Aguilar Cruz María del Rosario
Solis Carrera Juan Antonio
Instrumentación del laboratorio clínico.
Uso del software NCSS

NCSS es un software estadístico que permite realizar gran variedad de procedimientos estadísticos. En este documento se tomarán en cuenta 6 procedimientos que son los más comunes para el uso de los datos obtenidos en el laboratorio clínico, se ejemplificara cada uno de ellos, así como los pasos a seguir para obtener los reportes.

Prueba t para una muestra.

La prueba t se utiliza para realizar inferencias sobre la población a partir de una muestra de esta.

Ejemplo.

Un investigador desea conocer si la media de una muestra es diferente de 130, de acuerdo con los datos obtenidos.

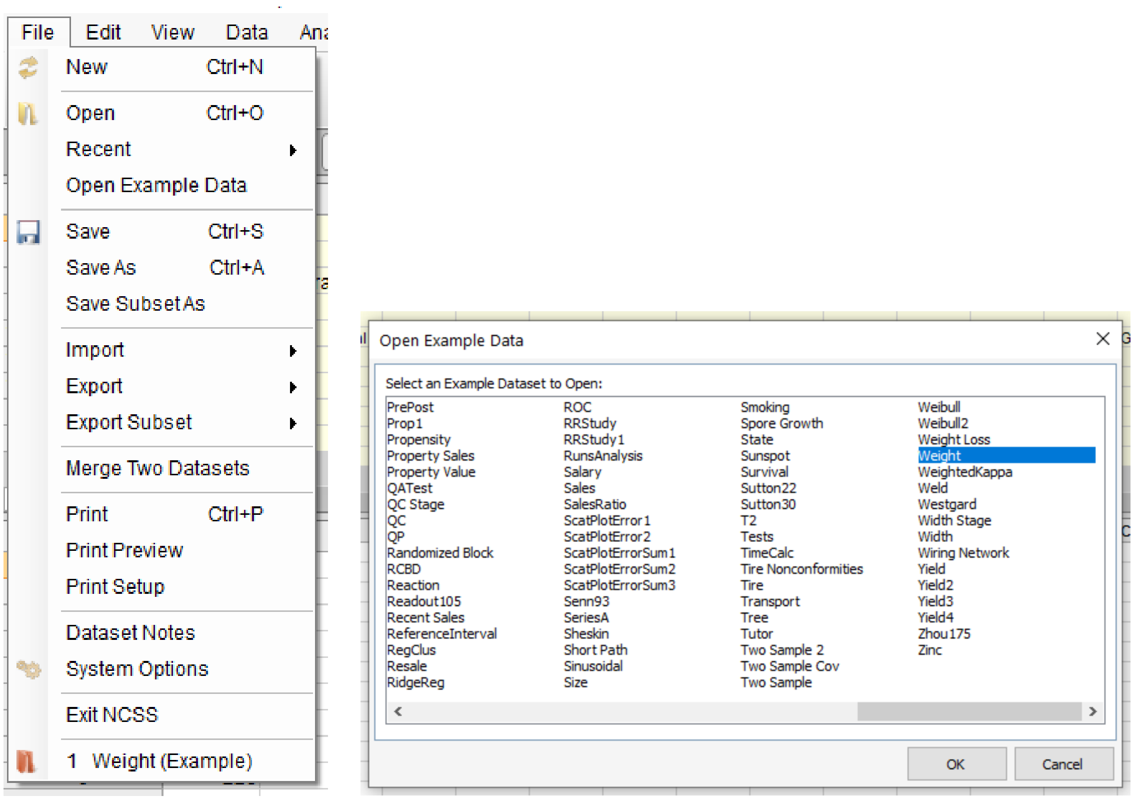


Fig. 1 Menús para cargar los datos del ejemplo

En la figura 1 se ilustra la localización de los menús para cargar los datos del ejemplo del peso para realizar la prueba t para una sola muestra.

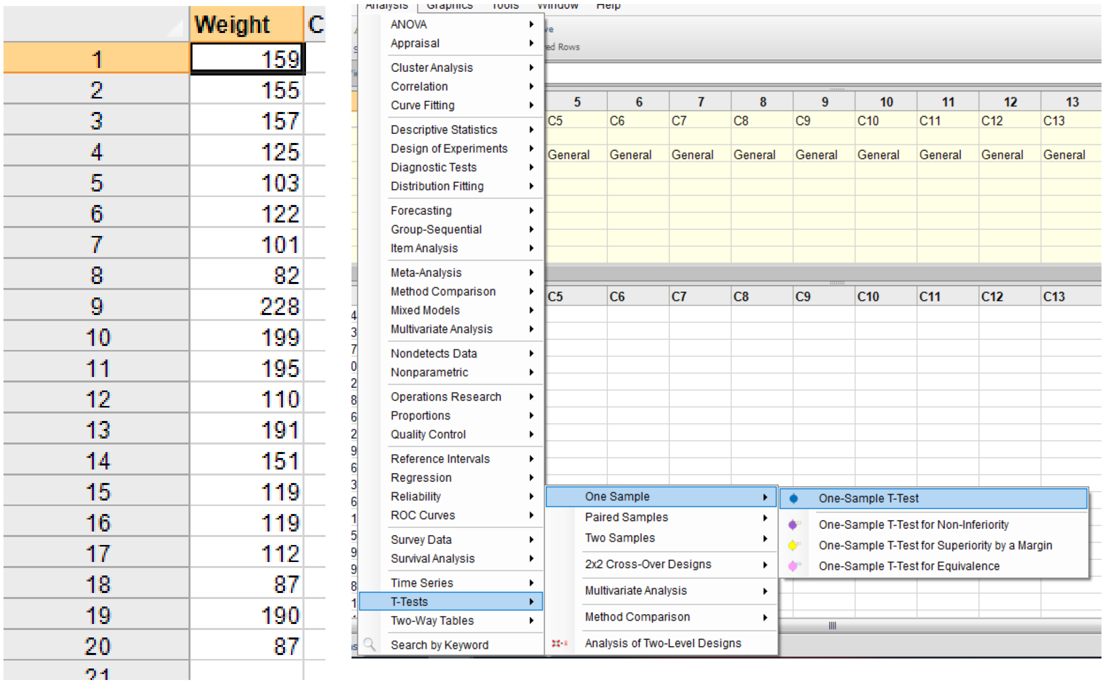
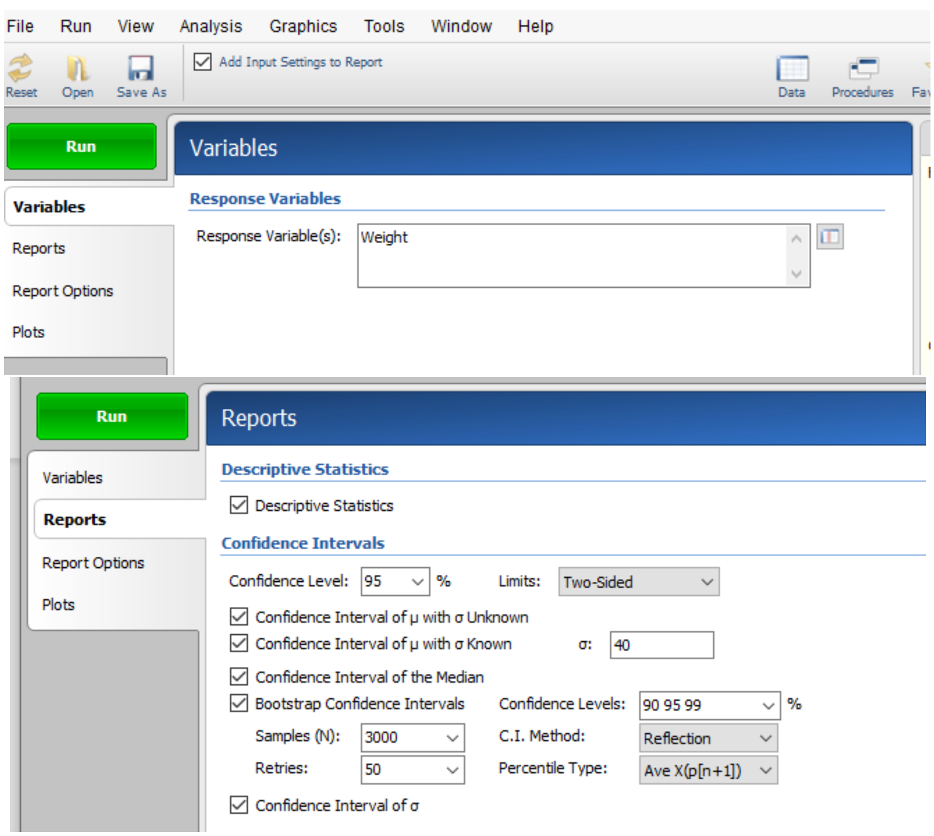


Fig. 2 Datos de la muestra de pesos y el menú en el que se selecciona la prueba t de una sola muestra.

Los datos cargados para realizar la muestra de la prueba t se muestran en la figura 2, así como la imagen del menú donde se encuentra la prueba t de una sola muestra.



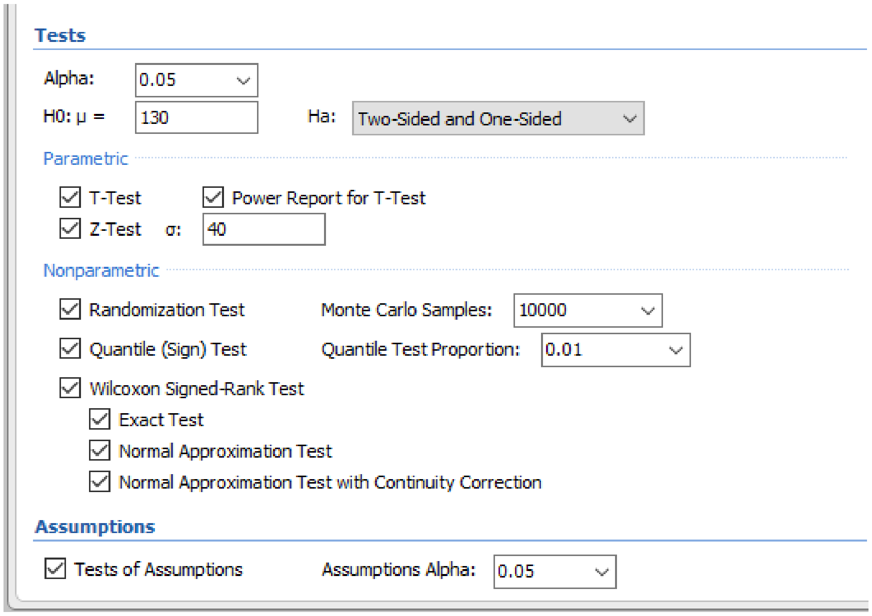


Fig. 3 Muestra de los ajustes posibles para la prueba t de una sola muestra.

Una vez seleccionado el procedimiento de la prueba t para una sola muestra, se debe cargar los datos para los cuales se va a realizar la prueba, como se desea conocer si la media es diferente a 130 se pone como hipótesis nula que la media sea igual a 130, obteniendo así los siguientes resultados.

One-Sample T-Test Report

Dataset C:\...\NCSS\NCSS 2019\Example Data\Weight.NCSS
Response Variable Weight

Descriptive Statistics

Variable	Count	Mean	Standard Deviation	Standard Error	Median
Weight	20	139.6	43.1221	9.642395	123.5

Two-Sided Confidence Interval of μ with σ Unknown

							95.0% C. I. of μ	
Variable	Count	Mean	Standard Deviation	Standard Error	T*	DF	Lower Limit	Upper Limit
Weight	20	139.6	43.1221	9.642395	2.0930	19	119.4182	159.7818

Two-Sided Confidence Interval of μ with σ Known

						95.0% C. I. of μ	
Variable	Count	Mean	σ	Standard Error	Z*	Lower Limit	Upper Limit
Weight	20	139.6	40	8.944272	1.9600	122.0695	157.1304

Two-Sided Confidence Interval of the Median

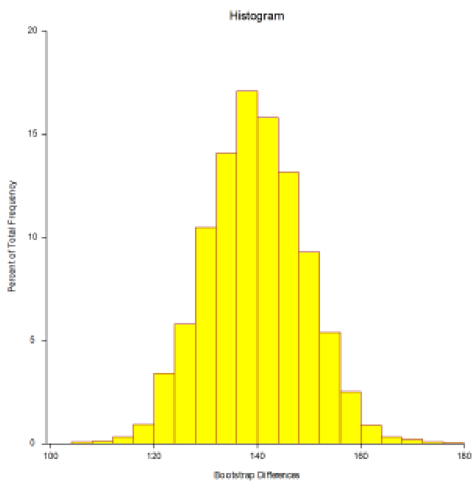
			95.0% C. I. of the Median	
Variable	Count	Median	Lower Limit	Upper Limit
Weight	20	123.5	110	159

Bootstrap Section

----------- Estimation Results -----------		----------- Bootstrap Confidence Limits ---------------		
Parameter	**Estimate**	**Conf. Level**	**Lower**	**Upper**
Mean				
Original Value	139.6000	90.00	124.1050	155.0475
Bootstrap Mean	139.6219	95.00	121.3000	157.6500
Bias (BM - OV)	0.0219	99.00	113.9508	163.8495
Bias Corrected	139.5781			
Standard Error	9.4109			

Confidence Limit Type = Reflection, Number of Samples = 3000.

Bootstrap Histograms Section



Two-Sided Confidence Interval of σ

			95.0% C. I. of σ	
Variable	**Count**	**Standard Deviation**	**Lower Limit**	**Upper Limit**
Weight	20	43.1221	32.79395	62.98292

One-Sample T-Test

Alternative Hypothesis	Mean	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.050?
μ ≠ 130	139.6	9.642395	0.9956	19	0.33195	No
μ < 130	139.6	9.642395	0.9956	19	0.83402	No
μ > 130	139.6	9.642395	0.9956	19	0.16598	No

Power for the One-Sample T-Test

This section assumes the population mean and standard deviation are equal to the sample values.

Alternative Hypothesis	N	μ	σ	Power (α = 0.05)	Power (α = 0.01)
μ ≠ 130	20	139.6	43.1221	0.15711	0.04833
μ < 130	20	139.6	43.1221	0.00458	0.00057
μ > 130	20	139.6	43.1221	0.24681	0.08081

One-Sample Z-Test

Alternative Hypothesis	Mean	σ	Standard Error	Z-Statistic	Prob Level	Reject H0 at α = 0.050?
μ ≠ 130	139.6	40	8.944272	1.0733	0.28313	No
μ < 130	139.6	40	8.944272	1.0733	0.85844	No
μ > 130	139.6	40	8.944272	1.0733	0.14157	No

Randomization Test (Two-Sided)

Alternative Hypothesis: The distribution center is not 130.
Number of Monte Carlo samples: 10000

Test	Prob Level	Reject H0 at α = 0.050?
Randomization Test	0.32530	No

Quantile (Sign) Test

This Quantile test is equivalent to the Sign test if the Quantile Proportion is 0.5.

Null Quantile (Q0)	Quantile Proportion	Number Lower	Number Higher	H1: Q ≠ Q0 Prob Level	H1: Q < Q0 Prob Level	H1: Q > Q0 Prob Level
130	0.01	11	9	0.000000	0.000000	1.000000

Wilcoxon Signed-Rank Test

Sum of Ranks (W)	Mean of W	Std Dev of W	Number of Zeros	Number Sets of Ties	Multiplicity Factor
126	105	26.77686	0	4	24

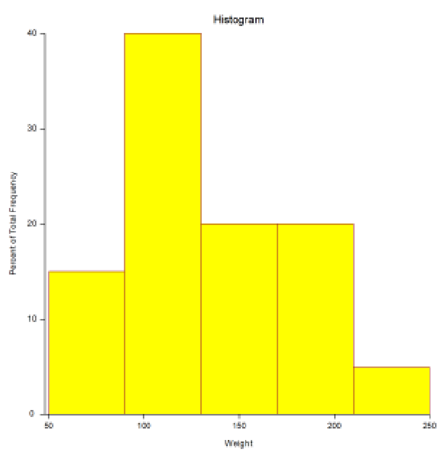
Test Type	Alternative Hypothesis	Z-Value	Prob Level	Reject H0 at α = 0.050?
Exact*	Median ≠ 130			
Exact*	Median < 130			
Exact*	Median > 130			
Normal Approximation	Median ≠ 130	0.7843	0.43289	No
Normal Approximation	Median < 130	0.7843	0.78356	No
Normal Approximation	Median > 130	0.7843	0.21644	No
Normal Approx. with C.C.	Median ≠ 130	0.7656	0.44392	No
Normal Approx. with C.C.	Median < 130	0.8029	0.78899	No
Normal Approx. with C.C.	Median > 130	0.7656	0.22196	No

* The Exact Test is provided only when there are no ties.

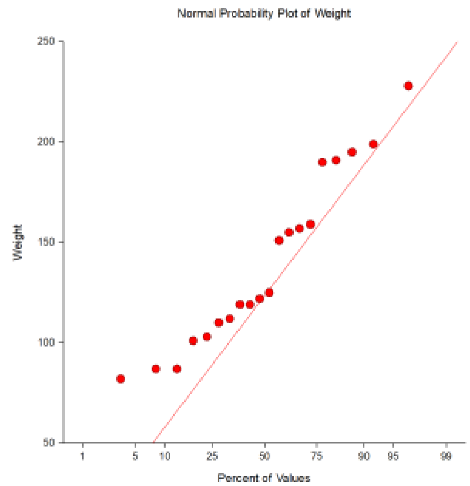
Tests of Assumptions

Assumption	Value	Prob Level	Decision (α = 0.050)
Shapiro-Wilk Normality	0.9298	0.153381	Cannot reject normality
Skewness Normality	1.0233	0.306143	Cannot reject normality
Kurtosis Normality	-0.9607	0.336698	Cannot reject normality
Omnibus Normality	1.9702	0.373401	Cannot reject normality

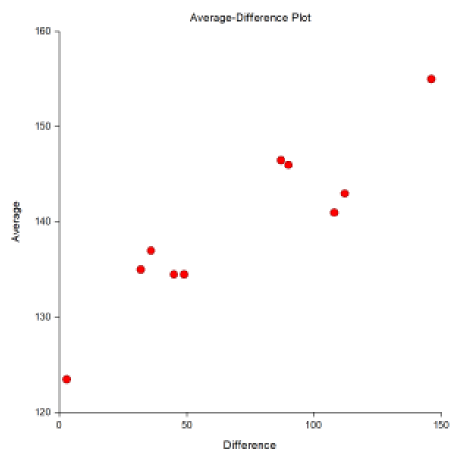
Histogram of Weight



Probability Plot of Weight



Average vs Difference Plot of Weight



El reporte anterior se observa como se realizan todos estadísticos, se tiene la media, la desviación estándar, la mediana, intervalos de confianza con el 95% para desviación estándar desconocida y conocida, histograma, pruebas de normalidad y la prueba t para la hipótesis nula, en este caso como se seleccionaron las 3 hipótesis alternas se obtiene le valor de probabilidad para cada una de ellas y en ningún caso se puede rechazar la hipótesis nula, por lo tanto se concluye que la media de la muestra es igual a 130.

Prueba t para muestras pareadas.

Este procedimiento sirve para realizar inferencias de las diferencias entre las medias de dos grupos a partir de dos muestras pareadas.

Ejemplo.

Se desea conocer la diferencia entre la llanta izquierda y la derecha.

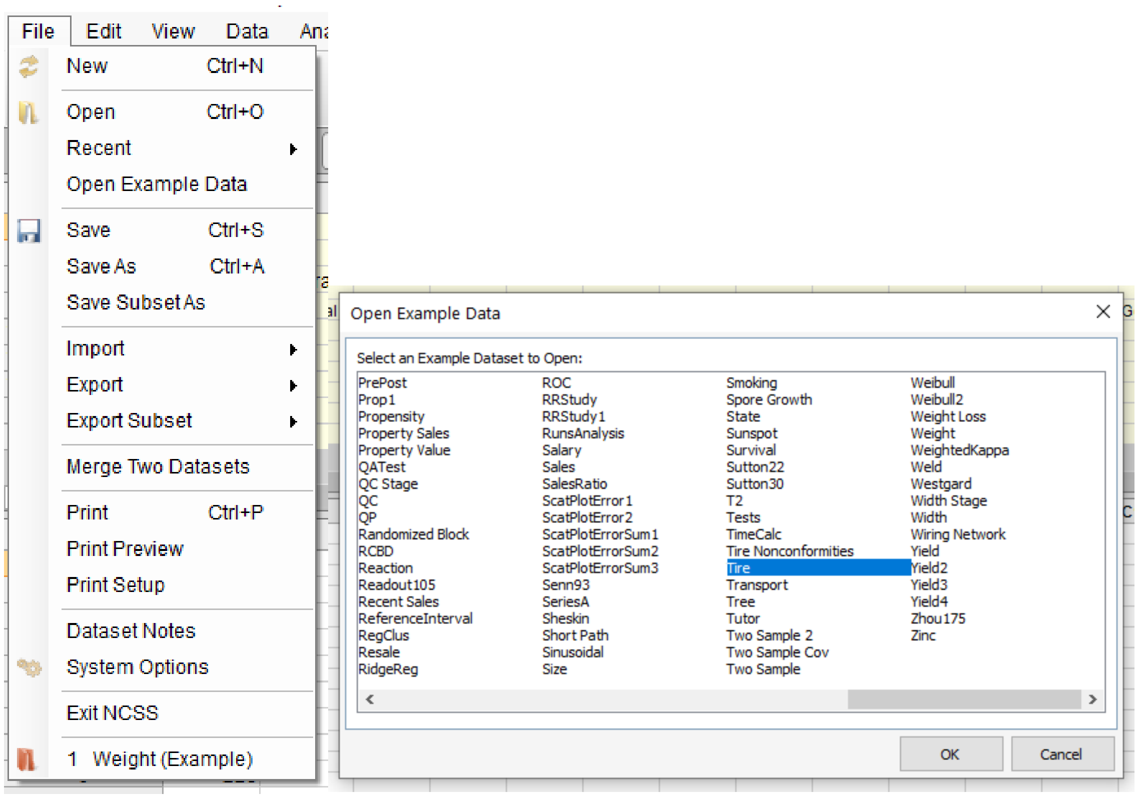
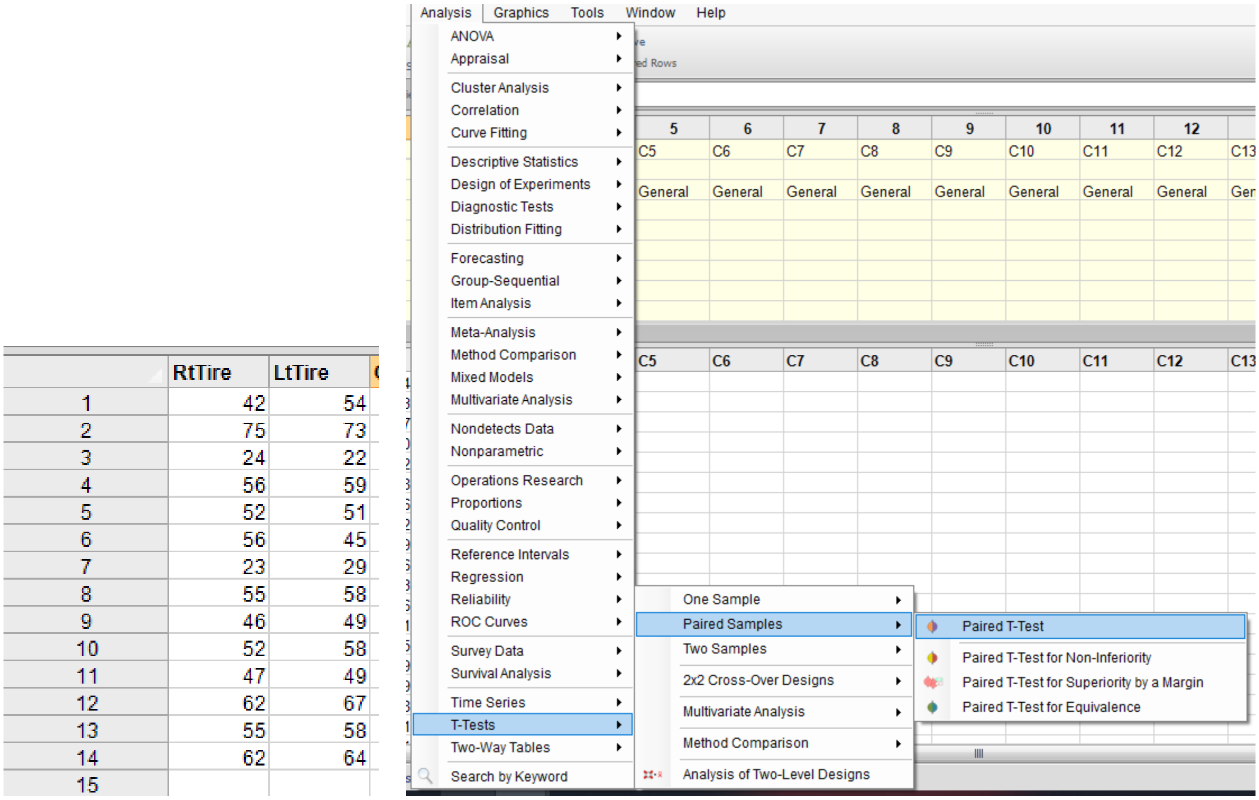


Fig. 4 Menús para cargar los datos del ejemplo para la prueba t para muestras pareadas.

Como se detalló anteriormente en el menú de file, se tiene la opción de abrir datos de ejemplos y a continuación se abre una ventana donde se encuentran todos los datos, para este procedimiento se utilizarán los datos de las llantas.



	RtTire	LtTire
1	42	54
2	75	73
3	24	22
4	56	59
5	52	51
6	56	45
7	23	29
8	55	58
9	46	49
10	52	58
11	47	49
12	62	67
13	55	58
14	62	64
15		

Fig. 5 Datos de las muestras de la llanta derecha e izquierda y el menú en el que se selecciona la prueba t de muestras pareadas.

Para este ejemplo se tienen dos muestras de 14 elementos como se observan en el episodio 5, así como se ilustra el menú de análisis en donde esta la prueba t para muestras pareadas.

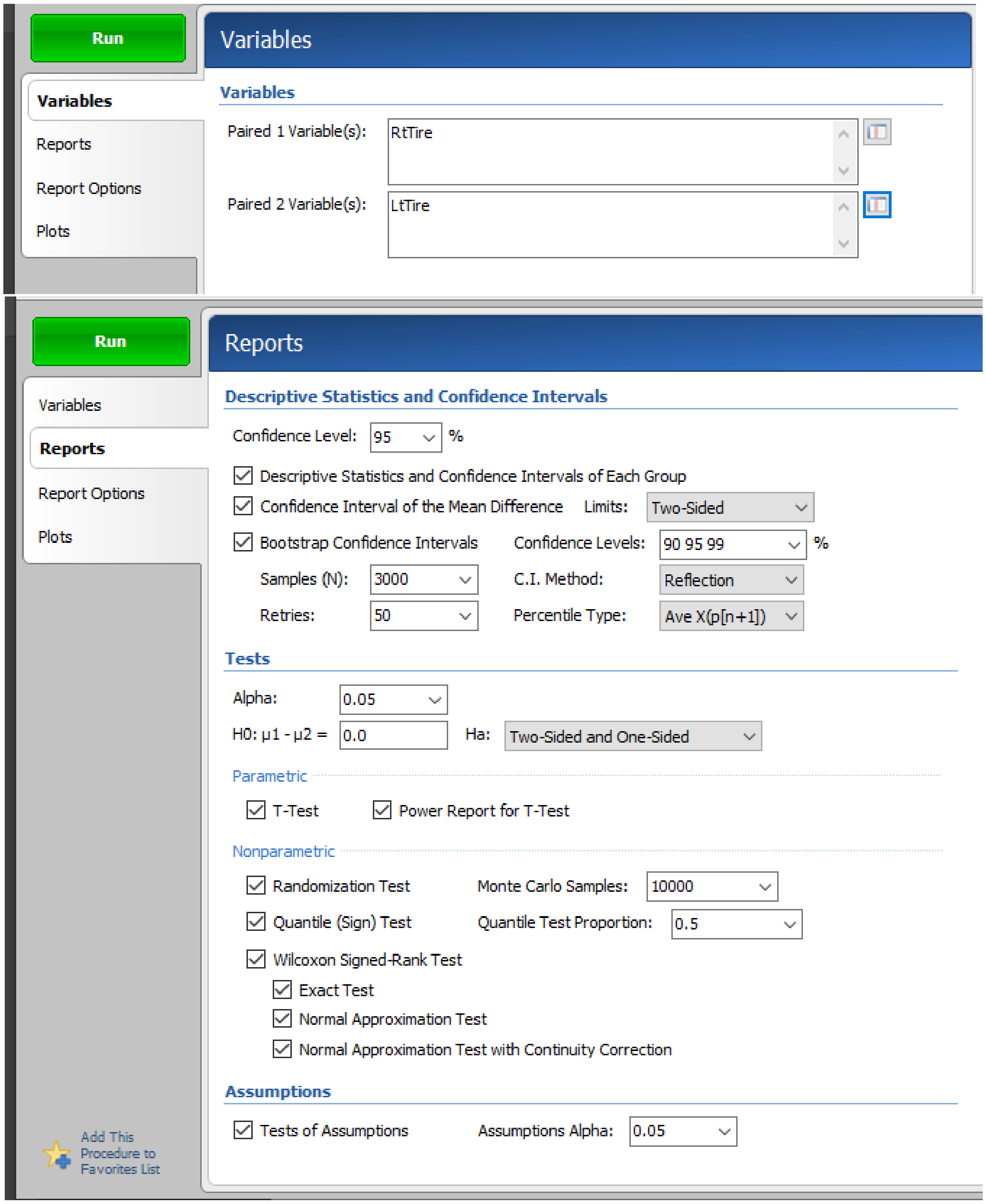


Fig. 6 Selección de las variables pareadas y los ajustes para correr la prueba t de muestras pareadas.

Como se observa en la figura 6 se cargan los datos de cada una de las variables pareadas, sin importar el orden, como se desea comparar las medias se pone como hipótesis nula que la diferencia entre ellas sea cero, además se seleccionan todas las demás opciones para poder documentar a que se refiere cada una de ellas.

Paired T-Test Report

Dataset	C:\Program Files (x86)\NCSS\NCSS 2019\Example Data\Tire.NCSS
Paired 1 Variable	RtTire
Paired 2 Variable	LtTire
Paired Difference	(RtTire) - (LtTire)

Descriptive Statistics

Variable	Count	Mean	Standard Deviation of Data	Standard Error of Mean	T*	95.0% LCL of Mean	95.0% UCL of Mean
RtTire	14	50.5	13.96011	3.730996	2.1604	42.43967	58.56033
LtTire	14	52.57143	13.7657	3.679038	2.1604	44.62335	60.51951

Two-Sided Confidence Interval of the Mean Difference

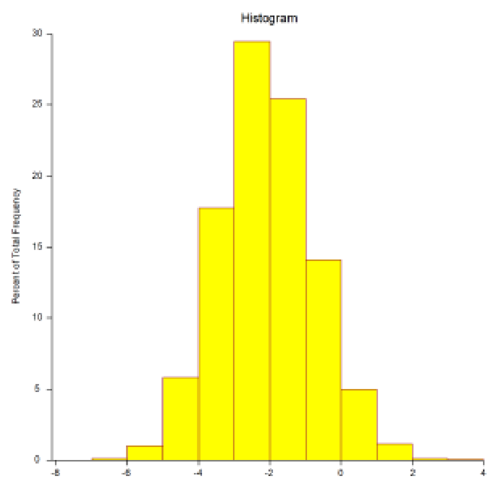
							95.0% C. I. of Mean Diff.	
Statistic	Count	Mean Difference	Standard Deviation	Standard Error	T*	DF	Lower Limit	Upper Limit
Mean Difference	14	-2.071429	5.225151	1.39648	2.1604	13	-5.088341	0.9454835

Bootstrap Section

----------- Estimation Results -----------		----------- Bootstrap Confidence Limits ---------------		
Parameter	Estimate	Conf. Level	Lower	Upper
Mean Difference				
Original Value	-2.0714	90.00	-4.3571	0.0714
Bootstrap Mean Difference	-2.0542	95.00	-4.7857	0.4286
Bias (BMD - OV)	0.0172	99.00	-5.7139	1.1425
Bias Corrected	-2.0887			
Standard Error	1.3287			

Confidence Limit Type = Reflection, Number of Samples = 3000.

Bootstrap Histograms Section



Paired-Sample T-Test

Alternative Hypothesis	Mean Difference	Standard Error	T-Statistic	DF	Prob Level	Reject H0 at α = 0.050?
Mean Diff. ≠ 0	-2.071429	1.39648	-1.4833	13	0.16182	No
Mean Diff. < 0	-2.071429	1.39648	-1.4833	13	0.08091	No
Mean Diff. > 0	-2.071429	1.39648	-1.4833	13	0.91909	No

Power for the Paired-Sample T-Test
This section assumes the population mean of paired differences and standard deviation of paired differences are equal to the sample values.

Alternative Hypothesis	N	μ	σ	Power (α = 0.05)	Power (α = 0.01)
Mean Diff. ≠ 0	14	-2.071429	5.225151	0.27964	0.10154
Mean Diff. < 0	14	-2.071429	5.225151	0.40555	0.16041
Mean Diff. > 0	14	-2.071429	5.225151	0.00112	0.00012

Randomization Test (Two-Sided)
Alternative Hypothesis: The distribution center of paired differences is not 0.
Number of Monte Carlo samples: 10000

Test	Prob Level	Reject H0 at α = 0.050?
Randomization Test	0.16030	No

Quantile (Sign) Test
This Quantile test is equivalent to the Sign test if the Quantile Proportion is 0.5.

Null Quantile (Q0)	Quantile Proportion	Number Lower	Number Higher	H1: Q ≠ Q0 Prob Level	H1: Q < Q0 Prob Level	H1: Q > Q0 Prob Level
0	0.5	10	4	0.179565	0.089783	0.971313

Wilcoxon Signed-Rank Test

Sum of Ranks (W)	Mean of W	Std Dev of W	Number of Zeros	Number Sets of Ties	Multiplicity Factor
21	52.5	15.84692	0	3	126

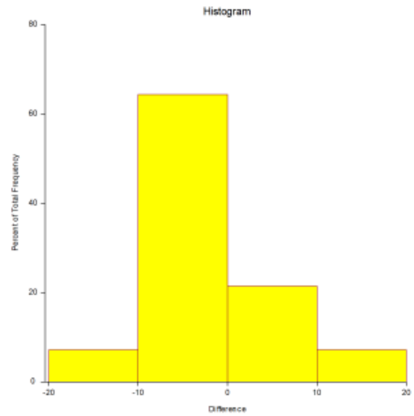
Test Type	Alternative Hypothesis	Z-Value	Prob Level	Reject H0 at α = 0.050?
Exact*	Median Diff. ≠ 0			
Exact*	Median Diff. < 0			
Exact*	Median Diff. > 0			
Normal Approximation	Median Diff. ≠ 0	1.9878	0.04684	Yes
Normal Approximation	Median Diff. < 0	-1.9878	0.02342	Yes
Normal Approximation	Median Diff. > 0	-1.9878	0.97658	No
Normal Approx. with C.C.	Median Diff. ≠ 0	1.9562	0.05044	No
Normal Approx. with C.C.	Median Diff. < 0	-1.9562	0.02522	Yes
Normal Approx. with C.C.	Median Diff. > 0	-2.0193	0.97827	No

* The Exact Test is provided only when there are no ties.

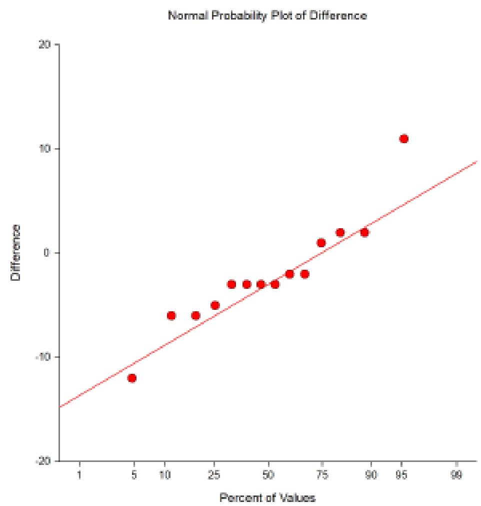
Tests of Assumptions

Assumption	Value	Prob Level	Decision (α = 0.050)
Shapiro-Wilk Normality	0.9103	0.159217	Cannot reject normality
Skewness Normality	1.3651	0.172212	Cannot reject normality
Kurtosis Normality	1.9065	0.056589	Cannot reject normality
Omnibus Normality	5.4982	0.063985	Cannot reject normality
Correlation Coefficient	0.929062		

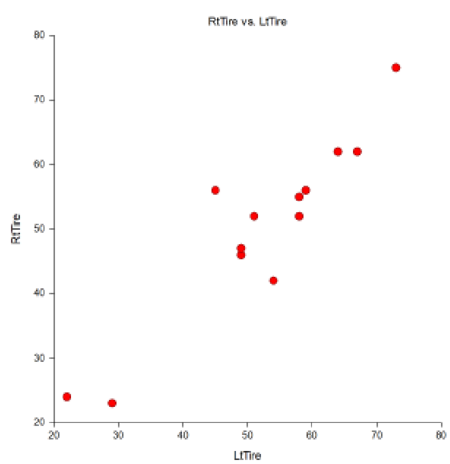
Histogram of Differences



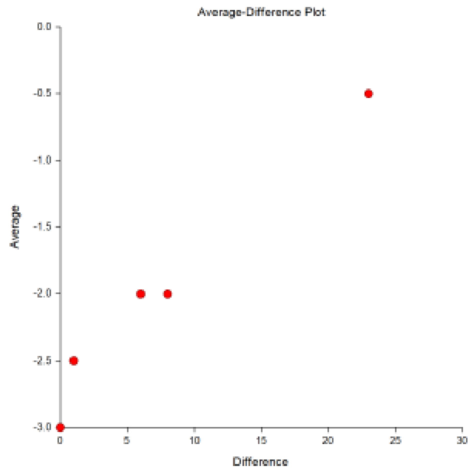
Probability Plot of Differences



Scatter Plot of Pairs



Average vs Difference Plot



En el reporte anterior se observan todos los datos que este nos proporciona de la diferencia de las muestras como inferencia de las poblaciones, para empezar se muestran los datos como la media, desviación estándar, mediana y error estándar de cada una de las muestras.

Posteriormente de ve el intervalo de confianza del 95% para la diferencia de las medias, en este caso se puede ver que el intervalo incluye al cero, por lo que sabemos que la hipótesis nula será confirmada indicando que las medias son iguales, posteriormente se encuentran los resultados de la prueba t donde se confirma la hipótesis nula, posteriormente se encuentra la potencia para cada una de las hipótesis alterna, la cual nos da información del nivel de confianza de estos resultados.

Además se encuentran los histogramas y otras gráficas de estos datos.

Análisis de varianza unidireccional

Este análisis se realiza para comparar más de dos grupos de datos para determinar si al menos uno de ellos es diferente en su media a los demás.

Ejemplo.

Se desea conocer la comparación entre el rendimiento de 3 tipos de maíz.

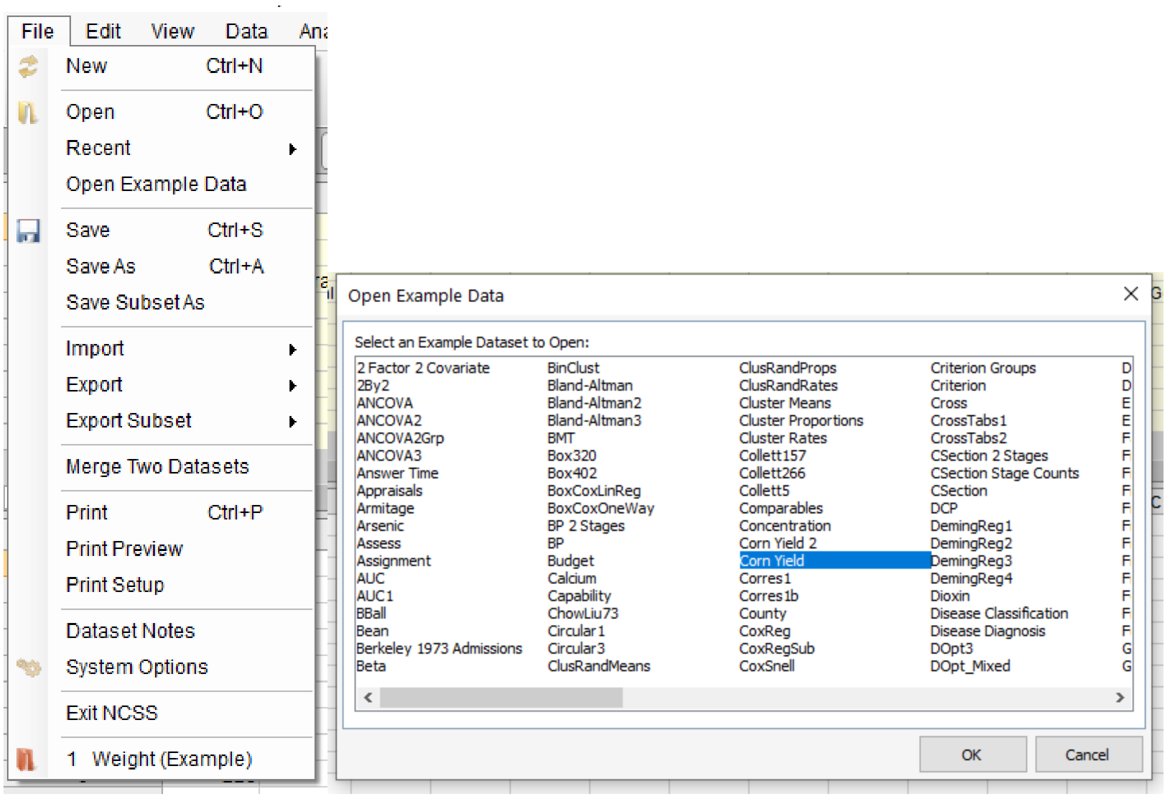
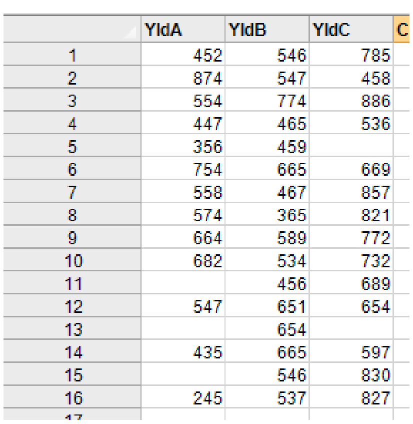


Fig. 7 Menú y nombre de los datos del ejemplo del análisis de la varianza unidireccional.

Se debe comenzar con cargar los datos del ejemplo dentro del menú file, abrir datos de ejemplo y cargar el archivo llamado Corn Yield como se muestra en la figura 7.



	YldA	YldB	YldC
1	452	546	785
2	874	547	458
3	554	774	886
4	447	465	536
5	356	459	
6	754	665	669
7	558	467	857
8	574	365	821
9	664	589	772
10	682	534	732
11		456	689
12	547	651	654
13		654	
14	435	665	597
15		546	830
16	245	537	827

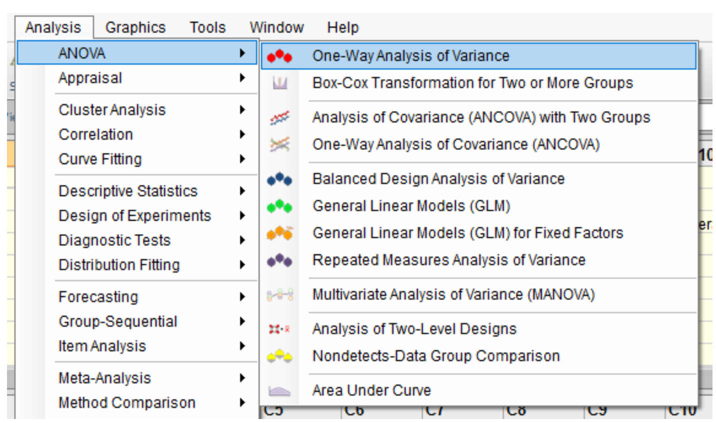
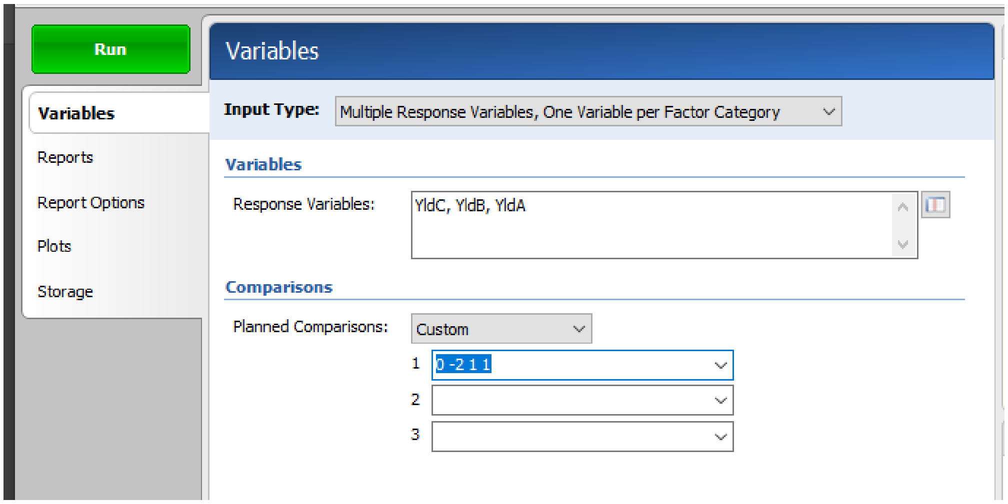


Fig. 8 Datos de ejemplos cargados y muestra del menú donde se encuentra el análisis.

Los datos son 3 tipos de maíz como se observa en la figura 8, posteriormente se debe abrir el menú del análisis de varianza unidireccional.



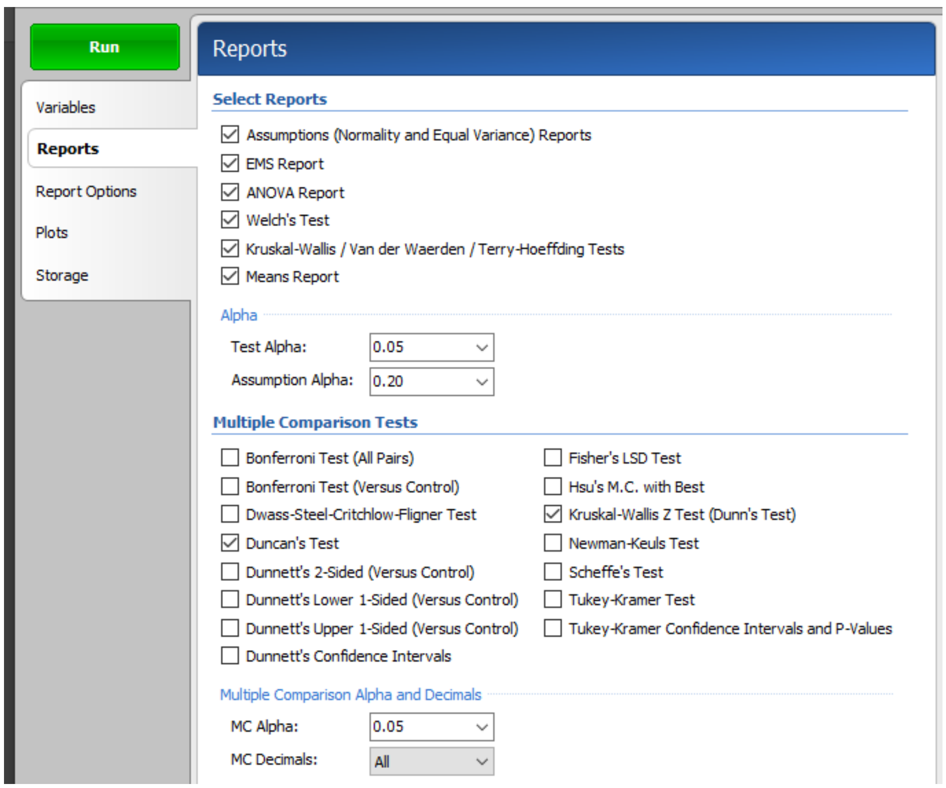


Fig. 9 Muestra de la configuración para llevar a cabo el análisis de las varianzas.

La figura 9 ilustra las configuraciones que se harán en este ejemplo para obtener el resultado deseado, comenzamos con las pruebas para validar la normalidad de los datos, posteriormente las pruebas para validar la igualdad de los datos y posteriormente marcamos las demás pruebas para poder obtener un análisis mas completo.

One-Way Analysis of Variance Report

Dataset C:\...\NCSS\NCSS 2019\Example Data\Corn Yield.NCSS
Response YldC,YldB,YldA

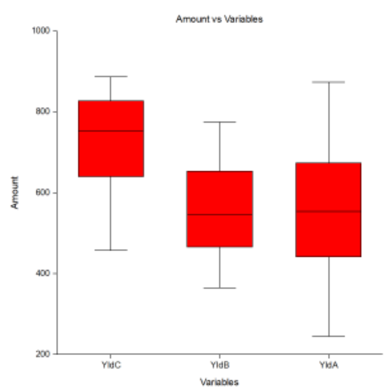
Tests of the Normality of Residuals Assumption

Normality Attributes	Test Value	Prob Level	Reject Normality? (α=0.20)
Skewness	-0.1787	0.85814	No
Kurtosis	0.4200	0.67447	No
Skewness and Kurtosis (Omnibus)	0.2084	0.90106	No

Tests of the Equality of Group Variances Assumption

Test Name	Test Value	Prob Level	Reject Equal Variances? (α=0.20)
Brown-Forsythe (Data - Medians)	1.0866	0.34711	No
Levene (Data - Means)	1.0789	0.34964	No
Conover (Ranks of Deviations)	3.0747	0.21495	No
Bartlett (Likelihood Ratio)	3.0198	0.22093	No

Box Plot Section



Expected Mean Squares Table

Model Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A (...)	2	Yes	σ^2	σ^2 + sA
Error	40	No		σ^2

Note: Expected Mean Squares are for the balanced cell-frequency case.

Analysis of Variance Table and F-Test

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Reject Equal Means? (α=0.05)	Power (α=0.05)
Between	2	268532.4	134266.2	7.4740	0.00175	Yes	0.92528
Within (Error)	40	718574.3	17964.36				
Adjusted Total	42	987106.6					
Total	43						

Welch's Test of Means Allowing for Unequal Variances

Model Term	Numerator DF	Denominator DF	F-Ratio	Prob Level	Reject Equal Means? (α=0.05)
Between Groups	2	24.27	8.0236	0.00211	Yes

Kruskal-Wallis One-Way ANOVA on Ranks

Hypotheses
H0: All medians are equal.
H1: At least two medians are different.

Test Results

Method	DF	Chi-Squared (H)	Prob Level	Reject H0? (α=0.05)
Not Corrected for Ties	2	11.2674	0.00358	Yes
Corrected for Ties	2	11.2708	0.00357	Yes

Number Sets of Ties	4
Multiplicity Factor	24

Group Detail

Group	Count	Sum of Ranks	Mean Rank	Z-Value	Median
YldC	14	437.50	31.25	3.3564	752
YldB	16	279.00	17.44	-1.8342	546
YldA	13	229.50	17.65	-1.4941	554

Normal Scores Tests

Hypotheses
H0: All group data distributions are the same.
H1: At least one group has observations that tend to be greater than those of the other groups.

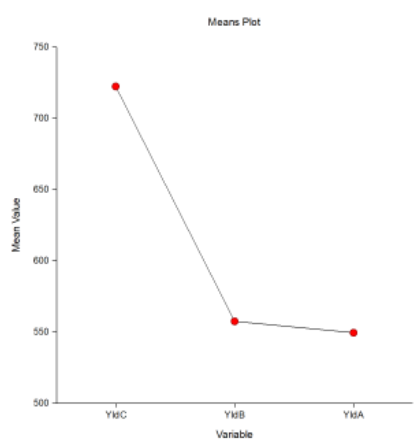
Results

Test	DF	Chi-Squared (H)	Prob Level	Reject H0? (α=0.20)
Terry-Hoeffding - Expected Normal Scores	2	10.9026	0.00429	Yes
Van der Waerden - Normal Quantiles	2	10.9819	0.00412	Yes

Descriptive Statistics

Group	Count (ni)	Mean	Effect	Median	Standard Deviation	Standard Error √(MSE/ni)
All	43	608.7209	609.7473			
A:						
YldC	14	722.3571	112.6099	752	127.8873	35.82134
YldB	16	557.5	-52.24725	546	104.6219	33.50779
YldA	13	549.3846	-60.36264	554	168.7629	37.17356

Plots of Means Section



Duncan's Multiple-Comparison Test

Response: YldC,YldB,YldA
Term A:

Alpha=0.050 Error Term=S(A) DF=40 MSE=17964.36

Group	Count	Mean	Different From Groups
YldC	14	722.3571	YldB, YldA
YldB	16	557.5	YldC
YldA	13	549.3846	YldC

Notes:
This report provides multiple comparison tests for all pairwise differences between the means. According to Hsu(1996, page 130), the specified family-wise error rate (alpha) is overstated and the Tukey-Kramer method is recommended instead.

Planned Comparison: A: Custom Comparison 1

Response: YldC,YldB,YldA
Term A:

Alpha=0.050 Error Term=S(A) DF=40 MSE=17964.36

Comparison Value=-565.6154 T-Value=7.3806 Prob>|T|=0.000000 Decision(0.05)=Reject
Comparison Std Error = 76.63526 Comparison Confidence Interval = -720.501 to -410.7297

Group	Comparison Coefficient	Count	Mean
YldC	0	14	722.3571
YldB	-2	16	557.5
YldA	1	13	549.3846

Notes:
This section presents the results of the custom planned comparisons.

Kruskal-Wallis Multiple-Comparison Z-Value Test (Dunn's Test)

Variable	YldC	YldB	YldA
YldC	0.0000	3.0063	2.8117
YldB	3.0063	0.0000	0.0462
YldA	2.8117	0.0462	0.0000

Regular Test: Medians significantly different if z-value > 1.9600
Bonferroni Test: Medians significantly different if z-value > 2.3940

En el reporte del análisis que nos regresa el software primero podemos observar que se comprueba la normalidad de los datos y viendo los diagramas de cajas de los datos podemos esperar que el grupo A, la cual tiene una media mayor que los otros dos grupos, sea diferente en el análisis posterior de la diferencia entre grupos.

En las pruebas de igualdad de medias y varianzas se rechaza la hipótesis nula que indica que son iguales, por lo que se comprueba lo esperado a partir del análisis de las graficas de cajas. Más adelante del reporte tenemos diferentes resultados siendo los que nos interesan, el del test de Duncan el cual nos dice si los grupos son diferentes y con que otros grupos son diferentes, en este caso tenemos que el grupo A es diferente con el grupo B y C, para los grupo B y C nos reporta diferencia con el grupo A.

Control de calidad

Genera barras x y graficas de control R para las variables, las cuales sirven para monitorear la media y la variación de los datos en el transcurso del tiempo.

Ejemplo.

Se desea realizar el análisis de los limites del control de calidad, por lo que se cargan datos de 50 subgrupos de 5.

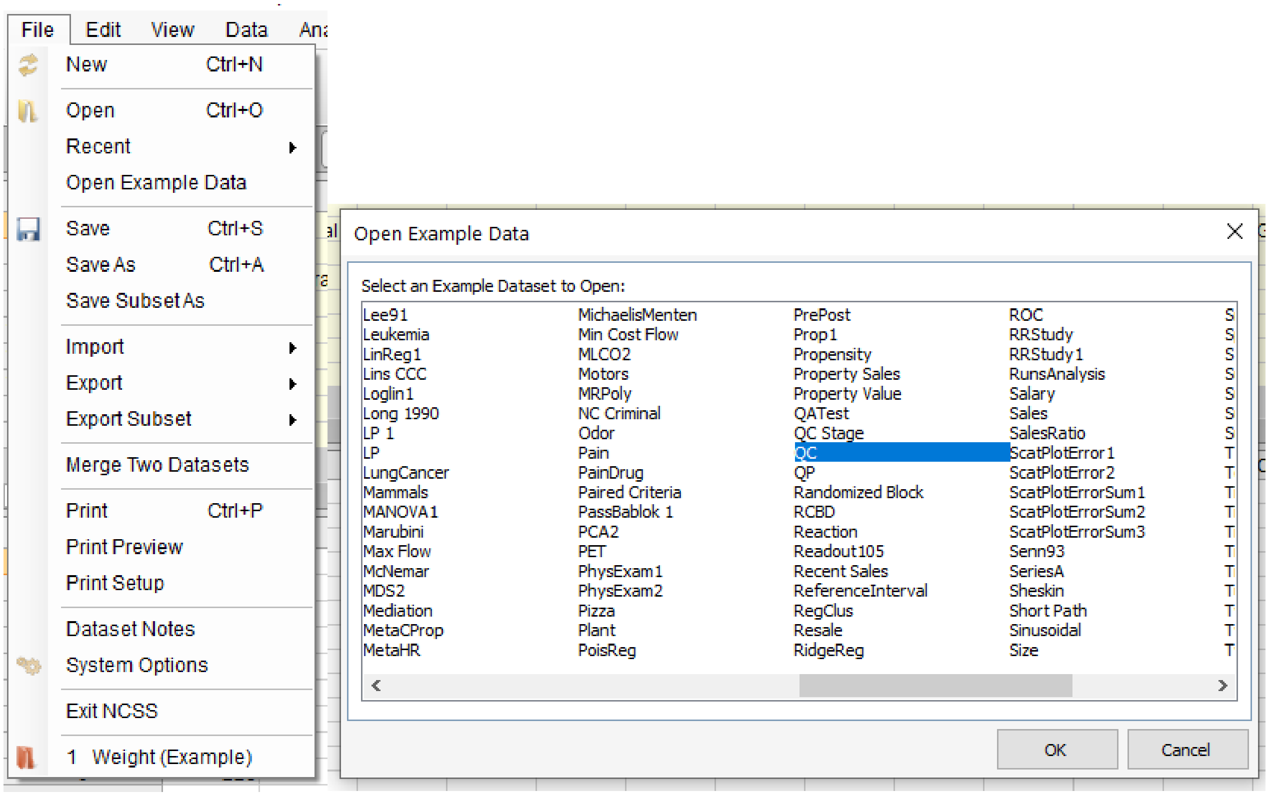


Fig. 10 Menú y nombre de los datos de ejemplo para control de calidad.

Para mostrar el resultado que nos regresa el software al realizar el procedimiento de control de calidad, se cargan los datos “QC” dentro del menú de abrir datos de ejemplo como se ve en la figura 10.

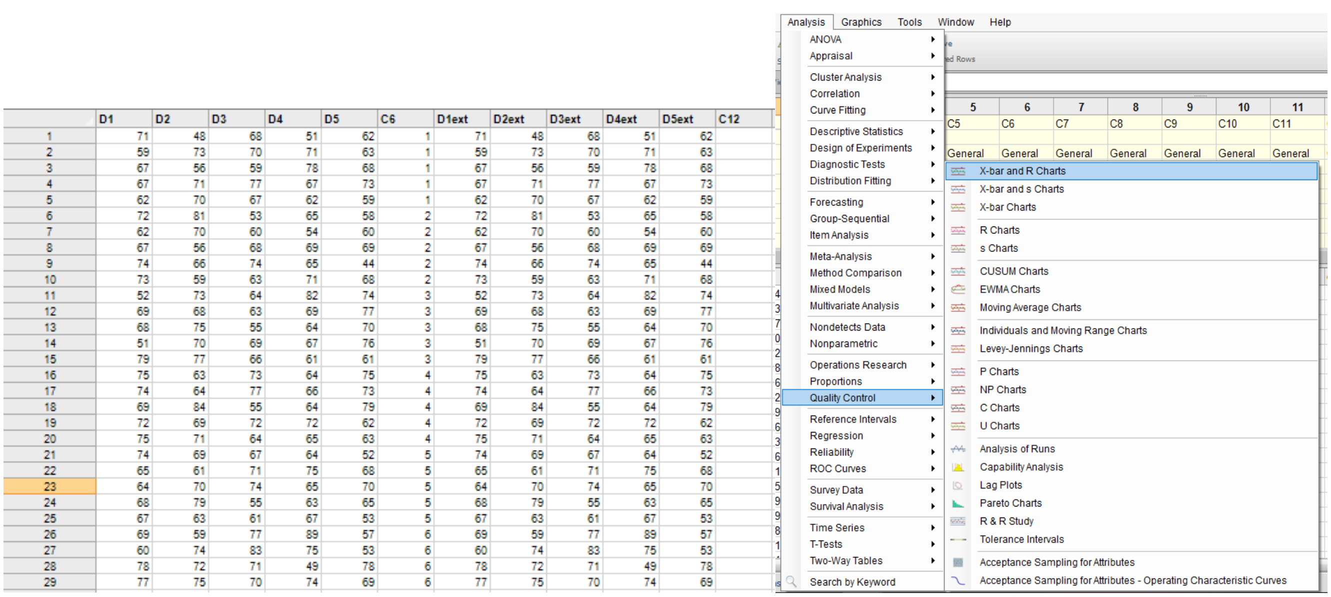


Fig. 11 Muestra de los datos de ejemplo y menú en donde se encuentra la opción de “x-bar and R-charts”.

Estos datos están cargados en la figura 11, como se observa son datos divididos en 5 grupos llamados D1, D2, D3, D4 y D5, así como la ilustración del menú a elegir para llevar a cabo este análisis el cual se encuentra dentro del menú "quality control".

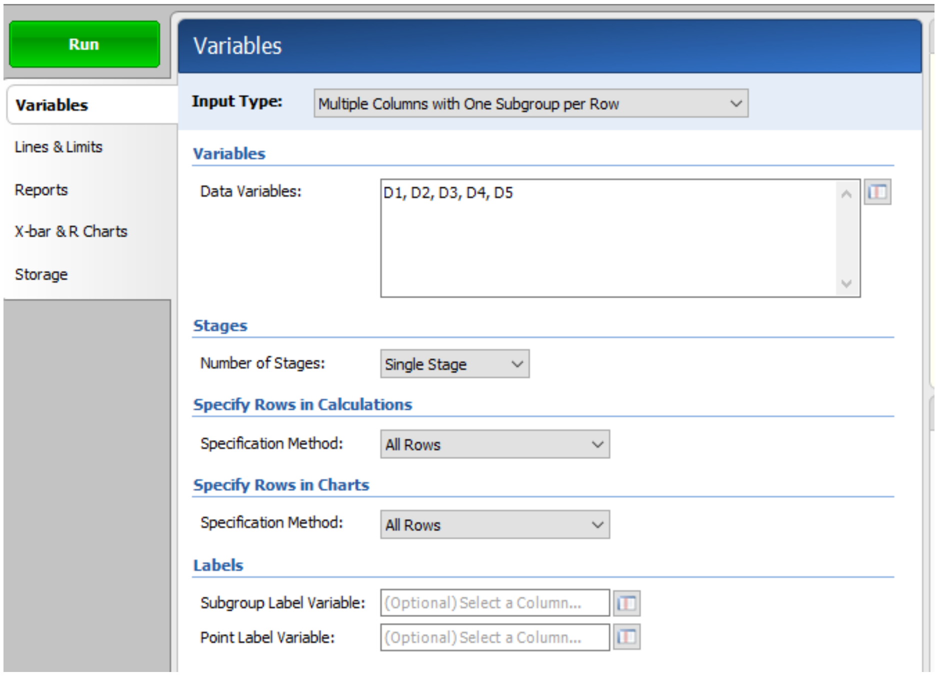


Fig. 12 Muestra de las opciones marcadas para el procedimiento.

Para llevar a cabo este procedimiento solo es necesario seleccionar los datos como se ve en la imagen 12.

Center Line Section for Subgroups 1 to 50

Number of Subgroups 50

Center Line	**Estimate**
Estimated Grand Average (X-bar-bar)	67.12
R-bar	18.14

Primary Control Limit Section for Subgroups 1 to 50

These limits are based on a subgroup of size 5.

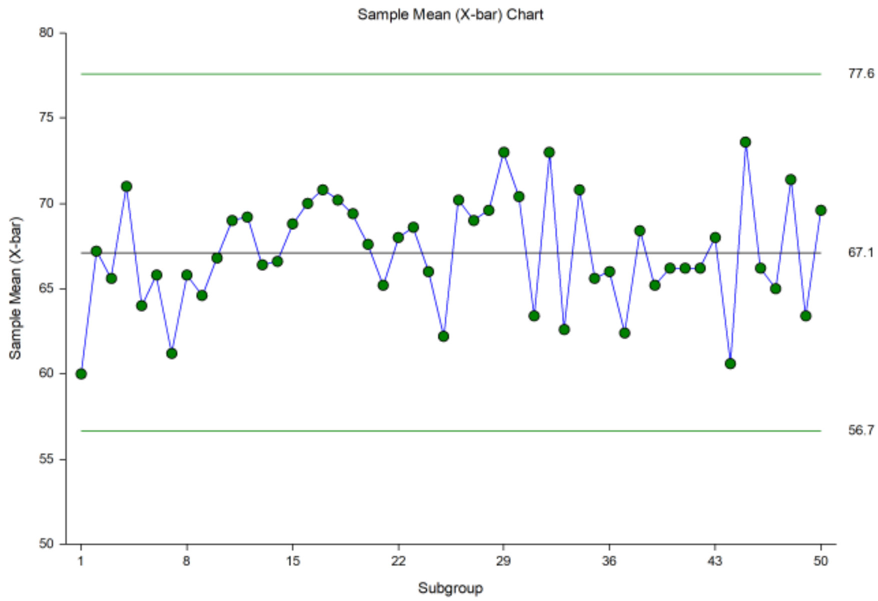
	Primary Control Limits	
Chart Type	**Lower**	**Upper**
X-bar	56.65682	77.58318
Range	0	38.35448

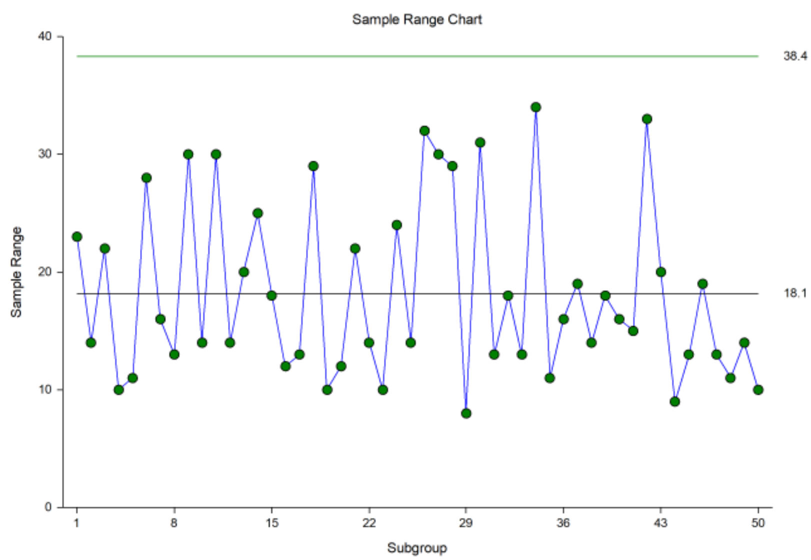
Sigma Estimation Section for Subgroups 1 to 50

Estimation Type	**Estimated Value**	**Estimated Sigma**
Ranges (R-bar)*	18.14	7.798796
Standard Deviations (s-bar)	7.365443	7.835698
Weighted Approach (s-bar)	7.902911	7.902911

* Indicates the estimation type used in this report.

Chart Section for Subgroups 1 to 50





Out-of-Control List for Subgroups 1 to 50

Subgroup	Mean	Range	Subgroup Label	Reason
30	70.4	31	30	Range: 4 of 5 in zone B or beyond

En el reporte mostrado por el software tenemos los valores medios de las dos gráficas, la x-bar la cual es la media de las muestras y la r-chart la cual es el rango de las muestras. Además de calcular la desviación estándar de los datos lo que nos da los limites de control.

Por último nos muestran las graficas de la media y del rango de los 50 subgrupos de 5 datos cada uno, así como están marcados la línea media y los limites de control tanto superior como inferior, dándonos además información de los subgrupos que no pasen el control, como en este caso el subgrupo 30 por tener 4 de 5 datos en la zona b o más abajo.

Bland-Altman

Este procedimiento es usado para comparar dos mediciones de la misma variable, es una técnica de comparación muy usada.

Ejemplo.

Se tienen dos grupos de mediciones de dos métodos que se llevaron a cabo en 100 personas, el primer método es muy invasivo y el segundo es un método mucho menos invasivo. Los ingenieros desean analizar si dos métodos concuerdan.

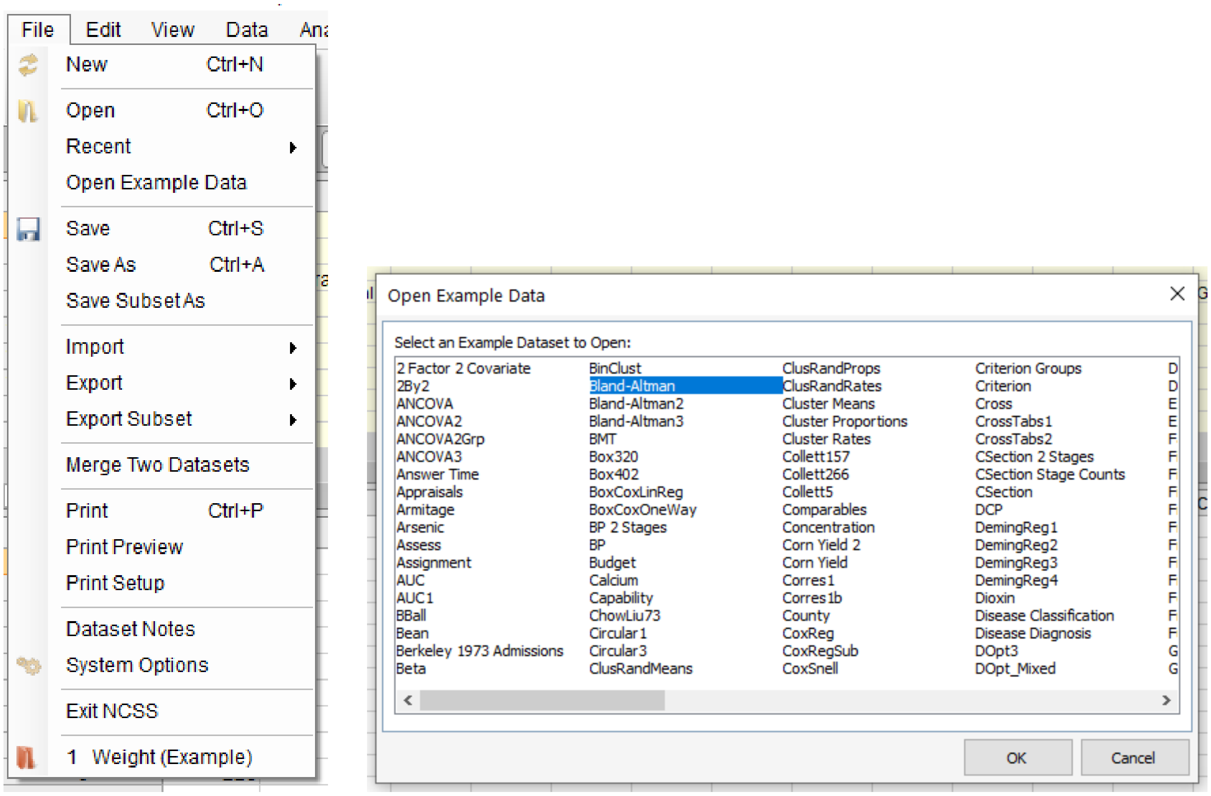


Fig. 13 Menú y nombre de los datos de ejemplo para el procedimiento de Bland-Altman.

Para este procedimiento se usaran los datos de ejemplo de nombre “Bland-Altman” que se cargaran desde el menú de abrir datos de ejemplos como se ilustra en la figura 13.

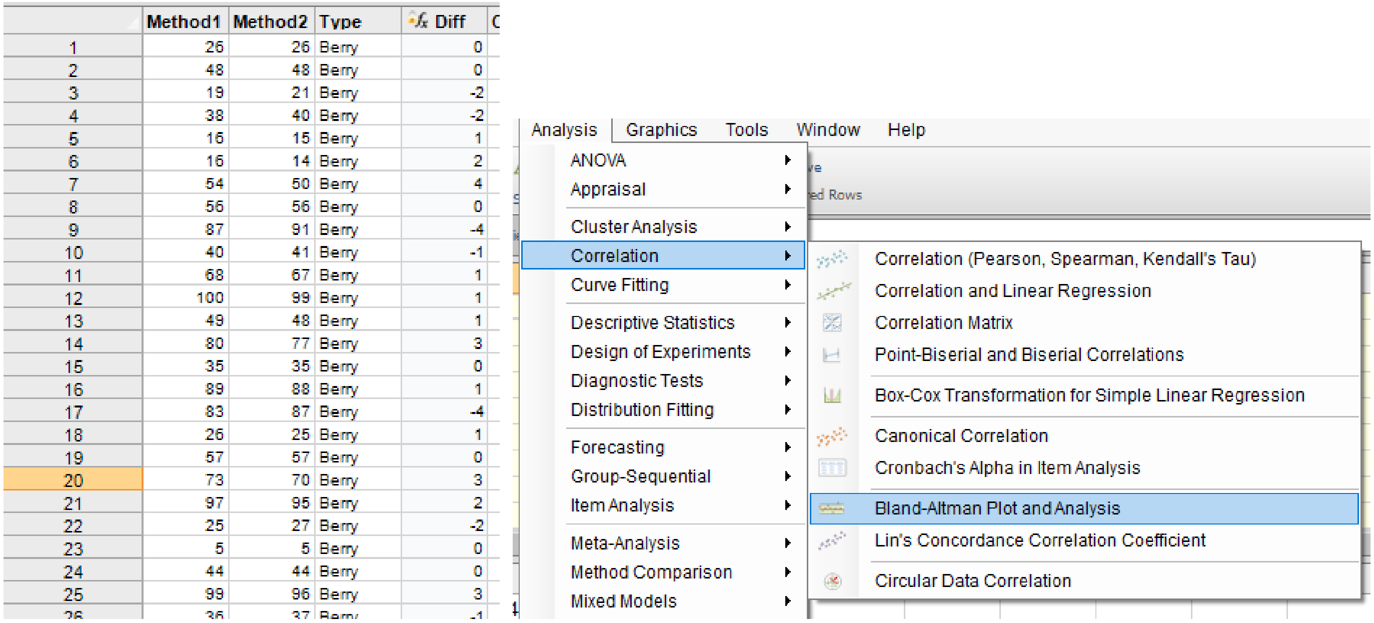


Fig. 14 Muestra de los datos de ejemplo y del menú al que hay que acceder para realizar el procedimiento Bland-Altman.

En este ejemplo se cargan lo datos de ejemplo de los 100 sujetos de estudio para cada uno de los métodos y posteriormente se selecciona el procedimiento de Bland-Altman como se ve en la figura 14.

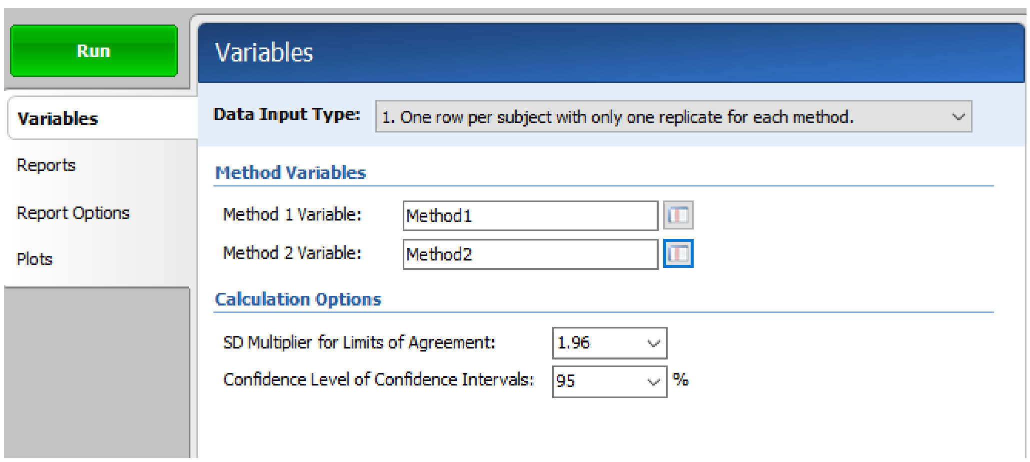
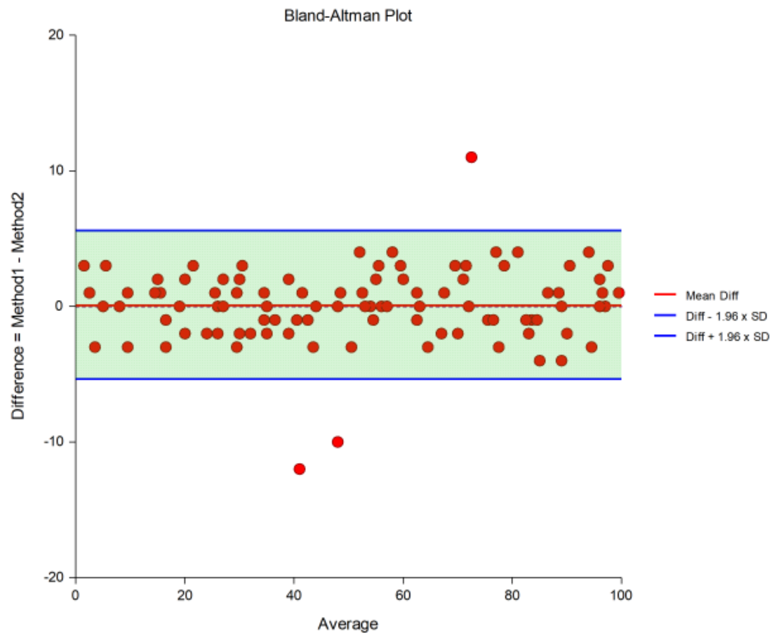


Fig. 15 Opciones marcadas para llevar a cabo este procedimiento.

Para este procedimiento solo es necesario cargar los datos de los dos métodos a comparar como se muestran en la figura 15, sin realizar cambios adicionales en las opciones de cálculo.



Descriptive Statistics

Variable	Count	Mean	Standard Deviation	95.0% LCL of Mean	95.0% UCL of Mean
Method1	100	50.72	28.30893	45.10289	56.3371
Method2	100	50.62	28.07701	45.04891	56.19109
Difference	100	0.1	2.787055	-0.4530122	0.6530122

Correlation Coefficient = 0.995147

Bland-Altman Analysis Report

Dataset C:\...\NCSS\NCSS 2019\Example Data\Bland-Altman.NCSS

Bland-Altman Analysis: Bias and Limits of Agreement for Method1 and Method2

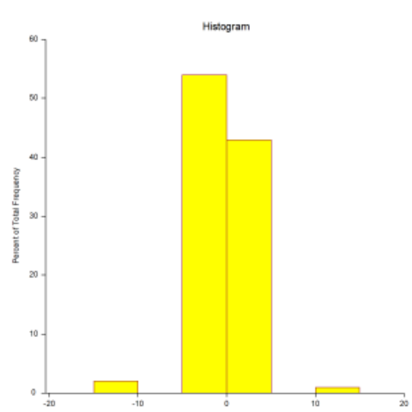
Limits of Agreement = Diff ± 1.96 x (Std Dev of Difference)

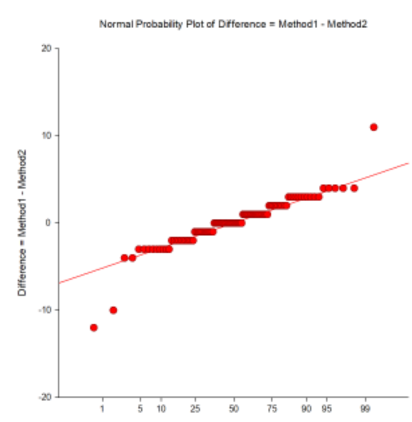
Parameter	Count	Value	Standard Deviation	95.0% LCL of Value	95.0% UCL of Value
Bias (Difference)	100	0.1	2.787055	-0.4530122	0.6530122
Lower Limit of Agreement	100	-5.362628	0.4778968	-6.310879	-4.414377
Upper Limit of Agreement	100	5.562628	0.4778968	4.614377	6.510879

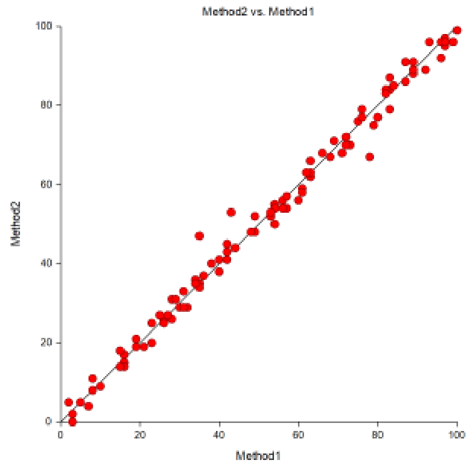
Test of Normality of Differences Assumption

Assumption	Value	Prob Level	Decision (α = 0.050)
Shapiro-Wilk	0.893	0.0000	Reject normality

Evaluation of Assumptions Plots







El reporte de este procedimiento nos regresa una grafica en donde en el eje horizontal esta la media de la diferencia de los datos y en el eje vertical esta la diferencia punto a punto de los datos, teniendo también limites tanto superior como inferior marcados con un 95% en el intervalo.

Nos regresa también los estadísticos de los dos métodos, como la media, la desviación estándar y los intervalos de confianza, así como el de la diferencia. Además del coeficiente de correlación.

Por último tenemos la prueba de normalidad de la diferencia, el histograma de esta y la gráfica de normalidad.

En este caso para el ejemplo podemos concluir que no existe diferencia significativa entre los dos métodos comparados.

Correlación y regresión lineal

La regresión lineal es una técnica que nos permite ajustar una recta a la relación que existe entre dos variables, la regresión lineal estima los valores de la pendiente y de la ordenada al origen de esta recta.

Una vez obtenidos la pendiente y la ordenada al origen se utiliza el coeficiente de correlación, el cual es un factor entre -1 y 1, donde el 0 se interpreta como independencia lineal y 1 es una correlación lineal perfecta entre dos variables.

Ejemplo.

Se realizará una regresión lineal para conocer la relación entre peso y altura, usando estos datos se desea predecir la altura para valores de peso de 90, 100, 150, 200 y 250.

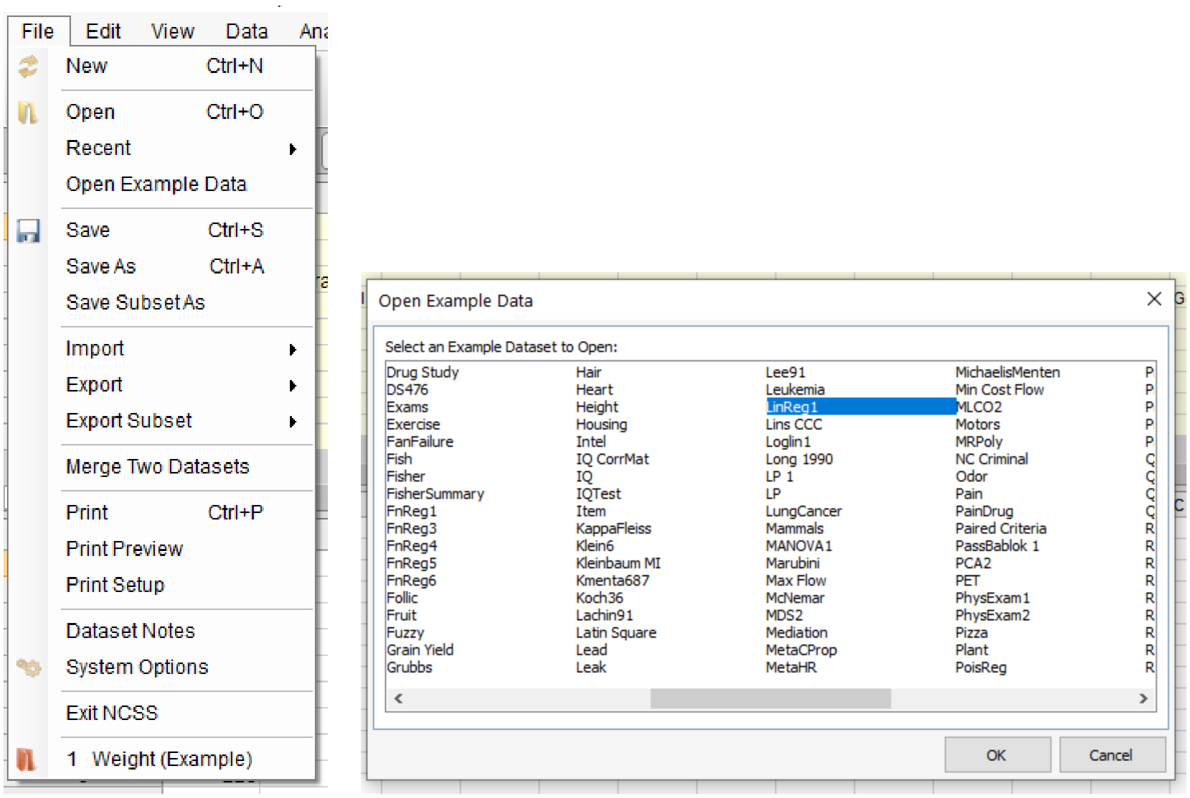


Fig. 16 Menú para cargar los datos del ejemplo para la regresión lineal.

Para obtener los datos de este ejemplo es necesario abrir el archivo de datos llamado ”LinReg1” en el menú de abrir datos de ejemplos.

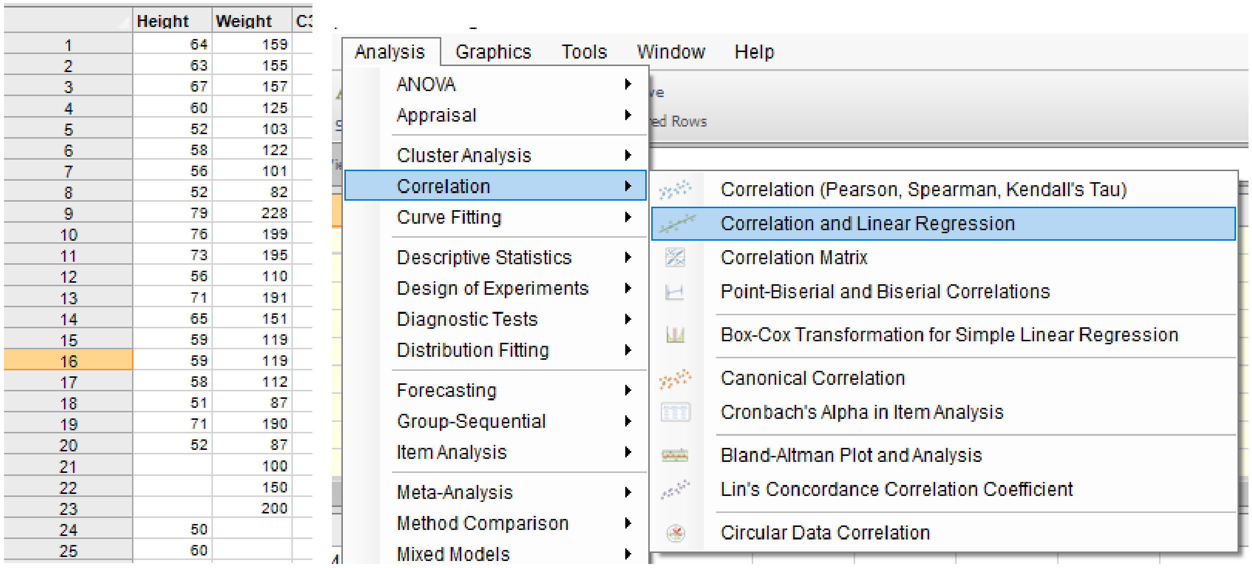


Fig. 17 Datos de altura y peso a los cuales se les hará la regresión lineal mediante el menú mostrado.

Los datos cargados son datos de peso y altura, mediante los cuales se obtendrá la regresión lineal y así poder predecir cuales son los valores de la altura para los datos de peso dados.

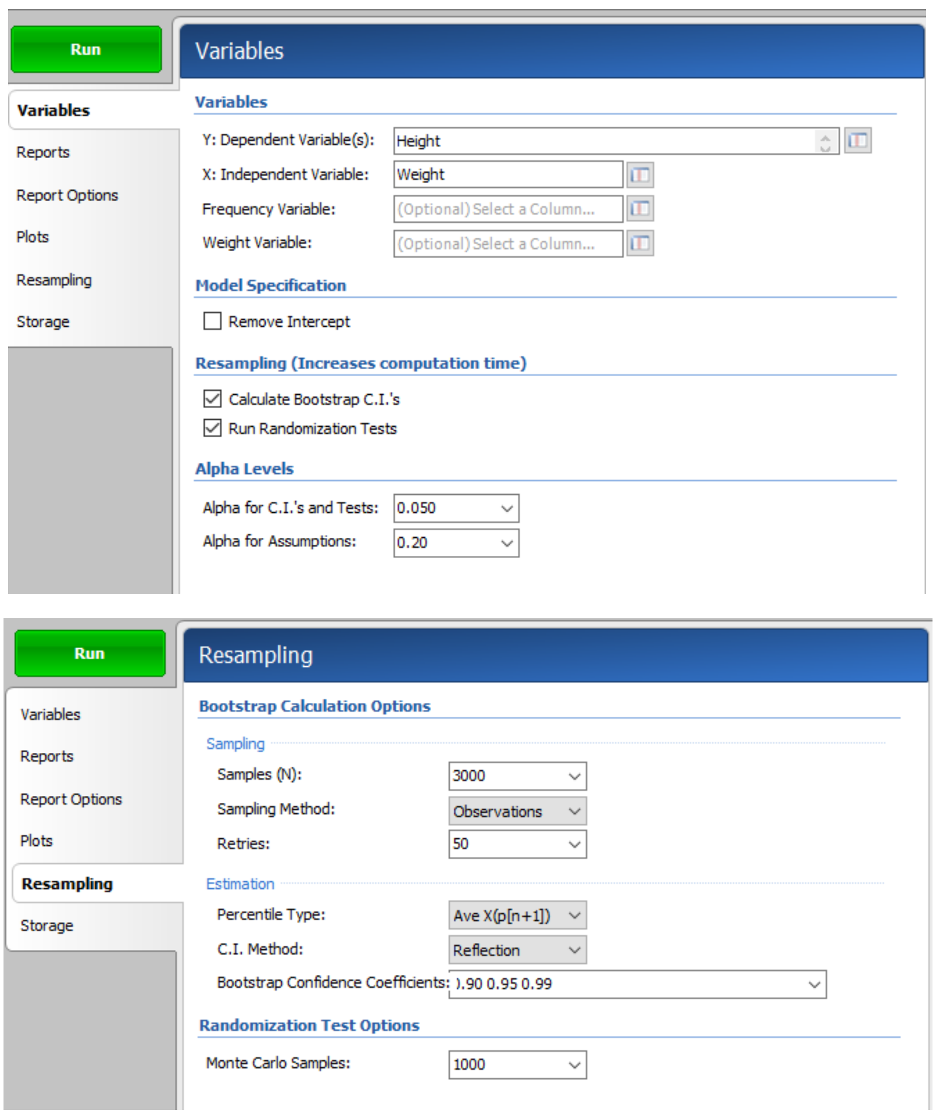


Fig. 18 Opciones marcadas para llevar a cabo la regresión lineal.

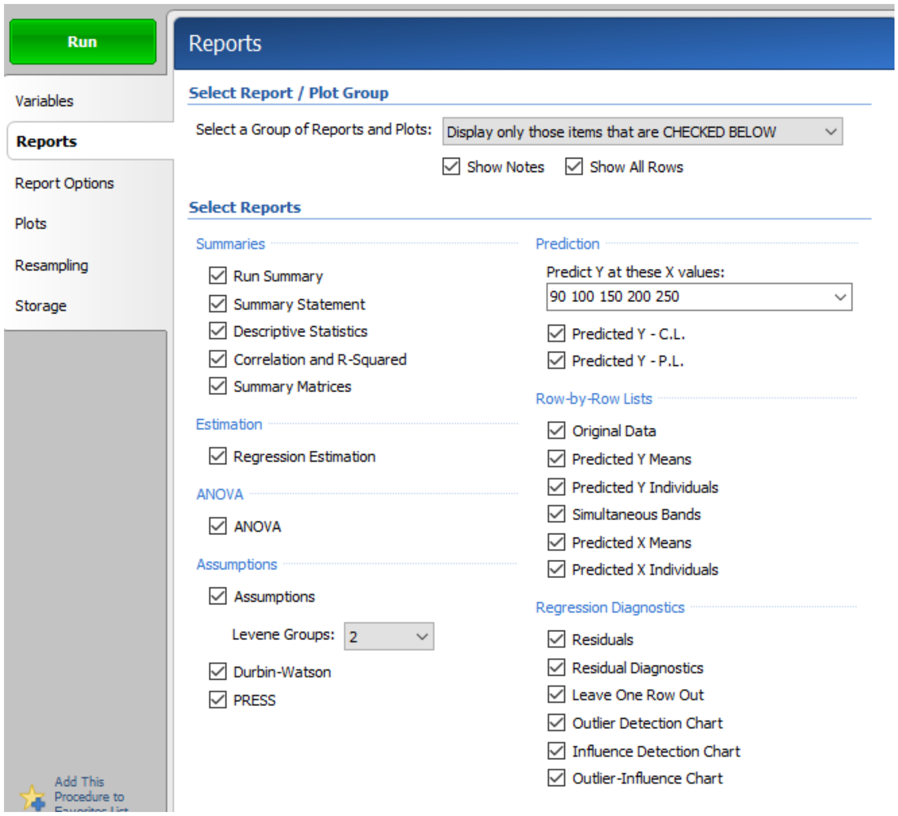
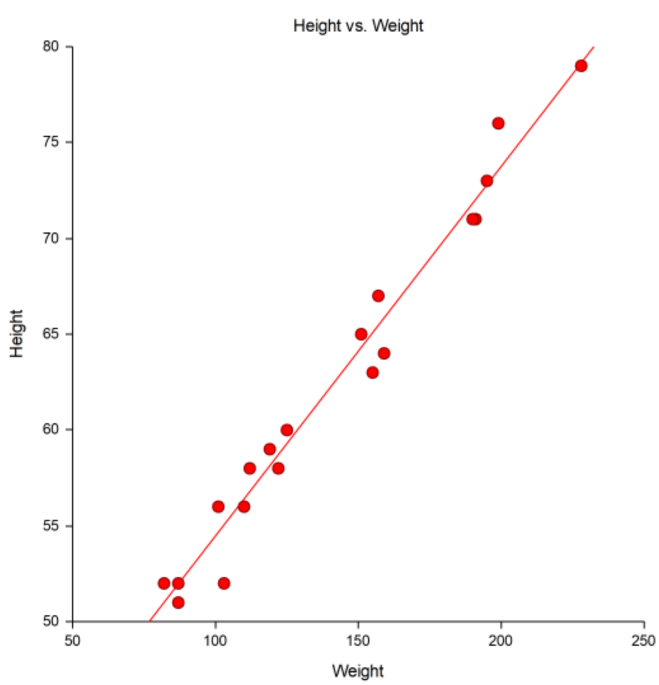


Fig. 19 Opciones marcadas que se desean obtener en el reporte.

Se cargan los datos tanto de la variable dependiente como la independiente, en este caso la independiente es el peso y la altura dependiente, además de marcar la interpolación mediante el remuestreo agregamos a las opciones los datos para los que queremos estimar el valor de la altura según el peso, como se observa en las imágenes 18 y 19.



Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Height	Rows Processed	26
Independent Variable	Weight	Rows Used in Estimation	20
Frequency Variable	None	Rows with X Missing	3
Weight Variable	None	Rows with Freq Missing	0
Intercept	35.1337	Rows Prediction Only	3
Slope	0.1932	Sum of Frequencies	20
R-Squared	0.9738	Sum of Weights	20.0000
Correlation	0.9868	Coefficient of Variation	0.0226
Mean Square Error	1.970176	Square Root of MSE	1.40363

Summary Statement

The equation of the straight line relating Height and Weight is estimated as: Height = (35.1337) + (0.1932) Weight using the 20 observations in this dataset. The y-intercept, the estimated value of Height when Weight is zero, is 35.1337 with a standard error of 1.0887. The slope, the estimated change in Height per unit change in Weight, is 0.1932 with a standard error of 0.0075. The value of R-Squared, the proportion of the variation in Height that can be accounted for by variation in Weight, is 0.9738. The correlation between Height and Weight is 0.9868.

A significance test that the slope is zero resulted in a t-value of 25.8679. The significance level of this t-test is 0.0000. Since 0.0000 < 0.0500, the hypothesis that the slope is zero is rejected.

The estimated slope is 0.1932. The lower limit of the 95% confidence interval for the slope is 0.1775 and the upper limit is 0.2089. The estimated intercept is 35.1337. The lower limit of the 95% confidence interval for the intercept is 32.8464 and the upper limit is 37.4209.

Descriptive Statistics Section

Parameter	Dependent	Independent
Variable	Height	Weight
Count	20	20
Mean	62.1000	139.6000
Standard Deviation	8.4411	43.1221
Minimum	51.0000	82.0000
Maximum	79.0000	228.0000

Regression Estimation Section

Parameter	Intercept B(0)	Slope B(1)
Regression Coefficients	35.1337	0.1932
Lower 95% Confidence Limit	32.8464	0.1775
Upper 95% Confidence Limit	37.4209	0.2089
Standard Error	1.0887	0.0075
Standardized Coefficient	0.0000	0.9868
T Value	32.2716	25.8679
Prob Level (T Test)	0.0000	0.0000
Prob Level (Randomization Test N =1000)		0.0010
Reject H0 (Alpha = 0.0500)	Yes	Yes
Power (Alpha = 0.0500)	1.0000	1.0000
Regression of Y on X	35.1337	0.1932
Inverse Regression from X on Y	34.4083	0.1984
Orthogonal Regression of Y and X	35.1076	0.1934

Notes:
The above report shows the least-squares estimates of the intercept and slope followed by the corresponding standard errors, confidence intervals, and hypothesis tests. Note that these results are based on several assumptions that should be validated before they are used.

Estimated Model
(35.1336680743148) + (0.193168566802902) * (Weight)

Bootstrap Section

----------- Estimation Results -----------		| ----------- Bootstrap Confidence Limits ---------------		
Parameter	**Estimate**	**| Conf. Level**	**Lower**	**Upper**
Intercept				
Original Value	35.1337	| 0.9000	33.5115	36.8293
Bootstrap Mean	35.1669	| 0.9500	33.2267	37.2536
Bias (BM - OV)	0.0333	| 0.9900	32.6224	38.0422
Bias Corrected	35.1004			
Standard Error	1.0170			
Slope				
Original Value	0.1932	| 0.9000	0.1813	0.2048
Bootstrap Mean	0.1929	| 0.9500	0.1790	0.2069
Bias (BM - OV)	-0.0003	| 0.9900	0.1730	0.2121
Bias Corrected	0.1935			
Standard Error	0.0071			
Correlation				
Original Value	0.9868	| 0.9000	0.9798	0.9968
Bootstrap Mean	0.9866	| 0.9500	0.9788	1.0000
Bias (BM - OV)	-0.0002	| 0.9900	0.9771	1.0000
Bias Corrected	0.9870			
Standard Error	0.0056			
R-Squared				
Original Value	0.9738	| 0.9000	0.9599	0.9935
Bootstrap Mean	0.9735	| 0.9500	0.9580	0.9999
Bias (BM - OV)	-0.0003	| 0.9900	0.9546	1.0000
Bias Corrected	0.9741			
Standard Error	0.0111			
Standard Error of Estimate				
Original Value	1.4036	| 0.9000	1.1674	1.8489
Bootstrap Mean	1.3206	| 0.9500	1.1220	1.9142
Bias (BM - OV)	-0.0831	| 0.9900	1.0277	2.0774
Bias Corrected	1.4867			
Standard Error	0.2066			
Orthogonal Intercept				
Original Value	35.1076	| 0.9000	33.4778	36.8147
Bootstrap Mean	35.1404	| 0.9500	33.1953	37.2441
Bias (BM - OV)	0.0329	| 0.9900	32.5952	38.0530
Bias Corrected	35.0747			
Standard Error	1.0222			
Orthogonal Slope				
Original Value	0.1934	| 0.9000	0.1814	0.2050
Bootstrap Mean	0.1931	| 0.9500	0.1792	0.2071
Bias (BM - OV)	-0.0003	| 0.9900	0.1731	0.2123
Bias Corrected	0.1936			
Standard Error	0.0071			
Predicted Mean and Confidence Limits of Height when Weight = 90.0000				
Original Value	52.5188	| 0.9000	51.8230	53.2869
Bootstrap Mean	52.5263	| 0.9500	51.7260	53.4824
Bias (BM - OV)	0.0075	| 0.9900	51.4481	53.7944
Bias Corrected	52.5114			
Standard Error	0.4516			

----------- Estimation Results -----------		| ----------- Bootstrap Confidence Limits ---------------		
Parameter	**Estimate**	**| Conf. Level**	**Lower**	**Upper**
Predicted Mean and Confidence Limits of Height when Weight = 100.0000				
Original Value	54.4505	| 0.9000	53.8395	55.1363
Bootstrap Mean	54.4551	| 0.9500	53.7407	55.3088
Bias (BM - OV)	0.0046	| 0.9900	53.5005	55.6201
Bias Corrected	54.4459			
Standard Error	0.4023			
Predicted Mean and Confidence Limits of Height when Weight = 150.0000				
Original Value	64.1090	| 0.9000	63.5965	64.6500
Bootstrap Mean	64.0992	| 0.9500	63.5016	64.7567
Bias (BM - OV)	-0.0097	| 0.9900	63.2687	64.9508
Bias Corrected	64.1187			
Standard Error	0.3191			
Predicted Mean and Confidence Limits of Height when Weight = 200.0000				
Original Value	73.7674	| 0.9000	72.8804	74.6451
Bootstrap Mean	73.7433	| 0.9500	72.6912	74.8443
Bias (BM - OV)	-0.0240	| 0.9900	72.2479	75.2493
Bias Corrected	73.7914			
Standard Error	0.5410			
Predicted Mean and Confidence Limits of Height when Weight = 250.0000				
Original Value	83.4258	| 0.9000	82.0218	84.8206
Bootstrap Mean	83.3874	| 0.9500	81.7238	85.1182
Bias (BM - OV)	-0.0384	| 0.9900	80.9432	85.9647
Bias Corrected	83.4642			
Standard Error	0.8570			
Predicted Value and Prediction Limits of Height when Weight = 90.0000				
Original Value	52.5188	| 0.9000	49.7692	55.4926
Bootstrap Mean	52.5769	| 0.9500	49.3257	56.0941
Bias (BM - OV)	0.0581	| 0.9900	48.4962	57.1348
Bias Corrected	52.4608			
Standard Error	1.7196			

Standard Error	1.7402			
Predicted Value and Prediction Limits of Height when Weight = 150.0000				
Original Value	64.1090	0.9000	61.4314	67.1047
Bootstrap Mean	64.1260	0.9500	60.9273	67.6675
Bias (BM - OV)	0.0171	0.9900	60.2085	68.6550
Bias Corrected	64.0919			
Standard Error	1.7294			
Predicted Value and Prediction Limits of Height when Weight = 200.0000				
Original Value	73.7674	0.9000	71.0545	76.7834
Bootstrap Mean	73.7814	0.9500	70.6354	77.4653
Bias (BM - OV)	0.0140	0.9900	69.8229	78.5260
Bias Corrected	73.7534			
Standard Error	1.7446			

Bootstrap Section (Continued)

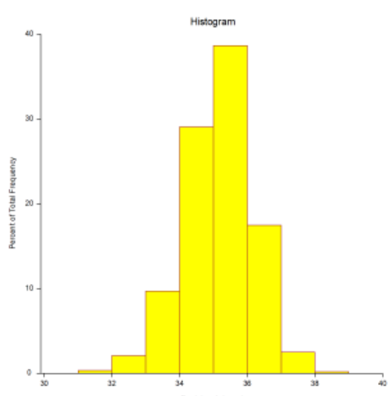
----------- Estimation Results -----------		----------- Bootstrap Confidence Limits ---------------		
Parameter	**Estimate**	**Conf. Level**	**Lower**	**Upper**
Predicted Value and Prediction Limits of Height when Weight = 250.0000				
Original Value	83.4258	0.9000	80.7299	86.3714
Bootstrap Mean	83.4500	0.9500	80.2993	86.8669
Bias (BM - OV)	0.0242	0.9900	79.5613	87.9526
Bias Corrected	83.4016			
Standard Error	1.7238			

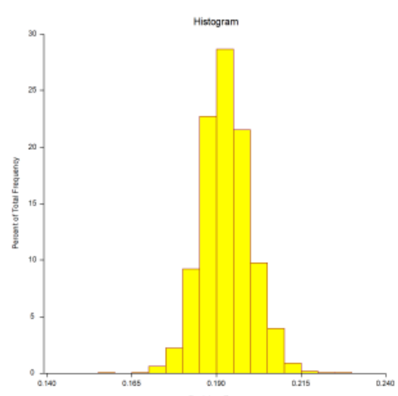
----------- Estimation Results -----------		----------- Bootstrap Confidence Limits ---------------		
Parameter	**Estimate**	**Conf. Level**	**Lower**	**Upper**
Predicted Value and Prediction Limits of Height when Weight = 250.0000				
Original Value	83.4258	0.9000	80.7299	86.3714
Bootstrap Mean	83.4500	0.9500	80.2993	86.8669
Bias (BM - OV)	0.0242	0.9900	79.5613	87.9526
Bias Corrected	83.4016			
Standard Error	1.7238			

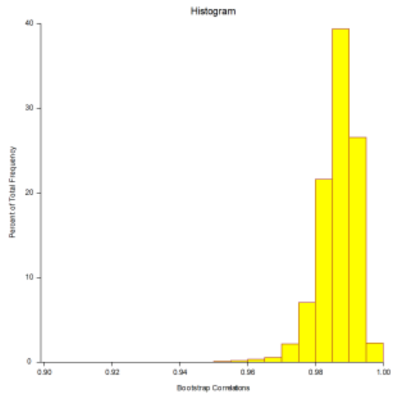
Sampling Method = Observation, Confidence Limit Type = Reflection, Number of Samples = 3000.

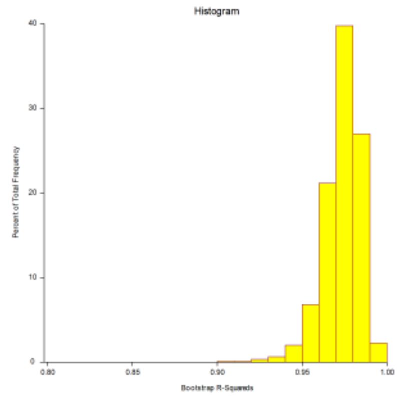
Notes:
The main purpose of this report is to present the bootstrap confidence intervals of various parameters. All gross outliers should have been removed. The sample size should be at least 50 and the sample should be 'representative' of the population it was drawn from.

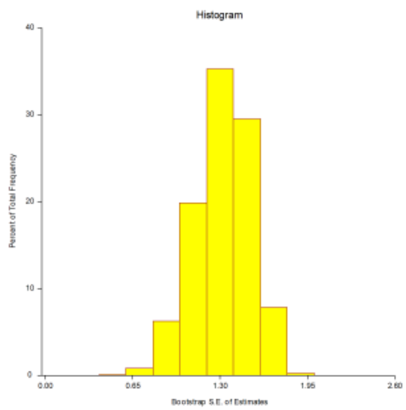
Bootstrap Histograms Section

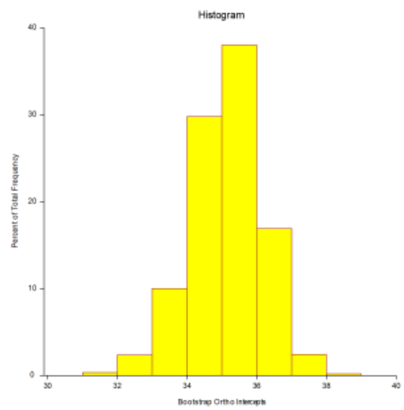


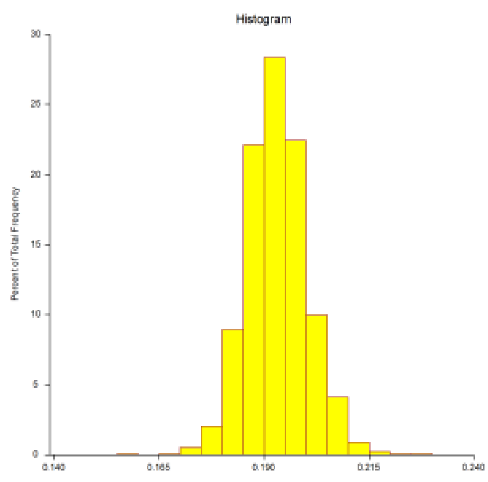



Histogram
Percent of Total Frequency
Bootstrap Correlations


Histogram
Percent of Total Frequency
Bootstrap R-Squareds


Histogram
Percent of Total Frequency
Bootstrap S.E. of Estimates


Histogram
Percent of Total Frequency
Bootstrap Ortho Intercepts


Histogram
Percent of Total Frequency
Bootstrap Ortho Slopes

Correlation and R-Squared Section

Parameter	Pearson Correlation Coefficient	R-Squared	Spearman Rank Correlation Coefficient
Estimated Value	0.9868	0.9738	0.9759
Lower 95% Conf. Limit (r dist'n)	0.9646		
Upper 95% Conf. Limit (r dist'n)	0.9945		
Lower 95% Conf. Limit (Fisher's z)	0.9662		0.9387
Upper 95% Conf. Limit (Fisher's z)	0.9949		0.9906
Adjusted (Rbar)		0.9723	
T-Value for H0: Rho = 0	25.8679	25.8679	18.9539
Prob Level for H0: Rho = 0	0.0000	0.0000	0.0000
Prob Level (Randomization Test N =1000)	0.0010		

Notes:
The confidence interval for the Pearson correlation assumes that X and Y follow the bivariate normal distribution. This is a different assumption from linear regression which assumes that X is fixed and Y is normally distributed.

Two confidence intervals are given. The first is based on the exact distribution of Pearson's correlation. The second is based on Fisher's z transformation which approximates the exact distribution using the normal distribution. Why are both provided? Because most books only mention Fisher's approximate method, it will often be needed to do homework. However, the exact methods should be used whenever possible.

The confidence limits can be used to test hypotheses about the correlation. To test the hypothesis that rho is a specific value, say r0, check to see if r0 is between the confidence limits. If it is, the null hypothesis that rho = r0 is not rejected. If r0 is outside the limits, the null hypothesis is rejected.

Spearman's Rank correlation is calculated by replacing the orginal data with their ranks. This correlation is used when some of the assumptions may be invalid.

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	77128.2	77128.2			
Slope	1	1318.337	1318.337	669.1468	0.0000	1.0000
Error	18	35.46317	1.970176			
Lack of Fit	16	34.96317	2.185198	8.7408	0.1074	
Pure Error	2	0.5	0.25			
Adj. Total	19	1353.8	71.25263			
Total	20	78482				

s = Square Root(1.970176) = 1.40363

Summary Matrices

Index	X'X 0	X'X 1	X'Y 2	X'X Inverse 0	X'X Inverse 1
0	20	2792	1242	0.6015912	-0.003951227
1	2792	425094	180208	-0.003951227	2.830392E-05
2 (Y'Y)			78482		
Determinant		706616			1.415196E-06

Variance - Covariance Matrix of Regression Coefficients

Index	VC(b) 0	VC(b) 1
0	1.185241	-0.007784612
1	-0.007784612	5.576369E-05

Tests of Assumptions Section

Assumption/Test	Test Value	Prob Level	Is the Assumption Reasonable at the 0.2000 Level of Significance?
Residuals follow Normal Distribution?			
Shapiro Wilk	0.9728	0.812919	Yes
Anderson Darling	0.2751	0.660885	Yes
D'Agostino Skewness	-0.9590	0.337543	Yes
D'Agostino Kurtosis	0.1205	0.904066	Yes
D'Agostino Omnibus	0.9343	0.626796	Yes
Constant Residual Variance?			
Modified Levene Test	0.0946	0.761964	Yes
Relationship is a Straight Line?			
Lack of Linear Fit F(16, 2) Test	8.7408	0.107381	No

No Serial Correlation?
Evaluate the Serial-Correlation report and the Durbin-Watson test if you have equal-spaced, time series data.

Notes:
A 'Yes' means there is not enough evidence to make this assumption seem unreasonable. This lack of evidence may be because the sample size is too small, the assumptions of the test itself are not met, or the assumption is valid.
A 'No' means the that the assumption is not reasonable. However, since these tests are related to sample size, you should assess the role of sample size in the tests by also evaluating the appropriate plots and graphs. A large dataset (say N > 500) will often fail at least one of the normality tests because it is hard to find a large dataset that is perfectly normal.

Normality and Constant Residual Variance:
Possible remedies for the failure of these assumptions include using a transformation of Y such as the log or square root, correcting data-recording errors found by looking into outliers, adding additional independent variables, using robust regression, or using bootstrap methods.

Serial Correlation of Residuals Section

Lag	Serial Correlation	Lag	Serial Correlation	Lag	Serial Correlation
1	0.1029	9	-0.2353	17	
2	-0.4127*	10	-0.0827	18	
3	0.0340	11	-0.0316	19	
4	0.2171	12	-0.0481	20	
5	-0.1968	13	0.0744	21	
6	-0.0194	14	0.0073	22	
7	0.2531	15		23	
8	-0.0744	16		24	

Notes:
Each serial correlation is the Pearson correlation calculated between the original series of residuals and the residuals lagged the specified number of periods. This feature of residuals is only meaningfull for data obtained sorted in time order. One of the assumptions is that none of these serial correlations is significant. Starred correlations are those for which |Fisher's Z| > 1.645 which indicates whether the serial correlation is 'large.'

If serial correlation is detected in time series data, the remedy is to account for it either by replacing Y with first differences or by fitting the serial pattern using a method such as that proposed by Cochrane and Orcutt.

Durbin-Watson Test For Serial Correlation

Parameter	Value	Did the Test Reject H0: Rho(1) = 0?
Durbin-Watson Value	1.6978	
Prob. Level: Positive Serial Correlation	0.2366	No
Prob. Level: Negative Serial Correlation	0.7460	No

Notes:
The Durbin-Watson test was created to test for first-order serial correlation in regression data taken over time. If the rows of your dataset do not represent successive time periods, you should ignore this test.

This report gives the probability of rejecting the null hypothesis of no first-order serial correlation. Possible remedies for serial correlation were given in the Notes to the Serial Correlation report, above.

PRESS Section

Parameter	From PRESS Residuals	From Regular Residuals
Sum of Squared Residuals	43.15799	35.46317
Sum of |Residuals|	24.27421	22.02947
R-Squared	0.9681	0.9738

Notes:
A PRESS residual is found by estimating the regression equation without the observation, predicting the dependent variable, and subtracting the predicted value from the actual value. The PRESS values are calculated from these PRESS residuals. The Regular values are the corresponding calculations based on the regular residuals.

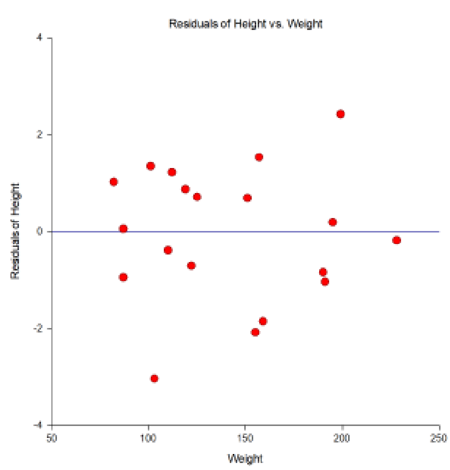
The PRESS values are often used to compare models in a multiple-regression variable selection. They show how well the model predicts observations that were not used in the estimation.

Predicted Values and Confidence Limits Section

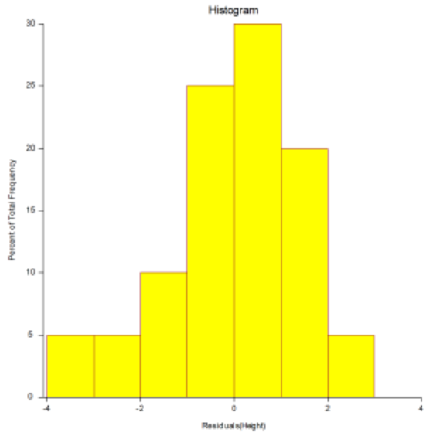
Weight (X)	Predicted Height (Yhat|X)	Standard Error of Yhat	Lower 95% Confidence Limit of Y|X	Upper 95% Confidence Limit of Y|X
90.0000	52.5188	0.4855	51.4989	53.5388
100.0000	54.4505	0.4312	53.5446	55.3565
150.0000	64.1090	0.3233	63.4297	64.7882
200.0000	73.7674	0.5495	72.6129	74.9218
250.0000	83.4258	0.8821	81.5725	85.2791

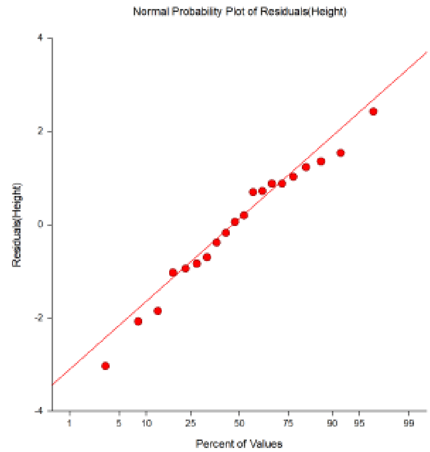
The confidence interval estimates the mean of the Y values in a large sample of individuals with this value of X. The interval is only accurate if all of the linear regression assumptions are valid.

Residual vs X Plots



Distributional Plots of Residuals





Original Data Section

Row	Weight (X)	Height (Y)	Predicted Height (Yhat|X)	Residual
1	159.0000	64.0000	65.8475	-1.8475
2	155.0000	63.0000	65.0748	-2.0748
3	157.0000	67.0000	65.4611	1.5389
4	125.0000	60.0000	59.2797	0.7203
5	103.0000	52.0000	55.0300	-3.0300
6	122.0000	58.0000	58.7002	-0.7002
7	101.0000	56.0000	54.6437	1.3563
8	82.0000	52.0000	50.9735	1.0265
9	228.0000	79.0000	79.1761	-0.1761
10	199.0000	76.0000	73.5742	2.4258
11	195.0000	73.0000	72.8015	0.1985
12	110.0000	56.0000	56.3822	-0.3822
13	191.0000	71.0000	72.0289	-1.0289
14	151.0000	65.0000	64.3021	0.6979
15	119.0000	59.0000	58.1207	0.8793
16	119.0000	59.0000	58.1207	0.8793
17	112.0000	58.0000	56.7685	1.2315
18	87.0000	51.0000	51.9393	-0.9393
19	190.0000	71.0000	71.8357	-0.8357
20	87.0000	52.0000	51.9393	0.0607
21	100.0000		54.4505	
22	150.0000		64.1090	
23	200.0000		73.7674	
24		50.0000		
25		60.0000		
26		70.0000		

This report provides a data list that may be used to verify whether the correct variables were selected.

Predicted Values and Confidence Limits of Means

Row	Weight (X)	Height (Y)	Predicted Height (Yhat|X)	Standard Error of Yhat	Lower 95% Conf. Limit of Y Mean|X	Upper 95% Conf. Limit of Y Mean|X
1	159.0000	64.0000	65.8475	0.3457	65.1212	66.5737
2	155.0000	63.0000	65.0748	0.3343	64.3725	65.7771
3	157.0000	67.0000	65.4611	0.3397	64.7475	66.1748
4	125.0000	60.0000	59.2797	0.3323	58.5817	59.9778
5	103.0000	52.0000	55.0300	0.4162	54.1557	55.9044
6	122.0000	58.0000	58.7002	0.3403	57.9854	59.4151
7	101.0000	56.0000	54.6437	0.4261	53.7484	55.5390
8	82.0000	52.0000	50.9735	0.5325	49.8548	52.0922
9	228.0000	79.0000	79.1761	0.7309	77.6404	80.7118
10	199.0000	76.0000	73.5742	0.5434	72.4326	74.7158
11	195.0000	73.0000	72.8015	0.5193	71.7106	73.8925
12	110.0000	56.0000	56.3822	0.3839	55.5757	57.1887
13	191.0000	71.0000	72.0289	0.4958	70.9872	73.0705
14	151.0000	65.0000	64.3021	0.3252	63.6189	64.9853
15	119.0000	59.0000	58.1207	0.3495	57.3864	58.8551
16	119.0000	59.0000	58.1207	0.3495	57.3864	58.8551
17	112.0000	58.0000	56.7685	0.3755	55.9797	57.5574
18	87.0000	51.0000	51.9393	0.5028	50.8830	52.9956
19	190.0000	71.0000	71.8357	0.4901	70.8061	72.8653
20	87.0000	52.0000	51.9393	0.5028	50.8830	52.9956
21	100.0000		54.4505	0.4312	53.5446	55.3565
22	150.0000		64.1090	0.3233	63.4297	64.7882
23	200.0000		73.7674	0.5495	72.6129	74.9218
24		50.0000				
25		60.0000				
26		70.0000				

The confidence interval estimates the mean of the Y values in a large sample of individuals with this value of X.
The interval is only accurate if all of the linear regression assumptions are valid.

Predicted Values and Prediction Limits

Row	Weight (X)	Height (Y)	Predicted Height (Yhat|X)	Standard Error of Yhat	Lower 95% Prediction Limit of Y|X	Upper 95% Prediction Limit of Y|X
1	159.0000	64.0000	65.8475	1.4456	62.8104	68.8845
2	155.0000	63.0000	65.0748	1.4429	62.0434	68.1062
3	157.0000	67.0000	65.4611	1.4441	62.4271	68.4952
4	125.0000	60.0000	59.2797	1.4424	56.2493	62.3101
5	103.0000	52.0000	55.0300	1.4640	51.9542	58.1058
6	122.0000	58.0000	58.7002	1.4443	55.6659	61.7346
7	101.0000	56.0000	54.6437	1.4669	51.5619	57.7255
8	82.0000	52.0000	50.9735	1.5012	47.8195	54.1275
9	228.0000	79.0000	79.1761	1.5825	75.8513	82.5009
10	199.0000	76.0000	73.5742	1.5051	70.4120	76.7364
11	195.0000	73.0000	72.8015	1.4966	69.6573	75.9458
12	110.0000	56.0000	56.3822	1.4552	53.3250	59.4394
13	191.0000	71.0000	72.0289	1.4886	68.9014	75.1564
14	151.0000	65.0000	64.3021	1.4408	61.2751	67.3291
15	119.0000	59.0000	58.1207	1.4465	55.0818	61.1597
16	119.0000	59.0000	58.1207	1.4465	55.0818	61.1597
17	112.0000	58.0000	56.7685	1.4530	53.7159	59.8212
18	87.0000	51.0000	51.9393	1.4910	48.8069	55.0717
19	190.0000	71.0000	71.8357	1.4867	68.7122	74.9592
20	87.0000	52.0000	51.9393	1.4910	48.8069	55.0717
21	100.0000		54.4505	1.4684	51.3656	57.5355
22	150.0000		64.1090	1.4404	61.0828	67.1351
23	200.0000		73.7674	1.5074	70.6005	76.9342
24		50.0000				
25		60.0000				
26		70.0000				

The prediction interval estimates the predicted value of Y for a single individual with this value of X.
The interval is only accurate if all of the linear regression assumptions are valid.

Working-Hotelling Simultaneous Confidence Band

Row	Weight (X)	Height (Y)	Predicted Height (Yhat|X)	Standard Error of Yhat	Lower 95% Conf. Band of Y Mean|X	Upper 95% Conf. Band of Y Mean|X
1	159.0000	64.0000	65.8475	0.3457	64.8036	66.8914
2	155.0000	63.0000	65.0748	0.3343	64.0654	66.0842
3	157.0000	67.0000	65.4611	0.3397	64.4353	66.4869
4	125.0000	60.0000	59.2797	0.3323	58.2764	60.2831
5	103.0000	52.0000	55.0300	0.4162	53.7732	56.2868
6	122.0000	58.0000	58.7002	0.3403	57.6727	59.7278
7	101.0000	56.0000	54.6437	0.4261	53.3568	55.9306
8	82.0000	52.0000	50.9735	0.5325	49.3655	52.5814
9	228.0000	79.0000	79.1761	0.7309	76.9688	81.3834
10	199.0000	76.0000	73.5742	0.5434	71.9333	75.2151
11	195.0000	73.0000	72.8015	0.5193	71.2334	74.3697
12	110.0000	56.0000	56.3822	0.3839	55.2229	57.5415
13	191.0000	71.0000	72.0289	0.4958	70.5316	73.5261
14	151.0000	65.0000	64.3021	0.3252	63.3201	65.2842
15	119.0000	59.0000	58.1207	0.3495	57.0652	59.1763
16	119.0000	59.0000	58.1207	0.3495	57.0652	59.1763
17	112.0000	58.0000	56.7685	0.3755	55.6347	57.9024
18	87.0000	51.0000	51.9393	0.5028	50.4210	53.4577
19	190.0000	71.0000	71.8357	0.4901	70.3558	73.3156
20	87.0000	52.0000	51.9393	0.5028	50.4210	53.4577
21	100.0000		54.4505	0.4312	53.1483	55.7527
22	150.0000		64.1090	0.3233	63.1326	65.0853
23	200.0000		73.7674	0.5495	72.1080	75.4268
24		50.0000				
25		60.0000				
26		70.0000				

Residual Section

Row	Weight (X)	Height (Y)	Predicted Height (Yhat|X)	Residual	Standardized Residual	Percent Absolute Error
1	159.0000	64.0000	65.8475	-1.8475	-1.3580	2.8867
2	155.0000	63.0000	65.0748	-2.0748	-1.5220	3.2933
3	157.0000	67.0000	65.4611	1.5389	1.1299	2.2968
4	125.0000	60.0000	59.2797	0.7203	0.5282	1.2004
5	103.0000	52.0000	55.0300	-3.0300	-2.2604	5.8270
6	122.0000	58.0000	58.7002	-0.7002	-0.5142	1.2073
7	101.0000	56.0000	54.6437	1.3563	1.0142	2.4220
8	82.0000	52.0000	50.9735	1.0265	0.7904	1.9741
9	228.0000	79.0000	79.1761	-0.1761	-0.1470	0.2229
10	199.0000	76.0000	73.5742	2.4258	1.8744	3.1918
11	195.0000	73.0000	72.8015	0.1985	0.1522	0.2719
12	110.0000	56.0000	56.3822	-0.3822	-0.2831	0.6825
13	191.0000	71.0000	72.0289	-1.0289	-0.7835	1.4491
14	151.0000	65.0000	64.3021	0.6979	0.5111	1.0737
15	119.0000	59.0000	58.1207	0.8793	0.6468	1.4903
16	119.0000	59.0000	58.1207	0.8793	0.6468	1.4903
17	112.0000	58.0000	56.7685	1.2315	0.9105	2.1232
18	87.0000	51.0000	51.9393	-0.9393	-0.7168	1.8418
19	190.0000	71.0000	71.8357	-0.8357	-0.6354	1.1770
20	87.0000	52.0000	51.9393	0.0607	0.0463	0.1167
21	100.0000		54.4505			
22	150.0000		64.1090			
23	200.0000		73.7674			
24		50.0000				
25		60.0000				
26		70.0000				

The residual is the difference between the actual and the predicted Y values. The formula is Residual = Y - Yhat. The Percent Absolute Error is the 100 |Residual| / Y.

Residual Diagnostics Section

Row	Weight (X)	Residual	RStudent	Hat Diagonal	Cook's D	MSEi
1	159.0000	-1.8475	-1.3931	0.0607	0.0595	1.8723
2	155.0000	-2.0748	-1.5845	0.0567	0.0696	1.8176
3	157.0000	1.5389	1.1392	0.0586	0.0397	1.9381
4	125.0000	0.7203	0.5173	0.0560	0.0083	2.0537
5	103.0000	-3.0300	*-2.5957	0.0879	0.2462	1.4939
6	122.0000	-0.7002	-0.5034	0.0588	0.0083	2.0554
7	101.0000	1.3563	1.0150	0.0922	0.0522	1.9669
8	82.0000	1.0265	0.7818	0.1439	0.0525	2.0137
9	228.0000	-0.1761	*-0.1429	0.2712	0.0040	2.0836
10	199.0000	2.4258	*2.0305	0.1499	0.3097	1.6789
11	195.0000	0.1985	0.1480	0.1369	0.0018	2.0834
12	110.0000	-0.3822	-0.2757	0.0748	0.0032	2.0768
13	191.0000	-1.0289	-0.7748	0.1248	0.0438	2.0149
14	151.0000	0.6979	0.5003	0.0537	0.0074	2.0558
15	119.0000	0.8793	0.6360	0.0620	0.0138	2.0376
16	119.0000	0.8793	0.6360	0.0620	0.0138	2.0376
17	112.0000	1.2315	0.9060	0.0716	0.0319	1.9900
18	87.0000	-0.9393	-0.7067	0.1283	0.0378	2.0265
19	190.0000	-0.8357	-0.6245	0.1219	0.0280	2.0393
20	87.0000	0.0607	0.0450	0.1283	0.0002	2.0858
21	100.0000			0.0944		
22	150.0000			0.0531		
23	200.0000			0.1533		
24						
25						
26						

Outliers are rows that are separated from the rest of the data. Influential rows are those whose omission results in a relatively large change in the results. This report lets you see both.

An outlier may be defined as a row in which |RStudent| > 2. A moderately influential row is one with a CooksD > 0.5. A heavily influential row is one with a CooksD > 1.

Leave One Row Out Section

Row	RStudent	DFFITS	Cook's D	CovRatio	DFBETAS(0)	DFBETAS(1)
1	-1.3931	-0.3540	0.0595	0.9615	0.0494	-0.1483
2	-1.5845	-0.3885	0.0696	0.9023	0.0228	-0.1337
3	1.1392	0.2842	0.0397	1.0279	-0.0284	0.1087
4	0.5173	0.1260	0.0083	1.1511	0.0739	-0.0414
5	* -2.5957	-0.8059	0.2462	0.6304	-0.6820	0.5292
6	-0.5034	-0.1258	0.0083	1.1564	-0.0800	0.0486
7	1.0150	0.3234	0.0522	1.0978	0.2781	-0.2188
8	0.7818	0.3205	0.0525	1.2202	0.3024	-0.2589
9	-0.1429	-0.0872	0.0040	* 1.5346	0.0646	-0.0787
10	* 2.0305	0.8525	0.3097	0.8542	-0.5244	0.6959
11	0.1480	0.0589	0.0018	1.2955	-0.0347	0.0470
12	-0.2757	-0.0784	0.0032	1.2010	-0.0617	0.0451
13	-0.7748	-0.2925	0.0438	1.1951	0.1635	-0.2265
14	0.5003	0.1192	0.0074	1.1506	0.0033	0.0312
15	0.6360	0.1635	0.0138	1.1403	0.1112	-0.0720
16	0.6360	0.1635	0.0138	1.1403	0.1112	-0.0720
17	0.9060	0.2515	0.0319	1.0988	0.1928	-0.1381
18	-0.7067	-0.2712	0.0378	1.2138	-0.2516	0.2118
19	-0.6245	-0.2327	0.0280	1.2201	0.1281	-0.1787
20	0.0450	0.0173	0.0002	1.2858	0.0160	-0.0135
21						
22						
23						
24						
25						
26						

Each column gives the impact on some aspect of the linear regression of omitting that row.

RStudent represents the size of the residual. DFFITS represents the change in the fitted value of a row. Cook's D summarizes the change in the fitted values of all rows. CovRatio represents the amount of change in the determinant of the covariance matrix. DFBETAS(0) and DFBETAS(1) give the amount of change in the intercept and slope.

Outlier Detection Chart

Row	Weight (X)	Residual	Standardized Residual	RStudent
1	159.0000	-1.8475	-1.3580	-1.3931
2	155.0000	-2.0748	-1.5220	-1.5845
3	157.0000	1.5389	1.1299	1.1392
4	125.0000	0.7203	0.5282	0.5173
5	103.0000	-3.0300	-2.2604	* -2.5957
6	122.0000	-0.7002	-0.5142	-0.5034
7	101.0000	1.3563	1.0142	1.0150
8	82.0000	1.0265	0.7904	0.7818
9	228.0000	-0.1761	-0.1470	-0.1429
10	199.0000	2.4258	1.8744	* 2.0305
11	195.0000	0.1985	0.1522	0.1480
12	110.0000	-0.3822	-0.2831	-0.2757
13	191.0000	-1.0289	-0.7835	-0.7748
14	151.0000	0.6979	0.5111	0.5003
15	119.0000	0.8793	0.6468	0.6360
16	119.0000	0.8793	0.6468	0.6360
17	112.0000	1.2315	0.9105	0.9060
18	87.0000	-0.9393	-0.7168	-0.7067
19	190.0000	-0.8357	-0.6354	-0.6245
20	87.0000	0.0607	0.0463	0.0450
21	100.0000			
22	150.0000			
23	200.0000			
24				
25				
26				

Outliers are rows that are separated from the rest of the data. Since outliers can have dramatic effects on the results, corrective action, such as elimination, must be carefully considered. Outlying rows should not be automatically be removed unless a good reason for their removal can be given.

An outlier may be defined as a row in which |RStudent| > 2. Rows with this characteristic have been starred.

Influence Detection Chart

Row	Weight (X)	DFFITS	Cook's D	DFBETAS(1)
1	159.0000	-0.3540	0.0595	-0.1483
2	155.0000	-0.3885	0.0696	-0.1337
3	157.0000	0.2842	0.0397	0.1087
4	125.0000	0.1260	0.0083	-0.0414
5	103.0000	-0.8059	0.2462	0.5292
6	122.0000	-0.1258	0.0083	0.0486
7	101.0000	0.3234	0.0522	-0.2188
8	82.0000	0.3205	0.0525	-0.2589
9	228.0000	-0.0872	0.0040	-0.0787
10	199.0000	0.8525	0.3097	0.6959
11	195.0000	0.0589	0.0018	0.0470
12	110.0000	-0.0784	0.0032	0.0451
13	191.0000	-0.2925	0.0438	-0.2265
14	151.0000	0.1192	0.0074	0.0312
15	119.0000	0.1635	0.0138	-0.0720
16	119.0000	0.1635	0.0138	-0.0720
17	112.0000	0.2515	0.0319	-0.1381
18	87.0000	-0.2712	0.0378	0.2118
19	190.0000	-0.2327	0.0280	-0.1787
20	87.0000	0.0173	0.0002	-0.0135
21	100.0000			
22	150.0000			
23	200.0000			
24				
25				
26				

Influential rows are those whose omission results in a relatively large change in the results. They are not necessarily harmful. However, they will distort the results if they are also outliers. The impact of influential rows should be studied very carefully. Their accuracy should be double-checked.
DFFITS is the standardized change in Yhat when the row is omitted. A row is influential when DFFITS > 1 for small datasets (N < 30) or when DFFITS > 2*SQR(1/N) for medium to large datasets.

Outlier & Influence Chart

Row	Weight (X)	R Student (Outlier)	Cooks D (Influence)	Hat Diagonal (Leverage)
1	159.0000	-1.3931	0.0595	0.0607
2	155.0000	-1.5845	0.0696	0.0567
3	157.0000	1.1392	0.0397	0.0586
4	125.0000	0.5173	0.0083	0.0560
5	103.0000	* -2.5957	0.2462	0.0879
6	122.0000	-0.5034	0.0083	0.0588
7	101.0000	1.0150	0.0522	0.0922
8	82.0000	0.7818	0.0525	0.1439
9	228.0000	-0.1429	0.0040	0.2712
10	199.0000	* 2.0305	0.3097	0.1499
11	195.0000	0.1480	0.0018	0.1369
12	110.0000	-0.2757	0.0032	0.0748
13	191.0000	-0.7748	0.0438	0.1248
14	151.0000	0.5003	0.0074	0.0537
15	119.0000	0.6360	0.0138	0.0620
16	119.0000	0.6360	0.0138	0.0620
17	112.0000	0.9060	0.0319	0.0716
18	87.0000	-0.7067	0.0378	0.1283
19	190.0000	-0.6245	0.0280	0.1219
20	87.0000	0.0450	0.0002	0.1283
21	100.0000			0.0944
22	150.0000			0.0531
23	200.0000			0.1533
24				
25				
26				

Outliers are rows that are separated from the rest of the data. Influential rows are those whose omission results in a relatively large change in the results. This report lets you see both.

Inverse Prediction of X Means

Row	Height (Y)	Weight (X)	Predicted Weight (Xhat|Y)	X-Xhat|Y	Lower 95% Conf. Limit of X Mean|Y	Upper 95% Conf. Limit of X Mean|Y
1	64.0000	159.0000	149.4360	9.5640	145.9832	153.0193
2	63.0000	155.0000	144.2591	10.7409	140.8441	147.7361
3	67.0000	157.0000	164.9664	-7.9664	161.1310	169.1387
4	60.0000	125.0000	128.7287	-3.7287	125.1181	132.1948
5	52.0000	103.0000	87.3141	15.6859	81.4894	92.4444
6	58.0000	122.0000	118.3750	3.6250	114.3947	122.0735
7	56.0000	101.0000	108.0214	-7.0214	103.5227	112.1007
8	52.0000	82.0000	87.3141	-5.3141	81.4894	92.4444
9	79.0000	228.0000	227.0884	0.9116	219.7388	235.5997
10	76.0000	199.0000	211.5579	-12.5579	205.2283	218.8430
11	73.0000	195.0000	196.0274	-1.0274	190.6564	202.1477
12	56.0000	110.0000	108.0214	1.9786	103.5227	112.1007
13	71.0000	191.0000	185.6737	5.3263	180.8886	191.0708
14	65.0000	151.0000	154.6128	-3.6128	151.0743	158.3507
15	59.0000	119.0000	123.5518	-4.5518	119.7777	127.1129
16	59.0000	119.0000	123.5518	-4.5518	119.7777	127.1129
17	58.0000	112.0000	118.3750	-6.3750	114.3947	122.0735
18	51.0000	87.0000	82.1372	4.8628	75.9418	87.5695
19	71.0000	190.0000	185.6737	4.3263	180.8886	191.0708
20	52.0000	87.0000	87.3141	-0.3141	81.4894	92.4444
21		100.0000				
22		150.0000				
23		200.0000				
24	50.0000		76.9604		70.3836	82.7054
25	60.0000		128.7287		125.1181	132.1948
26	70.0000		180.4969		175.9821	185.5549

This confidence interval estimates the mean of X in a large sample of individuals with this value of Y. This method of inverse prediction is also called 'calibration.'

Inverse Prediction of X Individuals

Row	Height (Y)	Weight (X)	Predicted Weight (Xhat|Y)	X-Xhat|Y	Lower 95% Prediction Limit of X|Y	Upper 95% Prediction Limit of X|Y
1	64.0000	159.0000	149.4360	9.5640	133.7858	165.2167
2	63.0000	155.0000	144.2591	10.7409	128.5906	159.9896
3	67.0000	157.0000	164.9664	-7.9664	149.3036	180.9662
4	60.0000	125.0000	128.7287	-3.7287	112.9365	144.3765
5	52.0000	103.0000	87.3141	15.6859	70.7003	103.2335
6	58.0000	122.0000	118.3750	3.6250	102.4436	134.0246
7	56.0000	101.0000	108.0214	-7.0214	91.9059	123.7175
8	52.0000	82.0000	87.3141	-5.3141	70.7003	103.2335
9	79.0000	228.0000	227.0884	0.9116	210.4214	244.9172
10	76.0000	199.0000	211.5579	-12.5579	195.2744	228.7969
11	73.0000	195.0000	196.0274	-1.0274	180.0432	212.7609
12	56.0000	110.0000	108.0214	1.9786	91.9059	123.7175
13	71.0000	191.0000	185.6737	5.3263	169.8391	202.1203
14	65.0000	151.0000	154.6128	-3.6128	138.9697	170.4553
15	59.0000	119.0000	123.5518	-4.5518	107.6957	139.1949
16	59.0000	119.0000	123.5518	-4.5518	107.6957	139.1949
17	58.0000	112.0000	118.3750	-6.3750	102.4436	134.0246
18	51.0000	87.0000	82.1372	4.8628	65.3728	98.1386
19	71.0000	190.0000	185.6737	4.3263	169.8391	202.1203
20	52.0000	87.0000	87.3141	-0.3141	70.7003	103.2335
21		100.0000				
22		150.0000				
23		200.0000				
24	50.0000		76.9604		60.0352	93.0537
25	60.0000		128.7287		112.9365	144.3765
26	70.0000		180.4969		164.7214	196.8156

This prediction interval estimates the predicted value of X for a single individual with this value of Y.
This method of inverse prediction is also called 'calibration.'

Este reporte tal vez sea el que más información nos regresa a partir de todo el análisis realizado, para comenzar nos da la grafica de los datos y de la recta ajustada a estos datos, lo que nos permite observar valores que estén muy alejados de esta línea y que tan bien se ajusta esta línea recta a los datos.

Después nos da la información de la regresión lineal realizada, de donde lo que nos importa es la pendiente y la ordenada al origen que en el caso de los datos del ejemplo obtuvimos:

Pendiente: 0.1932
Ordena al origen: 35.1337

Tenemos la estadística de cada una de las variables utilizadas para la regresión lineal, como la media, desviación estándar, máximo y mínimo.

El reporte además incluye datos como coeficiente de correlación, el error cuadrático medio, los intervalos de confianza de los datos obtenidos, para este ejemplo nos interesa conocer cuales son los valores altura que esta regresión lineal nos da para los siguientes pesos:

PESO	ALTURA
90	52.5188
100	54.4505
150	64.1090
200	73.7674
250	83.4258