student's t-test is named after **William Sealy Gosset**, who introduced the t-distribution in 1908 in an article published under the pseudonym "Student".
  - The paired samples t-test.
- ✓ The independent or unpaired samples t-test is used to determine whether the means of two unrelated groups are the same.
  - This test requires that the subjects in each group be distinct,
  - Ensuring that the observations in one group do not influence or directly relate to those in the other group.

EXPERIMENTS USING A TWO-SAMPLE T-TEST

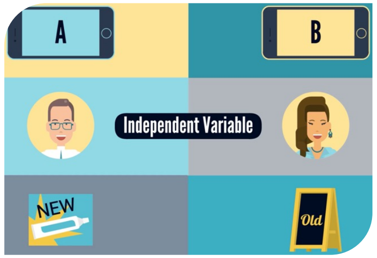
✓ Experiments using a two-sample t-test typically divide the sample into two distinct groups:
- The experimental group receives a specific treatment e.g. a promotional coupon or a modified version of a website.
- The control group receives no treatment e.g. they do not get coupons or see the modified website, but instead interact with the existing version.

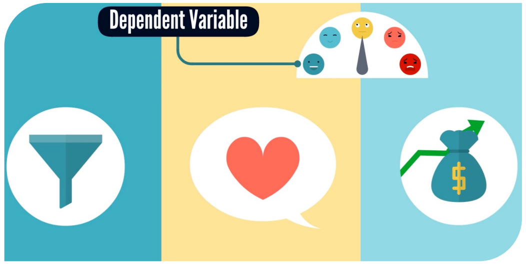


➔ Because the two groups are unrelated, we obtain two independent samples.

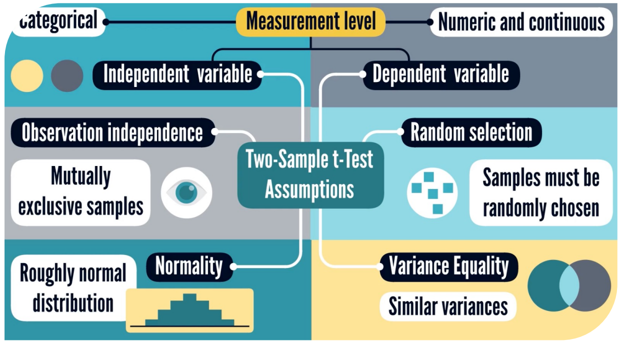
TYPES OF VARIABLE INVOLVED

- ✓ The two-sample t-test involves 2 variables.
  - The independent variable, which defines and distinguishes the 2 groups,
    - Ex.: *Design A* versus *Design B*, men versus women, the new versus the old display…
  - The dependent variable, which represents the measurement of interest,
    - Such as the conversion rate, brand affinity, the sales volume, etc.



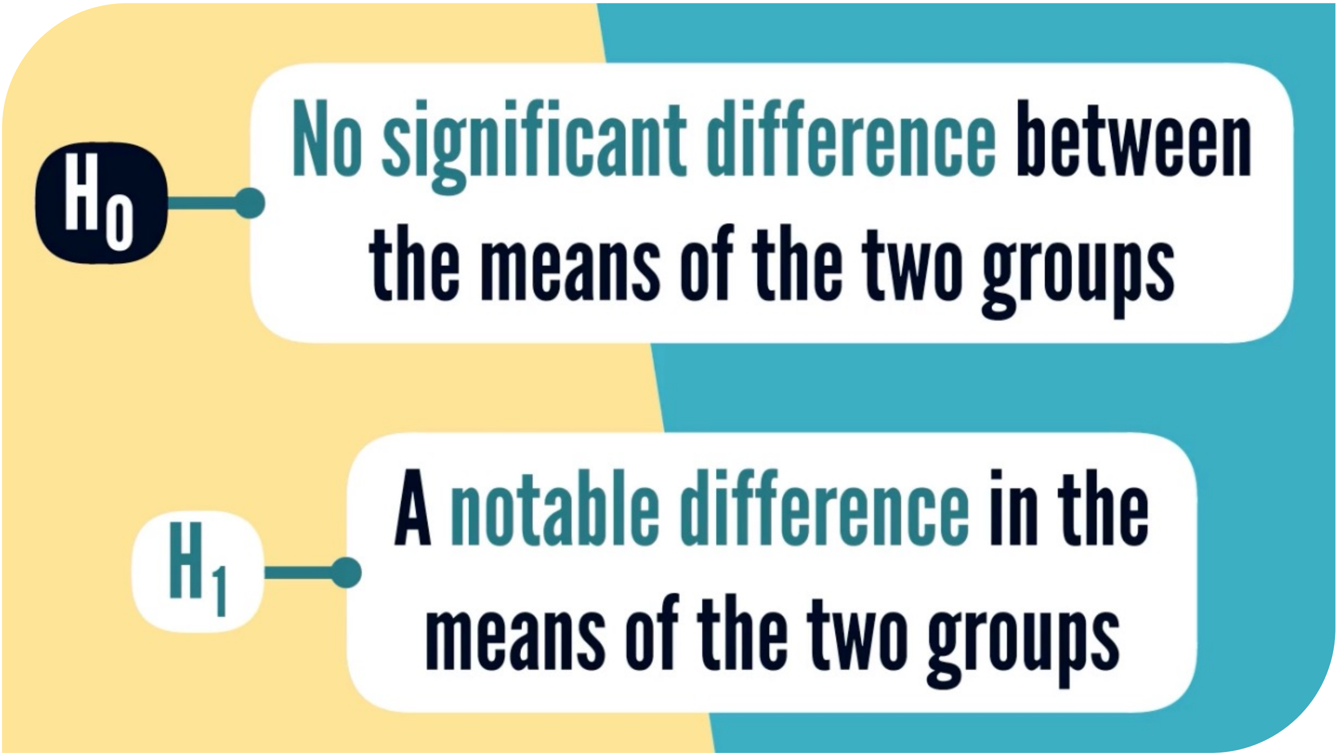


TWO-SAMPLE T-TEST ASSUMPTIONS



- ✓ **Measurement Level:**
  - The IV should be categorical, either nominal or ordinal, with exactly 2 categories.
  - The DV should be numeric and continuous in both samples.
- ✓ **Independence of Observations:** The samples must be mutually exclusive, with no overlapping values between them.
- ✓ **Random Selection:** The samples should be randomly selected to ensure that the test results can be generalized to the population.
- ✓ **Normality:** The distributions of the dependent variable in both sample and population should be approximately normal.
- ✓ **Equality of Variances:** The variances of the two samples should be similar.

HYPOTHESES OF THE TWO-SAMPLE T-TEST



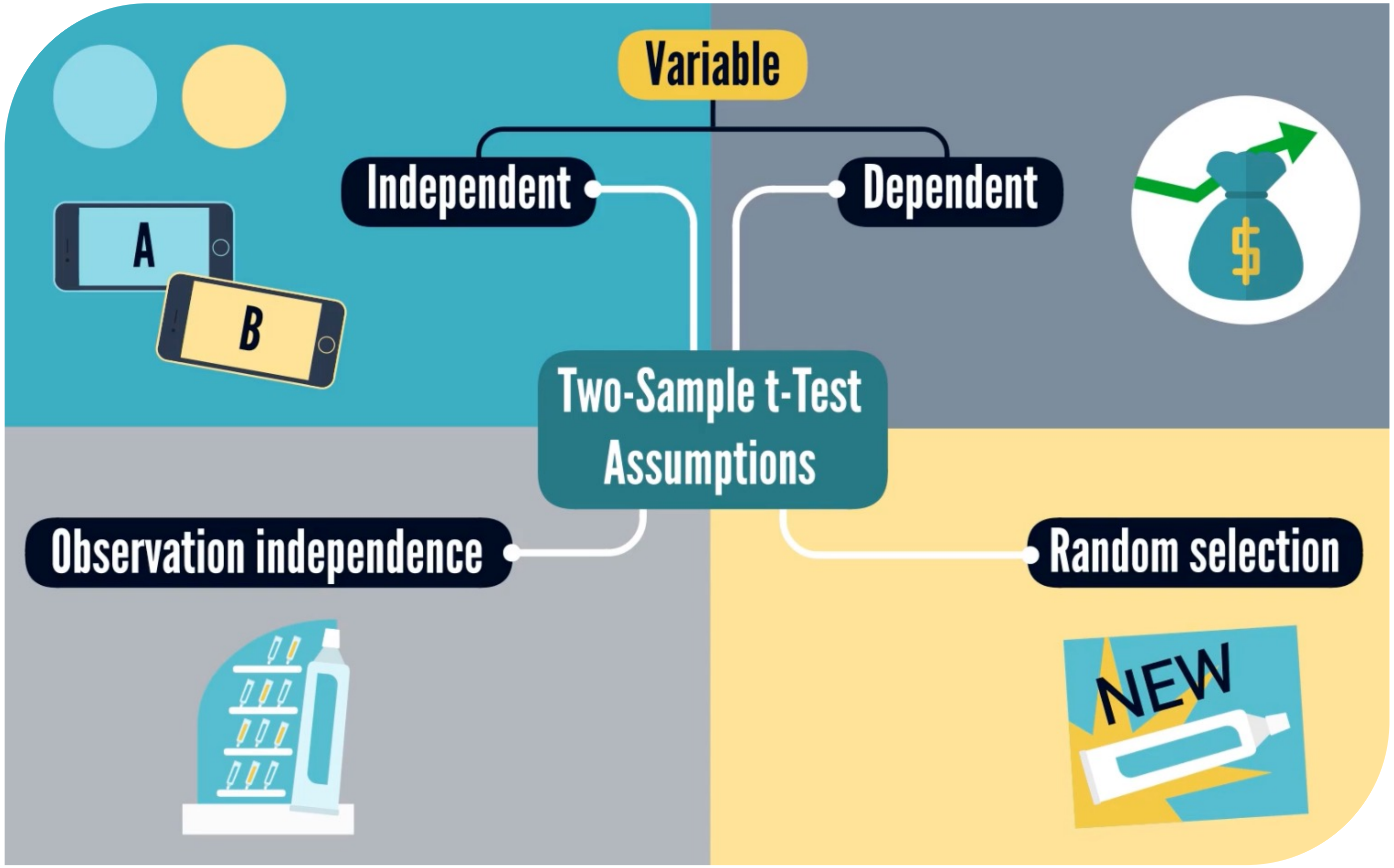
IN CARMEN'S CASE STUDY

- ✓ The vendor assured Carmen that the new in-store (IV) displays were more visually appealing than the previous ones, leading her to believe they would increase sales (DV).

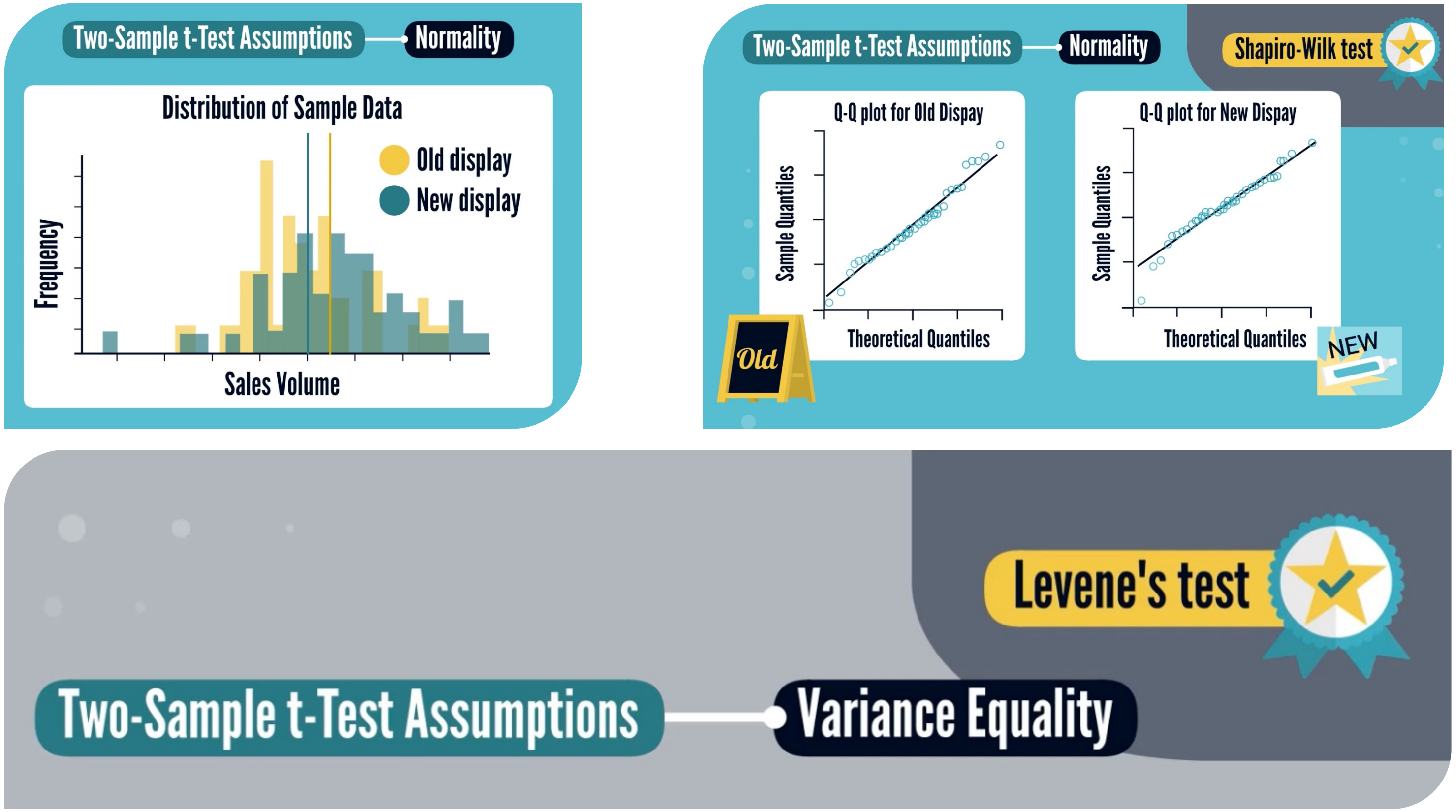
- ○ She sought statistical evidence to support this belief, which would justify their widespread adoption throughout the company's retail network.
- ○ To this end, a one-tailed alternative hypothesis could be used:
  the difference between the mean sales volumes for the new and old displays is greater than zero.

SUMMARY

VERIFICATION OF THE ASSUMPTIONS



NORMALITY AND VARIANCE ASSUMPTIONS

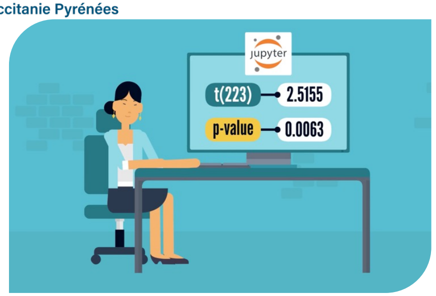


HYPOTHESES OF THE TEST

$$H_0: \mu(\text{Sales\_Vol}_{\text{New}}) - \mu(\text{Sales\_Vol}_{\text{Old}}) = 0$$

$$H_1: \mu(\text{Sales\_Vol}_{\text{New}}) - \mu(\text{Sales\_Vol}_{\text{Old}}) > 0$$

CONCLUSION OF THE STUDY




```python
from scipy.stats import shapiro
import matplotlib.pyplot as plt
import statsmodels.api as sm

pd.options.display.float_format = "{:.2f}".format
```

```python
# Read case data
df = pd.read_excel("4. Case 3 - Two-sample t-test.xlsx")
```

```python
df.head()
```

	Store Id	Display Type	Sales Volume
0	3846186	Old_display_type	2038.31
1	1083410	Old_display_type	2017.29
2	4278951	Old_display_type	1746.47
3	6670048	Old_display_type	2061.78
4	5054220	Old_display_type	2253.76

```python
# Create arrays with units sold for 'Old_display_type' and 'New_display_type'
data_o = df[df["Display Type"] =="Old_display_type"]["Sales Volume"].values
data_n = df[df["Display Type"] =="New_display_type"]["Sales Volume"].values
```

```python
# Perform Shapiro-Wilk test on both variables
print("Old display: " + str(shapiro(data_o)))
print("New display: " + str(shapiro(data_n)))
```

```
Old display: ShapiroResult(statistic=0.9745517373085022, pvalue=0.3508000373840332)
New display: ShapiroResult(statistic=0.9707477688789368, pvalue=0.2481728047132492)
```

EQUALITY OF VARIANCE TEST

```python
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import levene

pd.options.display.float_format = "{:.2f}".format
```

```python
# Read case data
df = pd.read_excel("4. Case 3 - Two-sample t-test.xlsx")
```

```python
df.head()
```

	Store Id	Display Type	Sales Volume
0	3846186	Old_display_type	2038.31
1	1083410	Old_display_type	2017.29
2	4278951	Old_display_type	1746.47
3	6670048	Old_display_type	2061.78
4	5054220	Old_display_type	2253.76

```python
# Create arrays with units sold for 'Old_display_type' and 'New_display_type'
data_o = df[df["Display Type"] =="Old_display_type"]["Sales Volume"].values
data_n = df[df["Display Type"] =="New_display_type"]["Sales Volume"].values
```

```python
# Levene's test centered at the mean
print(stats.levene(data_o, data_n, center='mean'))
```

TOW-SAMPLE T-TEST

```python
df.head()
```

	Store Id	Display Type	Sales Volume
0	3846186	Old_display_type	2038.31
1	1083410	Old_display_type	2017.29
2	4278951	Old_display_type	1746.47
3	6670048	Old_display_type	2061.78
4	5054220	Old_display_type	2253.76

```python
# Create arrays with units sold for 'Old_display_type' and 'New_display_type'
data_o = df[df["Display Type"] =="Old_display_type"]["Sales Volume"].values
data_n = df[df["Display Type"] =="New_display_type"]["Sales Volume"].values
```

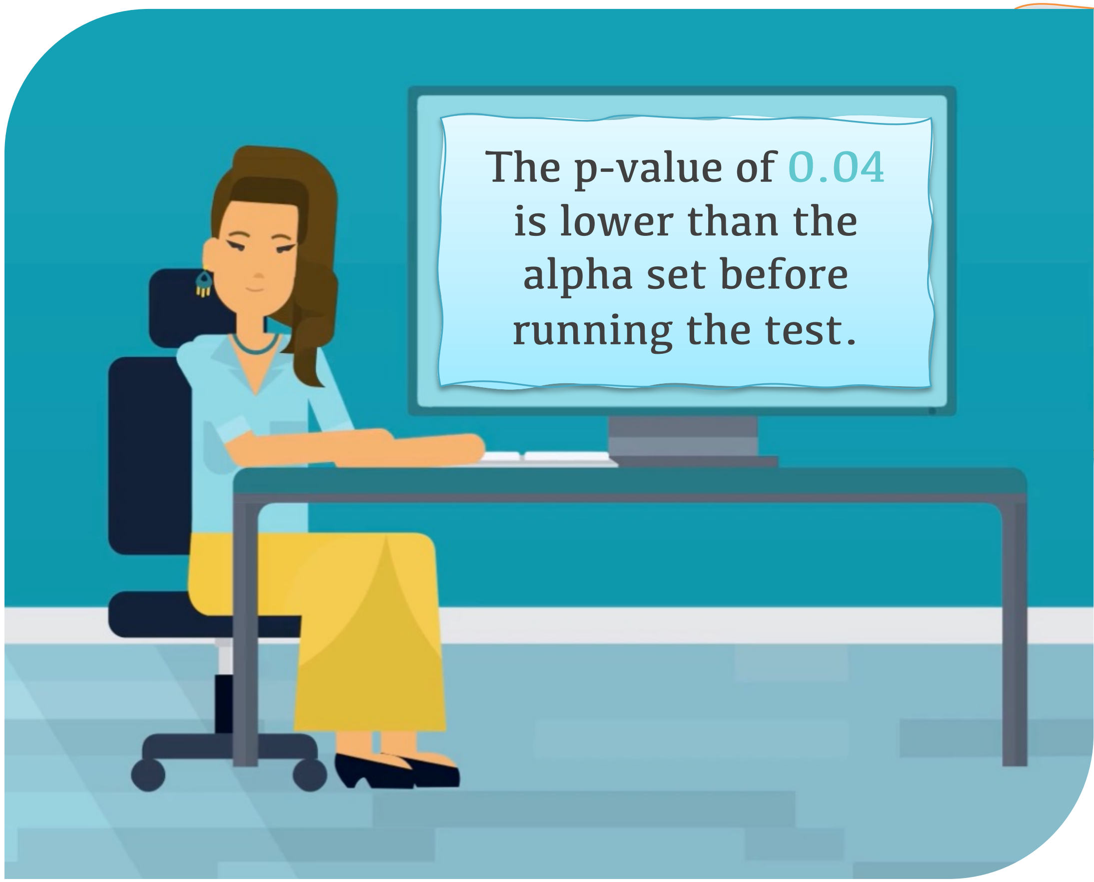
```python
# Mean sales for old and new display types
print("Mean Old: " + str(np.mean(data_o)))
print("Mean New: " + str(np.mean(data_n)))
```

```
Mean Old: 2002.4714000000001
Mean New: 2099.6049999999996
```

```python
# Run the two-sample t-test
stats.ttest_ind(data_n, data_o, equal_var=True, alternative = "greater")
```

```
Ttest_indResult(statistic=1.7693132318181237, pvalue=0.0399761880616222)
```

THE RESULT OF THE TEST



- Based on this result, we can reject the null hypothesis.
- Although the mean sales volume for the new in-store display is only slightly higher than for the old one.
- The results indicate that the difference is statistically significant at the 0.05 level.