Cátedra Nissan

-PROTHIUS-

Modelos de Organización en Obra y Empresa: Metodología para la selección y gestión de proyectos. Fase III.

Rubén Rami

D-18/2011

Departamento de Organización de Empresas

Universidad Politécnica de Cataluña

Publica:

Universitat Politècnica de Catalunya
www.upc.edu

Edita:

Cátedra Nissan
www.nissanchair.com
director@nissanchair.com

Fase 3. Análisis Causa Raíz

MOOE

CATEDRA NISSAN

DOE Departament d'Organització d'Empreses

Objetivo sesión

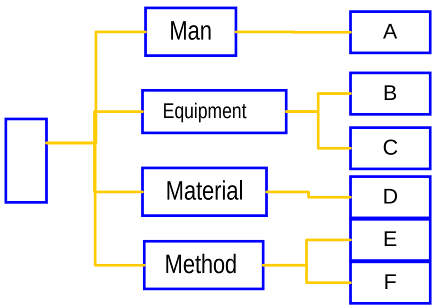
- Definir objetivo y expectativas de la fase 3: Análisis Causa Raíz
- Conocer herramientas más usuales en la fase 3
- Práctica con las herramientas

Fase 3: Análisis Causa Principal

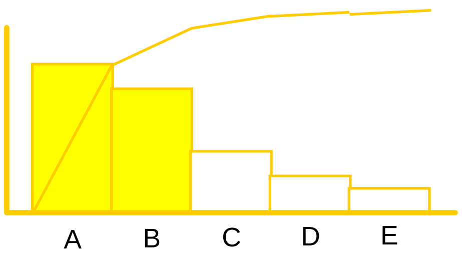
Objetivo	Identificar las causas principales del problema, y clarificar su relación con el 1er métrico

Expectativas de esta fase:

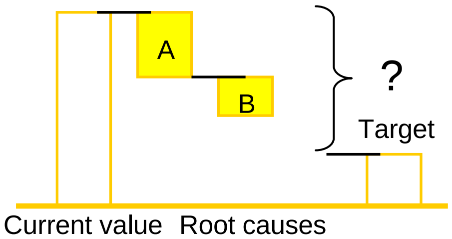
- Identificación de las causas



- Causa Raíz o Principal



- Impacto en el 1er Métrico



Identificar la Causa Raíz

Ejemplo

EFECTO → CAUSAS 1 → CAUSAS 2

NO SE CONSIGUEN OBJETIVOS CALIDAD

- DISEÑO DEFECTUOSO
  1. Falta de validación física
  2. Modificaciones de última hora
- VALIDACIÓN DEFICIENTE
  1. Falta de formación herramientas FV
  2. Falta prestaciones herramientas FV
- ADIESTRAMIENTO DEFICIENTE
  1. Falta de validación física
  2. Planificación deficiente

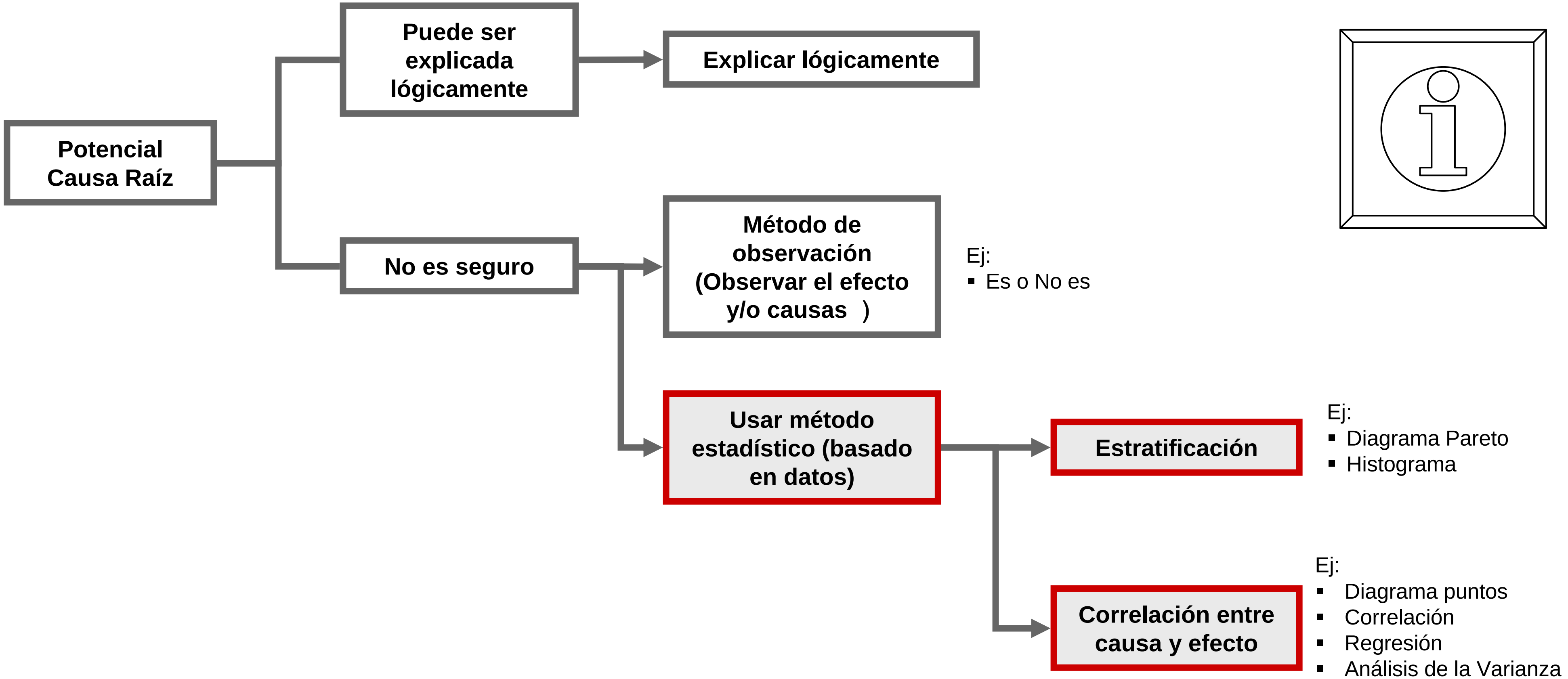
Establecer la hipótesis desplegando los efectos hacia las causas principales que los provocan

Obtención de datos para validar la hipótesis

#	Método	Ejemplos
1	Datos obtenidos diariamente	Rutina de gestión de datos
2	Datos obtenidos en el taller	Asignando un observador para evaluar un determinado resultado
3	Entrevistas con clientes o conducir un cuestionario	Entrevistas sobre productos o servicios
4	Realización de una prueba, test o simulación	Diseño de experimentos

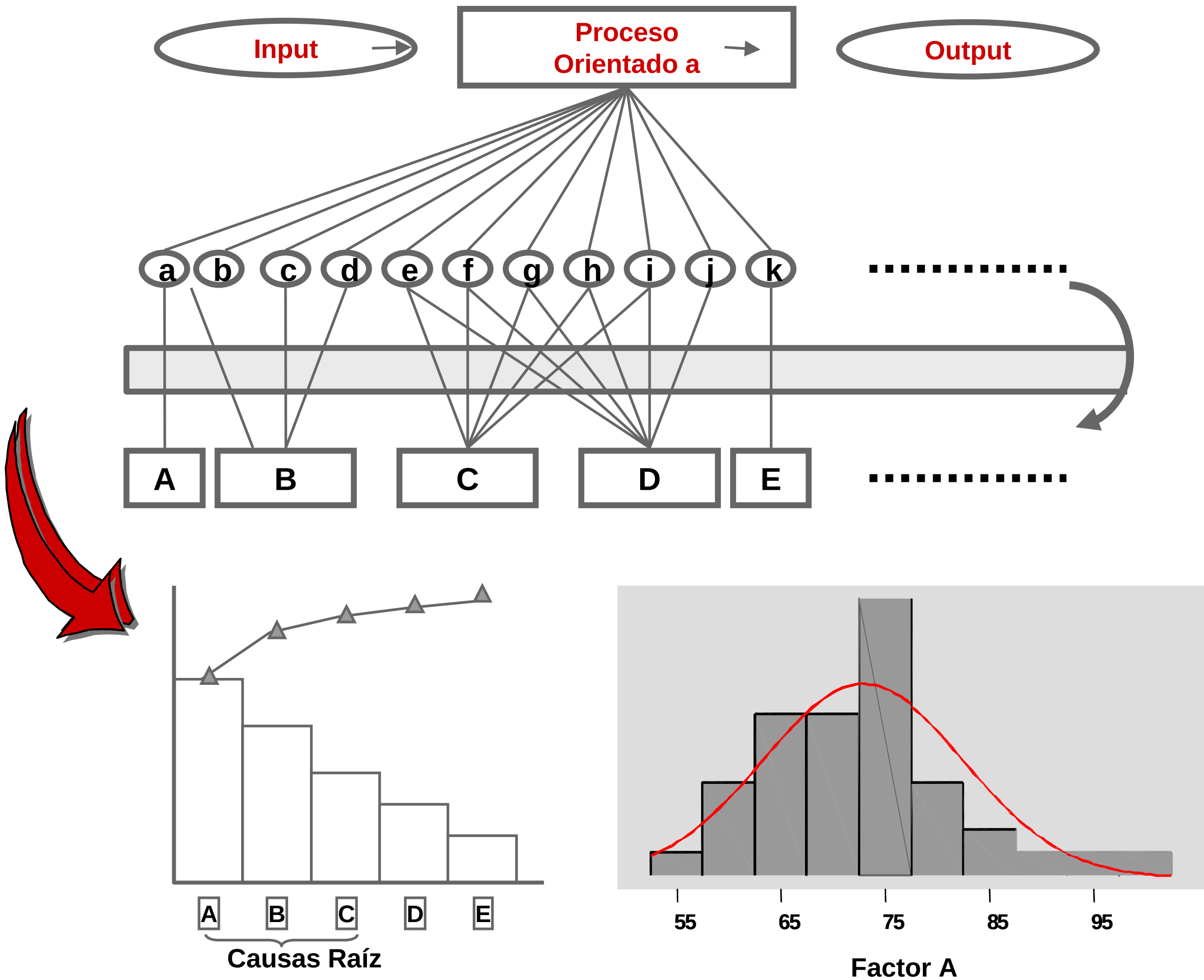
Es crucial obtener datos para validar tanto las causas como los efectos

Chequear la hipótesis



Determinar una aproximación cuantificable para especificar la causa raíz

Determinar los factores que causan el problema para especificar la causa raíz



Fase 2

Fase 3

Identificar todas las causas

Ej:

- **Brainstorming**
- **Diagrama afinidad**
- **Diagrama causa-efecto**
- **Diagrama interrelación**
- **Árbol lógico**

Especificar causa raíz

Ej análisis cuantificable:

- **Pareto**
- **Histograma**
- **Regresión**

Cuantificar siempre que sea posible los valores cualitativos

Desarrollar la relación entre la causa y el efecto

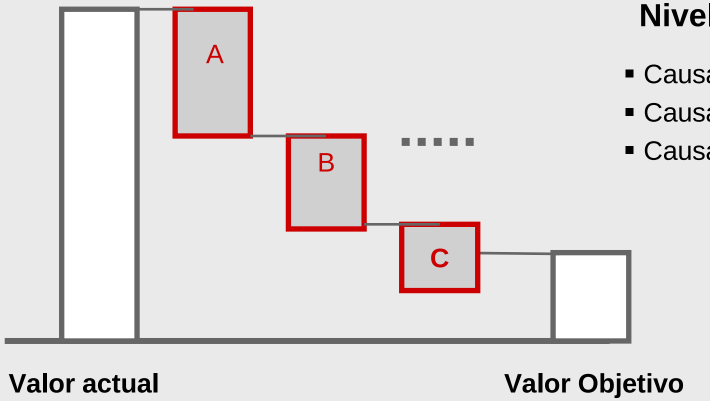
Desarrollar la relación entre el 1er métrico y las causas principales con la intención de estimar el impacto y objetivo del 1er métrico.

$$f(x) = ax_A + bx_B + cx_C + d$$

1st Métrico

Causas principales

Clarificar relación entre 1er métrico y causa raíz



Nivel de mejora

- Causa A: 50%
- Causa B: 30%
- Causa C: 25%

Estimar impacto

Ej:. Análisis cuantificable
- Diagrama pareto
- Gráficos
- Experiencia del equipo

Fase 3. Herramientas

ISHIKAWA

LOGIC TREE

DIAGRAMA INTERRELACION

QFD

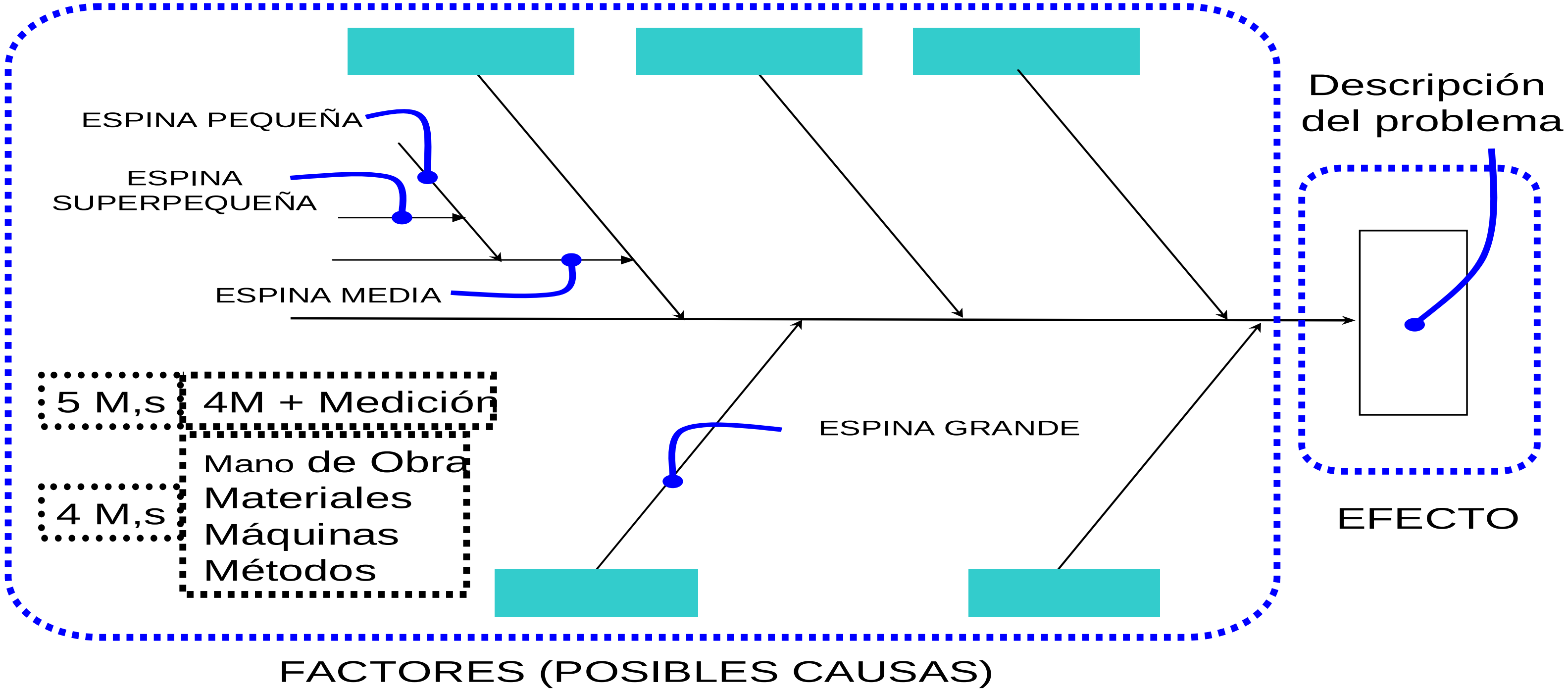
CATEDRA NISSAN
UPC

DOE Departament d'Organització d'Empreses

Ishikawa

DIAGRAMA CAUSA - EFECTO

- Es la representación organizada de la relación entre el EFECTO o problema y los posibles FACTORES que le afectan, estructurado en forma de "raspa de pescado".



Ishikawa

Ejemplo de Diagrama de Causa y Efecto.

Principales reglas:

- Recoger todos los factores posibles.
- Los factores deben expresarse en términos concretos.
- Los factores deben ser posibles causas, no contramedidas.
- Es muy conveniente clasificar los factores en las 4 M de la Producción.
- Las "espinas" de cada rama deben mantenerse paralelas, para facilitar su visualización

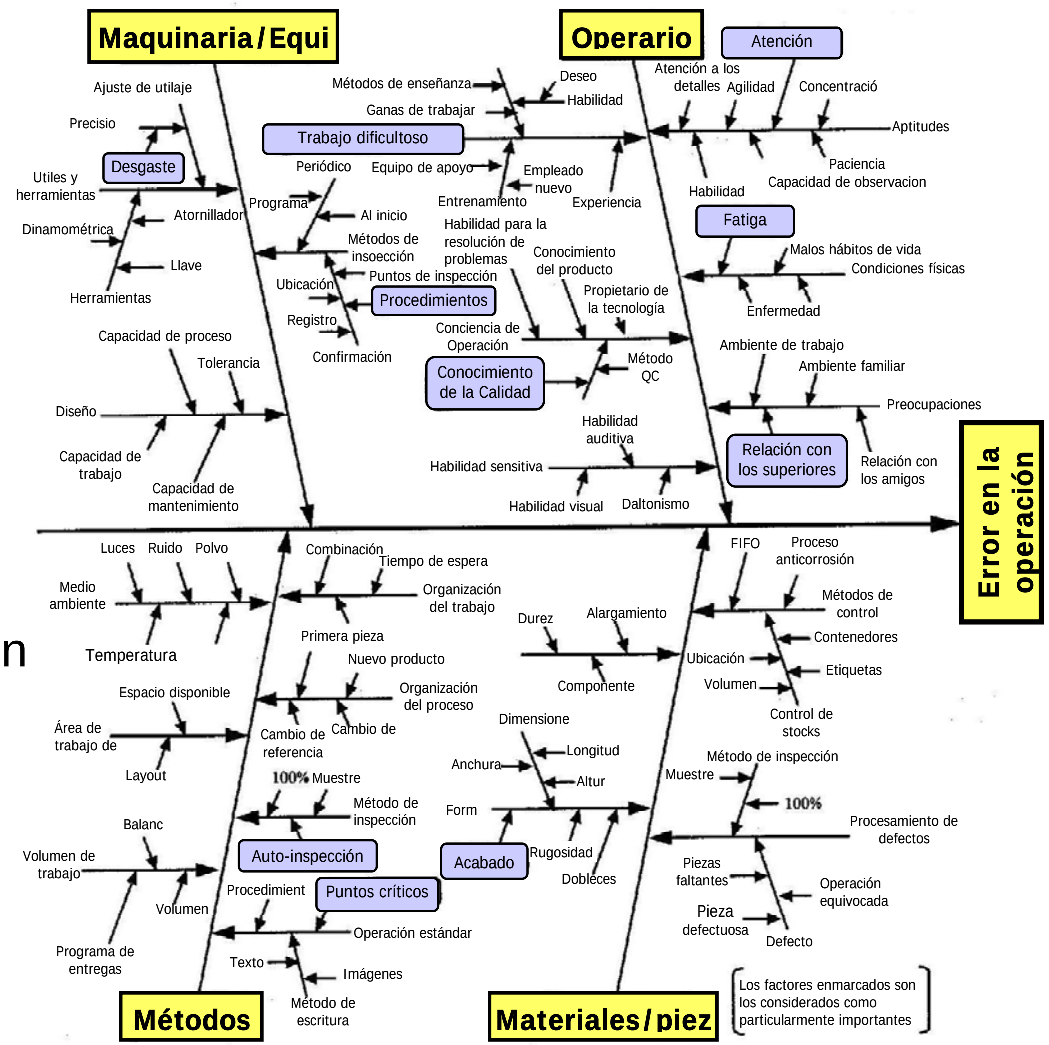
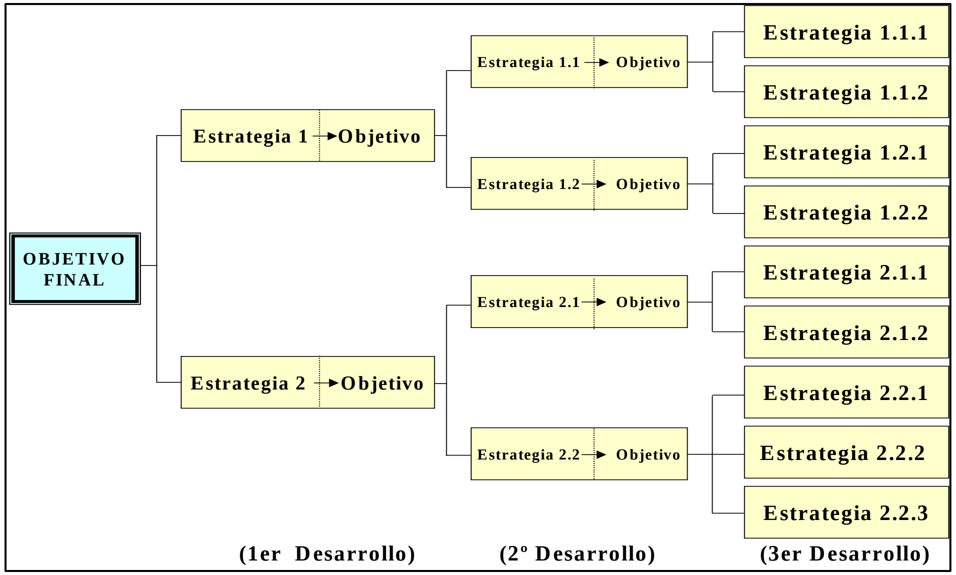


Diagrama de Árbol

Es un diagrama que representa los efectos y las diferentes causas encadenadas que los producen, ramificándose en forma de árbol.

Básicamente sirve para:

- Clarificar las causas con respecto a los efectos.
- Desarrollar estrategias y políticas para cumplir los objetivos.
- Descomponer un objetivo principal en sus elementos básicos.



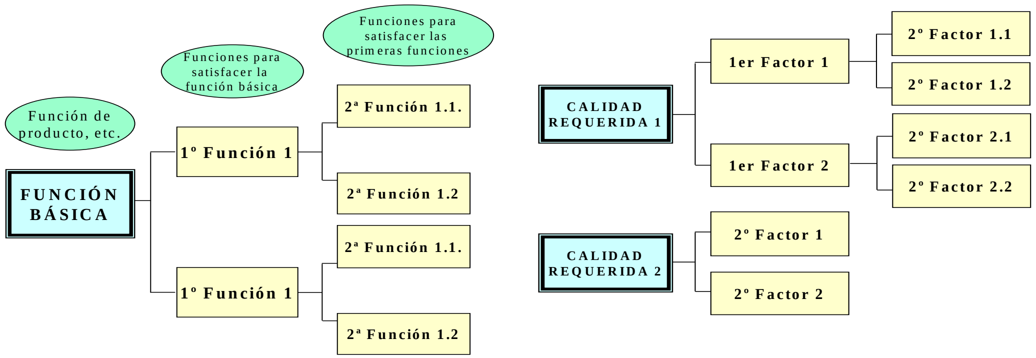


Diagrama de Árbol

Etapas para la elaboración de un diagrama de árbol:

- **PASO 1: Elegir el tema**

Describir el tema de forma clara y concreta, utilizando la pregunta ¿qué hay que hacer para reducir, mejorar, eliminar,…? y escribirlo en una etiqueta roja a la izquierda.

- **PASO 2: Indicar claramente las premisas**

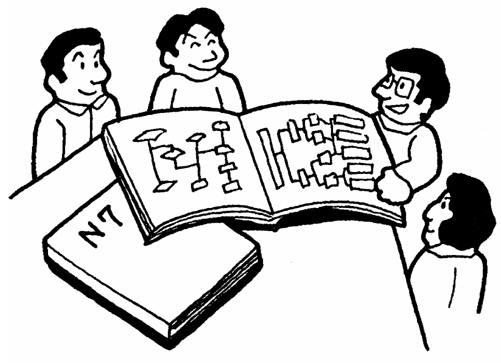
Identificar las premisas y escribirlas en etiquetas.

- **PASO 3: Identificar estrategias del primer nivel**

De todas las estrategias de primer nivel, elegir de 2 a 4 y escribirlas en etiquetas.

- **PASO 4: Colocar las etiquetas en el papel**

Las estrategias de primer nivel se colocan a la derecha del tema, en forma de árbol.



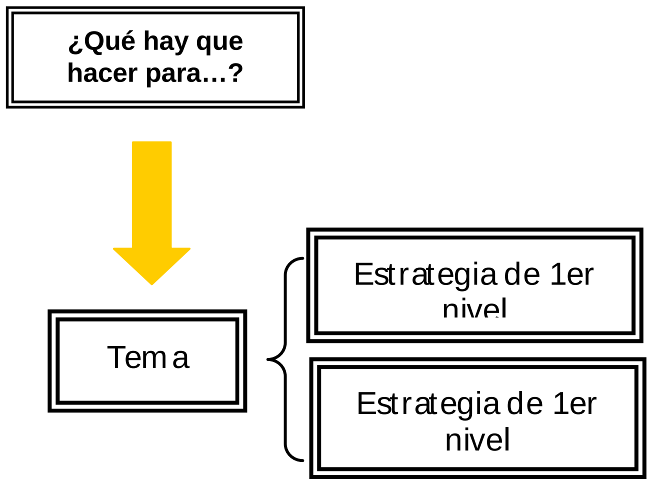
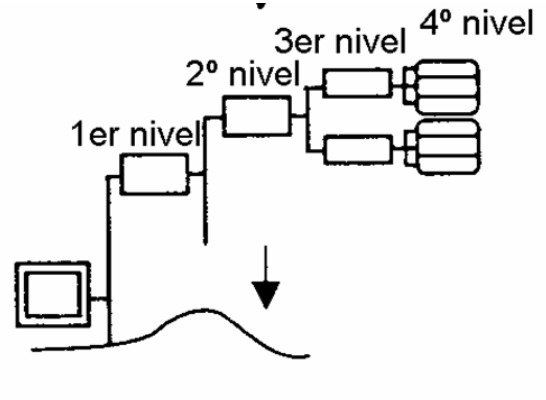


Diagrama de Árbol

Etapas para la elaboración de un diagrama de árbol (continuación):



- **PASO 5: Identificar estrategias de segundo nivel**

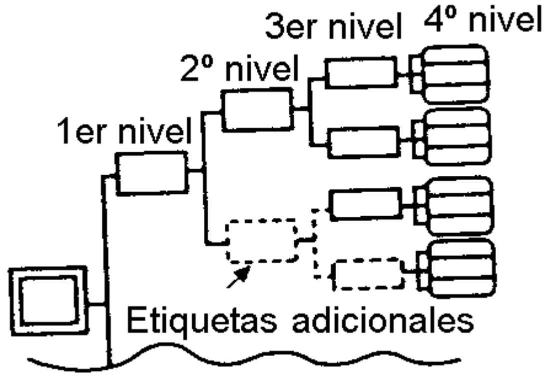
Las estrategias de primer nivel se convierten en objetivos y se identifican las estrategias de segundo nivel.

- **PASO 6: Desglosar todavía más las estrategias**

De la misma forma que en el paso 5

- **PASO 7: Revisar la conexión de las estrategias**

Tener en cuenta todos los niveles hasta llegar al tema

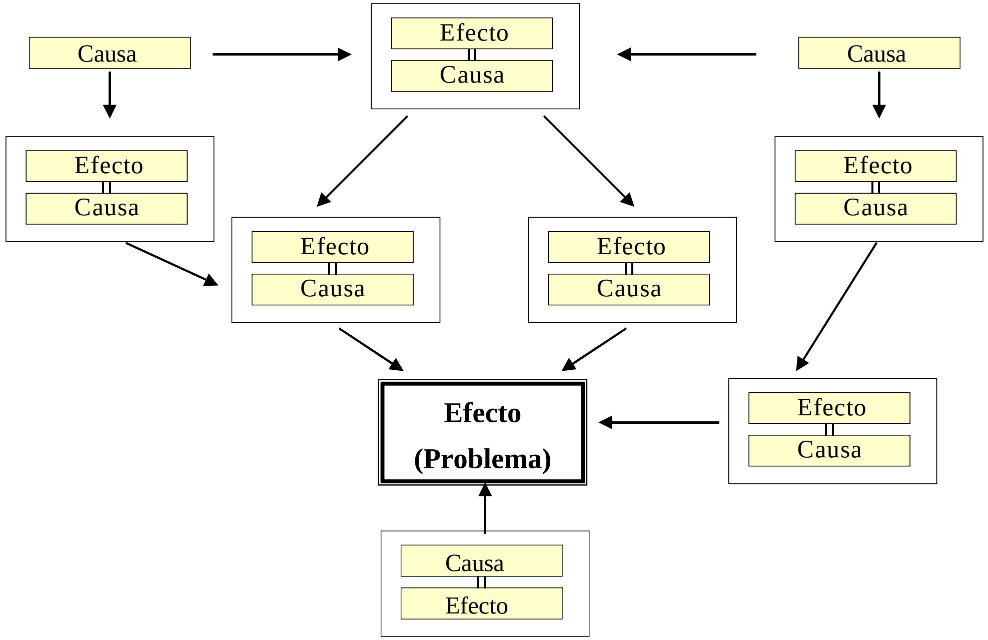


- **PASO 8: Ordenar las etiquetas y añadir estrategias**
- **PASO 9: Anotar datos como fecha, lugar, etc**

Diagrama de Interrelación

Se utiliza para resolver problemas en los que intervienen causas y efectos mezclados, en base a:

- Explorar todos los factores que puedan estar relacionados con el problema y listarlos, indicando las posibles conexiones.
- Confirmar las relaciones de causa – efecto entre los factores
- Existen varios modelos:
- Relación entre causas y efectos.
- Relación entre objetivos y estrategias / acciones para conseguir los objetivos.



Es conveniente aplicar esta herramienta en grupos de 6 a 10 personas.

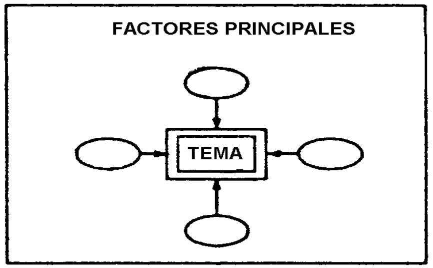
Diagrama de Interrelación

Etapas para la elaboración de un diagrama de relaciones:

- **PASO 1: Elegir el tema (problema)**

Escribir el tema de forma clara y concreta, utilizando la pregunta ¿por qué…? Y colocarlo en el centro de un papel grande en la pared.

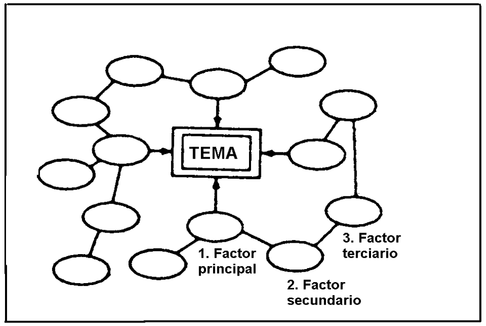
- **PASO 2: Confirmar el tema**



Confirmar entre todo el grupo antecedentes, premisas, etc.

- **PASO 3: Identificar los factores principales que causan el tema (problema)**

Identificar los factores principales y colocarlos dentro de un óvalo alrededor del tema. Colocar flechas de causalidad de forma provisional (con lápiz).



- **PASO 4: Desglosar las causas**

Identificar las posibles causas de los factores mediante al análisis "¿por qué?, ¿por qué?", conectándolos con flechas provisionales.

Diagrama de Interrelación

Etapas para la elaboración de un diagrama de relaciones (continuación):

- **PASO 5: Revisar las relaciones causales**

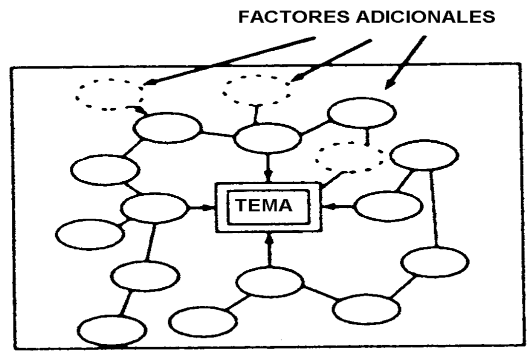
Una vez establecidas las causas, revisarlas en conjunto y confirmar las relaciones provisionales.

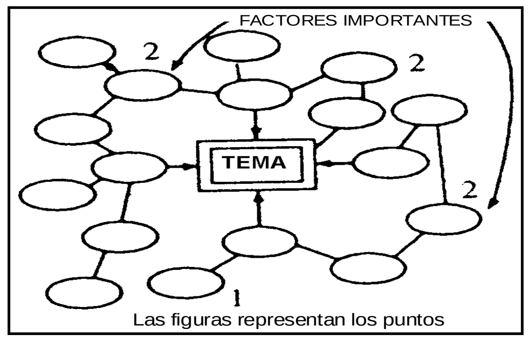
- **PASO 6: elegir los factores más influyentes**

Utilizar consenso o sistemas de votación, por ejemplo, cada asistente puntúa 5 de las causas con 3, 2 y 1 puntos y se priorizan.

- **PASO 7: Escribir la conclusión en la parte superior derecha**

La conclusión debe ser consensuada ente el equipo y se debe escribir de forma concisa. Indicar también los miembros del equipo, la fecha, el lugar, etc.





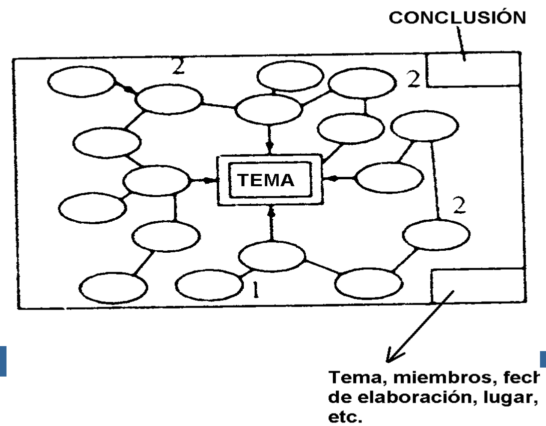


Diagrama de Interrelación

Herramientas de soporte para realizar el diagrama de relaciones:

BRAINSTORMING

Para identificar los factores. Respetar lar reglas del Brainstorming:

- No criticar
- Aceptar positivamente las ideas libremente expresadas
- Generar el máximo número de ideas
- Comprender las ideas de los demás (escuchar)

SISTEMA DE LOS 5 ¿POR QUÉ?

Para relación entre causas y efectos:

Efecto → Factor 1 → Factor 2 → Factor 3 → Factor 4 → Causa verdadera

¿Por qué? ¿Por qué? ¿Por qué? ¿Por qué? ¿Por qué?

QFD (Quality Function Deployment)

- El método QFD ha sido creado con el fin de traducir y dar respuesta a los deseos o exigencias de los clientes, en parámetros apropiados que sean entendibles por los técnicos y empleados de las empresas, todo bajo el prisma o paraguas que supone el nuevo concepto de Gestión de Calidad, tal y como lo define el doctor K. Ishikawa:

" *La gestión de la Calidad consiste en desarrollar, concebir y fabricar en los plazos previstos los productos y servicios más económicos, más útiles y más satisfactorios para el consumidor*."

Finalidad del QFD

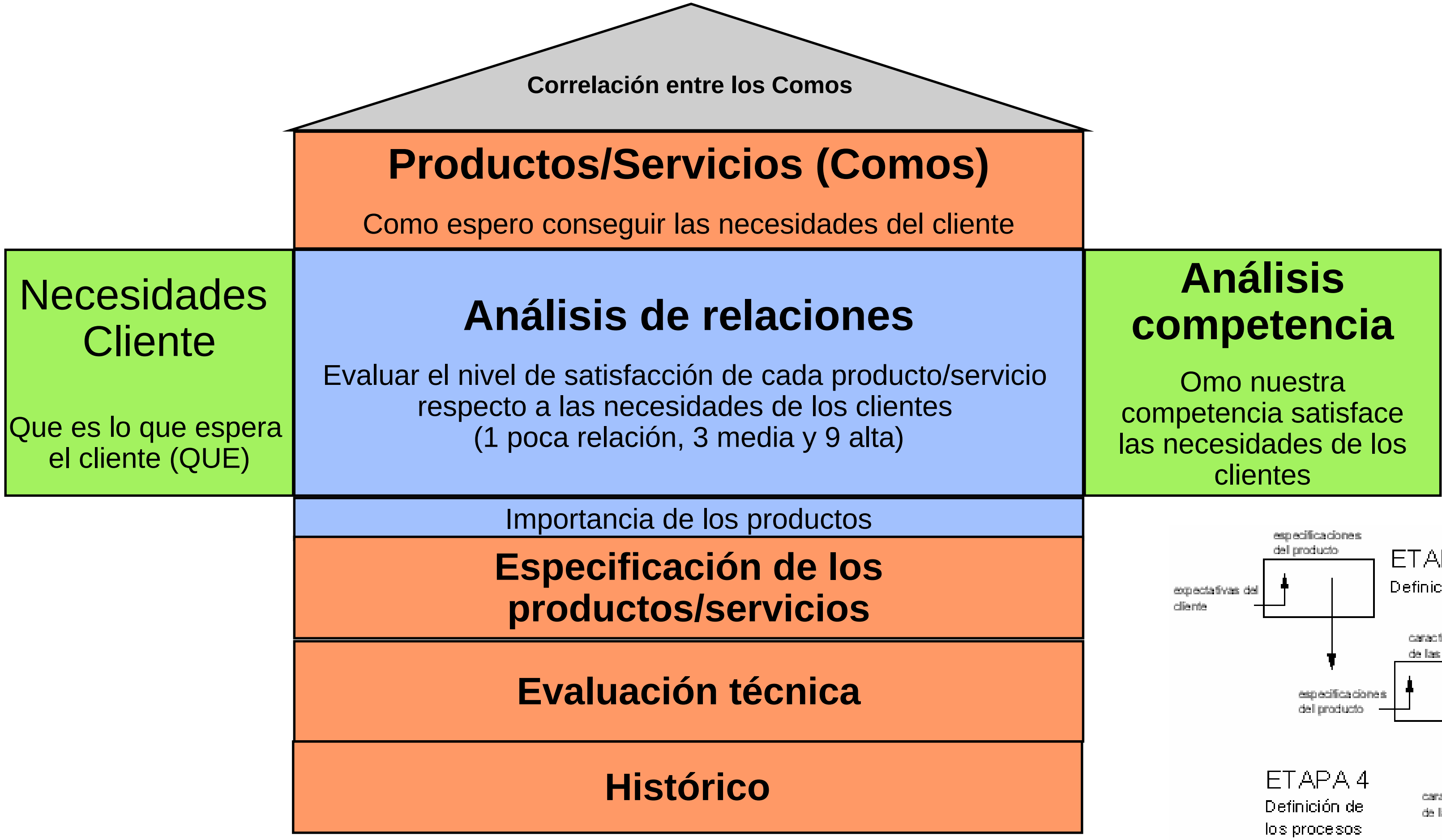
- 1) La obtención de una calidad de diseño excelente, mediante la conversión de las necesidades del cliente en características de calidad adecuadas, sin omisiones ni elementos superfluos.

- 2) El despliegue sistemático de dichas características mediante la búsqueda de unas funciones técnicas:
  - de los componentes del producto
  - de los componentes del proceso

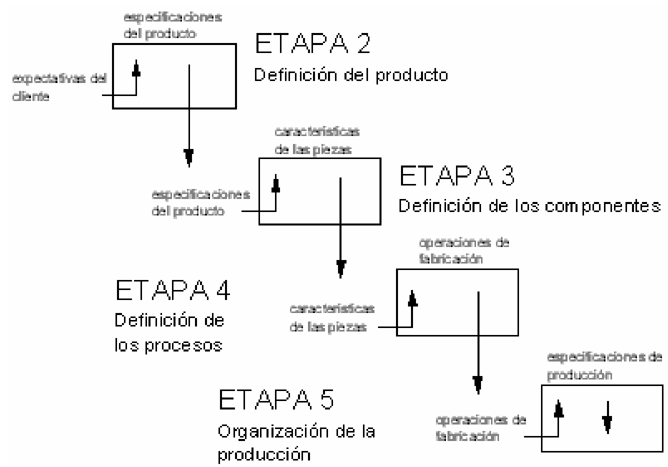
- 3) que estén lo mas correlacionadas posible con las necesidades del cliente.

  Este es el lado técnico del QFD. Para llegar a este punto, el QFD parte del principio según el cual las expectativas del cliente, todas pero "sin omisiones ni elementos superfluos" son las que permiten programar el producto adecuado en el momento adecuado.

Construcción del QFD



Ejemplo



Ejercicio 8

- **Diseña los factores principales para poder conseguir un resultado positivo en la próxima presentación a periodistas de nuestro nuevo vehículo.**
- **Utiliza la herramienta QFD, y construye la casa de la Calidad 1er nivel)**

Tiempo: 90 min (ejercicio) + 50 min (todas las presentaciones)

UPC CATEDRA NISSAN

DOE Departament d'Organització d'Empreses

Fase 3. Herramientas

F TEST

T TEST

ANOVA

CORRELACION

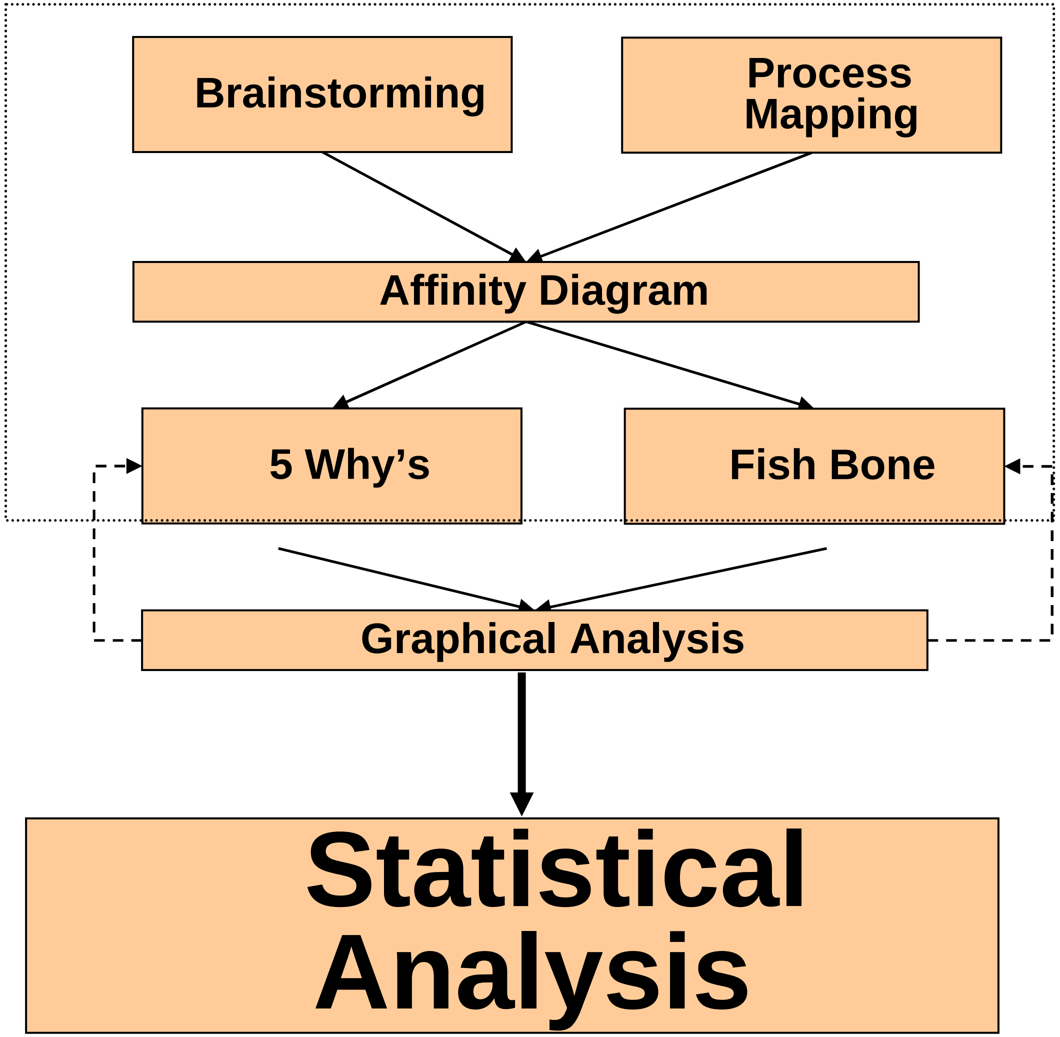
Departament
d'Organització
d'Empreses

Root cause analysis : Tools used

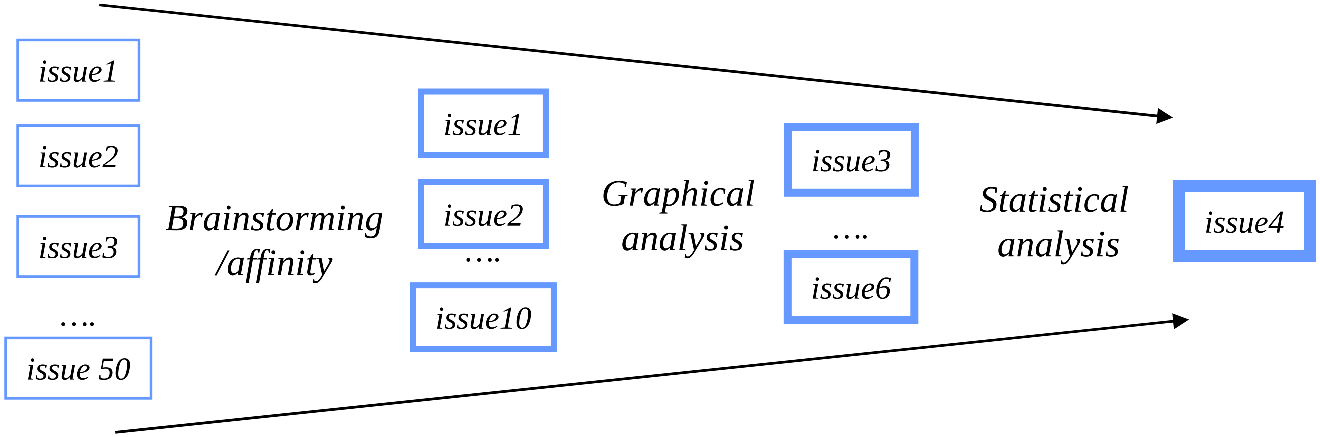
Identify potential root causes

Narrow down the number of factors

Confirm statistically the most important factors



Statistical Analysis



1 -Techniques such as **brainstorming, process mapping**…etc, allow us to focus on areas of concern by sharing expertise among the team members and therefore to prioritize the improvements according to the most important root causes.

2- **Graphical analysis** allows us to :

- Either « confirm » some root causes : « we can see it » and not only « we think it is an important root cause »
- Or focus the areas to brainstorm even before starting the analysis (for example, stratification).

The problem with graphical analysis is that it may be misleading (i.e take the wrong decision having read a graph) or not explicit enough for people to make the good decision

3- **Statistical analysis** allows us to confirm or reject a hypothesis based on the most important causes, whilst understanding that there is a risk of being wrong in the decision making process.

Because statistical analysis is time consuming and sometimes rather complex, it should be done ONLY after pre selection of the potential vital few root causes.

Root cause analysis : Tools used

Statistical Analysis : Y = f(X)

1st metrics

1. **Y is continuous, X is discrete:**
   - **T test to check the difference between the two means; (X has two levels),**
   - **F test to check difference on variance,**
   - **Anova to test multiple means difference (X has more than two levels)**

2. **Y is continuous, X is continuous:**
   - **Correlation and regression (one X)**
   - **Multi regressions (more than one continuous X)**

From ANOVA or regression, calculate the weight of each factor (X)

3. **Y is discrete, X is discrete:**
   - **CHI2 (or proportions)**

In each case, use graphs, test hypothesis, test normality and check conclusions with experts

Mean Test (also called t-Test)

What does it do?

- **When handling two sets of sample data, use this test to see if the mean value of each population group is equal (or not).**

When do you use it?

- **This test can be used to check if a factor has a real effect on the mean during the root cause analysis phase.**
- **It is also very useful to validate the impact of the improvement and demonstrate the difference in the mean between prior and after the change.**

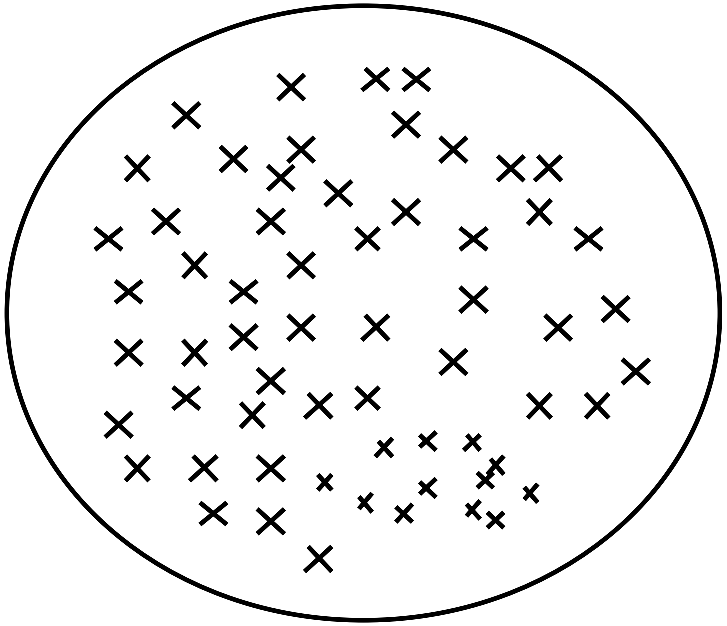
Limitation

- **As we are testing the mean, we need Normal distribution**

One sample t-test

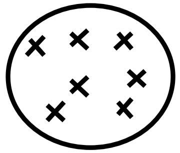
Population

N, m is known, but s is unknown



Sample :

$n, \bar{x}_1$



What do we want to know ?

- **Is the sample pertinent for the population from a mean perspective?**
- **Or, can we say that $\bar{X}_1$ = m ?**

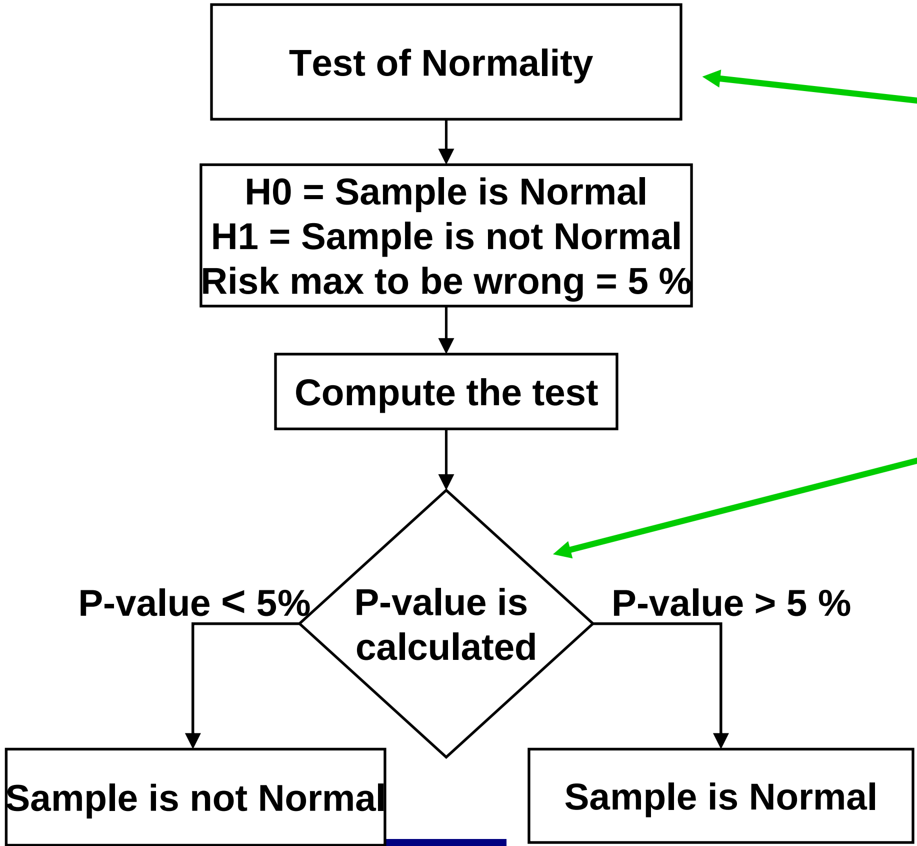
How can we check that ?

1. **Check Normality of the sample**
2. **Compute the 1-sample t-test**

1 sample t-test

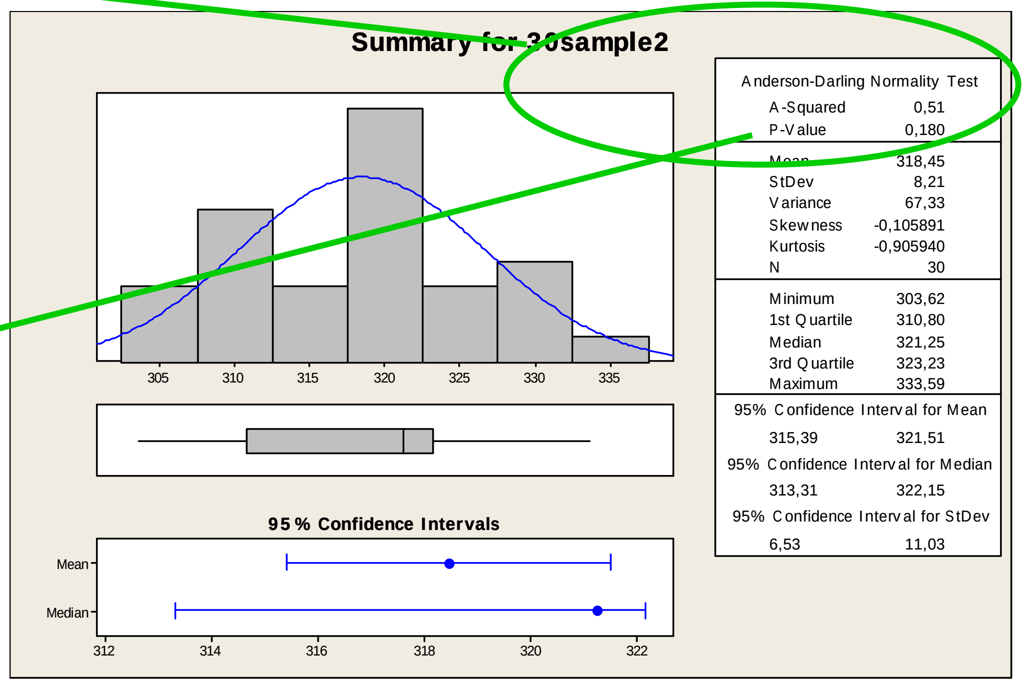
How can we check that ?

1. <u>Check Normality of the sample</u>
2. Compute the 1-sample t-test



In MINITAB…

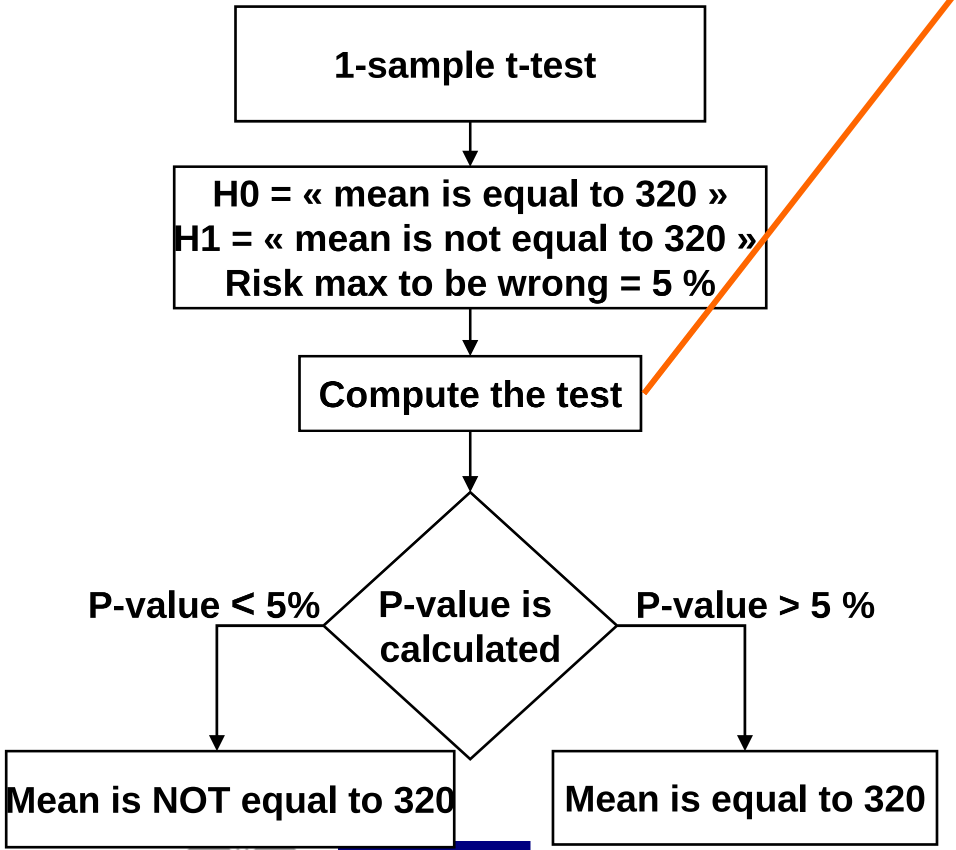
➢ **Stats>Basic Statistics>Display Descriptive Statistics**

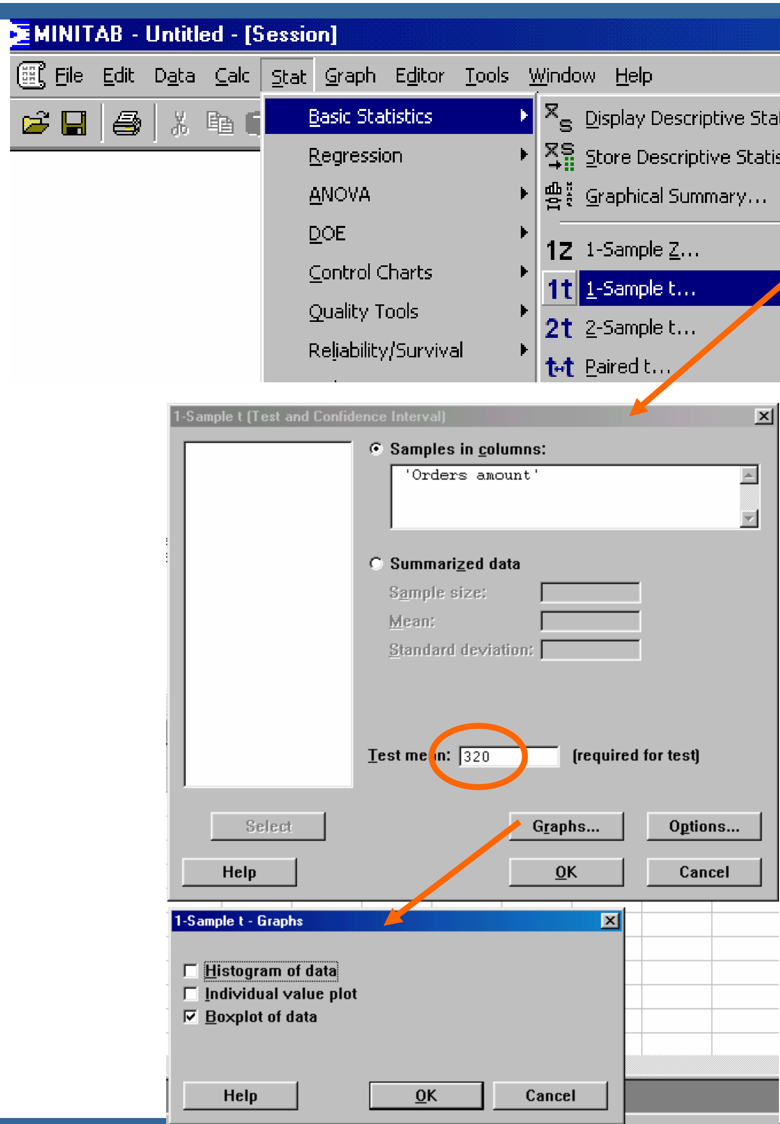


1 sample t-test

How can we check that ?

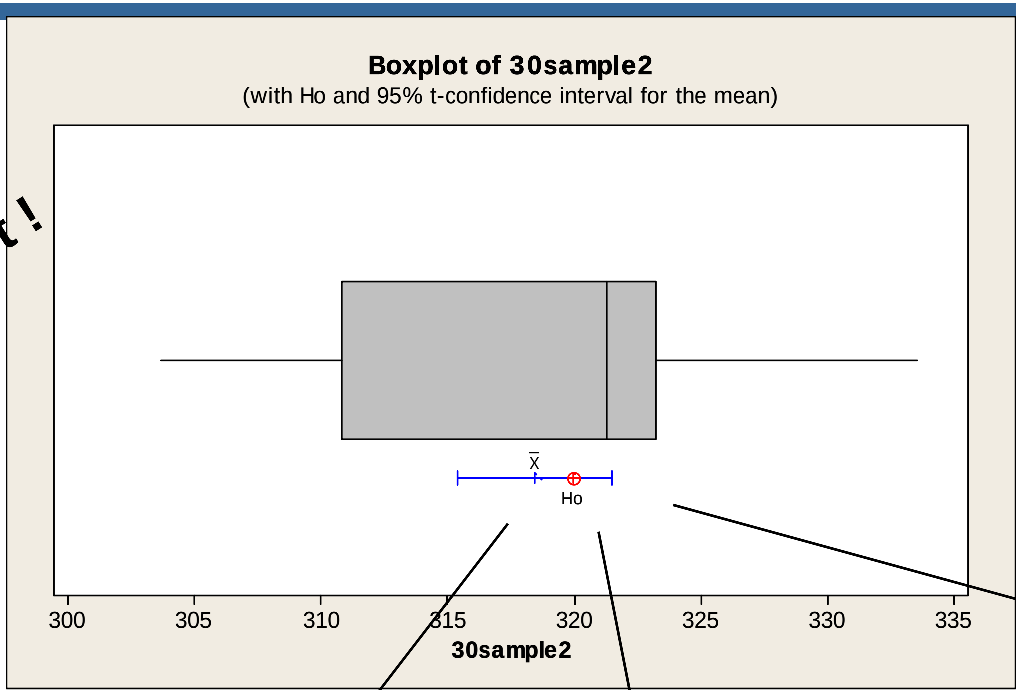
1. Check Normality of the sample
2. <u>Compute the 1-sample t-test</u>





1 sample t-test: application

Graphical result first !



Confidence interval on the mean

X = Computed value of the mean

HO: mean=320cm

Just looking at the graph, we can see that the 95% confidence interval overlaps 320cm so we can already ensure we have a good chance that trial1 is relevant to the original population from a mean perspective.

1 sample t-test: application

Statistical result second !

In addition, Minitab gives the calculations on the session window

One-Sample T: 30sample2

```
Test of mu = 320 vs not = 320

Variable     N     Mean   StDev  SE Mean         95% CI               T      P
30sample2   30  318,450   8,205    1,498  (315,386; 321,514)  -1,03  0,309
```

Value of t (from the t distribution)

Corresponding probability for 20 data points

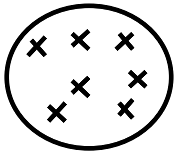
Here we can say that we have 30% chance of being wrong by assuming that the mean of « 30sample2» differs from 320 and therefore, we will consider that the calculated mean (318,45) is not different than the total population. Practically, we will consider this sample of 30 data as a good representative of the total population.

If p-value (also called a) is more than 5%, we will say that the mean is not different

2 sample t-test

We are now looking at 2 set of data and we want to check if they are identical in term of means

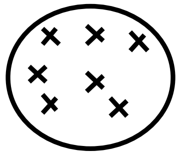
Sample 1



$n_1, \bar{x}_1$

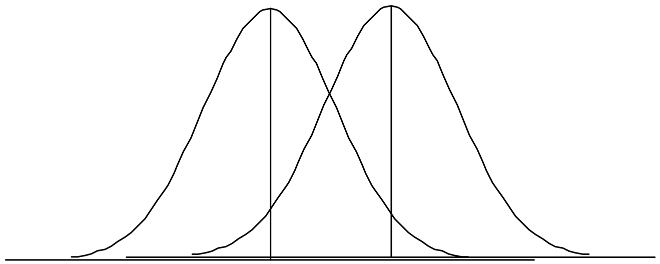
Before – After
Machine A – Machine B
UK – Spain
Dealer X – Dealer W
Etc…

Sample 2



$n_2, \bar{x}_2$

Can we consider that $\overline{X1}$ and $\overline{X2}$ are "identical"?

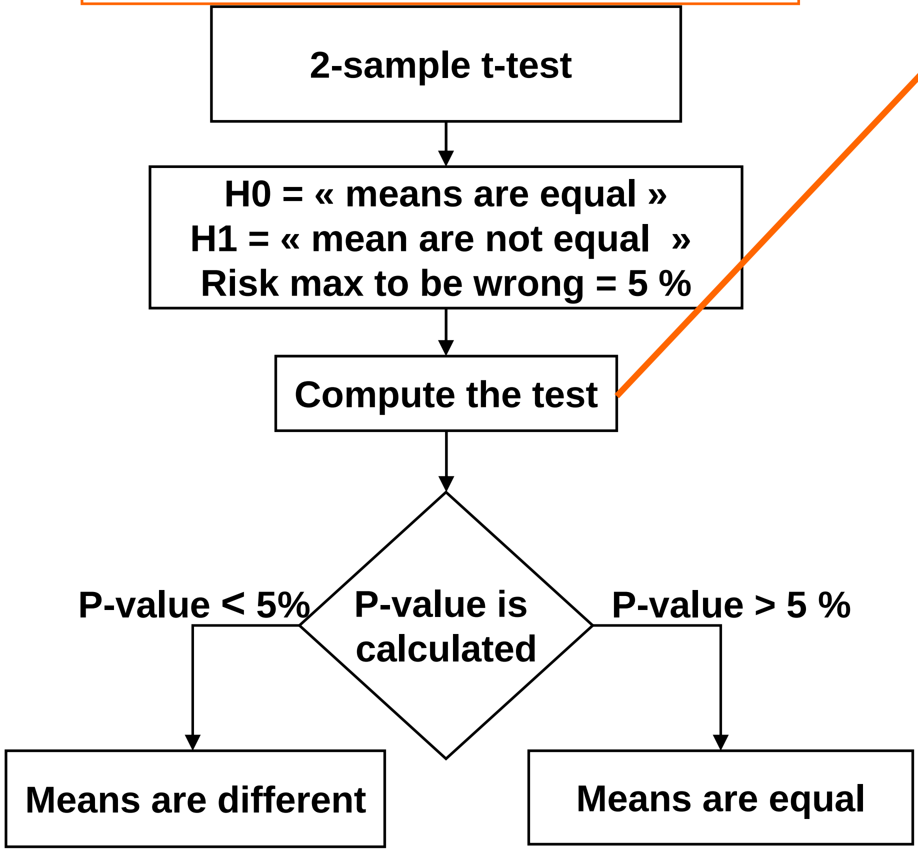


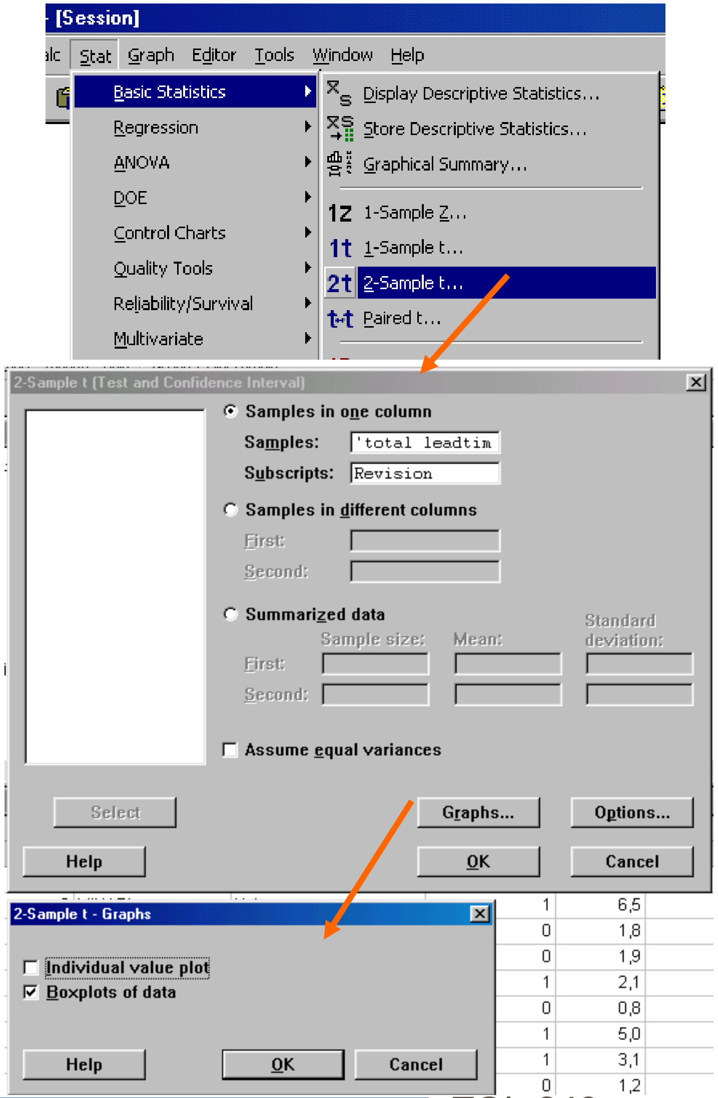
$\bar{X}_1$ $\bar{X}_2$

To answer this question we use a " two sample t-test"

2 sample t-test: application

***Warning** : this calculation is only valid if each distribution is normal*

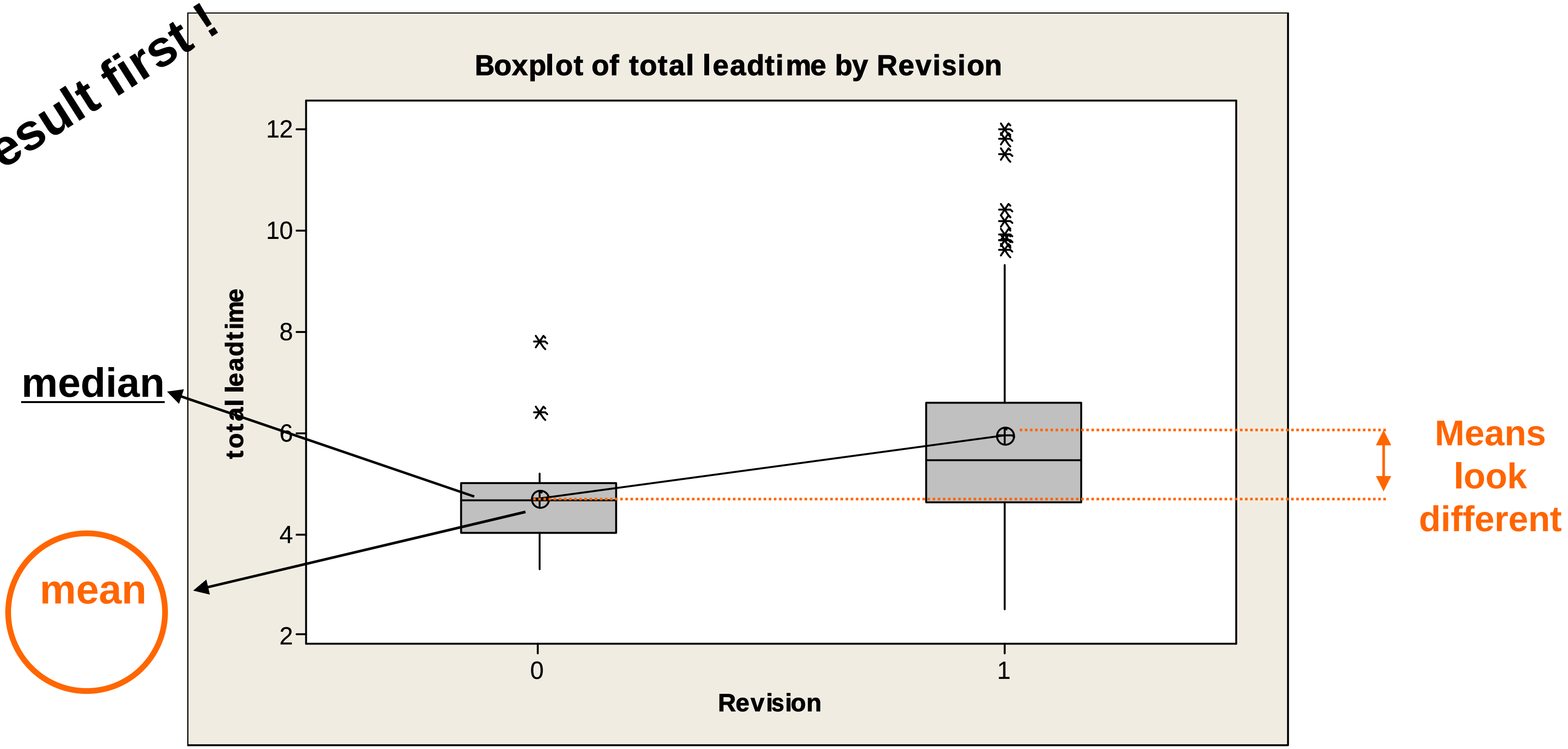




2 sample t-test: application

Total Lead time = f(Revisions)

Graphical result first !



Just looking at the graph it appears that the two means are different but there is also quite an overlap: are they really different therefore?
Let's look at the statistical result to have the right answer…

2 sample t-test: application

Statistical result second !

In addition, Minitab gives the calculations on the session window

```
Two-sample T for total leadtime
Revision      N   Mean   StDev   SE Mean
0            20   4,70    1,00      0,22
1           100   5,94    1,95      0,20

Difference = mu (0) - mu (1)
Estimate for difference:  -1,24400
95% CI for difference:  (-1,84081; -0,64719)
T-Test of difference = 0 (vs not =): T-Value = -4,18  P-Value = 0,000  DF = 52
```

Value of t (from the t distribution)

Calculated P-value = actual risk

In that case, there is no risk (P-value=0) to say that the 2 means are different

Redo the exercise with PIC of revision and Total Lead time

T-test is easy to use… make sure your data are normal, and check the practical significance of your result !

Variance Test (F-Test)

What does it do?

- When handling two sets of sample data, use this test to see if the variance of each population group is equal (or not).

When do you use it?

- This test can be used to check if a factor as a real effect on the Standard Deviation during the root cause analysis phase.
- It is also very useful to validate the impact of the improvement and demonstrate the reduction in process variation between prior and after a change is implemented.

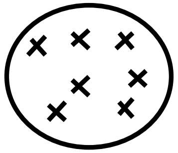
Limitation

- As we are testing the Std Deviation, we need Normal distribution

F-test

We are now looking at 2 set of data and we want to check if they are identical in term of Std Dev.

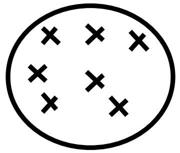
Sample 1



$n_1, \bar{x}_1, \sigma_1$

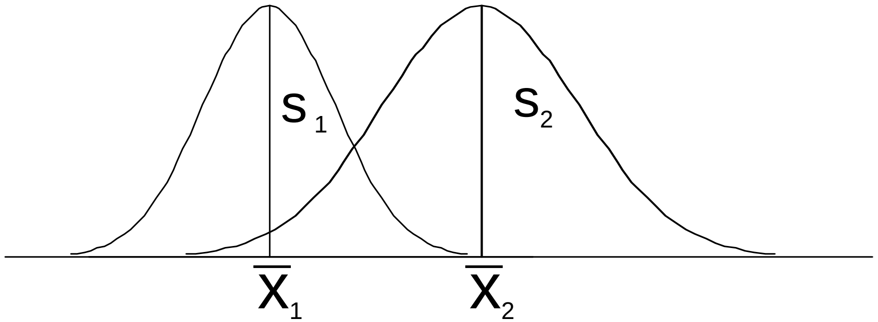
Before – After
Machine A – Machine B
UK – Spain
Dealer X – Dealer W
Etc…

Sample 2



$n_2, \bar{x}_2, \sigma_2$

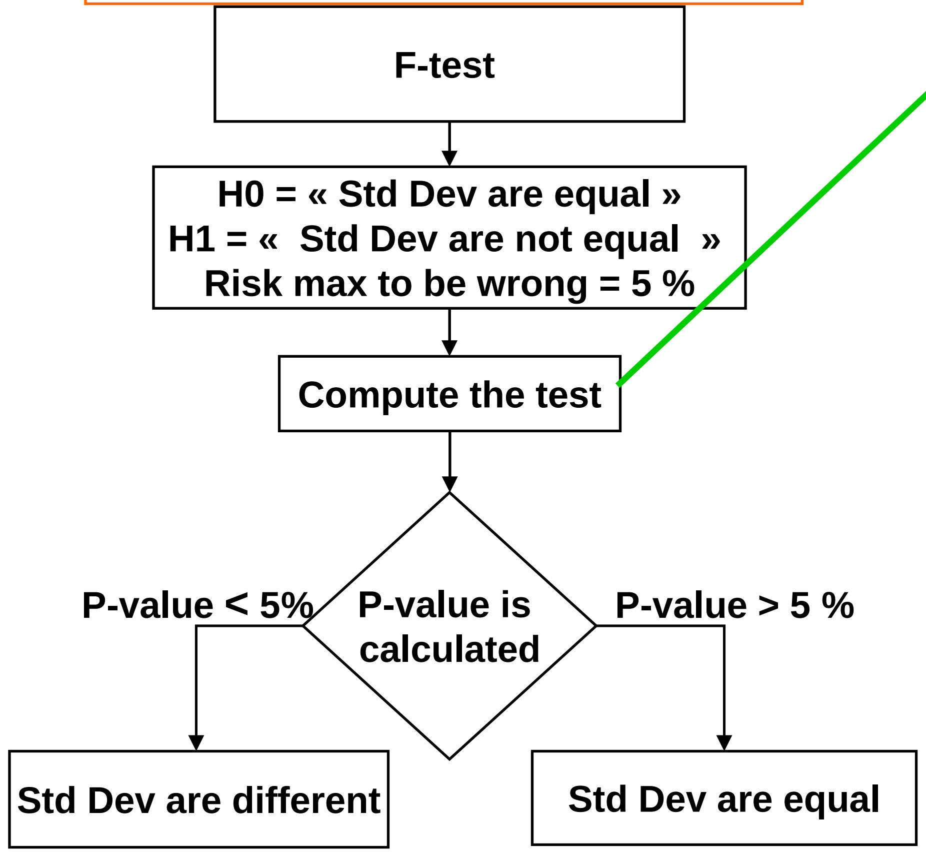
Do my two samples have the same standard deviation?

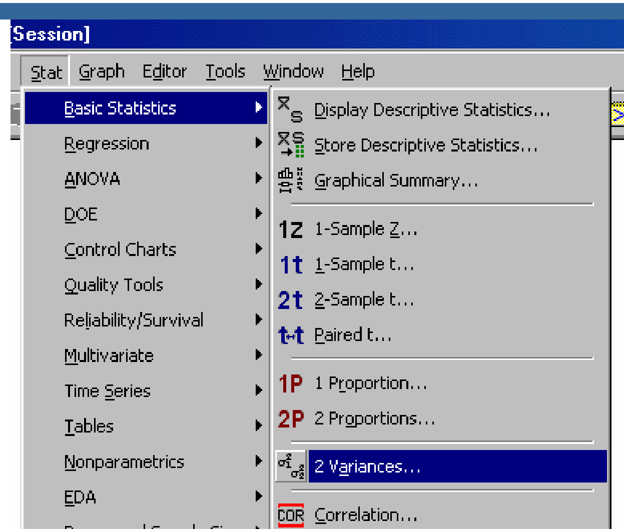


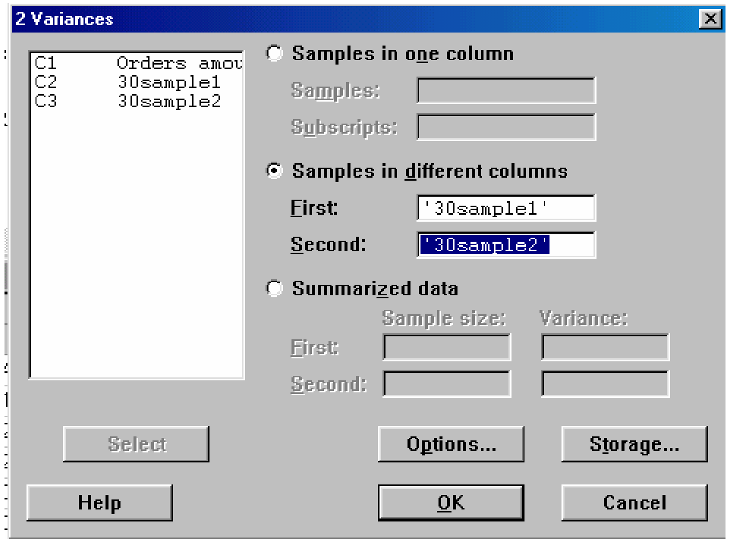
To answer this question we use a “ F-test”

F-Test: application

Warning ***: this calculation is only valid if each distribution is normal***





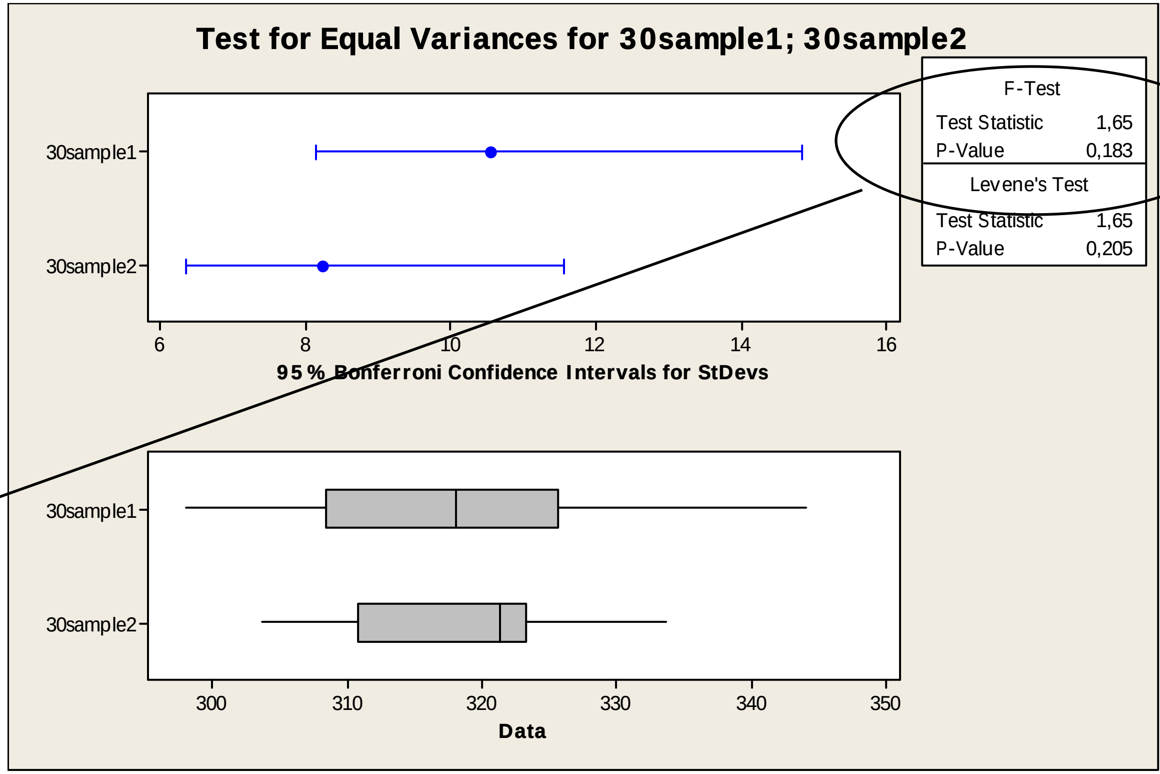


In this example, the samples are in two different columns

F-Test: application

Graphical and Statistical result at once !

Confidence interval for each case



F test Calculation : Because the p-value is over 5%, we can conclude that there is no difference of variance between the two samples

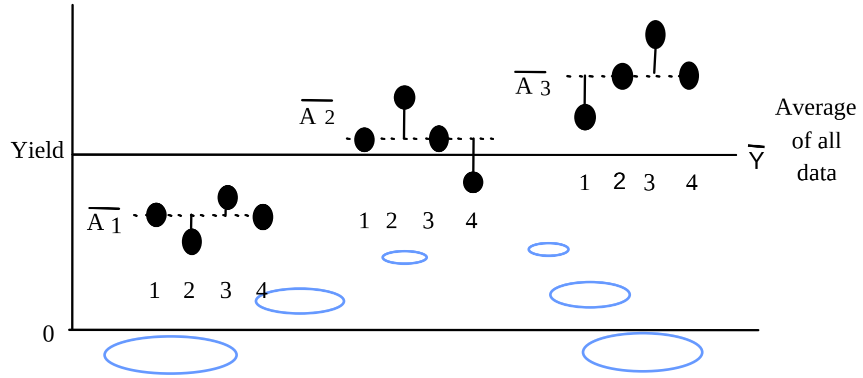
Total Lead time = f(Revisions)

The basic principles of ANOVA (1/3)

We collected sales data from 3 different salesmen. The actual margin is calculated and reported in the following chart.

Obtained data and data plot

Salesmen	A1	A2	A3
Margin	57	62	63
	56	64	65
	58	62	67
	57	60	65
Sum	228	248	260



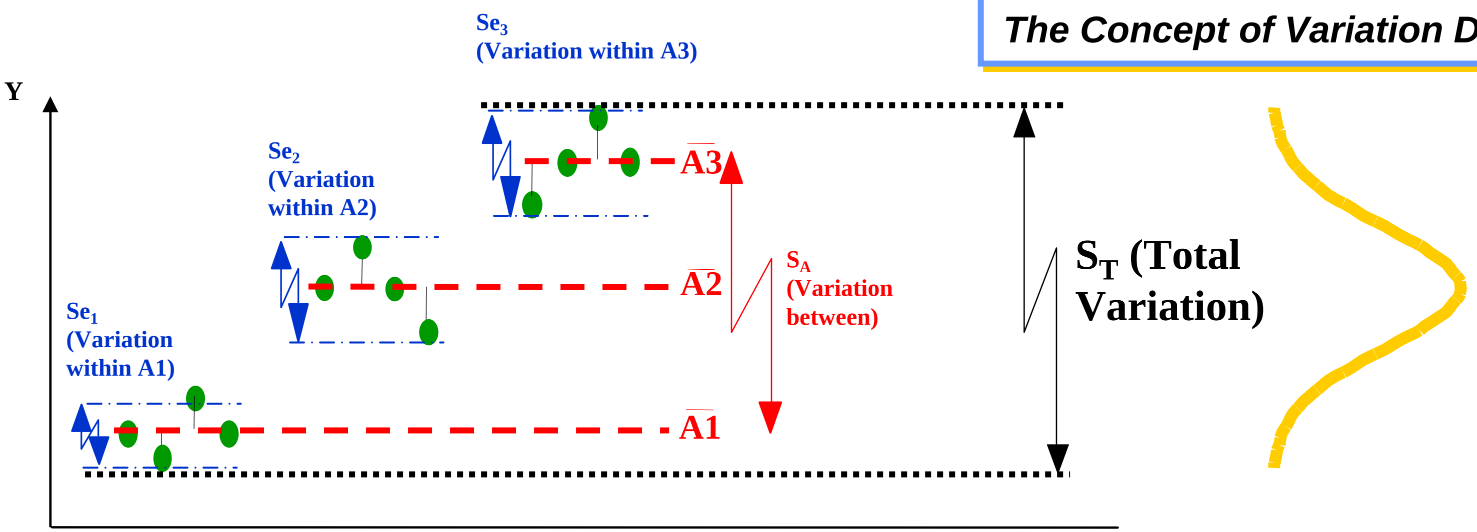
Can we conclude from these results that the salesman influence the average output?

Or is it ordinary variation of margin within several orders ?

To answer this question we use the ANOVA test

The basic principles of ANOVA (2/3)

The Concept of Variation Decomposition



S_T (Total Variation) = S_A (variation between salesmen) + S_e (variation within each salesman)

What do we want to know ?

- **Do I see actual changes in the mean when I change the sales man ?**
- **Is A1 really different from A2, from A3 and is A2 different from A3 ?**
- **Is the "local" variation (with one sales man) smaller than the variation seen when we change the sales man ?**
- **Is $S_e << S_A$?**

How can we check that ?

F-Test !!!

The basic principles of ANOVA (3/3)

MINITAB will do all the calculation

S_T (total variation) = S_A (variation of factor) + S_e (variation of error)

ANOVA

Source	DF	SS	MS	F	P
Factor	2	130.67	65.33	32.67	0.000
Error	9	18.00	2.00		
Total	11	148.67			

<0.05

F-Test !!!

These results indicate that the P value of the factor "salesmen" is less than 0.05. It can be concluded that "statistically speaking, the salesmen has a significant influence on the margin."

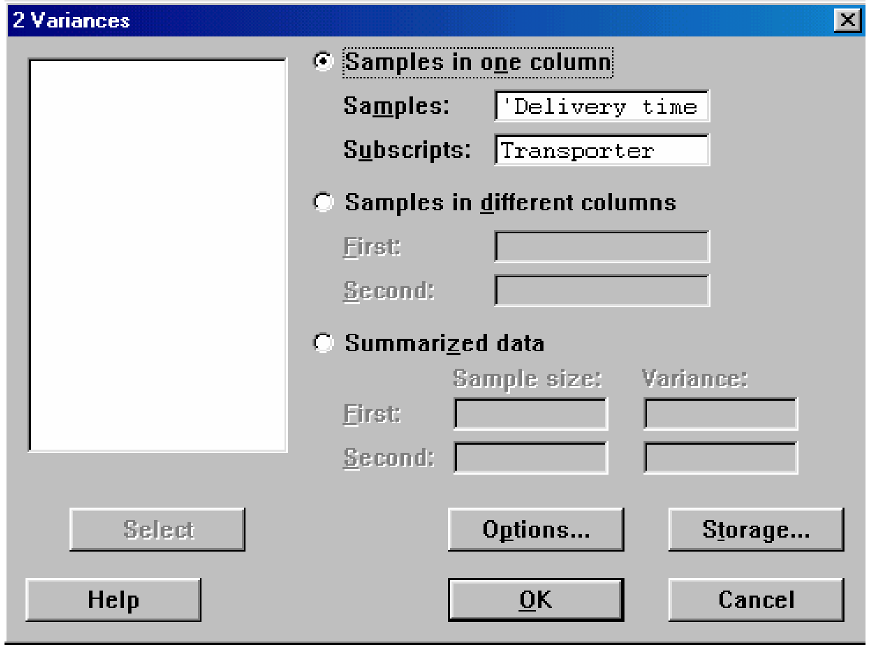
Warning : this computation is valid only if each distribution is normal and if the variances of each sample are equal

Application : One-way Layout ANOVA

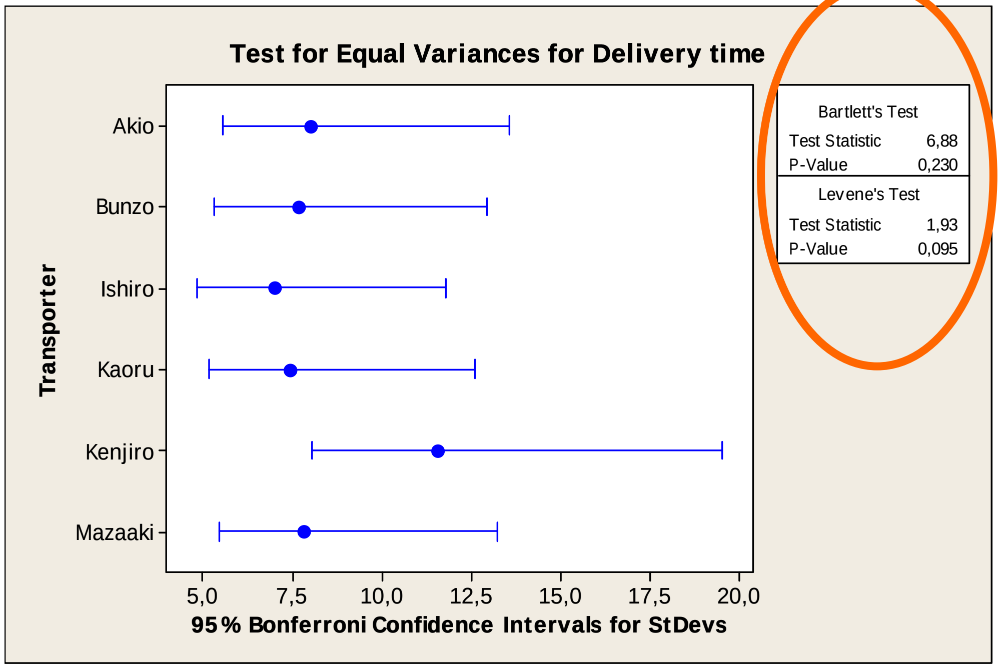
We would like to know if the "Delivery time" is linked with the Transporter.

Question: has the factor « Transporter » a significant influence on the total variability of "Delivery time"?

First lets check that the population is normal and then check that the variance of the population is the same for each level of the factor « Transporter » :

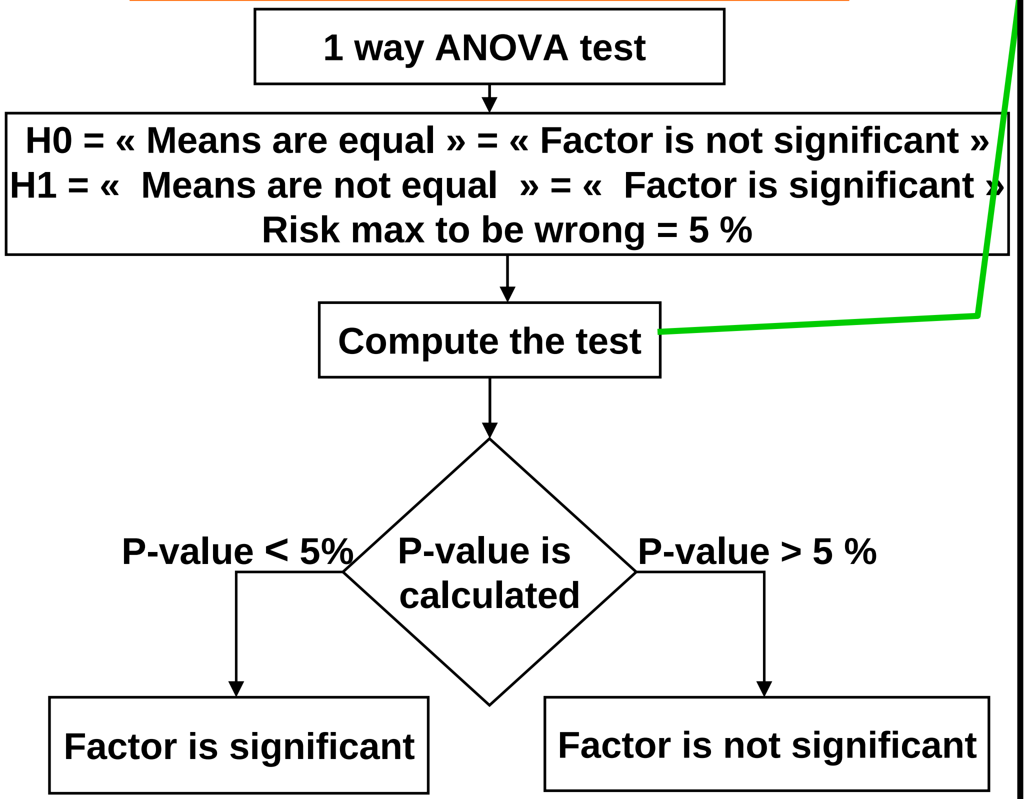


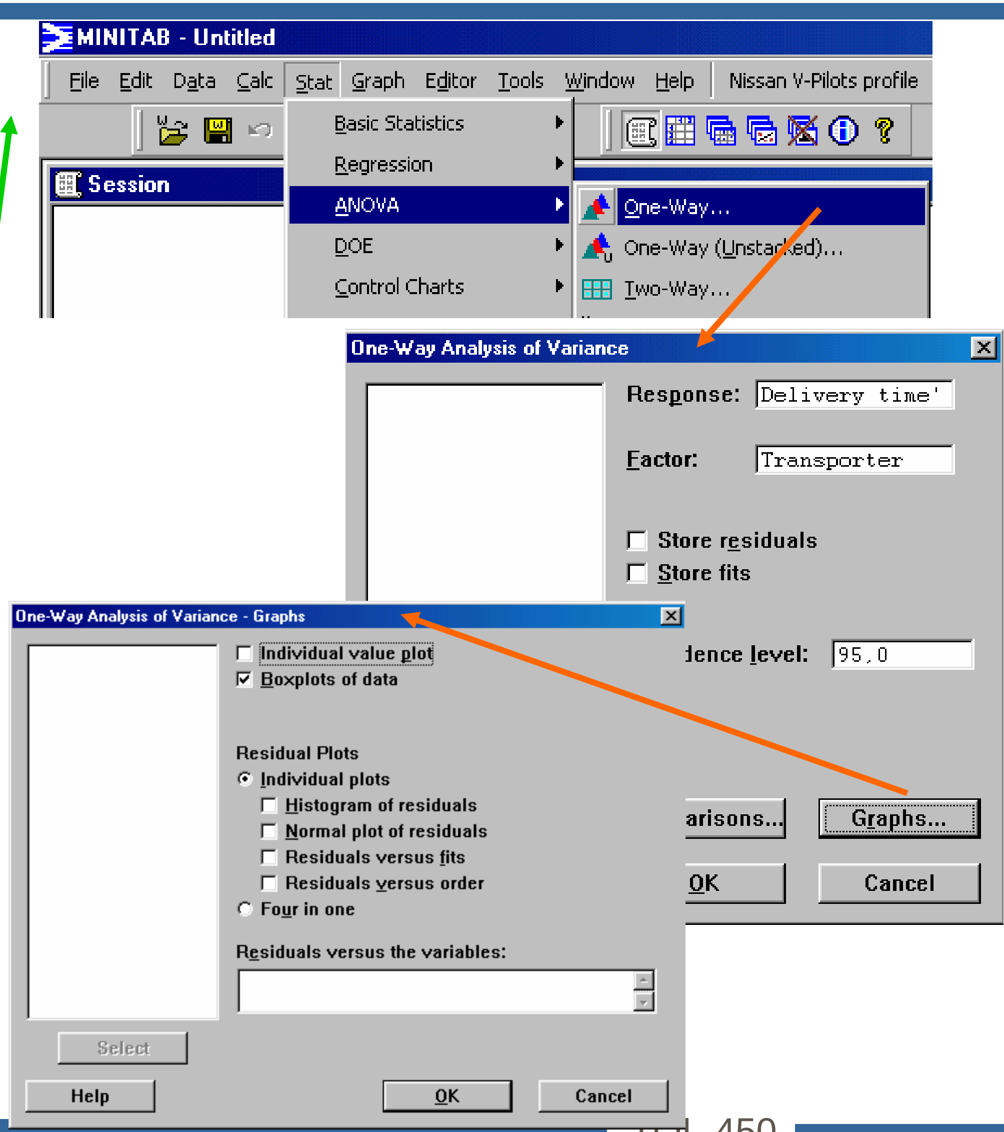
P-value??



One-way Layout ANOVA

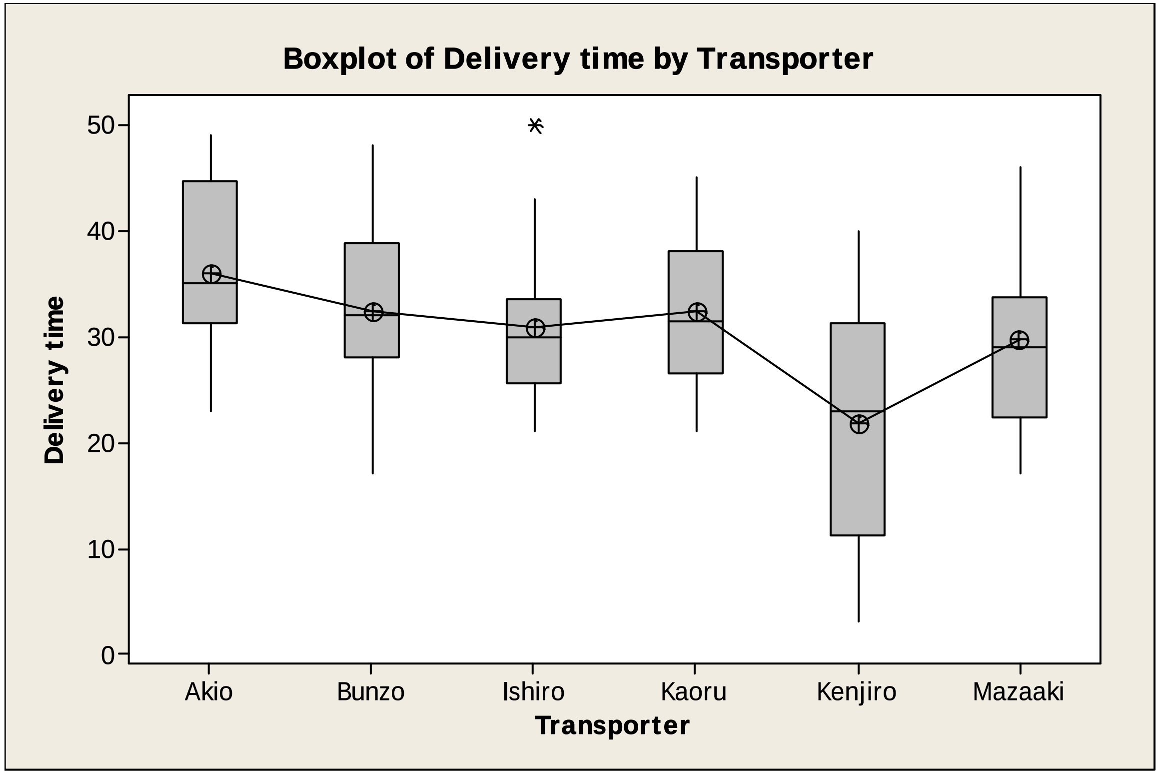
***Warning* : this calculation is only valid if each distribution is normal and has the same variance**





Application : One-way Layout ANOVA

Graphical result first !



From the box plot, we can already deduce that Kenjiro is in someway different (from a mean stand point) from the other transporters.

Application : One-way Layout ANOVA

Statistical result second !

The factor « transporter » has 6 levels therefore 5 degrees of freedom…
The famous « n-1 »

The « Sum of the Squares » represents the total distance between the mean at each level of the factor and the overall mean

The « Mean Square » is the variance « between » levels, it is calculated by SS/DF (=2272/5)

```
One-way ANOVA: Delivery time versus Transporter

Source        DF       SS      MS      F       P
Transporter    5   2272,7   454,5   6,46   0,000
Error        114   8015,3    70,3
Total        119  10288,0
```

The «p value» gives the confidence in the significance of the factor, in this case, the overall variability is truly influenced by the factor « transporter »

There are 120 data points therefore 119 degrees of freedom (and so only 119-5 = 114 DF for the error)

The F is the ratio between the mean square of the factor (454) divided by the mean square of the error (70)

In this example, we are sure (p=0.0%) that the factor « transporter » is significant in the overall observed variability, but this factor is not the only one as it explains only 2272/10288 = 22%.

Application : One-way Layout ANOVA

Minitab also calculates other data…

```
S = 8,385   R-Sq = 22,09%   R-Sq(adj) = 18,67%
```

22.0% of the total variability is explained by the factor « trial »

```
 Individual 95% CIs For Mean Based on Pooled StDev
Level     N    Mean   StDev     +---------+---------+---------+---------
Akio     20  36,000   8,020                             (-----*-----)
Bunzo    20  32,300   7,672                       (-----*-----)
Ishiro   20  30,850   6,976                    (-----*------)
Kaoru    20  32,400   7,451                       (-----*-----)
Kenjiro  20  21,800  11,542     (-----*------)
Mazaaki  20  29,650   7,829                 (-----*------)
                                +---------+---------+---------+---------
                             18,0      24,0      30,0      36,0
Pooled StDev = 8,385
```

No overlap of the means between Kenjiro and Mazaaki

Conclusion: the factor « transporter » is significant in the overall variability observed in the delivery time. Graphically we can see that Kenjiro is the one which creates the main difference.

Conclusion and traps : One-way Layout ANOVA

CONCLUSION :

ANOVA is one of the most popular tools for statistical root cause analysis : Y= f(X).

It not only confirms the influence of the factor X in the overall variability (S) of Y but it also indicates the weight of the factor.

FIVE COMMON TRAPS :

1. **If the population is abnormal, calculations may lead to the wrong decision,**
2. **If the variances are not homogeneous, calculations may lead to the wrong decision,**
3. **If each level of the factor does not contain at least 5 data points, calculations may lead to the wrong decision,**
4. **If the factor is continuous (such as pressure, time..) and not discrete, calculations may lead to wrong decision.**
5. **If the factor looks graphically significant but the p- value is not good (for example 20%), it is may be due to the sample size being too small; in that case, add more data and re-check.**

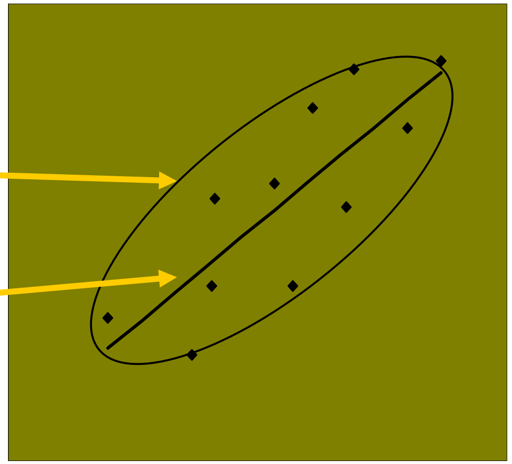
3.6) Correlation Analysis and Regression Analysis

1. What Does It Do?

- **Scatter diagrams are used to describe the relationship of two quantitative variables. Correlation analysis is conducted to measure the correlation of the two variables.**
- **Regression analysis is conducted if there is correlation between the two quantitative variables. Regression analysis determines the equation of the two variables' linear relationship.**

Correlation analysis: Measures the strength of correlation.

Regression analysis: Draws a straight line in the scatter diagram.



Correlation Analysis and Regression Analysis

2. Where Do You Use It?

- **Investigate whether there is a correlation between the variables with correlation analysis. Then investigate further for cause-and-effect relationships.**
- **Use regression analysis where there is a correlation between two variables. See the following Examples.**
  - **When you want to control output variables with input variables**

Correlation Coefficient

"Correlation coefficient: R" is a statistic that describes the strength of a linear relationship between two data.

Correlation coefficient: r can take values between –1 and +1. Correlation gets stronger as the absolute value of R gets closer to 1.

Also, "Correlation coefficient: R>0" indicates "positive correlation"
"Correlation coefficient: R<0" indicates "negative correlation"
"Correlation coefficient: R=0" indicates "no correlation"

Guidelines for Correlation coefficient: R

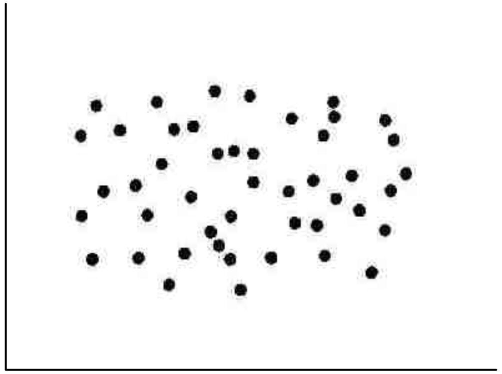
-1 ← -0.8 — -0.2 — 0 — 0.2 — 0.8 → 1

Strong Negative Correlation	Moderate Negative Correlation	Weak or no Correlation	Moderate Positive Correlation	Strong Positive Correlation

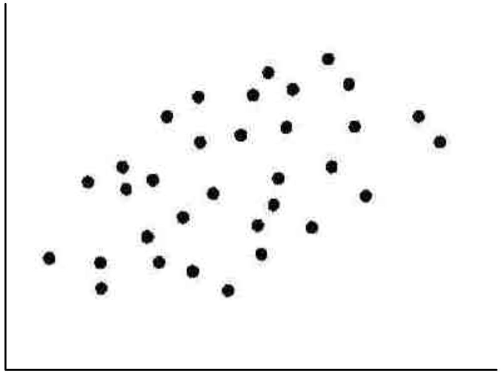
Scatter Diagram

Correlation coefficient and scatter diagrams are shown below

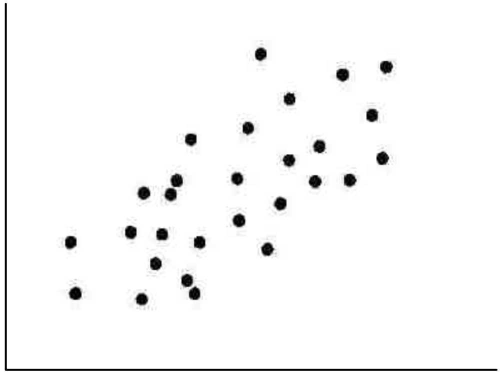
No correlation
Correlation coefficient=0



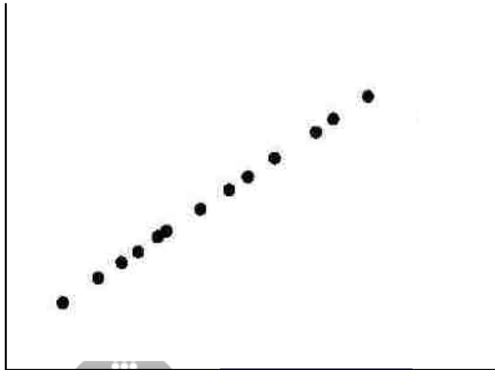
Moderate positive correlation
Correlation coefficient=0.6



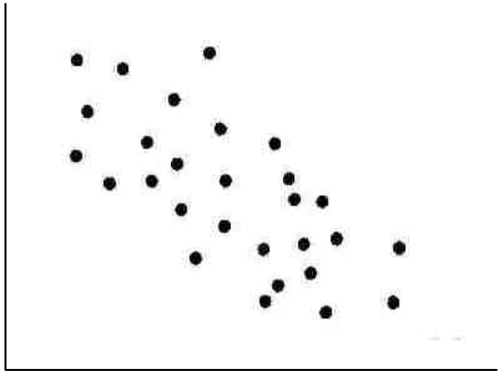
Strong positive correlation
Correlation coefficient=0.8



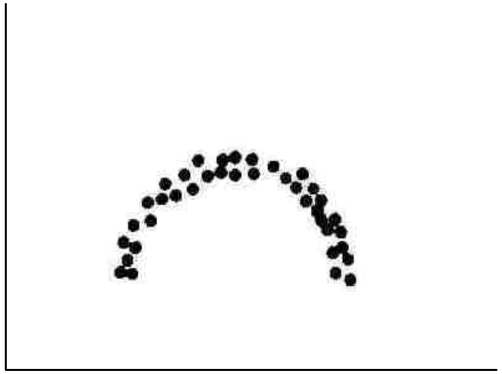
Perfect positive correlation
Correlation coefficient=1.0



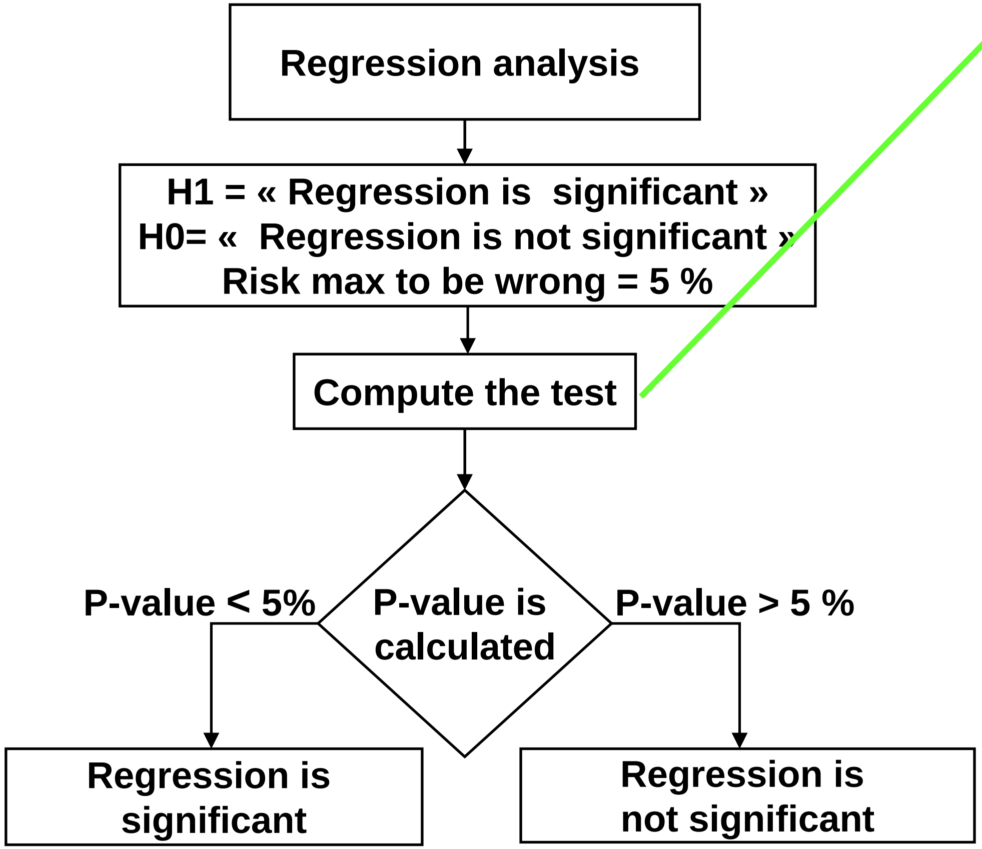
Strong negative correlation
Correlation coefficient=-0.8

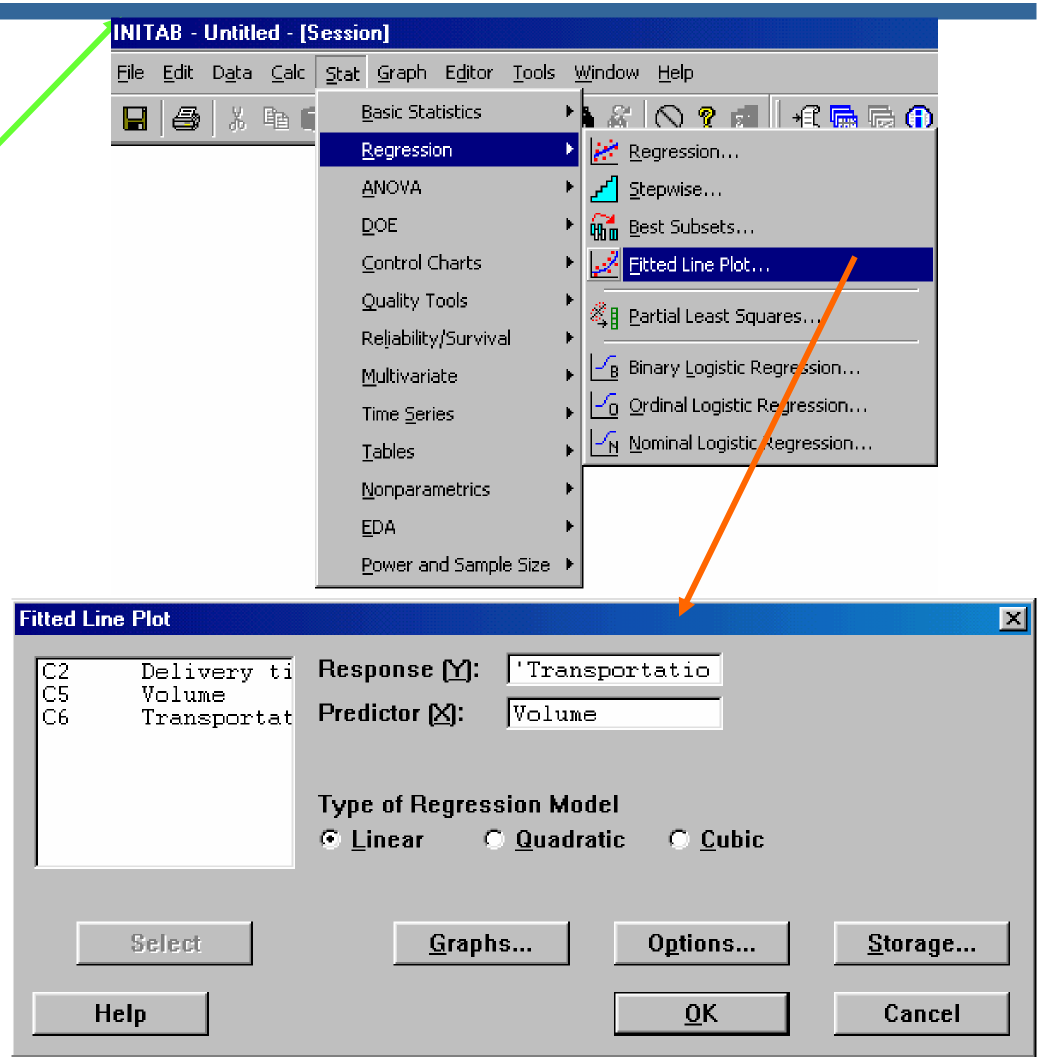


Non-linear correlation
Correlation coefficient=0



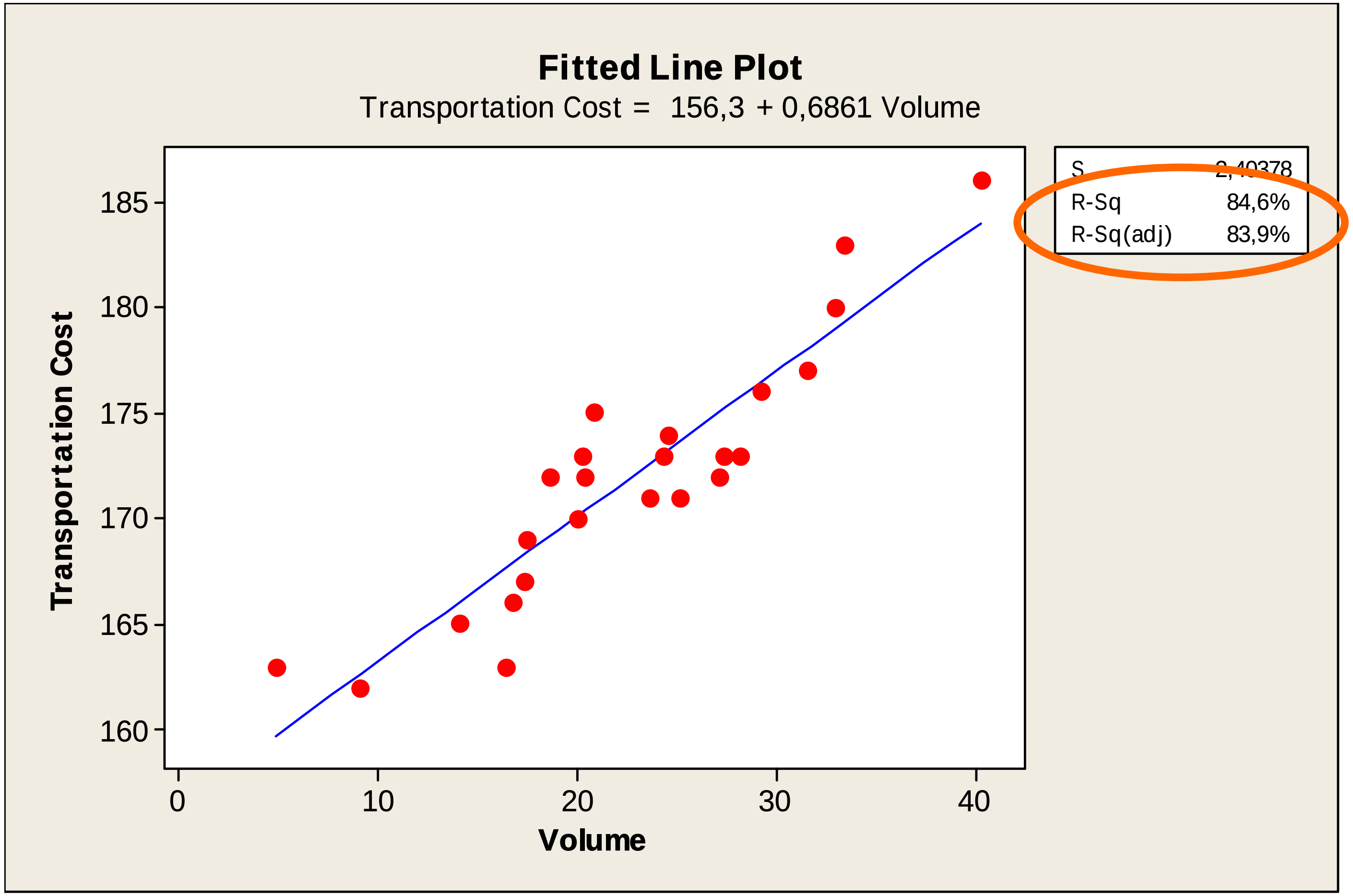
Correlation Analysis and Regression Analysis





Correlation Analysis and Regression Analysis

Graphical result first !



We can see a trend and the relationship is very strong… confirmed by a r² factor of 83.9 % (adjusted for sample size)

Correlation Analysis and Regression Analysis

Statistical result second !

Regression Analysis: Transportation Cost versus Volume

The regression equation is

Transportation Cost = 156,3 + 0,6861 Volume

S = 2,40378   R-Sq = 84,6%   R-Sq(adj) = 83,9%

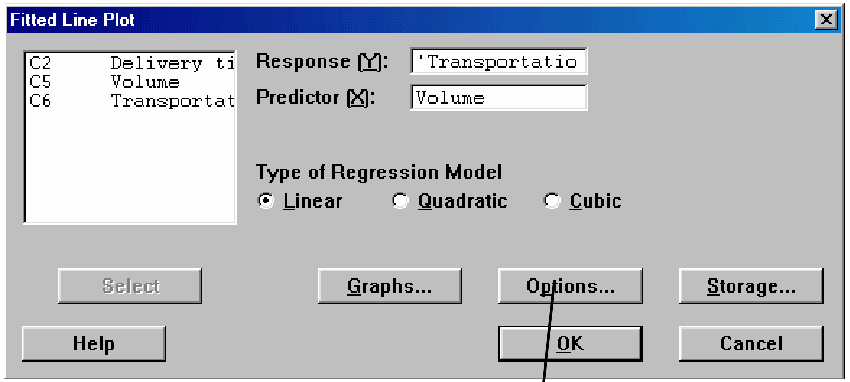
Analysis of Variance

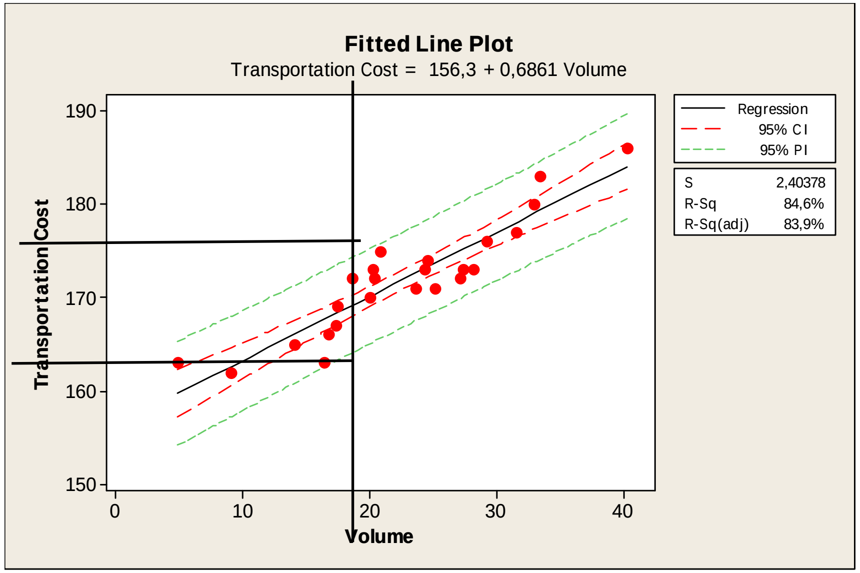
Source	DF	SS	MS	F	P
Regression	1	698,714	698,714	120,92	0,000
Error	22	127,120	5,778		
Total	23	825,833			

P-value is less than 5%, then the regression is significant

The R^2 factor indicates the strength of the relationship between weight and size

Correlation Analysis and Regression Analysis

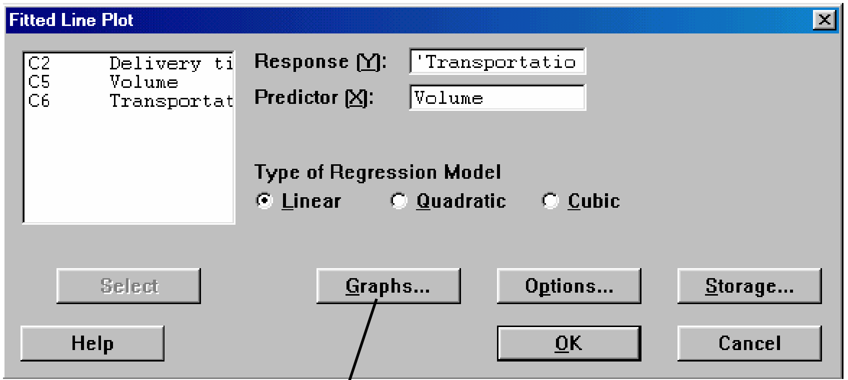




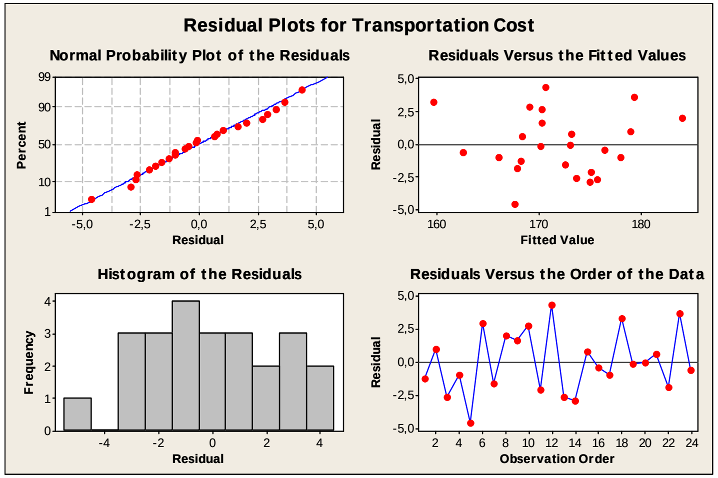
Option: display confidence intervals
Allows to better assess the confidence on the regression line and for individual data point.
For example, if the volume is "20", there is 95% chance to have a transportation cost between
165 and 175.

Use of confidence intervals

Correlation Analysis and Regression Analysis



To display residual analysis: this allows to check for special pattern which may lead to wrong interpretation of the correlation or to suspect hidden factor.



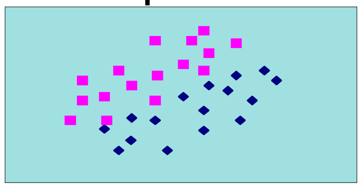
Use of residual analysis

Tips for Conducting Correlation Analysis and Regression Analysis (1/2)

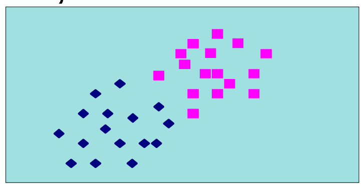
(1)Examine the data pictorially with scatter diagrams before correlation analysis and regression analysis. The points to check when you are examining scatter diagrams are:

- **Is there any need for stratification?** — **Diagrams a, b**
- **Are there any non-linear relationships?** — **Diagram d**
- **Are there any dislocated points?** — **Diagrams e, f**

a)

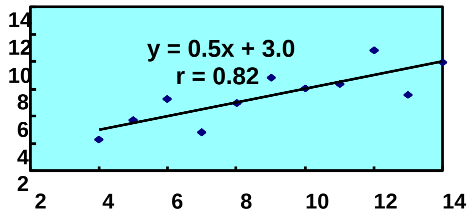


b)

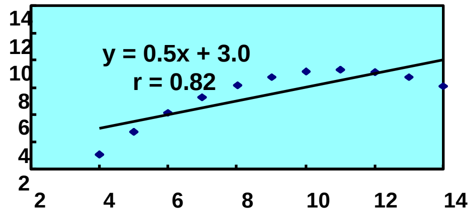


Always **draw scatter diagrams. There could be several different cases with the same correlation coefficient and regression equation, shown below.**

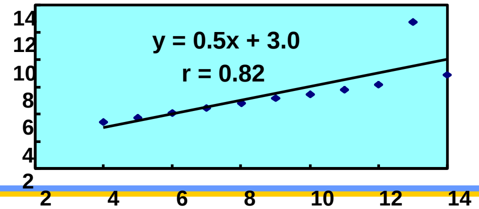
c)



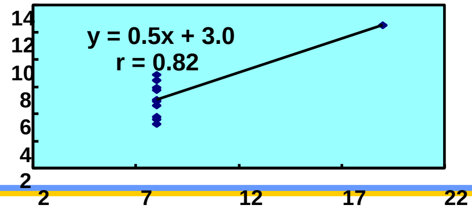
d)



e)



f)



Tips for Conducting Correlation Analysis and Regression Analysis (2/2)

Common pitfalls

(1) Skipping scatter diagrams and going onto correlation analysis and regression analysis.

(2) Jumping to the conclusion that there is no correlation, because the correlation efficient is small.

- ➔ **There are cases in which there seems to be no or weak correlation, when in fact there is a strong one. This happens when the analysis scope is too small.**

(3) Jumping to the conclusion that there is a cause-and–effect relationship, because there is a correlation.

- ➔ **A strong correlation between two variables does not necessarily mean that there is a cause-and-effect relationship. Judging whether there is a cause-and-effect relationship requires additional analysis.**

(4) Extending the regression line and predicting values outside the scope (extrapolation).

- ➔ **There is no guarantee that the regression line extends beyond the data obtained. You will need technical examination for extrapolation.**

Ejercicio 9

- **Análisis Gráfico: Analiza la relación entre descuento y margen entre las regiones y productos.**
- **Justifica las conclusiones**

(Fichero: análisis gráfico.xls)

DOE Departament d'Organització d'Empreses

Tiempo: 15 min (ejercicio) + 20 min (todas las presentaciones)

Ejercicio 10

- **t & F Test: Analiza la resultados de los costes de garantía de dos países**
- **Justifica las conclusiones**

Departament
d'Organització
d'Empreses

(Fichero:t y F Test.xls)

Tiempo: 15 min (ejercicio) + 20 min (todas las presentaciones)

Ejercicio 11

- **ANOVA: Evalúa el impacto en la media de diferentes modificaciones realizadas sobre una máquina**
- **Justifica las conclusiones**

(Fichero: ANOVA.xls)

Tiempo: 15 min (ejercicio) + 20 min (todas las presentaciones)

Departament d'Organització d'Empreses

Ejercicio 12

- **Correlación y regresión: Explica la relaciones entre coste y volumen del transporte determinado**
- **Justifica las conclusiones**

(Fichero: correlación y regresión.xls)

Tiempo: 15 min (ejercicio) + 20 min (todas las presentaciones)

CATEDRA NISSAN UPC

DOE Departament d'Organització d'Empreses

Ejercicio 13

- **Analiza los datos de carga las baterías, y comenta si justifican o no la causa raíz**

(Fichero: Pepe car baterias.xls)

Tiempo: 20 min (ejercicio) + 20 min (todas las presentaciones)

UPC CATEDRA NISSAN

DOE Departament d'Organització d'Empreses

Ejercicio 14

- **Justifica cual es la causa o causas principales de la falta de precisión en el plan de ventas de la compañía**

(Fichero: precisión ventas.xls)

UPC CATEDRA NISSAN

DOE Departament d'Organització d'Empreses

Tiempo: 20 min (ejercicio) + 20 min (todas las presentaciones)