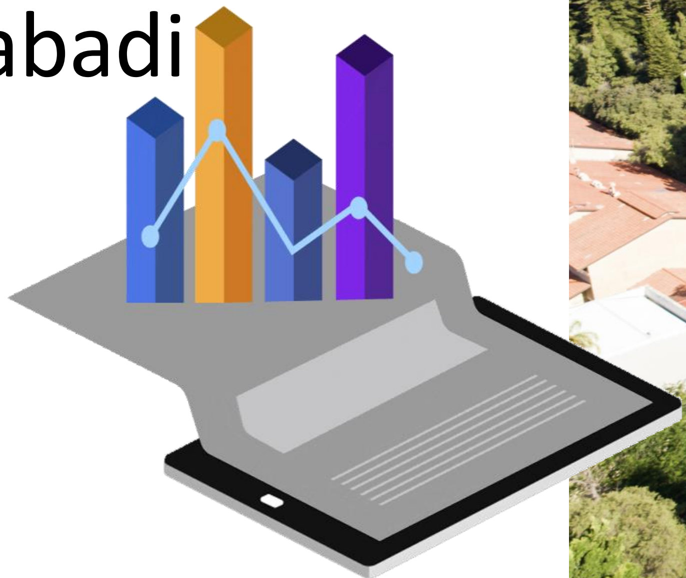
THE UNIVERSITY OF WESTERN AUSTRALIA

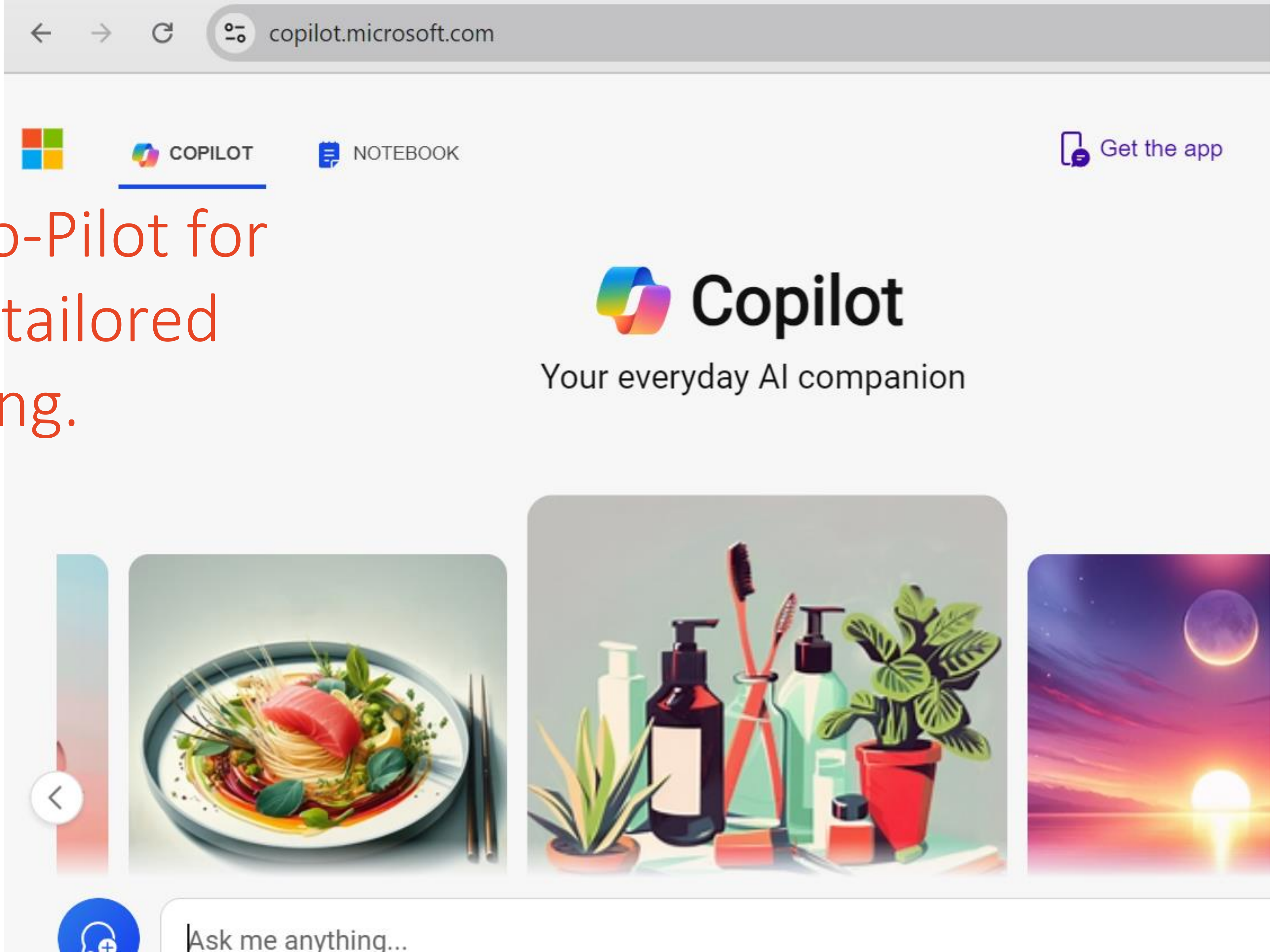
Week 5: Diagnostic Analytics in Supply Chains (SCs)

Presented by:

Dr. Mehdi Rajabi Asadabadi



Ask Co-Pilot for more tailored learning.



Five Key Questions in Supply Chain Analytics

- What is my plan? (The Goals)
- What is my present position (where am I)? (Descriptive Analytics)
- What are the variances? What caused them? (Diagnostic Analytics)
- What are the trends? What are the forecasts? (Predictive Analytics)
- What actions are required? (Prescriptive Analytics)

Types of Analytics

?

Prescriptive Analytics

Predictive Analytics

Descriptive Analytics *(Analysis)*

Added Value

Complexity and Cost

Diagnostic Analytics

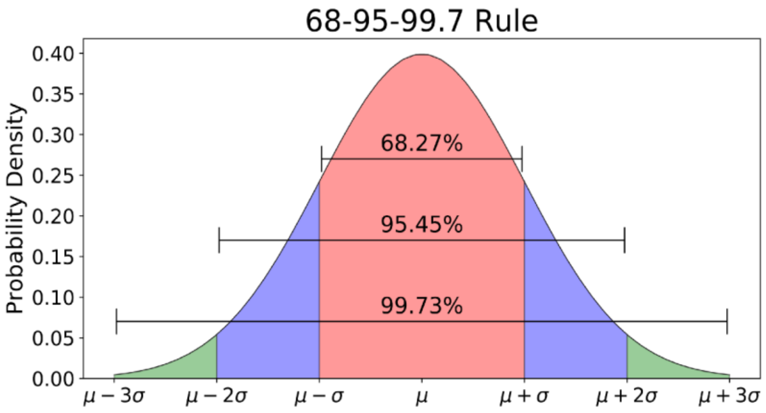
When do we need diagnostic analytics?	**Where someone claims something** and you want to see if it is correct to make a decision.
	Where you notice some possible meaningful relations between data that can assist making decisions
	Where there are anomalies and you do not understand why.

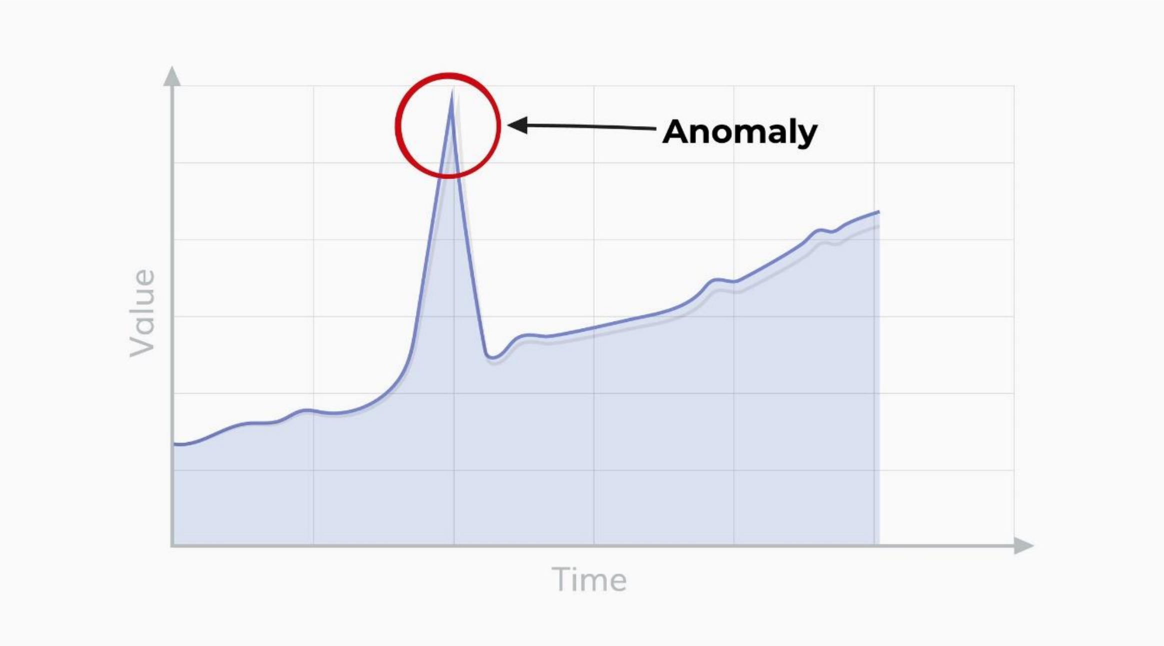
Diagnostic Analytics

When do we need diagnostic analytics?

- **Anomalies**

  - ❑ Where there are **sudden changes in data, their patterns or trends**
  - ❑ Where some **specific data items are beyond what you expected**



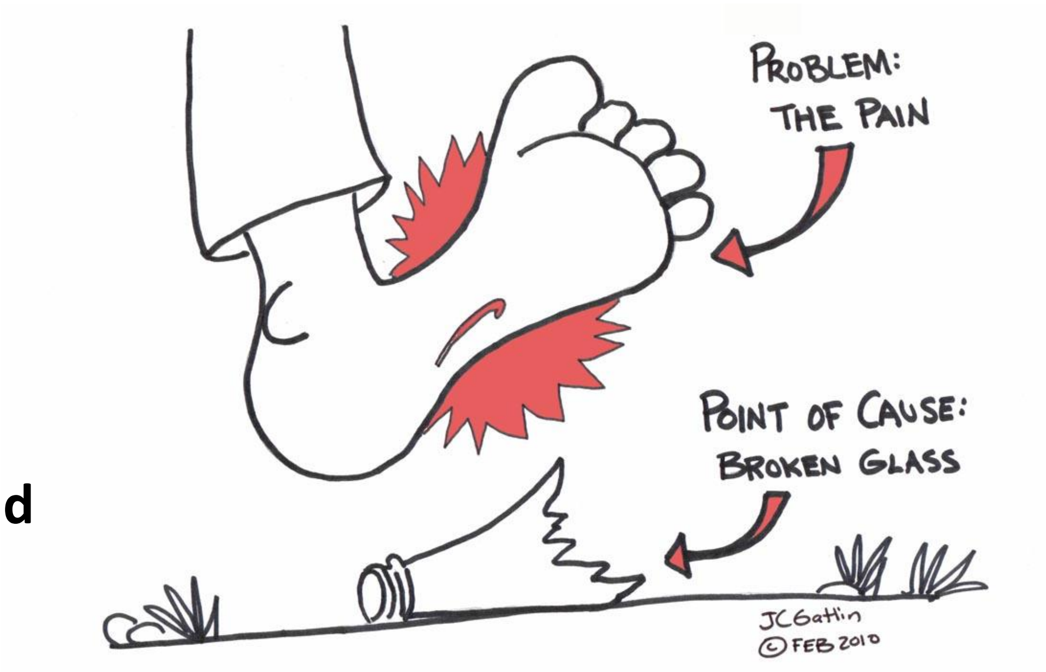


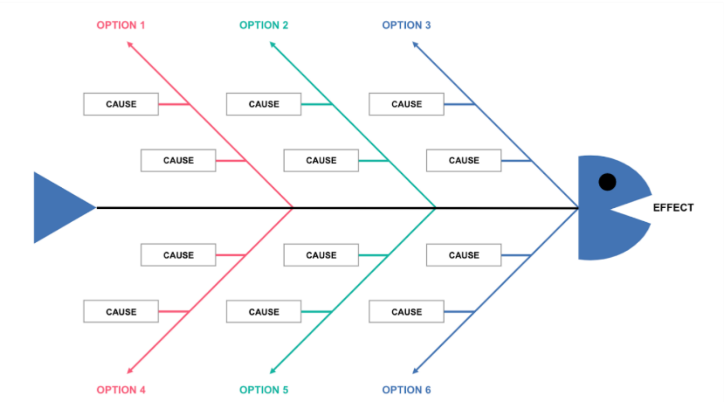
Diagnostic Analytics

Diagnostic analytics

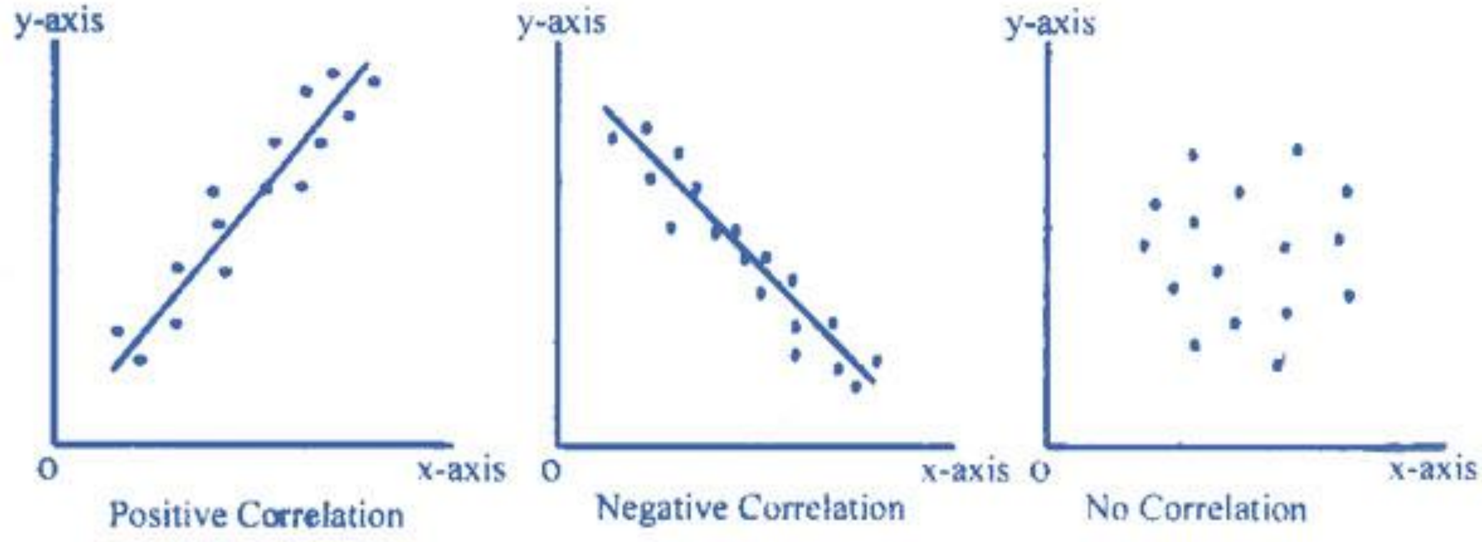
- **What to do**

  - ❑ **Drill into the analytics (discovery) and detect patterns**
  - ❑ **Determine relationships (finding the cause)**





Correlation



A scenario:

- Your organisation has commenced a long-term relationship with a new supplier.
- You notice that the employee's performance is not so satisfactory in this supplier's organisation.
- In a long-term relationship, organisations often try to improve the performance of their suppliers.
- So, you set a training course for the employees of this supplier and now you want to do some further analysis of their performance.

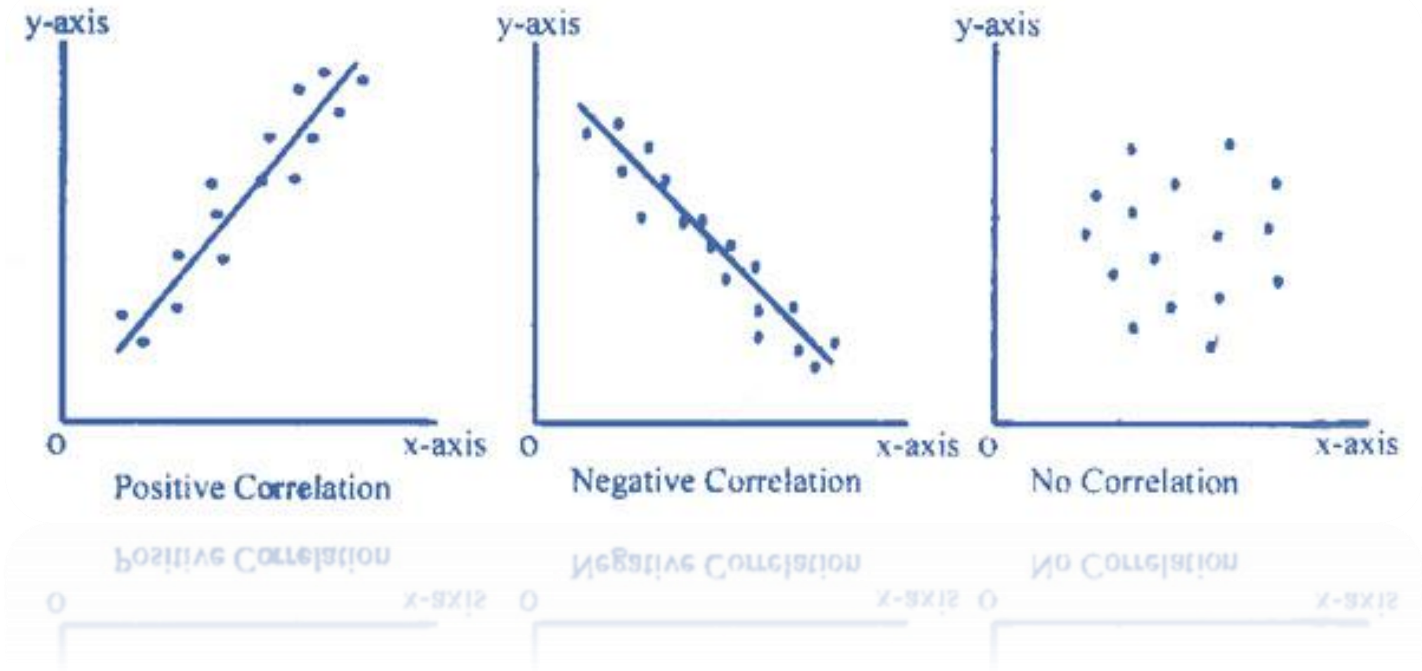
Example : Find out if employees' salary is related to their performance in 2024

Steps:

- Copy employees' salary and performance column into a new sheet.
- Select correlation from data analysis tool.
- Select both columns and press ok
- Note the coefficient of correlation to find the strength of relationship

Correlation

1.Perfect: If the value is near ± 1, then it said to be a perfect correlation: as one variable increases, the other variable tends to also increase (if positive) or decrease (if negative).
2.High degree: If the coefficient value lies between ± 0.50 and ± 1, then it is said to be a strong correlation.
3.Moderate degree: If the value lies between ± 0.30 and ± 0.49, then it is said to be a medium correlation.
4.Low degree: When the value lies below ± .29, then it is said to be a small correlation.



What does this mean?

Can it be used to do prescriptive analytics?

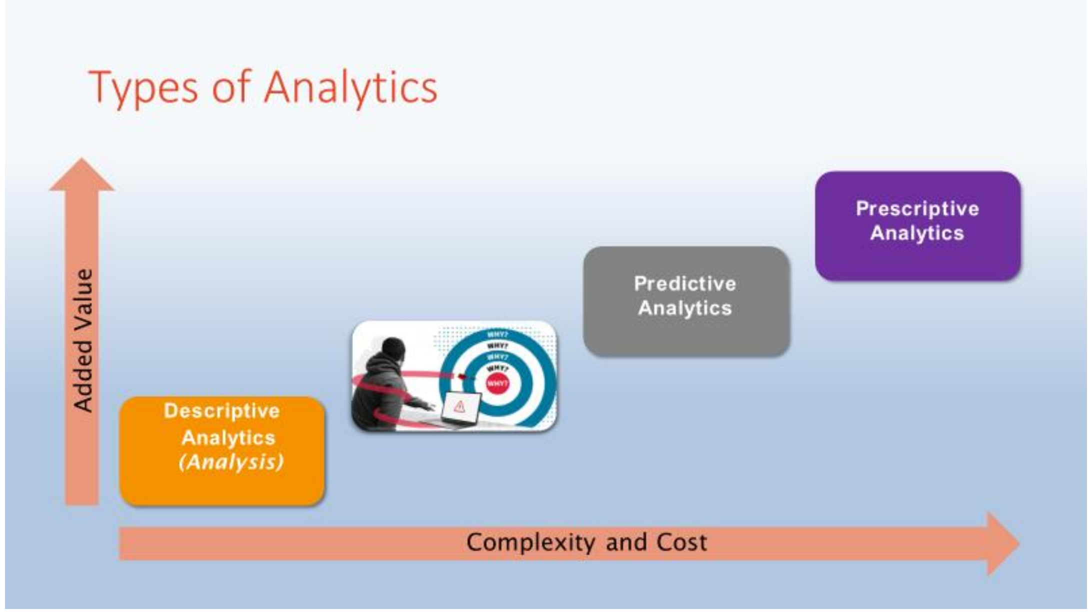

hypothesis
science
research
analysis
experiment
lab
technology
education
test
beaker
method
social
reaction
modern
ideology
industry
survey
chemicals
pattern
equipment
chemistry
resources
expertise
conception
group
data
drug
proof
teach
evidence
professor
innovate
medicine
biology
examine
study
therapy
observation
platform
understanding
pharmaceutical
health
analyzing
liquid
secure
future
information
conclusion
fluid
forensics
chemist
abstract
laboratory
sample
medical
innovation
patience
model
methodology
scientific
syndrome
philosophy
microscope
tube
develop
discovery
design
development
biotechnology
testing
knowledge
measure
learn
web
solution
reflection
healthcare
measurement
work
concept
theory
pharmacy
analyze

HYPOTHESIS?

Hypothesis Testing is to see whether a claim is reliable

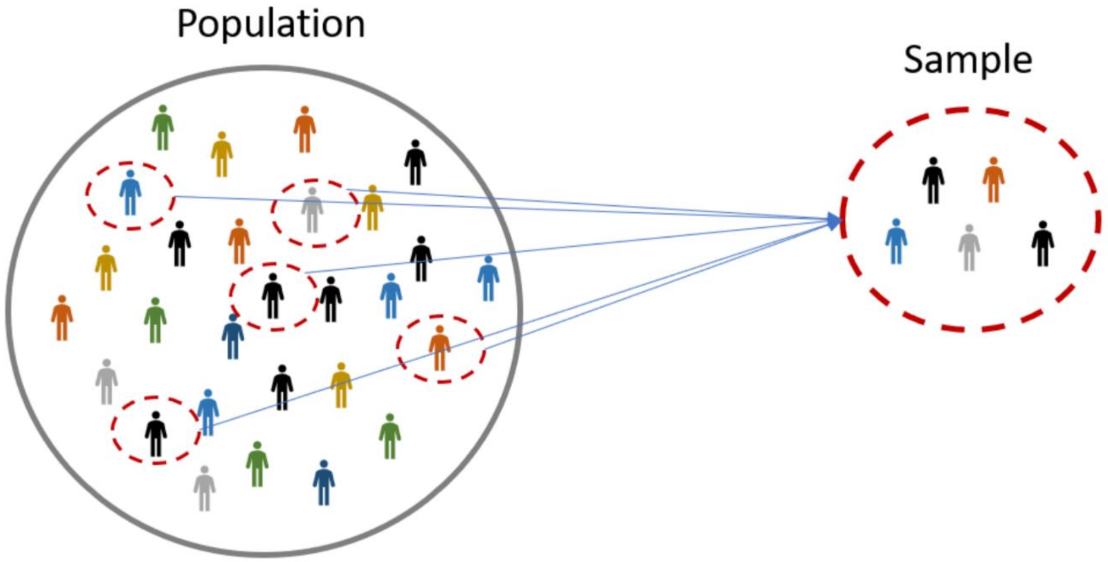
This will not only be useful in diagnostic analytics, but also if you are doing **a research study** and want to see if your claims are reliable.

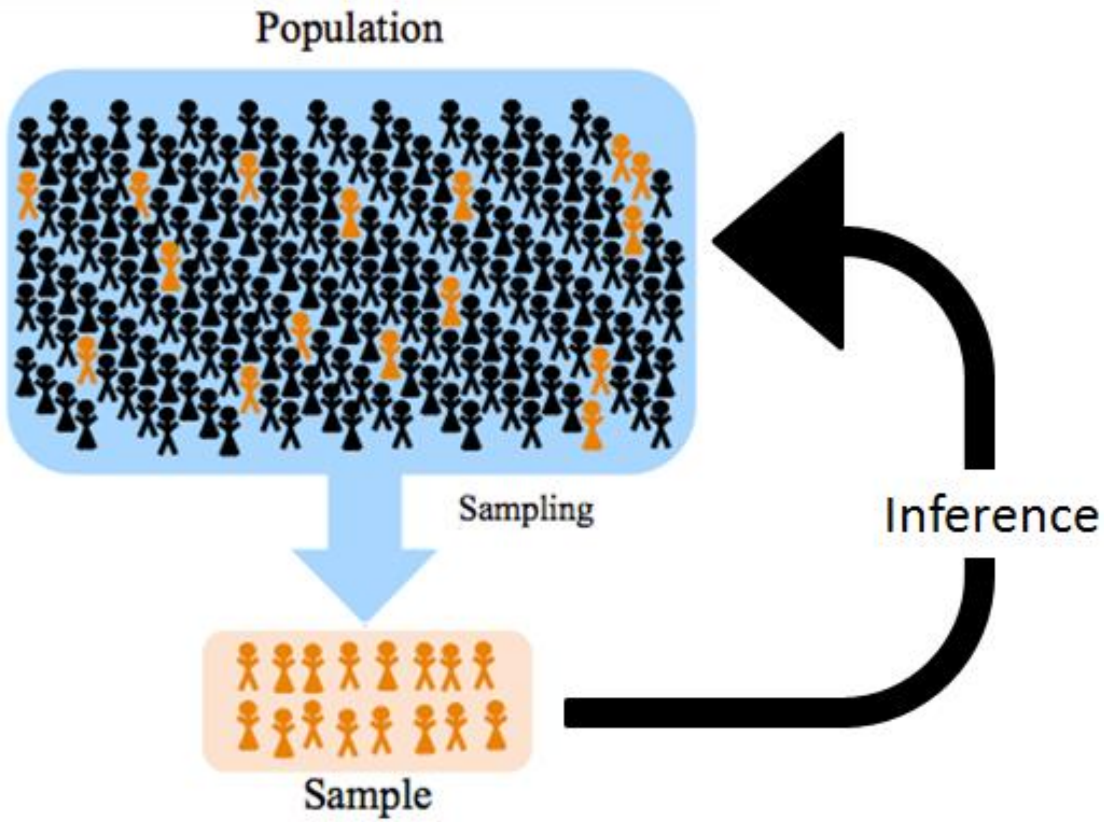
This will also be quite useful for **those who are planning to transition to pursue a PhD** degree



Often, we only have access to a sample of all data, so if we find some evidence, can we generalise that to the entire population?

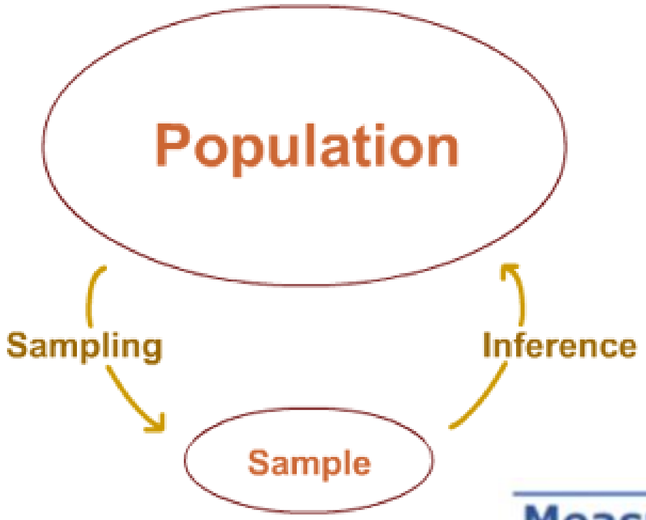
Sample vs. Population

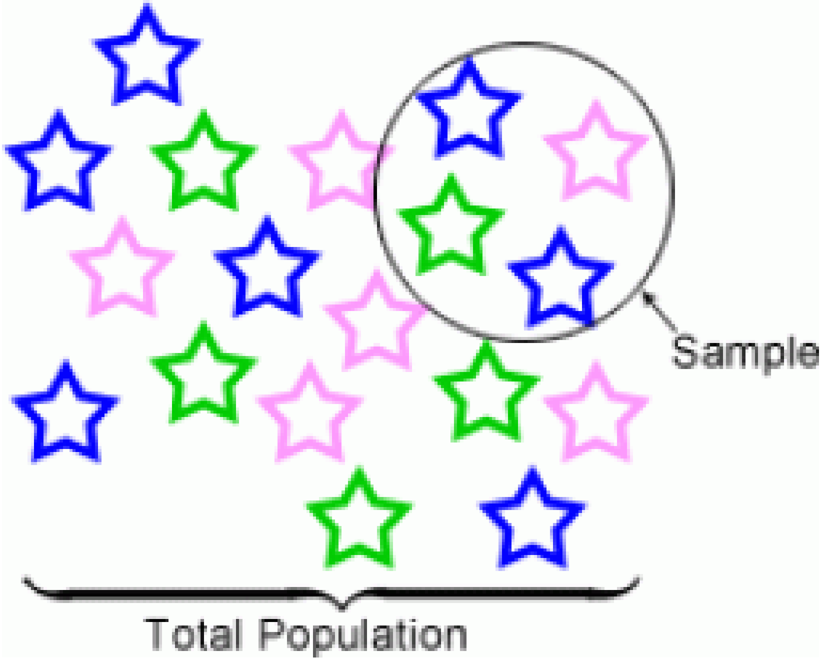




Sample vs. Population

Statistic vs Parameter



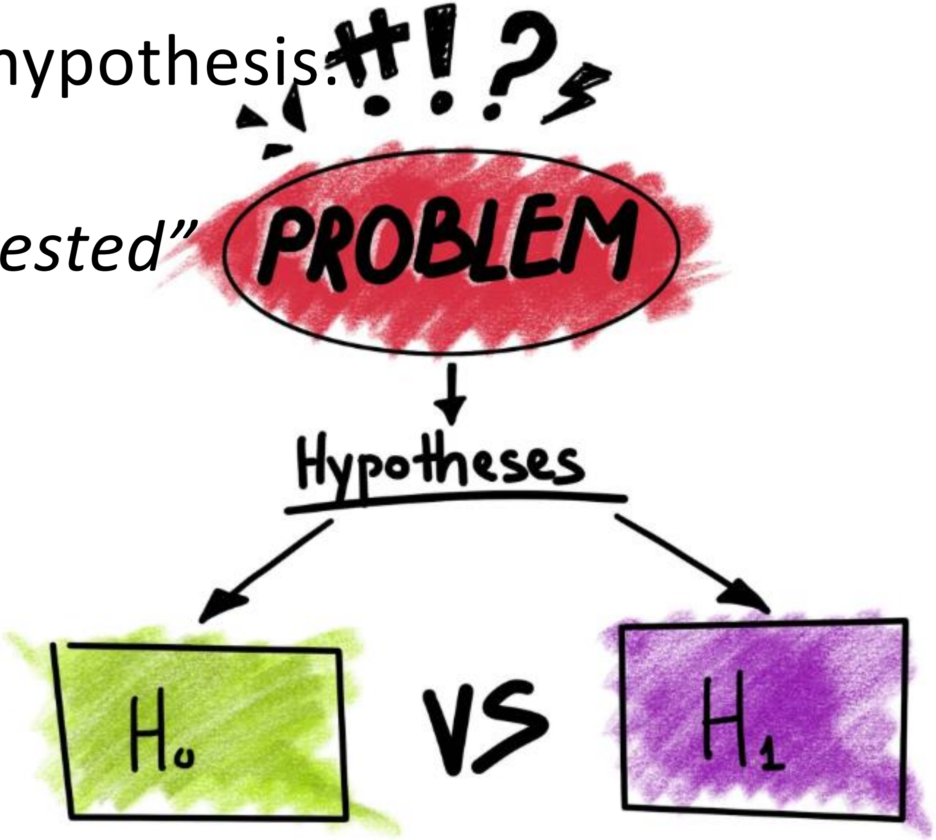


Measurement	**Sample** *Statistic*	**Population** *Parameter*
Size	n	N
Mean	$\bar{X}$	μ
Variance	s^2	σ^2
Standard Deviation	s	σ

Hypothesis Testing

Informal, but easy to digest definition of a hypothesis:

"A hypothesis is an idea/claim that can be tested"



Steps to test a hypothesis:

- Formulate a hypothesis
- Find the right test
- Run the test and make a decision

Someone says the average salary of government employees in Australia is smaller than $140,000 (claim). You have a sample saying it is $139,000. Can you reject this claim?

Hypotheses: H0 and H1 (or Ha)

Our alternative Hypothesis (H1 or Ha):

is the claim that we like. It is the one that is pleasing, basically that is our research claim.

For H1 :

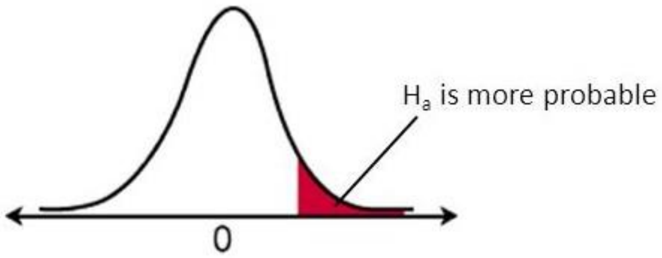
Allowed Signs: “ ≠ “ or “>“ or “<“

Null Hypothesis (H_0):

is the one that we want to reject (if we have enough evidence, meaning that P value is less that 0.05).

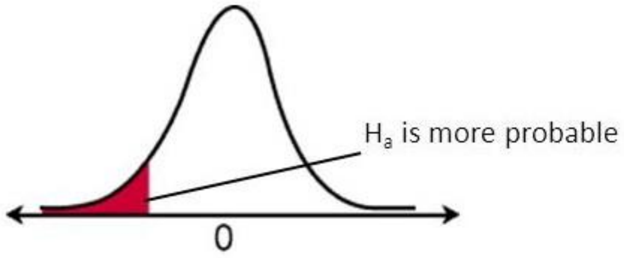
For H_0 :
Allowed Signs: “=“ or “>=“ or “<=“

One tailed vs. two tailed



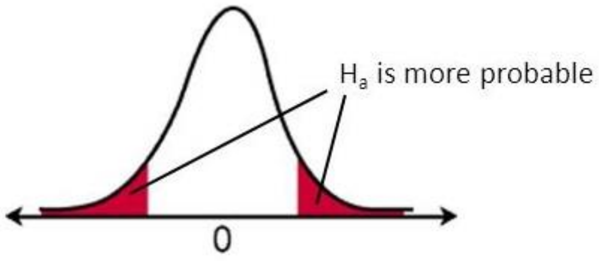
Right-tail test

H_a: $\mu >$ value



Left-tail test

H_a: $\mu <$ value

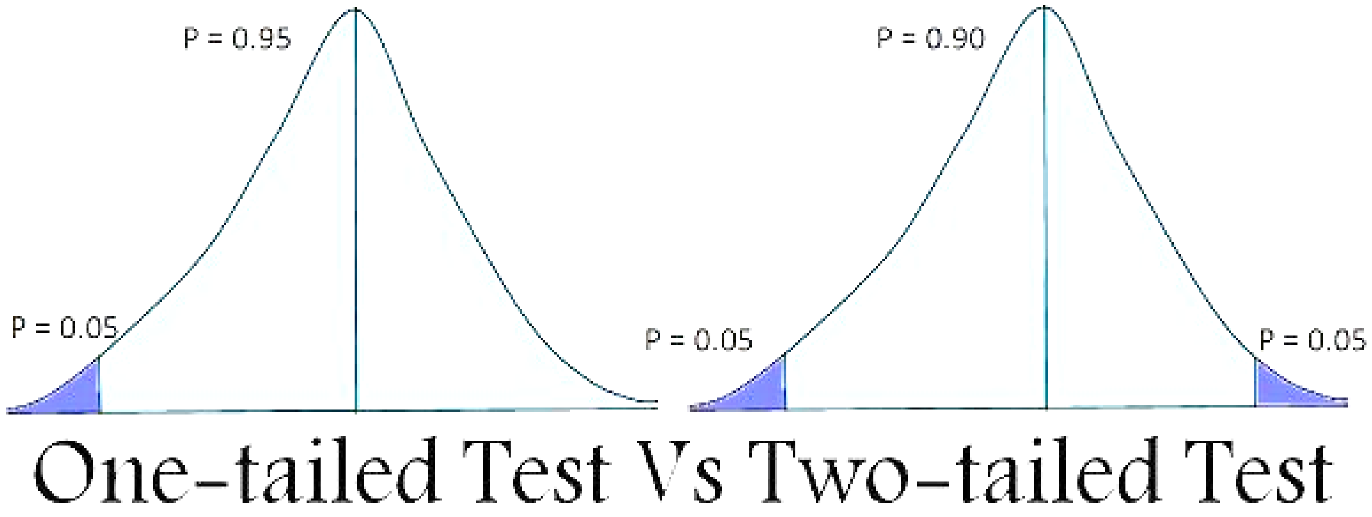


Two-tail test

H_a: $\mu \neq$ value

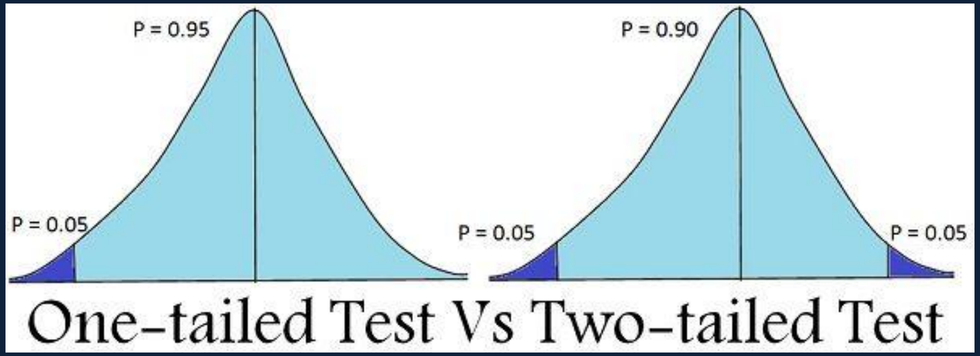
Diagnostic Analytics

Null Hypothesis, the one that we want to reject, is the opposite of H1



H_0 : Female salaries <= Male salary

H_1 : Female salaries > Male salaries

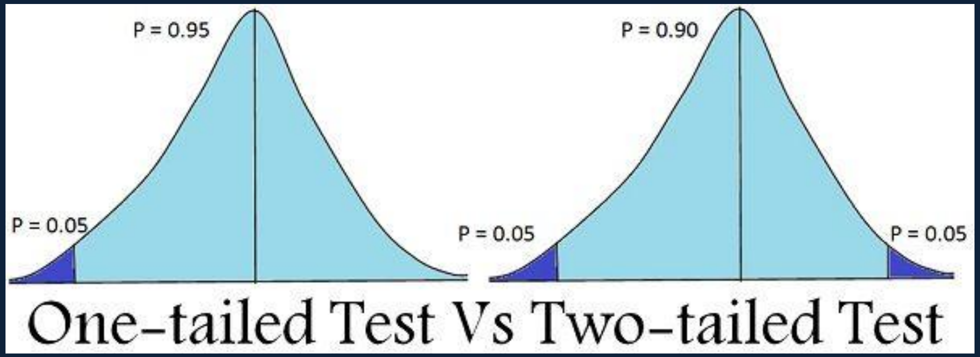


Diagnostic Analytics

We do not have access to data of all government job data.
We got a sample, and the sample has the average of:

- 100,000 dollars as the average of female salaries and
- 98,000 as the average of male salaries

Can we reject the idea that female salaries are smaller than male salaries?



All we can do in hypothesis testing is to reject the null hypothesis. To reject a null hypothesis, we compute P-value, and it needs to be smaller than 5% or 0.05.

P-value is the chance of your result being just a coincidence in your specific sample. This value needs to be too small (smaller than 0.05). Otherwise, you fail to reject H0.

P-value of 0.05 means if we get 100 samples, the results will be supportive of rejecting the null hypothesis in 95 samples.

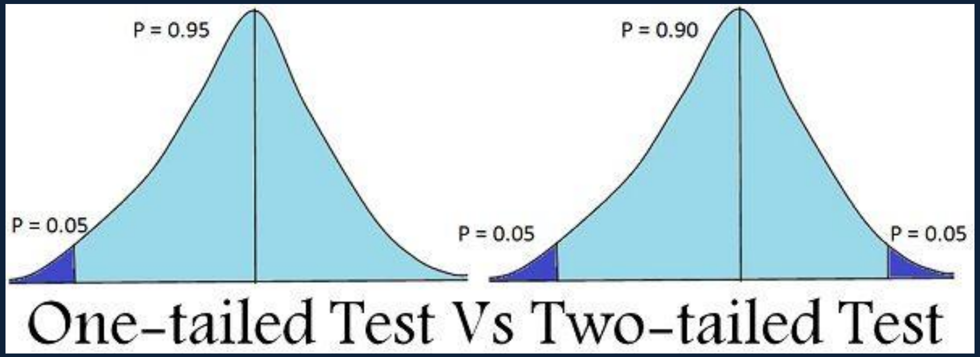
Failing to reject does not mean that the null hypothesis is correct (or wrong). It means that **even though you have evidence against the H0, it is not strong enough to reject H0**. So, you cannot generalise your claim. It might be just a coincidence!

Steps to reject a null hypothesis

1. **Check whether data is normal** (to see which test to use) (Excel does this: Data/Data Analytics/Descriptive Statistics)

2. **Do the test** (e.g. Data/Data Analysis/t-Test), and see if you find evidence (if you find no evidence, you fail to reject the null hypothesis – no further action)

3. **If you find evidence, you need to see if whether that evidence is strong enough** to reject the null hypothesis (P-value needs to be smaller than 0.05 – P-value comes with your test results)



We do not have access to data of all government job data. We got a sample, and the sample has the average of:

- 100,000 dollars as the average of female salaries and
- 98,000 as the average of male salaries

Is this evidence strong enough to reject the null hypothesis?

Rejecting the null hypothesis in this case means that females are paid more than males. You can propose to the CEO that this can potentially be a reason for male low job satisfaction rate, and they might want to investigate further or do something about it.

H_0 : Female salaries <= Male salary

H_1 : Female salaries > Male salaries

Examples of Parametric and non-Parametric Tests

	Parametric	Non-parametric	Example
Comparing two related samples	t-test	Wilcoxon signed rank test	Performance of a group before & after
Comparing two unrelated samples vs. a variable	t-test	Mann-Whitney U-test	Gender (M/F) vs. Job satisfaction level (Likert Scaled)
Comparing three or more related samples with one variable	ANOVA	Friedman test	Role at the company vs. Salary (low, medium, high)
Comparing three or more samples with unrelated variables	ANOVA	Kruskal-Wallies H-test	Assigned projects vs. Job satisfaction level (Likert Scale)
Comparing unrelated categories	None	Chi-square test	Company award (e.g. yes, no) vs. Performance level (high, low)
Comparing two independent ranks	Pearson product-moment correlation	Spearman rank-order test	Salary vs. Job satisfaction level (e.g. Likert Scale)

Normal Distribution

Other Distributions

Examples of Parametric and non-Parametric Tests

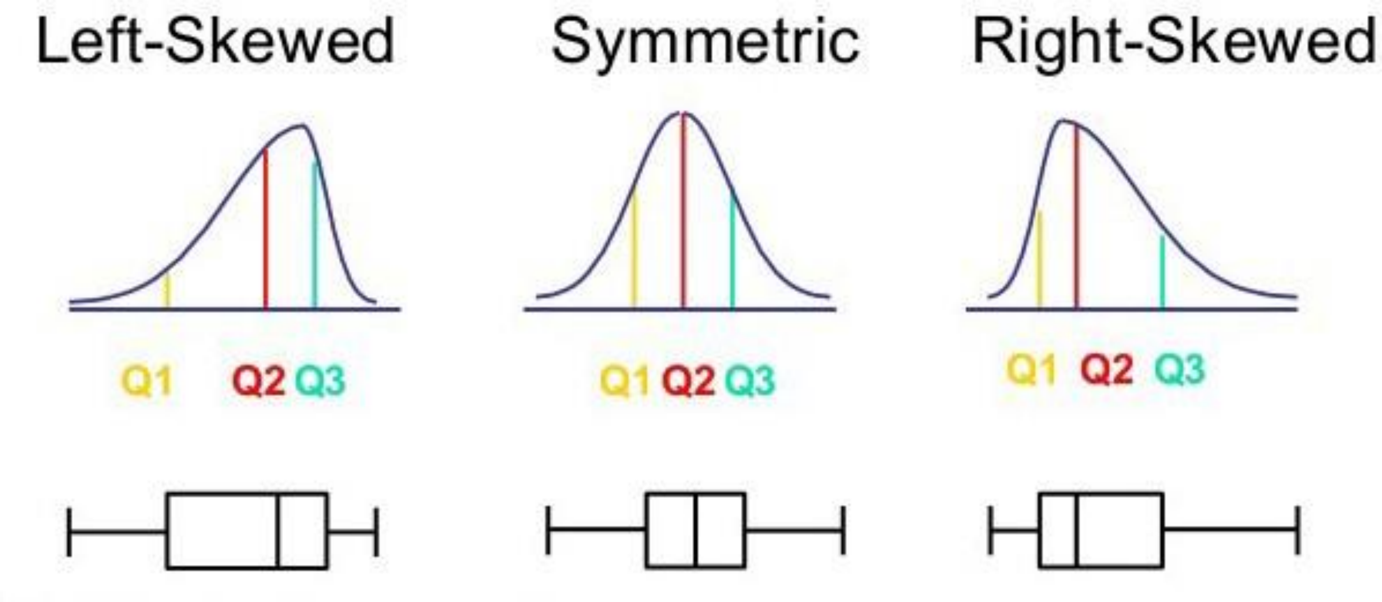
	Parametric	**Non-parametric**	**Example**
Comparing two related samples	t-test	Wilcoxon signed rank test	Performance of a group before & after
Comparing two unrelated samples vs. a variable	t-test	Mann-Whitney U-test	Gender (M/F) vs. Job satisfaction level (Likert Scaled)
Comparing three or more related samples with one variable	ANOVA	Friedman test	Role at the company vs. Salary (low, medium, high)
Comparing three or more samples with unrelated variables	ANOVA	Kruskal-Wallies H-test	Assigned projects vs. Job satisfaction level (Likert Scale)
Comparing unrelated categories	None	Chi-square test	Company award (e.g. yes, no) vs. Performance level (high, low)
Comparing two independent ranks	Pearson product-moment correlation	Spearman rank-order test	Salary vs. Job satisfaction level (e.g. Likert Scale)

How to find out if your data is normal?

Skewness

Kurtosis

Normal Distribution Skewness



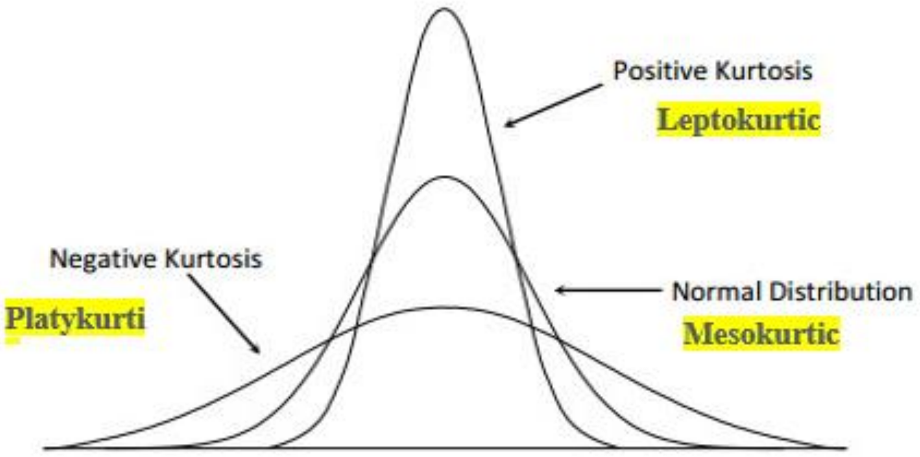
- Good if it is within -1 and 1
- Acceptable between -2 and 2

If two far from zero, we use nonparametric tests

Normal Distribution Kurtosis

- Good if it is within -2 to 2
- Acceptable within -3 and 3

If it is beyond these, we use non-parametric tests



Examples of Parametric and non-Parametric Tests

	Parametric	Non-parametric	Example
Comparing two related samples	t-test	Wilcoxon signed rank test	Performance of a group before & after
Comparing two unrelated samples vs. a variable	**t-test**	Mann-Whitney U-test	**Gender (M/F) vs. Job satisfaction level (Likert Scaled)**
Comparing three or more related samples with one variable	ANOVA	Friedman test	Role at the company vs. Salary (low, medium, high)
Comparing three or more samples with unrelated variables	ANOVA	Kruskal-Wallies H-test	Assigned projects vs. Job satisfaction level (Likert Scale)
Comparing unrelated categories	None	Chi-square test	Company award (e.g. yes, no) vs. Performance level (high, low)
Comparing two independent ranks	Pearson product-moment correlation	Spearman rank-order test	Salary vs. Job satisfaction level (e.g. Likert Scale)

t-Test: Two-Sample Assuming Unequal Variances

	Salary	*Salary*
Mean	104420	110347.8261
Variance	530534285.7	604339226
Observations	50	92
Hypothesized Mean	0	
df	106	
t Stat	-1.430174571	
P(T<=t) one-tail	0.077803996	
t Critical one-tail	1.659356034	
P(T<=t) two-tail	0.155607992	
t Critical two-tail	1.982597262	

Claim: Female salaries are smaller than Male Salaries

- Can you formulate this on your Excel sheet:

H0: Average of Female Salaries >= Average of male Salaries
H1: Average of Female Salaries < Average of male Salaries

t-Test: Two-Sample Assuming Unequal Variances		
	Female Salary	*Male Salary*
Mean	104420	110347.8261
Variance	530534285.7	604339226
Observations	50	92
Hypothesized Mean	0	
df	106	
t Stat	-1.430174571	
P(T<=t) one-tail	0.077803996	
t Critical one-tail	1.659356034	
P(T<=t) two-tail	0.155607992	
t Critical two-tail	1.982597262	

Second: Create a dummy column next to the 'working hours'

Because, Excel does not do t-Test on one sample

CL	CM
Working Hours	Dummy data
35	0
38	0
42	0
52	0
56	0
60	
35	
40	
50	
42	
35	
38	

ANOVA

t-Test was for two sets/groups/columns of data. An **ANOVA** test is a way to find out difference between more than two testing groups. In other words, they help you to figure out if you need to reject the null hypothesis.

Examples of testing more than two groups:

A group of psychiatric patients are trying three different therapies: counselling, medication and biofeedback. You want to see if one therapy is better than the others.

A manufacturer has three different processes to make light bulbs. They want to know if one process is better than the other.

Students from different colleges take the same exam. You want to see if one college outperforms the other.

CEO changed and assigned new managers to different production lines

Claim: The employee satisfactory levels of some lines have dropped.

Do we have evidence? Is it strong enough to reject the null hypothesis?

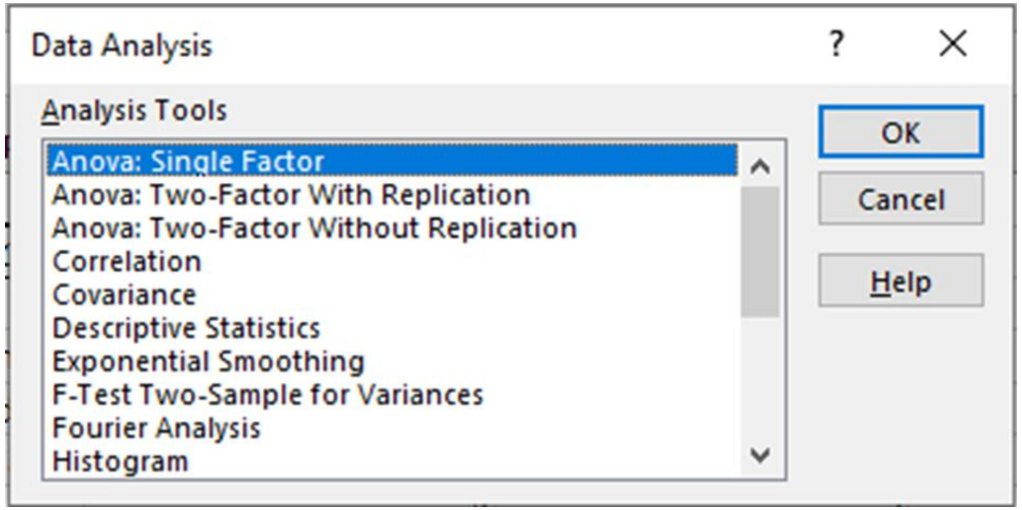
Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Line 1	23	76	3.304348	0.948617
Line 2	14	41	2.928571	0.532967
Line 3	45	177	3.933333	1.109091
Line 4	49	140	2.857143	1.416667
Line 5	12	44	3.666667	1.69697

ANOVA

Source of Varia	SS	df	MS	F	P-value	F crit
Between (	30.94499	4	7.736247	6.539083	7.55E-05	2.437265
Within Gr	163.2648	138	1.183078			
Total	194.2098	142				



Why did we select single factor!?

1. Why do certain products have consistently low ending inventory?
   You can check 'Ending Inventory' and check if high sales volume is a factor.
2. Which products frequently experience stock replenishment?
   Analyse the frequency of non-zero 'Received Inventory' entries for each product.
3. Are there any patterns of inventory shortage leading to stockouts?
   Correlate low 'Ending Inventory' levels with instances of stockouts (derived from Sales Data).
4. What causes fluctuations in inventory levels for specific stores?
   Compare inventory levels with sales data and external factors like promotions or seasonal trends.
5. Investigate the -relationship between inventory levels and discounts offered.
   Analyze if higher 'Discount Applied' (from Sales Data) correlates with lower 'Ending Inventory'.

Examples of Diagnostic Analytics in Inventory Data

1. Why do some suppliers have lower quality ratings?
   Investigate factors contributing to lower quality scores among suppliers.
2. Are there relations between cost and quality rating.
   Determine if higher costs correlate with better quality.
3. Why do certain stores have higher waste levels?
   Correlate waste levels with other factors like delivery delays or transportation costs.
4. Determine the cause of any significant spikes in transportation costs.
   Examine the months with high transportation costs and cross-reference with external events.
5. What factors contribute to the variance in risk factor scores?
   Explore the variation in risk factors to understand underlying causes.
6. Why are sale levels fluctuating; can it be related to discounts?
   Assess the impact of discounts on sales volumes.
   Compare 'Units Sold' with 'Discount Applied' to see correlation between discounts and sales.
7. Analyze if higher-priced products tend to sell less.
   Correlate 'Sale Price' with 'Units Sold' across products.
8. Claim: sales volumes are affected by store location.
   Compare 'Units Sold' across 'Store Name' and analyze differences.

Examples of Diagnostic Analytics in Suppliers And Sales Data