# *SÉPTIMA Y OCTAVA CLASE*

Analytics AoZ

1. Inferencia Estadística – Estimación Puntual.
2. Estimación Interválica.
3. Pruebas de Hipótesis y tipos de errores.
4. Pruebas Chi-cuadrado
5. Test en R
   - Ejemplo de t-student una muestra, dos muestras y muestras parejas.
6. **Diseño experimental – ANOVA**
7. Pruebas no Paramétricas
   - Wilcoxon
   - U de Mann-Whitney
   - K de Kruskall-Wallis
8. Casos aplicados de Geología de los puntos 1 al 7 en Rstudio.

**REPASO5 y REPASO6** : EJERCICIO PARA AFIANZAR LO APRENDIDO

# GUIA PARA CALSIFICAR TEST DE HIPÓTESIS CON VARIABLES DE RESPUESTA CONTINUA

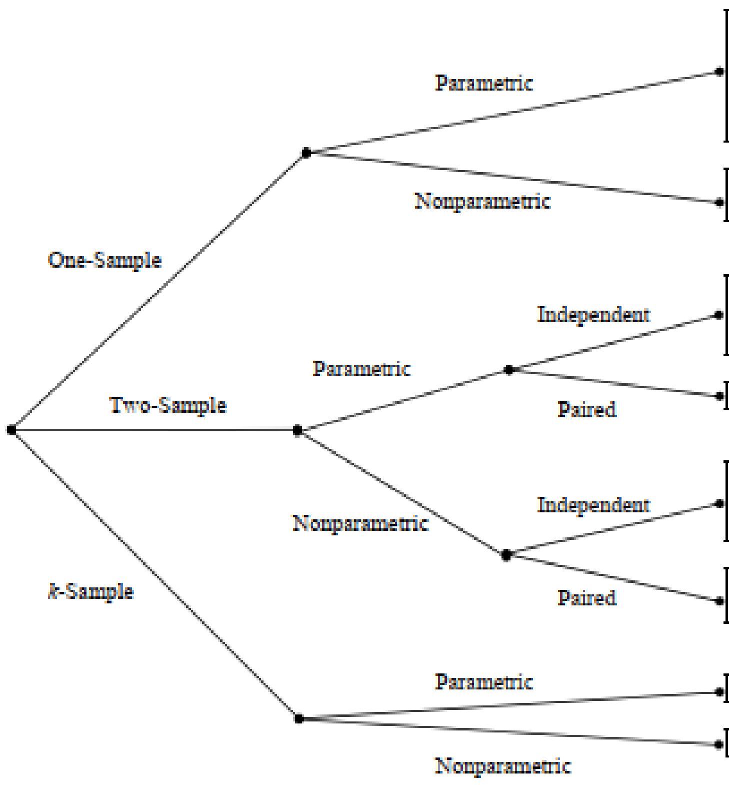
**Table 4.1.** Guide to the classification of some hypothesis tests with continuous response variables.

[-, not applicable]

| Parametric | Nonparametric | Permutation |
|---|---|---|
| | Two independent data groups (chap. 5) | |
| Two-sample *t*-test | Rank-sum test (two-sample Wilcoxon; Mann-Whitney test) | Two-sample permutation test |
| | Matched pairs of data (chap. 6) | |
| Paired *t*-test | Signed-rank test, sign test | Paired permutation test |
| | Three or more independent data groups (chap. 7) | |
| Analysis of variance | Kruskal-Wallis test | One-way permutation test |
| | Three or more dependent data groups (chap. 7) | |
| Analysis of variance without replication | Friedman test, aligned-rank test | - |
| | Two-factor group comparisons (chap. 7) | |
| Two-factor analysis of variance | Brunner-Dette-Munk (BDM) test | Two-factor permutation test |
| | Correlation between two continuous variables (chap. 8) | |
| Pearson's $r$ (linear correlation) | Spearman's $\rho$ or Kendall's $\tau$ (monotonic correlation) | Permutation test for Pearson's $r$ |
| | Model of relation between two continuous variables (chaps. 9 and 10) | |
| Linear regression | Theil-Sen line | Bootstrap of linear regression |

Referencia: Statical Methods in Water Resource 2020 , USGS

# ÁRBOL DE DECISIÓN PARA EL USO DE TES ESTADÍSTICOS PARAMÉTRICOS Y NO PARAMÉTRICOS




**Decision Tree for Selecting the Specific Method**

| Test | Population Parameter | Distributional Assumption | Section |
|---|---|---|---|
| t-Test and CI | *Mean* | Normal | 3.2.1.1 |
| Stratified t-Test | *Mean* | Normal | 3.2.1.3 |
| Chen Test | *Mean* | Right-Skewed | 3.2.1.4 |
| Land's CI Method | *Mean* | Lognormal | 3.2.1.5 |
| Test for a Proportion and CI | *Proportion* | | 3.2.1.6 |
| Sign Test | *Median* | None | 3.2.2.1 |
| Wilcoxon Signed Ranks Test | *Median /Mean* | Symmetric | 3.2.2.2 |
| t-Test and CI (equal variances) | *Diff in Means* | Normal | 3.3.1.1.1 |
| t-Test and CI (unequal variances) | *Diff in Means* | Normal | 3.3.1.1.2 |
| Test for Proportions and CI | *Diff in Props* | | 3.3.1.1.3 |
| Paired t-Test | *Diff in Means* | Normal | 3.3.1.2.1 |
| Wilcoxon Rank Sum Test | *Diff in Means* | Same Variance | 3.3.2.1.1 |
| Quantile Test | Right-Tail | | 3.3.2.1.2 |
| Slippage Test | Right-Tail | | 3.3.2.1.3 |
| Sign Test | *Median* | None | 3.3.2.2.1 |
| Wilcoxon Signed Ranks Test | *Median /Mean* | Symmetric | 3.3.2.2.2 |
| Dunnett's Test | *Mean* | | 3.4.1.1 |
| Kruskal-Wallis Test | *Mean* | | 3.4.1.2 |

EPA QA/G-9S 47 DRAFT FINAL September 2004

Referencia: EPA QA/G-9S

# ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

dicotómicas

politómicas

**Chi cuadrado**

**Chi cuadrado**

Prueba exacta de Fisher

Analytics AoZ

# ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

Cualitativa dicotómica

**t student**

Wilcoxon-Mann-Withney

Cualitativa politómica

**Anova**

Kruskal Wallis

Leandro Huayanay Falconi

# ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

Gráficos

politómicas

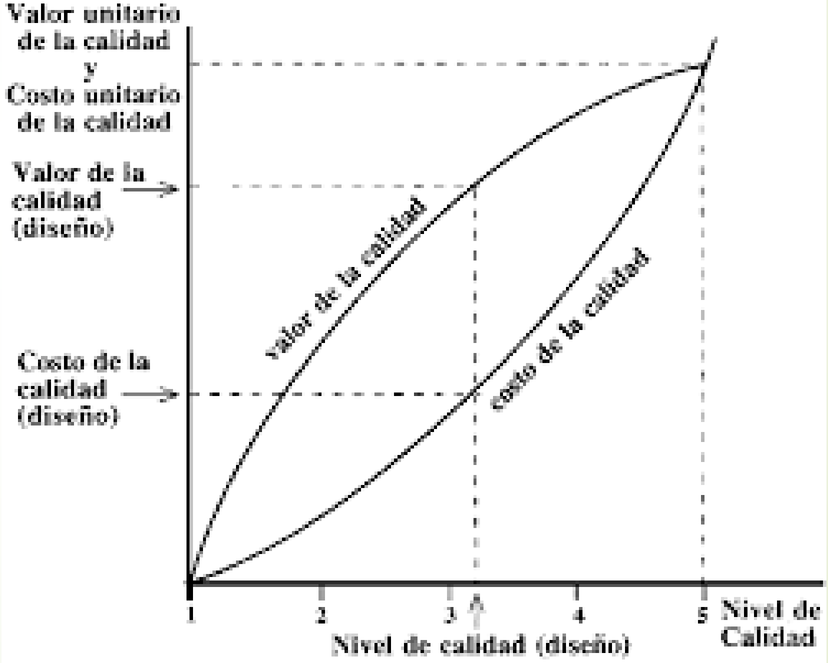
**Grafico de dispersión**

**Correlación de Pearson**

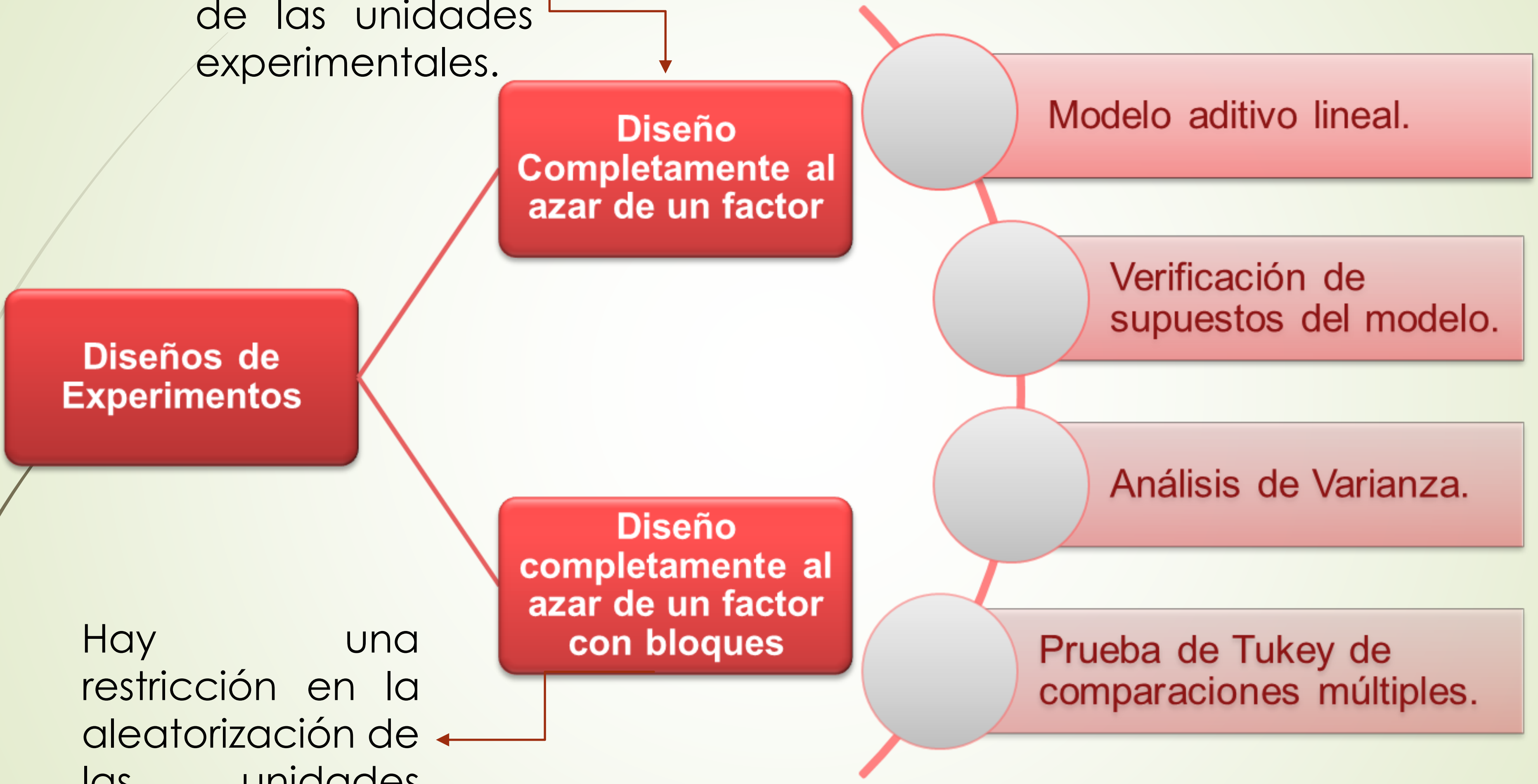
**Regresión lineal**

Sobrevida

# Análisis de Datos Experimentales



No hay restricciones en la aleatorización de las unidades experimentales.




Hay una restricción en la aleatorización de las unidades experimentales.

En estadística, la principal preocupación es el estudio de la variabilidad de la variable respuesta.

# Conceptos Básicos

Analytics AoZ

El término "*experimento*" se refiere a la creación y preparación de lotes de prueba que verifiquen la validez de las hipótesis establecidas sobre las causas de un determinado problema o defecto, objeto de estudio.

En un experimento, el "*experimentador*" escoge ciertos factores para su estudio, los altera deliberadamente de forma controlada y después, observa el efecto resultante.

El "*diseño*" comprende la forma de aplicar los tratamientos a las unidades experimentales y mediante un modelo estadístico se cuantifica la variación debido a factores controlables y no controlables.

# Conceptos Básicos

**Factor (o tratamiento):** Es una variable de interés y los valores que asume son denominados niveles. Esta variable manipula el investigador para estudiar sus efectos, también se llama variable independiente.

**Variable Respuesta:** Es la variable de estudio, aquella cuyos cambios se desean estudiar, también se llama variable dependiente.

**Niveles del Factor:** Viene a ser cada una de categorías, valores o formas específicas del factor (o Tratamiento). Los niveles puede ser de tipo cualitativo o cuantitativo.

**Unidad Experimental:** Viene a ser la unidad expuesta al tratamiento.

# Diseño completamente aleatorio de un factor

Este es el diseño en el cual los tratamientos se asignan al azar entre las unidades experimentales

Este diseño de es aplicable solo cuando las unidades experimentales son homogéneas.

La variable respuesta depende de la influencia de un único factor, bajo el supuesto de que el experimento ha sido completamente aleatorio

# EL MODELO ADITIVO LINEAL:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

$$\text{Siendo: } \quad j = 1, \cdots, k \qquad i = 1, \cdots, n_j$$

donde:

* $y_{ij}$ es la **variable respuesta**,
* $\mu$ es la **media global**,
* $\alpha_j$ es el **efecto del j-ésimo tratamiento** y
* $\varepsilon_{ij}$ es el ***error aleatorio o residual***

Analytics AoZ

# Supuestos del Modelo

1. El valor esperado de los residuales es: $E(\varepsilon_{ij}) = 0$.
2. La varianza de los residuales es $V(\varepsilon_{ij}) = \sigma^2$ (Homocedasticidad)
3. La residuales se ajustan a una Distribución Normal: $\varepsilon_{ij} \sim N(0, \sigma^2)$ (Normalidad)
4. Los errores aleatorios o residuales $\varepsilon_{ij}$ son independientes

Los datos de las muestras se disponen de acuerdo a la siguiente tabla:

| Observaciones | Muestra o Tratamiento | | | | Gran Total |
|---|---|---|---|---|---|
| | 1 | 2 | ... | k | |
| 1 | $y_{11}$ | $y_{12}$ | ... | $y_{1k}$ | |
| 2 | $y_{21}$ | $y_{22}$ | | $y_{2k}$ | |
| ... | ... | ... | ... | ... | |
| $n_j$ | $y_{n_1 2}$ | $y_{n_2 2}$ | | $y_{n_k k}$ | |
| Total Columna | $y_{*1}$ | $y_{*2}$ | ... | $y_{*k}$ | $y_{**}$ |
| Tamaño de muestra | $n_1$ | $n_2$ | ... | $n_k$ | $n$ |

# Cuadro de análisis de varianza (ANOVA)

| Fuente de Variación | Grados de Libertad GL | Suma de Cuadrados SC | Cuadrados Medios CM | F |
|---|---|---|---|---|
| **Entre grupos** | k – 1 | $SCTrat = \sum_{j=1}^{k} \frac{y_{*j}^2}{n_j} - \frac{y_{**}^2}{n}$ | $CMTrat = \frac{SCTrat}{k-1}$ | $F_c = \frac{CMTrat}{CME}$ |
| **Dentro de los grupos (error)** | n – k | SCE = SCTotal –SCTrat | $CME = \frac{SCE}{n-k}$ | |
| Total | n –1 | $SCTotal = \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{y_{**}^2}{n}$ | | |

# Prueba de normalidad de los residuales

Analytics AoZ

### 1. Planteamiento de hipótesis

- $H_0$ : Los residuales se ajustan a la distribución normal
- $H_1$ : Los residuales no se ajustan a la distribución normal

### 2. Nivel de significación

- Los valores pueden estar de $0 \leq \alpha \leq 1$

### 3. Estadística de prueba

- Estadístico de la prueba **Anderson-Darling**

### 4. Valor crítico

- Determinar el p-valor

### 5. Regla de decisión

- $H_0$ se rechaza sí: p-valor < $\alpha$, caso contrario se no se rechaza $H_0$

# Prueba de **homocedasticidad** de los residuales

Analytics AoZ

1. **Planteamiento de hipótesis**
   - $H_0$ : Los residuales presentan homocedasticidad
   - $H_1$ : Los residuales no presentan homocedasticidad
2. **Nivel de significación**
   - Los valores pueden estar de $0 \leq \alpha \leq 1$
3. **Estadística de prueba**
   - Estadístico de la prueba de **Bartlett**
4. **Valor crítico**
   - Determinar el p-valor
5. **Regla de decisión**
   - $H_0$ se rechaza sí: p-valor < $\alpha$, caso contrario se no se rechaza $H_0$

# Prueba de hipótesis principal de DCA

Analytics AoZ

## 1. Planteamiento de hipótesis

$H_0 : \mu_1 = \mu_2 = ... = \mu_k$

$H_1$ : Al menos un $\mu_i \neq \mu_j$

## 2. Nivel de significación

Los valores pueden estar de $0 \leq \alpha \leq 1$

## 3. Estadística de prueba

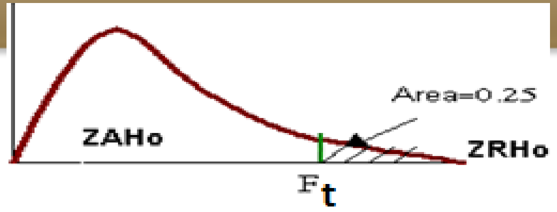
$$F_c = \frac{CMTrat}{CME}$$

## 4. Distribución de la estadístico de prueba

Distribución F de k-1 y n-k grados de libertad

## 5. Valor crítico

Región crítica a la derecha $\mathbf{F}_{(k-1;\ n-k,\ 1-\alpha)}$



## 6. Regla de decisión

- Si p-valor < $\alpha$, entonces $H_0$ se rechaza, caso contrario no se rechaza
- Si $F > F_{(k-1;\ n-k,\ 1-\alpha)}$, entonces $H_0$ se rechaza, caso contrario no se rechaza

## 7. Decisión y conclusión

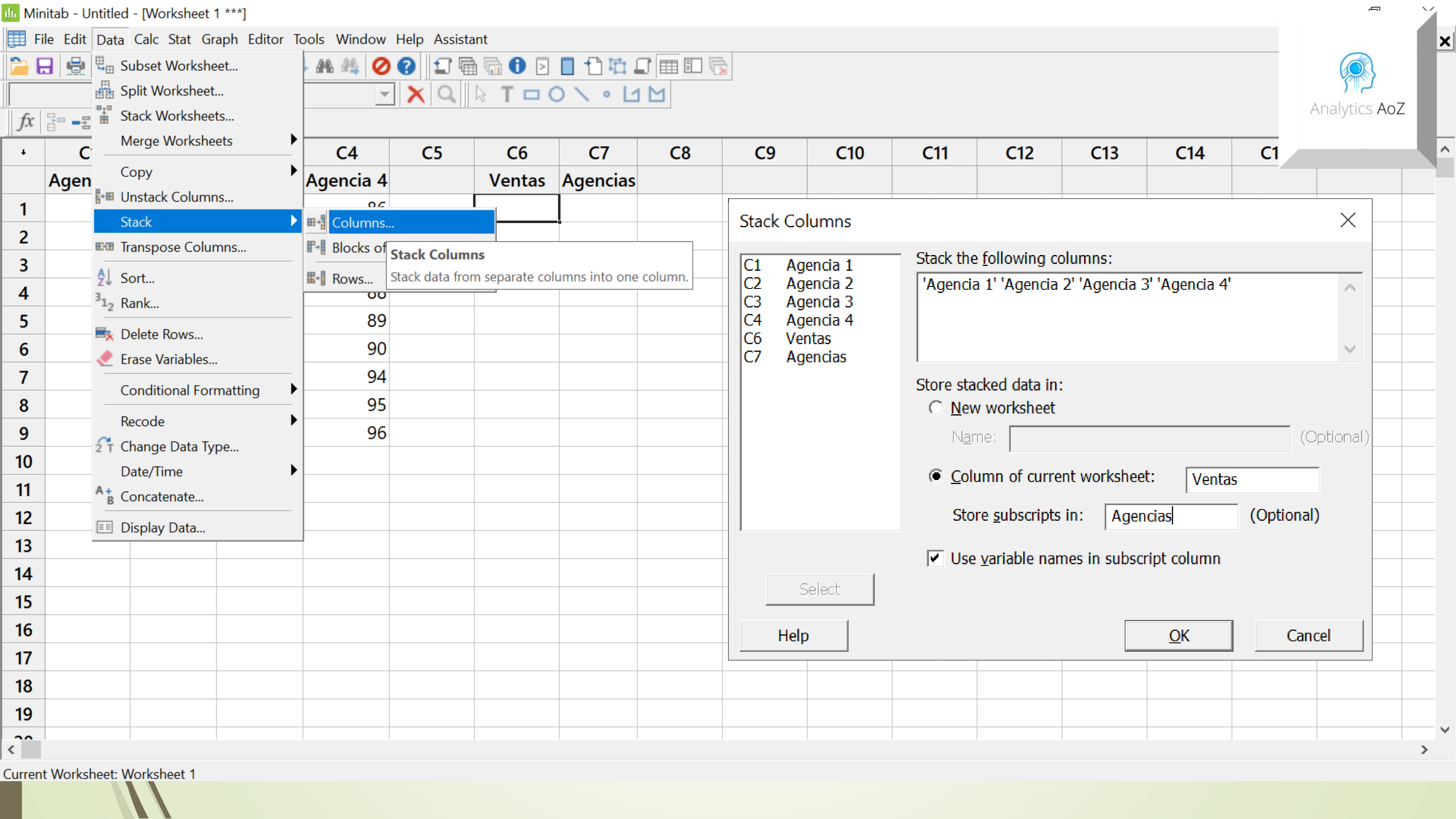
# Comparaciones Múltiples

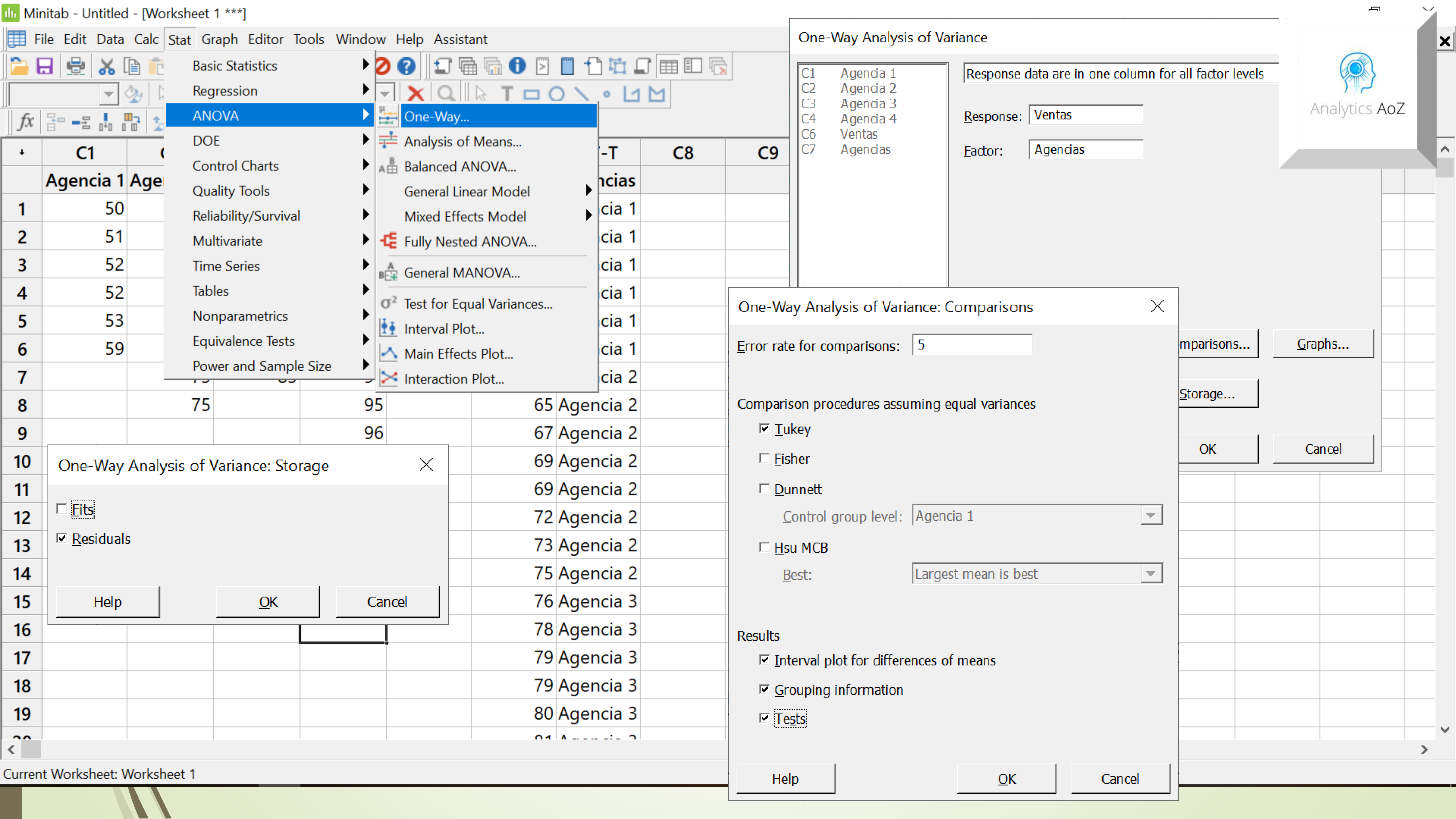
Si la hipótesis nula de la prueba ANOVA ha sido rechazada, se debe determinar cual o cuales de las medias son diferentes y el método que se utilizará es el de Tukey.

# Ejemplo de diseño completamente aleatorio de un factor

Una empresa comercializadora tiene cuatro agencias, las cuales cuentan con su propio equipo de vendedores. Con el objeto de incrementar las ventas se sometió a todos los vendedores a un nuevo programa de entrenamiento, y luego a cada vendedor le fue asignada aleatoriamente una zona. Al final de la primera semana posterior al entrenamiento, se registraron sus ventas. ¿Cree usted que se puede afirmar que las ventas en promedio son diferentes en las cuatro agencias?

| Agencia 1 | Agencia 2 | Agencia 3 | Agencia 4 |
|---|---|---|---|
| 50 | 62 | 76 | 86 |
| 51 | 65 | 78 | 87 |
| 52 | 67 | 79 | 87 |
| 52 | 69 | 79 | 88 |
| 53 | 69 | 80 | 89 |
| 59 | 72 | 81 | 90 |
| | 73 | 83 | 94 |
| | 75 | | 95 |
| | | | 96 |


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Subset Worksheet...
Split Worksheet...
Stack Worksheets...
Merge Worksheets
Copy
Unstack Columns...
Stack
Columns...
Blocks of
Rows...
Stack Columns
Stack data from separate columns into one column.
Transpose Columns...
Sort...
Rank...
Delete Rows...
Erase Variables...
Conditional Formatting
Recode
Change Data Type...
Date/Time
Concatenate...
Display Data...
Analytics AoZ
C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14
Agencia 4 Ventas Agencias
89
90
94
95
96
Stack Columns
C1 Agencia 1
C2 Agencia 2
C3 Agencia 3
C4 Agencia 4
C6 Ventas
C7 Agencias
Stack the following columns:
'Agencia 1' 'Agencia 2' 'Agencia 3' 'Agencia 4'
Store stacked data in:
New worksheet
Name: (Optional)
Column of current worksheet: Ventas
Store subscripts in: Agencias (Optional)
Use variable names in subscript column
Select
Help
OK
Cancel
Current Worksheet: Worksheet 1


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
Equivalence Tests
Power and Sample Size
One-Way...
Analysis of Means...
Balanced ANOVA...
General Linear Model
Mixed Effects Model
Fully Nested ANOVA...
General MANOVA...
Test for Equal Variances...
Interval Plot...
Main Effects Plot...
Interaction Plot...
C1
Agencia 1
C8
C9
50
51
52
52
53
59
75
95
96
65 Agencia 2
67 Agencia 2
69 Agencia 2
69 Agencia 2
72 Agencia 2
73 Agencia 2
75 Agencia 2
76 Agencia 3
78 Agencia 3
79 Agencia 3
79 Agencia 3
80 Agencia 3
One-Way Analysis of Variance: Storage
Fits
Residuals
Help
OK
Cancel
One-Way Analysis of Variance
C1 Agencia 1
C2 Agencia 2
C3 Agencia 3
C4 Agencia 4
C6 Ventas
C7 Agencias
Response data are in one column for all factor levels
Response: Ventas
Factor: Agencias
Graphs...
Storage...
OK
Cancel
Analytics AoZ
One-Way Analysis of Variance: Comparisons
Error rate for comparisons: 5
Comparison procedures assuming equal variances
Tukey
Fisher
Dunnett
Control group level: Agencia 1
Hsu MCB
Best: Largest mean is best
Results
Interval plot for differences of means
Grouping information
Tests
Help
OK
Cancel
Current Worksheet: Worksheet 1

Minitab - Untitled - [Worksheet 1 ***]

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Analytics AoZ

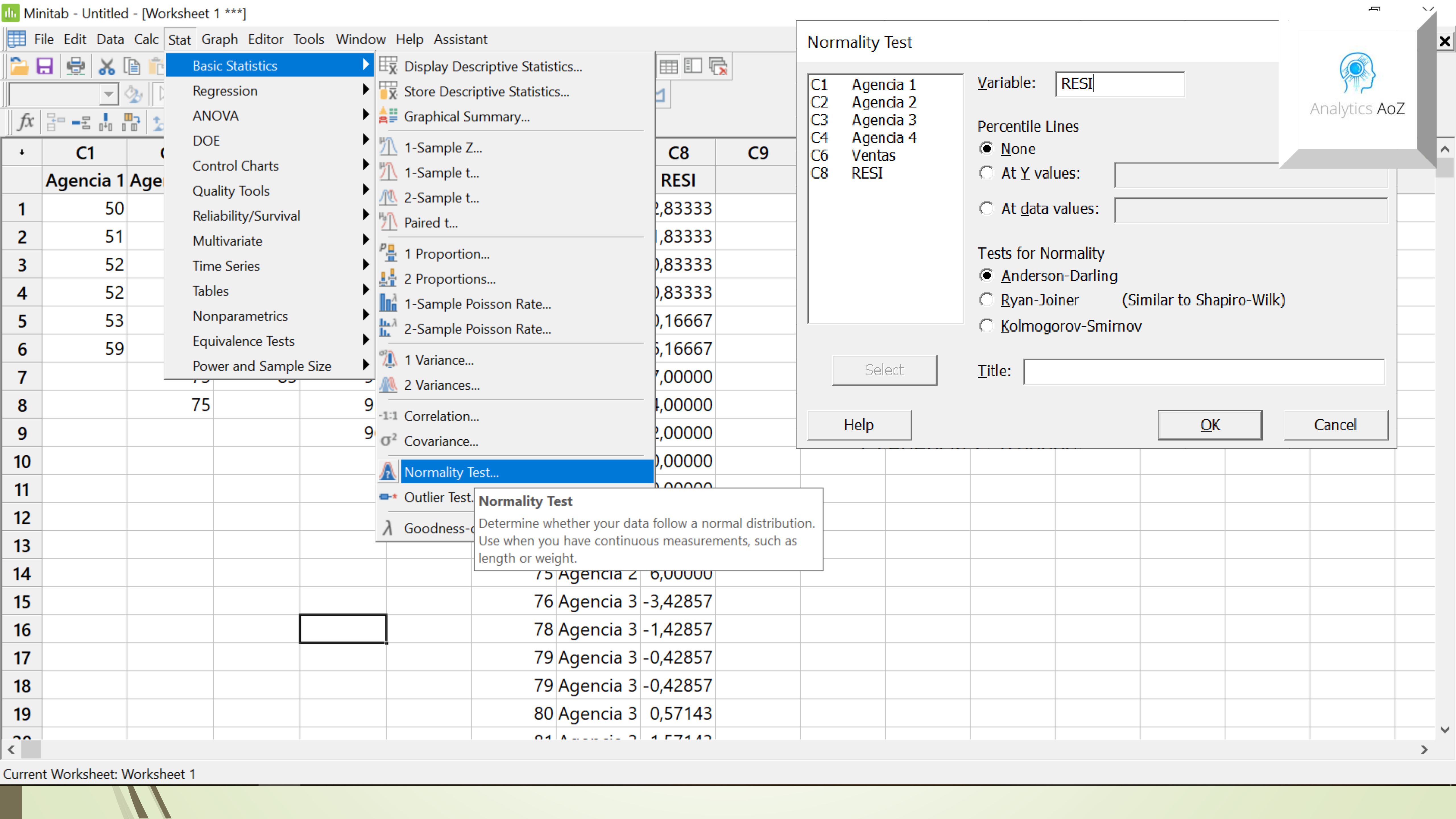
| | C1 | C2 | C3 | C4 | C5 | C6 | C7-T | C8 |
|---|---|---|---|---|---|---|---|---|
| | Agencia 1 | Agencia 2 | Agencia 3 | Agencia 4 | | Ventas | Agencias | RESI |
| 1 | 50 | 62 | 76 | 86 | | 50 | Agencia 1 | -2,83333 |
| 2 | 51 | 65 | 78 | 87 | | 51 | Agencia 1 | -1,83333 |
| 3 | 52 | 67 | 79 | 87 | | 52 | Agencia 1 | -0,83333 |
| 4 | 52 | 69 | 79 | 88 | | 52 | Agencia 1 | -0,83333 |
| 5 | 53 | 69 | 80 | 89 | | 53 | Agencia 1 | 0,16667 |
| 6 | 59 | 72 | 81 | 90 | | 59 | Agencia 1 | 6,16667 |
| 7 | | 73 | 83 | 94 | | 62 | Agencia 2 | -7,00000 |
| 8 | | 75 | | 95 | | 65 | Agencia 2 | -4,00000 |
| 9 | | | | 96 | | 67 | Agencia 2 | -2,00000 |
| 10 | | | | | | 69 | Agencia 2 | 0,00000 |
| 11 | | | | | | 69 | Agencia 2 | 0,00000 |
| 12 | | | | | | 72 | Agencia 2 | 3,00000 |
| 13 | | | | | | 73 | Agencia 2 | 4,00000 |
| 14 | | | | | | 75 | Agencia 2 | 6,00000 |
| 15 | | | | | | 76 | Agencia 3 | -3,42857 |
| 16 | | | | | | 78 | Agencia 3 | -1,42857 |
| 17 | | | | | | 79 | Agencia 3 | -0,42857 |
| 18 | | | | | | 79 | Agencia 3 | -0,42857 |
| 19 | | | | | | 80 | Agencia 3 | 0,57143 |

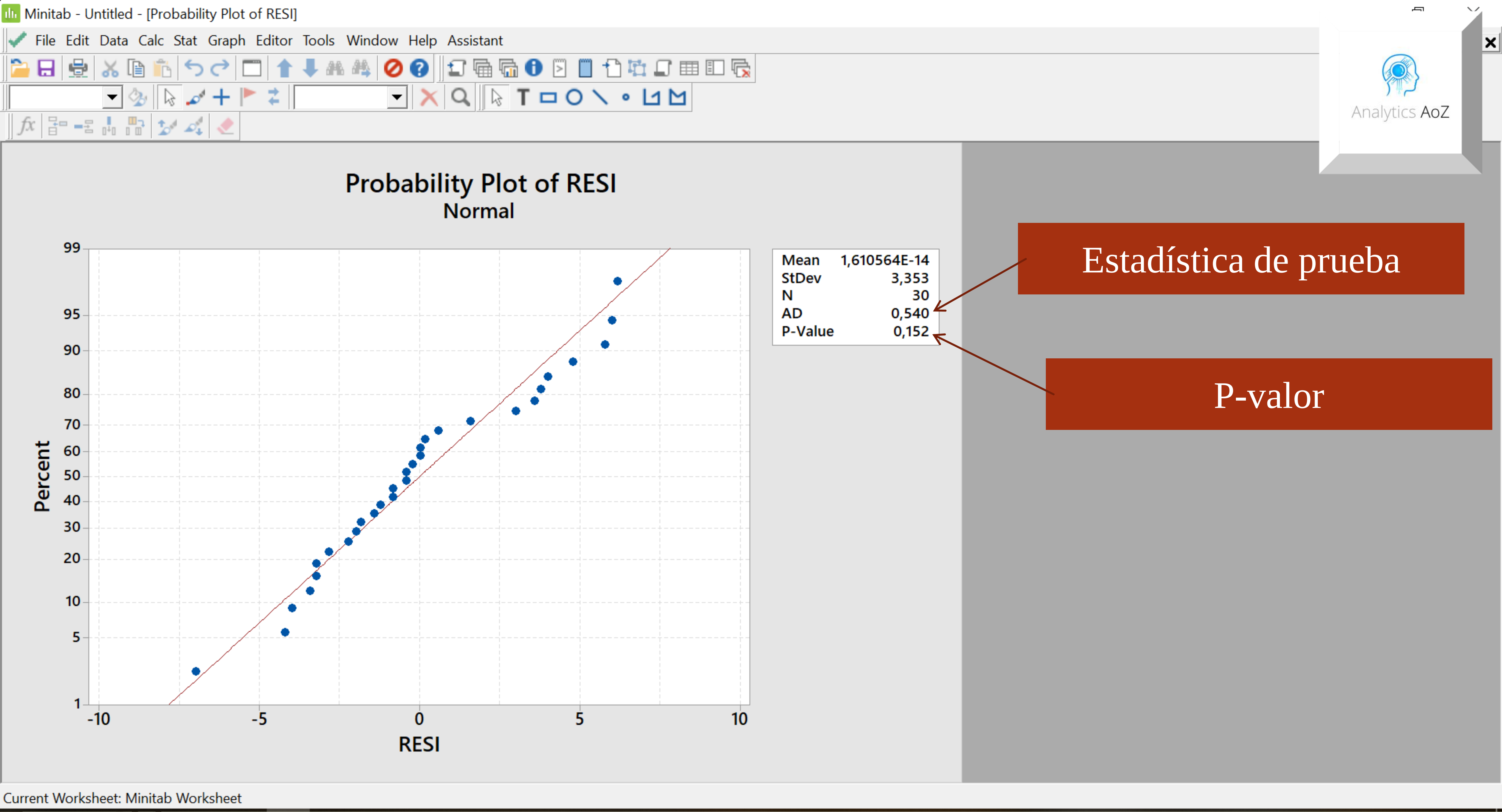
Lo que se debe hacer es probar los supuestos del modelo; es esto:

Primero.- Prueba de normalidad

Segundo.- Prueba de homocedasticidad

Current Worksheet: Worksheet 1


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
Equivalence Tests
Power and Sample Size
Display Descriptive Statistics...
Store Descriptive Statistics...
Graphical Summary...
1-Sample Z...
1-Sample t...
2-Sample t...
Paired t...
1 Proportion...
2 Proportions...
1-Sample Poisson Rate...
2-Sample Poisson Rate...
1 Variance...
2 Variances...
Correlation...
Covariance...
Normality Test...
Outlier Test...
Goodness-
Normality Test
Determine whether your data follow a normal distribution. Use when you have continuous measurements, such as length or weight.
Normality Test
C1 Agencia 1
C2 Agencia 2
C3 Agencia 3
C4 Agencia 4
C6 Ventas
C8 RESI
Variable: RESI
Percentile Lines
None
At Y values:
At data values:
Tests for Normality
Anderson-Darling
Ryan-Joiner (Similar to Shapiro-Wilk)
Kolmogorov-Smirnov
Select
Title:
Help
OK
Cancel
Analytics AoZ
C1
Agencia 1
50
51
52
52
53
59
C8
RESI
C9
75 Agencia 2 6,00000
76 Agencia 3 -3,42857
78 Agencia 3 -1,42857
79 Agencia 3 -0,42857
79 Agencia 3 -0,42857
80 Agencia 3 0,57143
Current Worksheet: Worksheet 1


Minitab - Untitled - [Probability Plot of RESI]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Probability Plot of RESI
Normal
Percent
RESI
99
95
90
80
70
60
50
40
30
20
10
5
1
-10
-5
0
5
10
Mean 1,610564E-14
StDev 3,353
N 30
AD 0,540
P-Value 0,152
Estadística de prueba
P-valor
Analytics AoZ
Current Worksheet: Minitab Worksheet

# Prueba de normalidad de los residuales

**1. Planteamiento de hipótesis**

- $H_0$ : Los residuales se ajustan a la distribución normal
- $H_1$ : Los residuales no se ajustan a la distribución normal

**2. Nivel de significación**

- Se asume $\alpha = 0,05$

**3. Estadística de prueba**

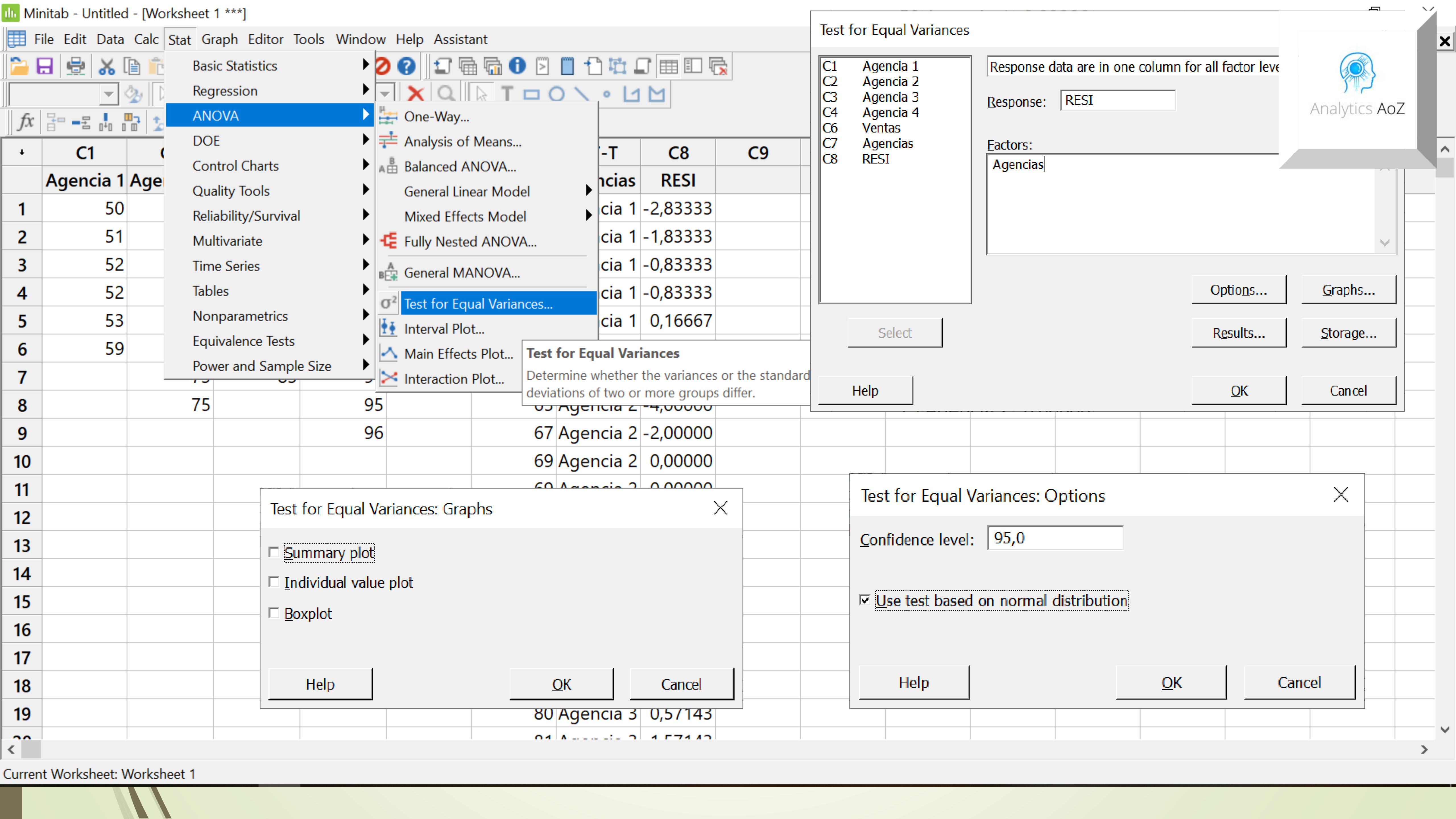
- Estadístico de la prueba Anderson-Darling  AD = 0,540

**4. Valor crítico**

- En este caso p-valor = 0,152

**5. Decisión / conclusión**

- $H_0$ no se rechaza. Los residuales se ajustan a una distribución normal


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
Equivalence Tests
Power and Sample Size
One-Way...
Analysis of Means...
Balanced ANOVA...
General Linear Model
Mixed Effects Model
Fully Nested ANOVA...
General MANOVA...
Test for Equal Variances...
Interval Plot...
Main Effects Plot...
Interaction Plot...
Test for Equal Variances
Determine whether the variances or the standard deviations of two or more groups differ.
C1 Agencia 1
C8 RESI
C9
Test for Equal Variances
C1 Agencia 1
C2 Agencia 2
C3 Agencia 3
C4 Agencia 4
C6 Ventas
C7 Agencias
C8 RESI
Response data are in one column for all factor leve
Response: RESI
Factors:
Agencias
Select
Options...
Graphs...
Results...
Storage...
Help
OK
Cancel
Analytics AoZ
Test for Equal Variances: Graphs
Summary plot
Individual value plot
Boxplot
Help
OK
Cancel
Test for Equal Variances: Options
Confidence level: 95,0
Use test based on normal distribution
Help
OK
Cancel
Current Worksheet: Worksheet 1

Minitab - Untitled - [Session]

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Analytics AoZ

Interval Plot of Ventas vs Agencias

# Test for Equal Variances: RESI versus Agencias

## Method

| | |
|---|---|
| Null hypothesis | All variances are equal |
| Alternative hypothesis | At least one variance is different |
| Significance level | α = 0,05 |

*Bartlett's method is used. This method is accurate for normal data only.*

## 95% Bonferroni Confidence Intervals for Standard Deviations

| Agencias | N | StDev | CI |
|---|---|---|---|
| Agencia 1 | 6 | 3,18852 | (1,77045; 10,5956) |
| Agencia 2 | 8 | 4,30946 | (2,56870; 11,0595) |
| Agencia 3 | 7 | 2,22539 | (1,28509; 6,3668) |
| Agencia 4 | 9 | 3,80058 | (2,32592; 8,9767) |

*Individual confidence level = 98,75%*

Estadística de prueba

## Tests

| Method | Test Statistic | P-Value |
|---|---|---|
| Bartlett | 2,58 | 0,461 |

P-valor

Current Worksheet: Worksheet 1

# Prueba de homocedasticidad de los residuales

Analytics AoZ

### 1. Planteamiento de hipótesis

- $H_0$ : Los residuales presentan homocedasticidad
- $H_1$ : Los residuales no presentan homocedasticidad

### 2. Nivel de significación

- Se asume $\alpha = 0{,}05$

### 3. Estadística de prueba

- Estadístico de la prueba de Bartlett = 2,58

### 4. Valor crítico

- En este caso p-valor = 0,461

### 5. Decisión / conclusión

- $H_0$ no se rechaza. Los residuales presentan homocedasticidad

# Ejemplo de diseño completamente aleatorio de un factor

Analytics AoZ

| | Agencia 1 | Agencia 2 | Agencia 3 | Agencia 4 | |
|---|---|---|---|---|---|
| | 50 | 62 | 76 | 86 | |
| | 51 | 65 | 78 | 87 | |
| | 52 | 67 | 79 | 87 | |
| | 52 | 69 | 79 | 88 | |
| | 53 | 69 | 80 | 89 | |
| | 59 | 72 | 81 | 90 | |
| | | 73 | 83 | 94 | |
| | | 75 | | 95 | |
| | | | | 96 | |
| sumas | 317 | 552 | 556 | 812 | 2237 |
| muestra | 6 | 8 | 7 | 9 | 30 |

$$SCTotal = \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}^2 - \frac{y_{**}^2}{n} = \left[ (50^2 + \cdots + 59^2) + \cdots + (86^2 + \ldots + 96^2) \right] - \frac{2237^2}{30} = 5779{,}37$$

$$SCTrat = \sum_{j=1}^{k} \frac{y_{*j}^2}{n_j} - \frac{y_{**}^2}{n} = \left( \frac{317^2}{6} + \ldots + \frac{812^2}{9} \right) - \frac{2237^2}{30} = 5453{,}26$$

$$SCE = SCTotal\text{-}SCTrat = 5779{,}37 - 5453{,}26 = 326{,}11$$

# Ejemplo de diseño completamente aleatorio de un factor

## Cuadro de análisis de varianza (ANOVA)

| Fuente de Variación | Grados de Libertad | Suma de Cuadrados | Cuadrados Medios | F |
|---|---|---|---|---|
| | GL | SC | CM | |
| Entre grupos | 3 | 5453,26 | 1817,75 | 144,93 |
| Dentro de grupos (error) | 26 | 326,11 | 12,54 | |
| Total | 29 | 5779,37 | | |

Minitab - Untitled - [Session]

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Analytics AoZ

# One-way ANOVA: Ventas versus Agencias

## Method

| | |
|---|---|
| Null hypothesis | All means are equal |
| Alternative hypothesis | Not all means are equal |
| Significance level | α = 0,05 |

*Equal variances were assumed for the analysis.*

## Factor Information

| Factor | Levels | Values |
|---|---|---|
| Agencias | 4 | Agencia 1; Agencia 2; Agencia 3; Agencia 4 |

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Agencias | 3 | 5453,3 | 1817,75 | 144,93 | 0,000 |
| Error | 26 | 326,1 | 12,54 | | |
| Total | 29 | 5779,4 | | | |

## Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 3,54153 | 94,36% | 93,71% | 92,57% |

## Means

Current Worksheet: Worksheet 1

# Prueba de hipótesis principal de DCA

Analytics AoZ

**1. Planteamiento de hipótesis**

$H_0$ : $\mu_1 = \mu_2 = ... = \mu_k$ / Las ventas en promedio son iguales en las cuatro agencias

$H_1$ : Al menos un $\mu_i \neq \mu_j$ / Las ventas en promedio es diferente en al menos una de las cuatro agencias

**2. Nivel de significación**

Se asume que $\alpha = 0,05$

**3. Estadística de prueba**

$$F_c = \frac{CMTrat}{CME} = \frac{1817,75}{12,54} = 144,93$$

**4. Distribución de la estadístico de prueba**

Distribución F de k-1 = 3 y n-k = 26 grados de libertad

**5. Valor crítico**

Región crítica a la derecha: $F_{(k-1;\ n-k,\ 1-\alpha)} = F_{(3;\ 26,\ 0,95)} = 2,975$

# Prueba de hipótesis principal de DCA

Analytics AoZ

## 6. Regla de decisión

- Si p-valor < $\alpha$, entonces $H_0$ se rechaza, caso contrario no se rechaza
  Como p-valor = 0,00 < 0,05, $H_0$ se rechaza
- Si F > $F_{(k-1;\ n-k,\ 1-\alpha)}$, entonces H se rechaza, caso contrario no se rechaza
  Como $F_c$ =144,93 > $F_{(k-1;\ n-k,\ 1-\alpha)}$ = 2,975 $H_0$ se rechaza

## 7. Decisión y conclusión

$H_0$ se rechaza.

Al menos un $\mu_i \neq \mu_j$ ; esto es, se puede afirmar que las ventas en promedio es diferente en al menos una de las cuatro agencias

Esto significa que se debe de determinar cual de la cuatro agencia genera la diferencia.

***Para ello es necesario desarrollar la prueba de Tuckey de comparación de medias.***

Minitab - Untitled - [Session]

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

## Tukey Pairwise Comparisons

### Grouping Information Using the Tukey Method and 95% Confidence

| Agencias | N | Mean | Grouping | | | |
|---|---|---|---|---|---|---|
| Agencia 4 | 9 | 90,22 | A | | | |
| Agencia 3 | 7 | 79,429 | | B | | |
| Agencia 2 | 8 | 69,00 | | | C | |
| Agencia 1 | 6 | 52,83 | | | | D |

*Means that do not share a letter are significantly different.*

Se puede concluir que :

$\mu_4 > \mu_3 > \mu_2 > \mu_1$

Por tanto, se puede afirmar que las cuatro agencias tiene en promedios ventas diferentes.
Asimismo, la mejor es la agencia 4 y la agencia 1 tiene menor rendimiento.

### Tukey Simultaneous Tests for Differences of Means

| Difference of Levels | Difference of Means | SE of Difference | 95% CI | T-Value | Adjusted P-Value |
|---|---|---|---|---|---|
| Agencia 2 - Agencia 1 | 16,17 | 1,91 | (10,92; 21,41) | 8,45 | 0,000 |
| Agencia 3 - Agencia 1 | 26,60 | 1,97 | (21,19; 32,00) | 13,50 | 0,000 |
| Agencia 4 - Agencia 1 | 37,39 | 1,87 | (32,27; 42,51) | 20,03 | 0,000 |
| Agencia 3 - Agencia 2 | 10,43 | 1,83 | (5,40; 15,46) | 5,69 | 0,000 |
| Agencia 4 - Agencia 2 | 21,22 | 1,72 | (16,50; 25,94) | 12,33 | 0,000 |
| Agencia 4 - Agencia 3 | 10,79 | 1,78 | (5,90; 15,69) | 6,05 | 0,000 |

*Individual confidence level = 98,91%*

$\mu_2 < \mu_1$

$\mu_3 < \mu_1$

$\mu_4 < \mu_1$

$\mu_3 < \mu_2$

$\mu_4 < \mu_2$

$\mu_4 < \mu_3$

### Tukey Simultaneous 95% CIs

### Interval Plot of Ventas vs Agencias



Current Worksheet: Worksheet 1

# ANOVA TEST IN RSTUDIO:

El test de una dirección análisis de varianza (ANOVA), además conocido como un factor ANOVA, es una extensión del test de dos muestras independientes t-test para comparar medias en situación donde hay más de dos grupos. En un *one-way ANOVA,* la data es organizada dentro de varios grupos base o un grupo simple variable (además llamado *factor variable)*.

**ANOVA test hypotheses**:

- Null hypothesis: the means of the different groups are the same
- Alternative hypothesis: At least one sample mean is not equal to the others.

Note that, if you have only two groups, you can use t-test. In this case the F-test and the t-test are equivalent.

# SUPUESTOS DEL TEST ANOVA:

- Las observaciones son obtenidas independientemente y aleatorias desde una población definida por niveles de factores.
- La data de cada nivel de factor está normalmente distribuida.
- Estas poblaciones normales tiene una varianza común. (Se puede usar el **Levenes´s test**)

Existen diferentes pruebas que permiten evaluar la distribución de la varianza. Todos ellos consideran como hipótesis nula que la varianza es igual entre los grupos y como hipótesis alternativa que no lo es. La diferencia entre ellos es el estadístico de centralidad que utilizan:

**F-test (razón de varianzas):** El F-test, también conocido como contraste de la razón de varianzas, contrasta la hipótesis nula de que dos poblaciones normales tienen la misma varianza. Es muy potente, detecta diferencias muy sutiles, pero es muy sensible a violaciones de la normalidad de las poblaciones. Por esta razón, no es un test recomendable si no se tiene mucha certeza de que las poblaciones se distribuyen de forma normal.

**La prueba de Levene:** Puede comparar 2 o más poblaciones, por permitir elegir entre diferentes estadísticos de centralidad :mediana (por defecto), media, media truncada. Esto es importante a la hora de contrastar la homocedasticidad dependiendo de si los grupos se distribuyen de forma normal o no.

**La prueba de Bartlett:** Permite contrastar la igualdad de varianza en 2 o más poblaciones sin necesidad de que el tamaño de los grupos sea el mismo. Es más sensible que el test de Levene a la falta de normalidad, pero si se está seguro de que los datos provienen de una distribución normal, es la mejor opción.

**En resumen**

Si se tiene seguridad de que las muestras a comparar proceden de poblaciones que siguen una distribución normal, son recomendables el F-test y el test de Bartlet, pareciendo ser el segundo más recomendable ya que el primero es muy potente pero extremadamente sensible a desviaciones de la normal. Si no se tiene la seguridad de que las poblaciones de origen son normales, se recomiendan el test de Leven utilizando la mediana o el test no paramétrico.

# ¿CÓMO TRABAJA EL ONE-WAY ANOVA TEST?

Asumimos que temenos tres grupos (A, B, C) para comparar:

Computamos la varianza común, la cual es llamada la varianza dentro muestras (S2 dentro) o varianza residual.
Computamos la varianza entre la media muestras como sigue:

- Computamos la media para cada grupo.
- Computamos la varianza entre las medias muestrales (S2 entre).
- Producimos el F-estadiístico como el ratio de S2entre/S2dentro.

Nota:
A menor ratio (ratio<1) indica que no existe diferencia significative entre las medias de las muestras que estan siendo comparadas. Sin embargo, a mayor ratio implica que la cantidad de variación entre grupos de medias son significativas.

# Ejemplo One-Way ANOVA

Analytics AoZ

El siguiente ejemplo trata de verificar si existen diferencias entre tres diversos grupos los cuales corresponden a tratamientos realizados y un control para verificación respecto a los pesos, los datos se muestran parcialmente a continuación:

| | weight | group |
|---|---|---|
| 19 | 4.32 | trt1 |
| 18 | 4.89 | trt1 |
| 29 | 5.80 | trt2 |
| 24 | 5.50 | trt2 |
| 17 | 6.03 | trt1 |
| 1 | 4.17 | ctrl |
| 6 | 4.61 | ctrl |
| 16 | 3.83 | trt1 |
| 12 | 4.17 | trt1 |
| 15 | 5.87 | trt1 |

VAMOS AL R!!

## REFERENCIAS


http://www.sthda.com/english/wiki/one-way-anova-test-in-r#what-is-one-way-anova-test

https://www.datanovia.com/en/lessons/anova-in-r/#two-way-independent-anova

https://rpubs.com/Joaquin_AR/219148

https://www.guru99.com/r-anova-tutorial.html


Analytics AoZ