

تقييم وحدة التدريب والدعم الفني بعمادة التطوير وضمان الجودة

الإحصاء الوصفي والاستدلالي باستخدام برمجية IBM SPSS statistics

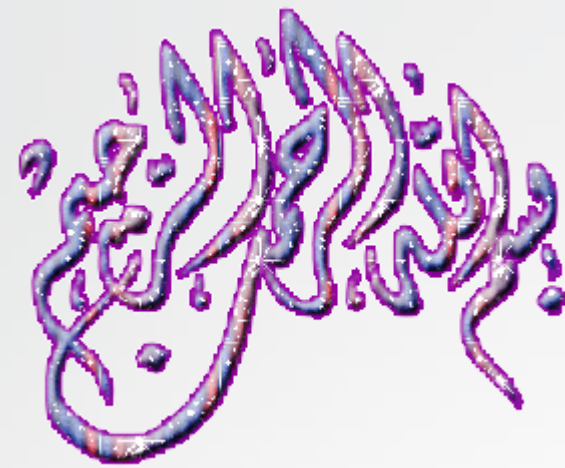
لأعضاء هيئة التدريس في الكليات الطبية

للمدرب:

أ.د. رشاد عبدالغني

وذلك ابتداءً من 2022/1/15م وحتى 2022/1/20م

قاعة السمнар - مبنى رئاسة الجامعة



Descriptive and Inferential Statistics Using IBM SPSS Statistics Software

Prepared by:

Dr. Rashad Abdul-Ghani, MPH, PhD (Associate Professor)





Descriptive Statistics Using IBM SPSS Statistics

Descriptive Statistics for Continuous Variables

➤ For **continuous variables**, we are interested in:

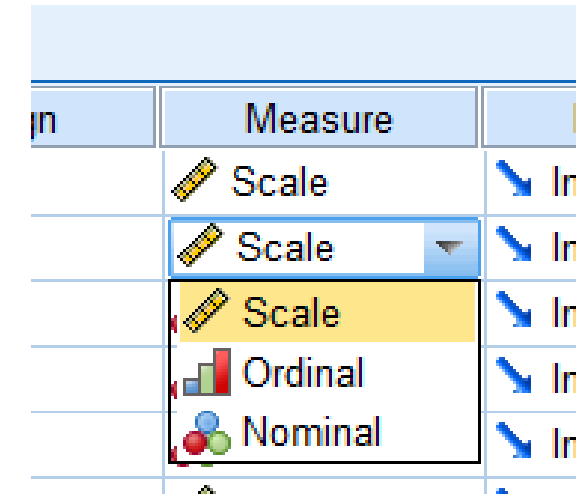
- ① Measures of **central tendency** → mean, median, mode
- ② Measures of **location** → quartiles, percentiles
- ③ Measures of **dispersion** and **spread** → standard deviation, variance, minimum, maximum, range, SE, outliers
- ④ Shape of **data distribution** → symmetry, skewness, kurtosis

Summary statistics to describe normal and skewed distributions

Statistic	Normal distribution	Skewed distribution
Centre	Mean (M)	Median (Mdn)
Spread	Standard deviation (SD) Variance (SD^2)	Inter-quartile (IQR) range
Precision	Standard error (SE) = SD/\sqrt{n} 95% confidence interval (95% CI) = $1.96 \times SE$	

➤ Continuous variables are measured at the **scale level**.

- **Scale variables** → with numeric values measured by an interval or ratio scale.



Measure	
Scale	In
Scale	In
Scale	In
Ordinal	In
Nominal	In

Interval scale

- A quantitative scale with no true zero and no equal intervals between neighboring points (e.g., temperature of 40°C is not as twice as 20°C).
- Examples: temperature, pH, and Likert scale

Ratio scale

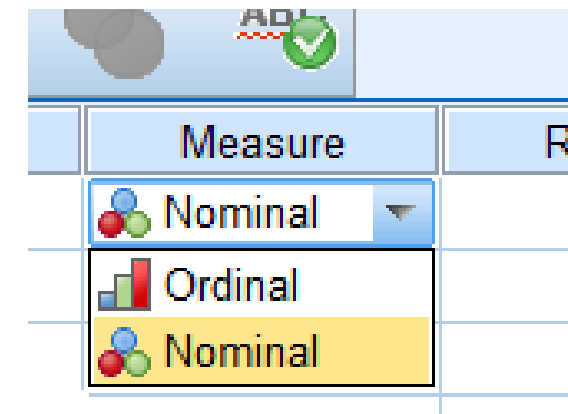
- A quantitative scale where there is a true zero and equal intervals between neighboring points.
- Examples: length, weight, income and age.

Descriptive Statistics for Categorical Variables

➤ For **categorical variables**, we are interested in:

- ① **Frequencies** → Number of cases in each category
- ② **Percentages** → Proportions of cases in each category
- ③ **Mode** → The most frequent category

➤ Categorical variables are measured at the **nominal** or **ordinal levels**.



Nominal

- Variables with categories that have no logical order or ranking.
- Example: “gender” with “male” and “female” categories.

Ordinal

- Variables with categories that have a logical order or ranking. and it is possible to measure
- Example: “degree of pain” with “mild”, “moderate” and “severe” categories.

Describing Data in SPSS

Menu

Analyze → Descriptive Statistics

→ Compare Means

Frequencies

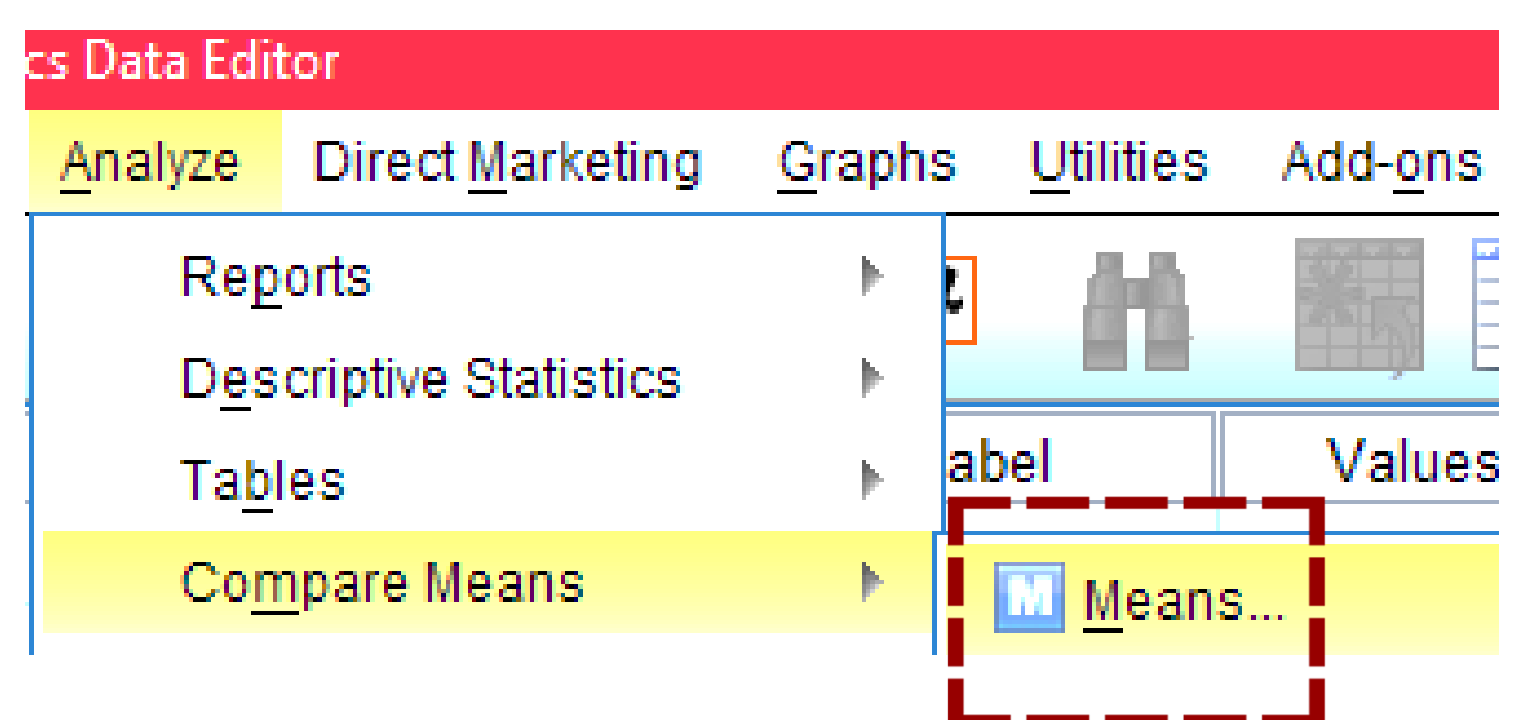
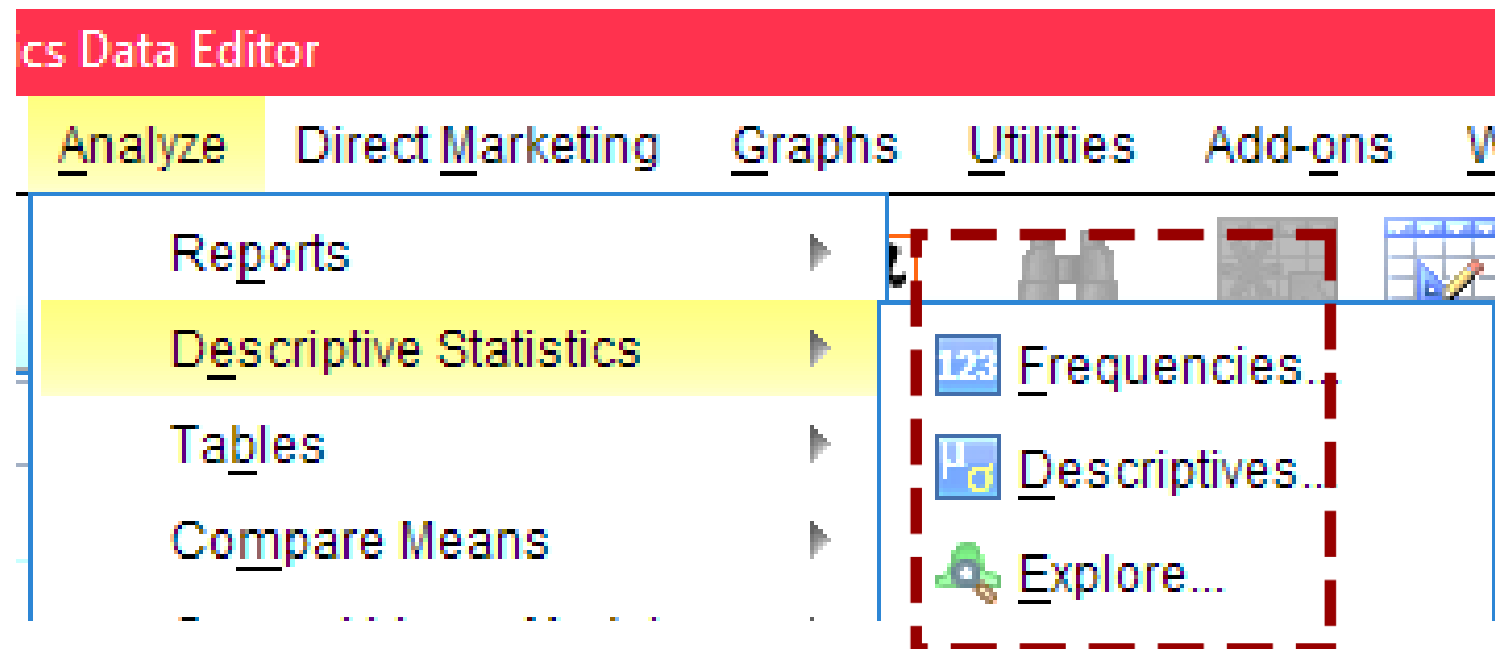
Descriptives

Explore

Means

Summary statistics for the entire sample

Descriptive statistics for sample subsets



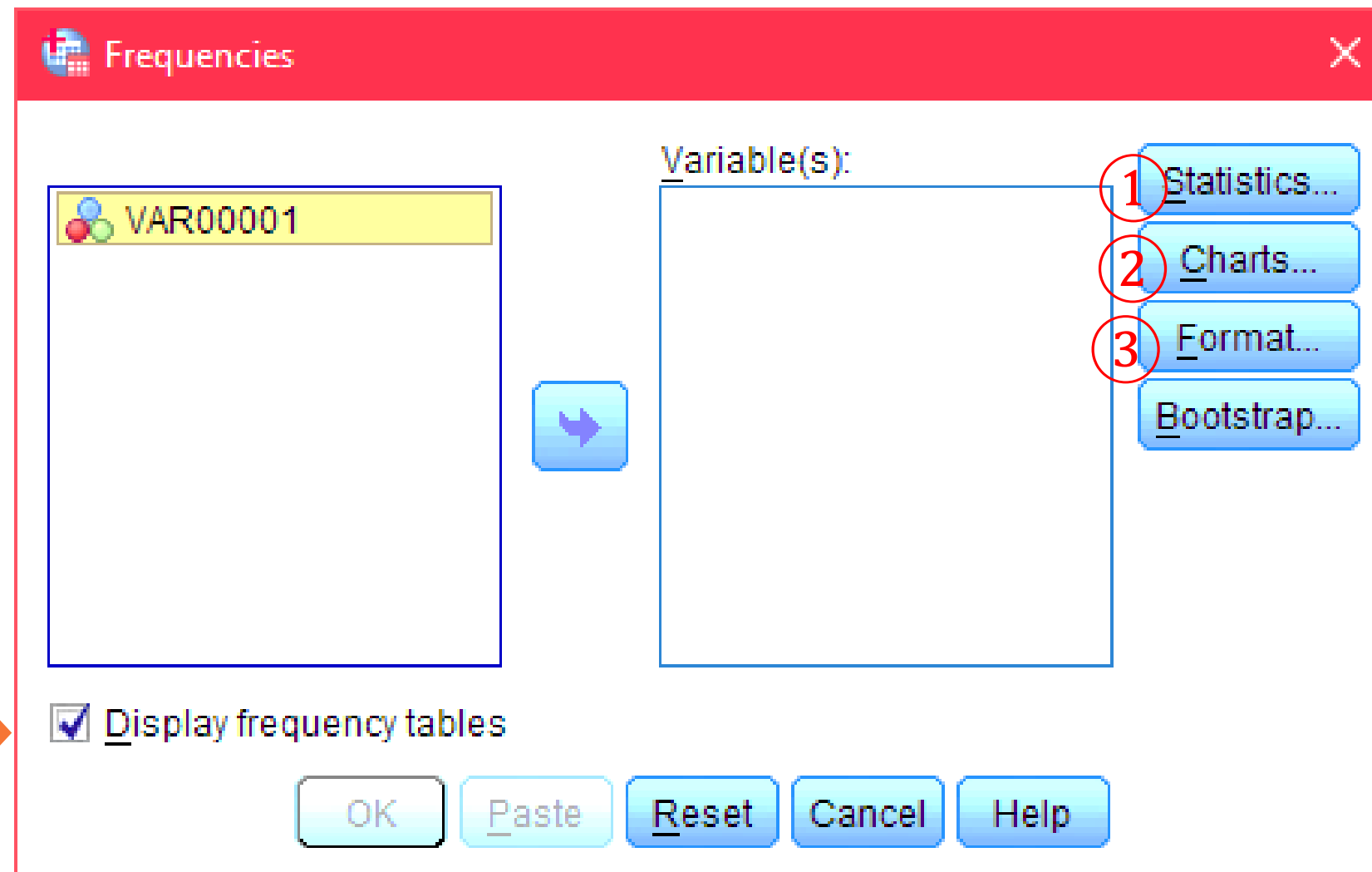


The *Frequencias*

Command

The *Frequencies* Command

- To produce **frequency distributions** of categorical variables (nominal or ordinal),
- To obtain measures of **central tendency** and **dispersion** as well as **distribution** of scale variables
- To obtain **percentile ranks of data** on an ordinal scale (not included in the Descriptives or Explore commands).



For categorical variables only

① Frequencies: Statistics

Percentile Values

- Quartiles
- Cut points for: 10 equal groups
- Percentile(s):

Central Tendency

- Mean
- Median
- Mode
- Sum

Dispersion

- Std. deviation
- Minimum
- Variance
- Maximum
- Range
- S.E. mean

Distribution

- Skewness
- Kurtosis

Values are group midpoints

Buttons: Add, Change, Remove, Continue, Cancel, Help

Callout: Not found in "Descriptives"

② Frequencies: Charts

Chart Type

- None
- Bar charts
- Pie charts
- Histograms:
 - Show normal curve on histogram

Chart Values

- Frequencies
- Percentages

Buttons: Continue, Cancel, Help

Bar/Pie charts → categorical variables
Histograms → scale variables

Practice

Use the **descriptives.sav** file to calculate **Frequencies** of the following variables: *height*, *weight* and *age*.



- In “**Frequencies**” window, move the variables to be analyzed into “**Variable(s)**” box
- Click “**Statistics**” button and select the appropriate options in “**Frequencies: Statistics**” window → Click the **Continue** button
- Click “**Charts**” button and select “**Histograms**” and tick “Show normal curve on histogram” option → Click **Continue** button.
- Click **OK** button.

The image displays two screenshots of the SPSS software interface. The left screenshot shows the "Frequencies" dialog box. On the left, a list of variables includes "Gender [gender]", "Weight six months a...", "Marital status [marital]", "Household income i...", and "Level of education [e...". A hand icon points to the "Variable(s)" list on the right, which contains "Height in inches [hei...", "Weight in pounds [w...", and "Age in years [age]". A blue arrow button is located between the two lists. Below the lists is a checkbox for "Display frequency tables" and a hand icon pointing to the "OK" button. On the right side of the dialog are buttons for "Statistics...", "Charts...", "Format...", and "Bootstrap...". A large black arrow points from the "Charts..." button to the right screenshot.

The right screenshot shows the "Frequencies: Charts" dialog box. Under the "Chart Type" section, the "Histograms:" radio button is selected, and the "Show normal curve on histogram" checkbox is checked. The "Chart Values" section has the "Frequencies" radio button selected. At the bottom are "Continue", "Cancel", and "Help" buttons.

Output

Frequencies

➔ [DataSet1] C:\Users\Rashad Abdul-Ghani\Desktop\SPSS WORKSHOP\decriptives.sav

Statistics

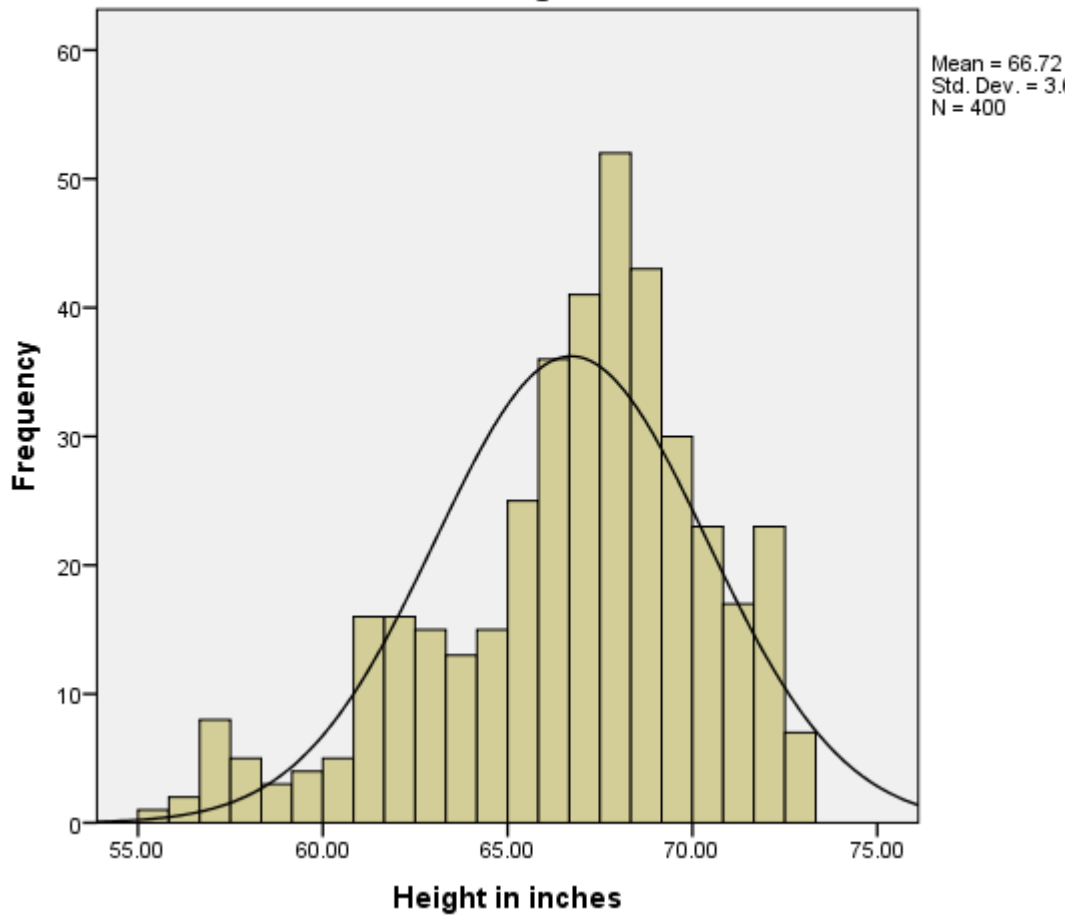
		Height in inches	Weight in pounds	Age in years
N	Valid	400	400	400
	Missing	0	0	0
Mean		66.7196	155.4151	42.08
Std. Error of Mean		.18353	.79530	.627
Median		67.3000	153.1450	41.50
Std. Deviation		3.67055	15.90601	12.534
Skewness		-.765-	.199	.267
Std. Error of Skewness		.122	.122	.122
Kurtosis		.284	-1.073-	-.623-
Std. Error of Kurtosis		.243	.243	.243
Percentiles	2.5	57.4005	129.3000	20.00
	25	64.8850	141.2000	32.00
	30	65.6260	144.1490	34.30
	50	67.3000	153.1450	41.50
	75	69.1925	168.3000	50.00

Output

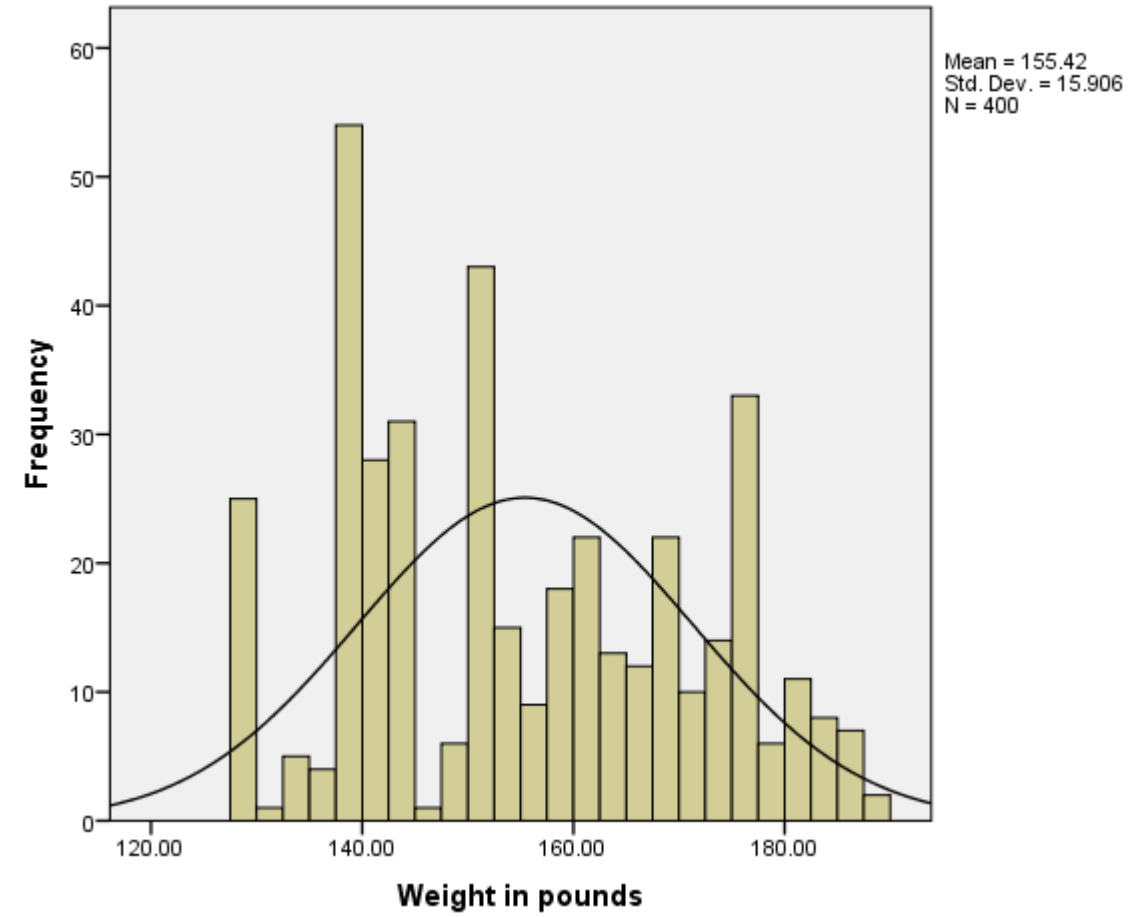


Histogram

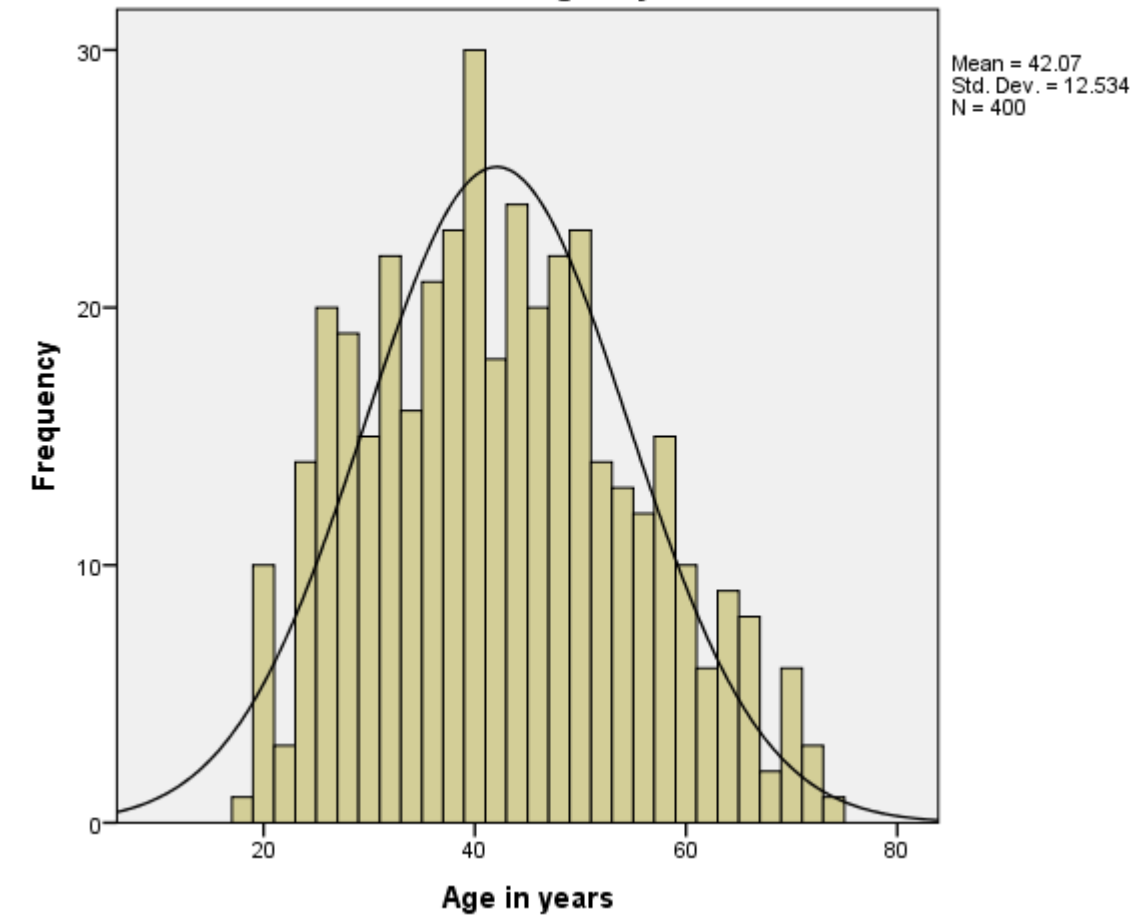
Height in inches



Weight in pounds



Age in years



Practice



Use the **descriptives.sav** file to calculate **Frequencies** for the following variables: *gender*, *marital status* and *level of education*.



- In “**Frequencies**” window, move the variables to be analyzed into “**Variable(s)**” box
→ Tick “**Display frequency tables**”
- Click “**Statistics**” button and choose Mode in “**Frequencies: Statistics**” window →
Click the **Continue** button
- Click “**Chart**” button and select “**Bar charts**” or “**Pie charts**” and choose
“**Percentage**” option from “**Chart Values**” → Click **Continue** button.
- Click **OK** button.

The image displays two overlapping dialog boxes from SPSS. The left dialog box is titled "Frequencies" and contains a list of variables on the left, a "Variable(s)" list on the right, and a "Display frequency tables" checkbox at the bottom left. The "Variable(s)" list includes "Gender [gender]", "Marital status [marital]", and "Level of education [e...]", with the latter highlighted. A hand icon points to the "Level of education" variable. The "Display frequency tables" checkbox is checked, and a hand icon points to it. The "OK" button is also highlighted with a hand icon. The right dialog box is titled "Frequencies: Charts" and shows "Chart Type" options: "None", "Bar charts" (selected), "Pie charts", and "Histograms" (with "Show normal curve on histogram" checked). The "Chart Values" section has "Frequencies" and "Percentages" (selected) options. A hand icon points to the "Bar charts" option. Both dialog boxes have "Continue", "Cancel", and "Help" buttons at the bottom.

Output



→ Frequencies

Statistics

		Gender	Marital status	Level of education
N	Valid	400	400	400
	Missing	0	0	0
Mode		1	1	2

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	207	51.8	51.8	51.8
	female	193	48.3	48.3	100.0
	Total	400	100.0	100.0	

Marital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	194	48.5	48.5	48.5
	Married	206	51.5	51.5	100.0
	Total	400	100.0	100.0	

Level of education

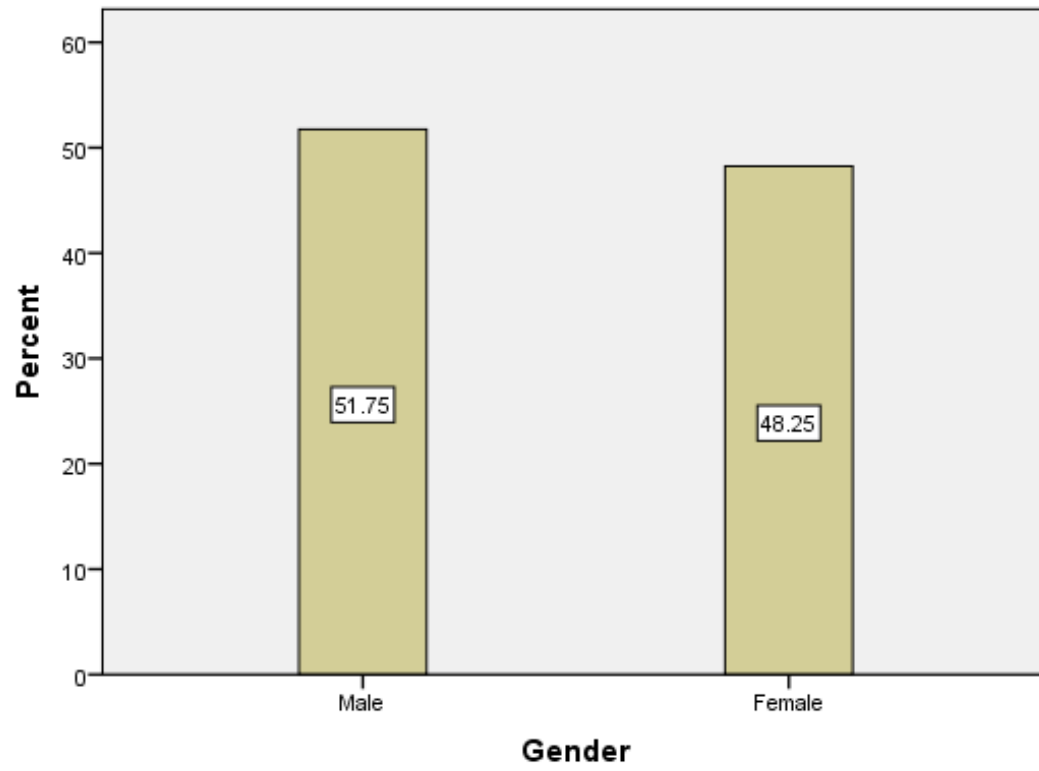
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Uneducated	84	21.0	21.0	21.0
	Read and write	123	30.8	30.8	51.8
	Primary education	76	19.0	19.0	70.8
	Secondary education	84	21.0	21.0	91.8
	University and post-graduate education	33	8.3	8.3	100.0
	Total	400	100.0	100.0	

Output

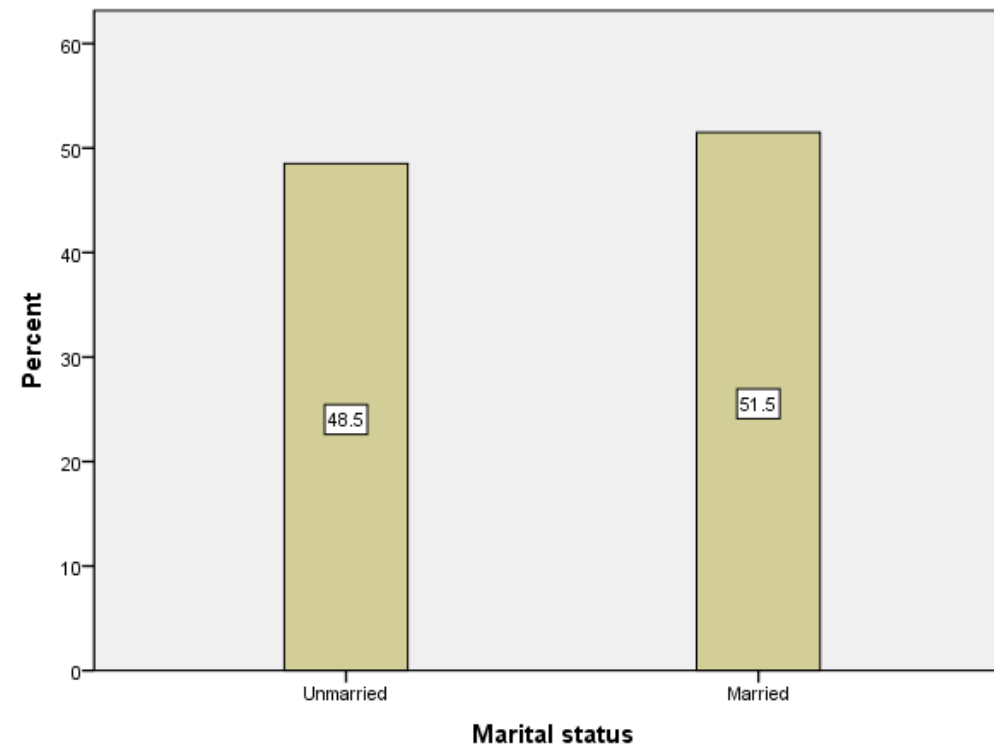


Bar Chart

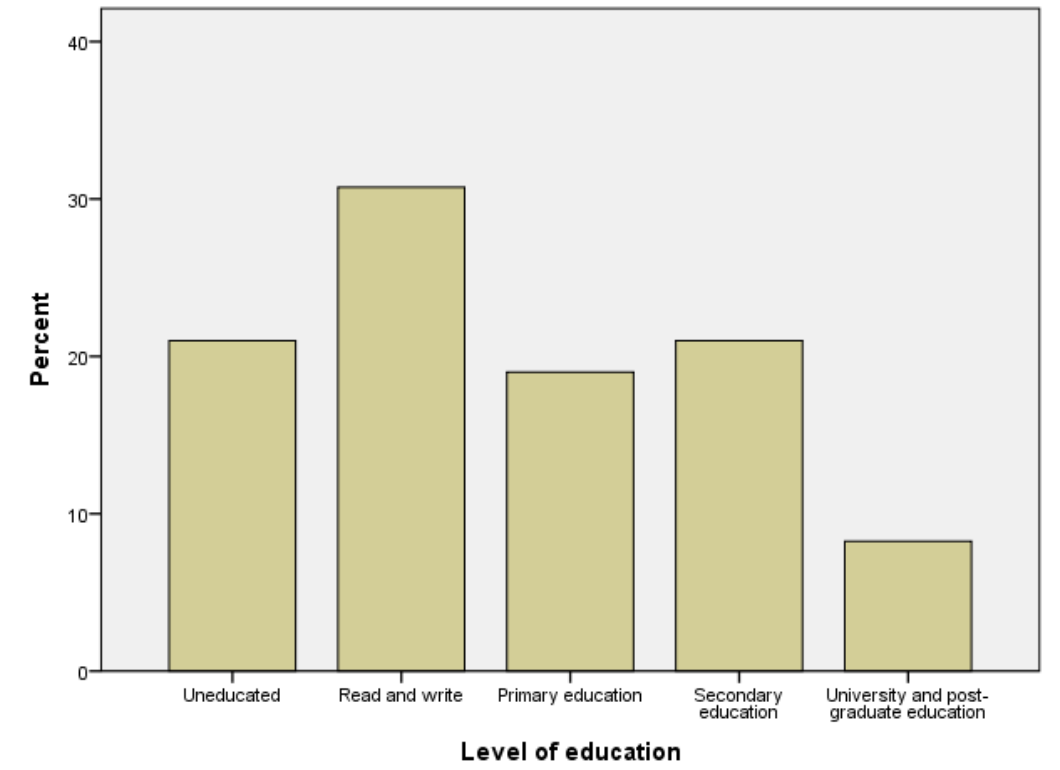
Gender



Marital status



Level of education

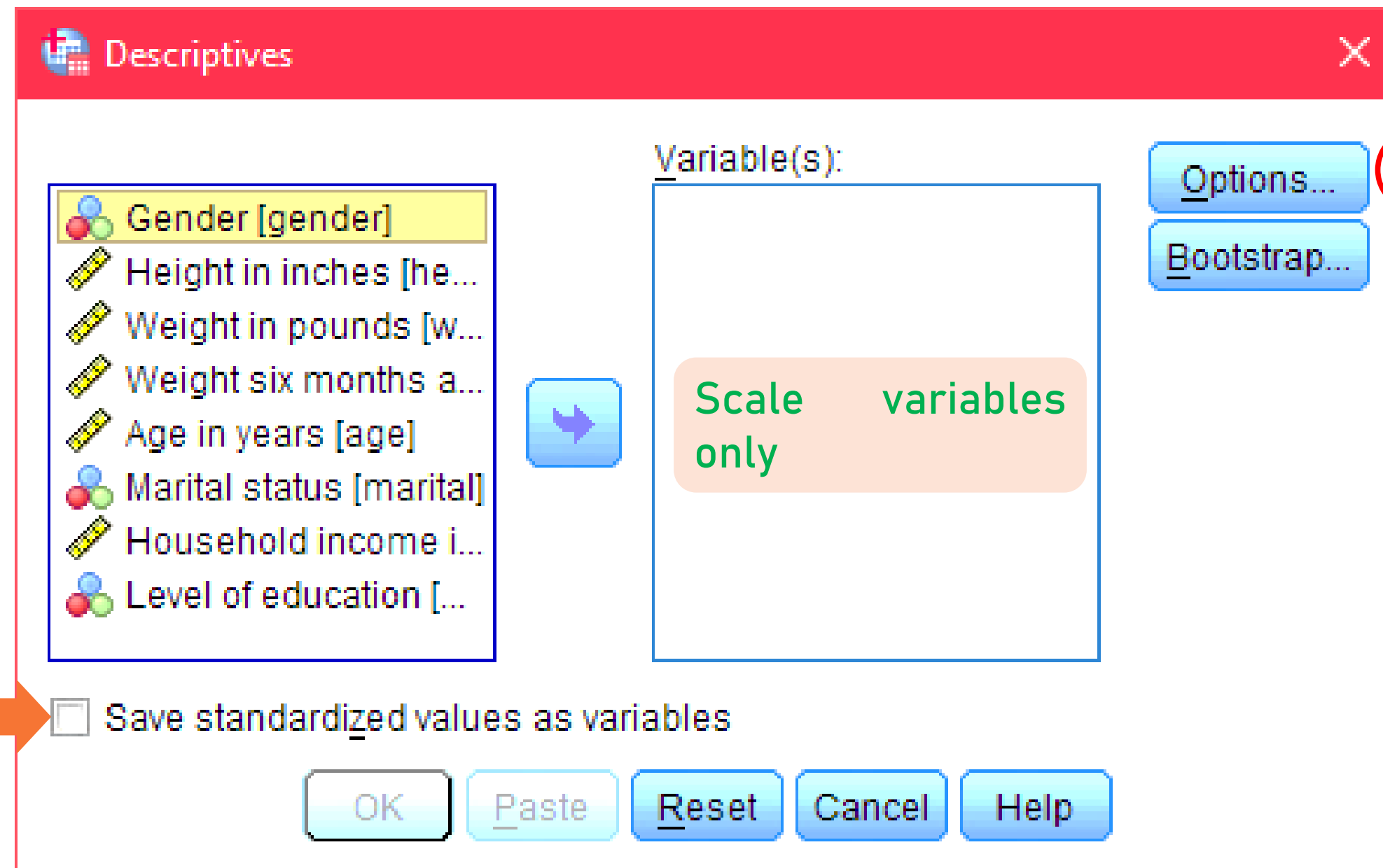




Descriptives

Descriptives

- To obtain quick summaries of scale variables, or to compare several scale variables.



To transform variables into a standard normal distribution

1
Statistics (no charts)

① Descriptives: Options

Mean Sum

Dispersion

Std. deviation Minimum

Variance Maximum

Range S.E. mean

Distribution

Kurtosis Skewness

Display Order

Variable list

Alphabetic

Ascending means

Descending means

Continue Cancel Help

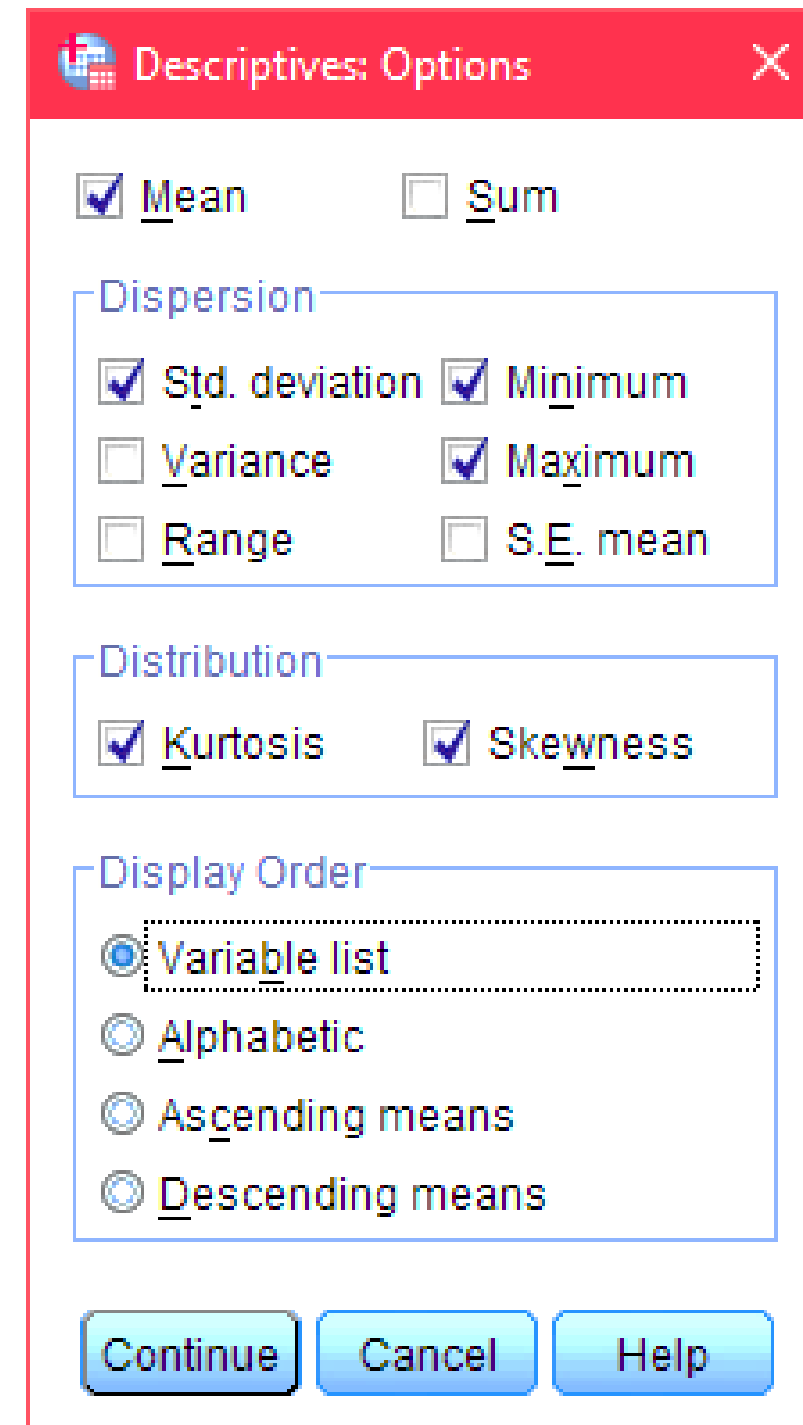
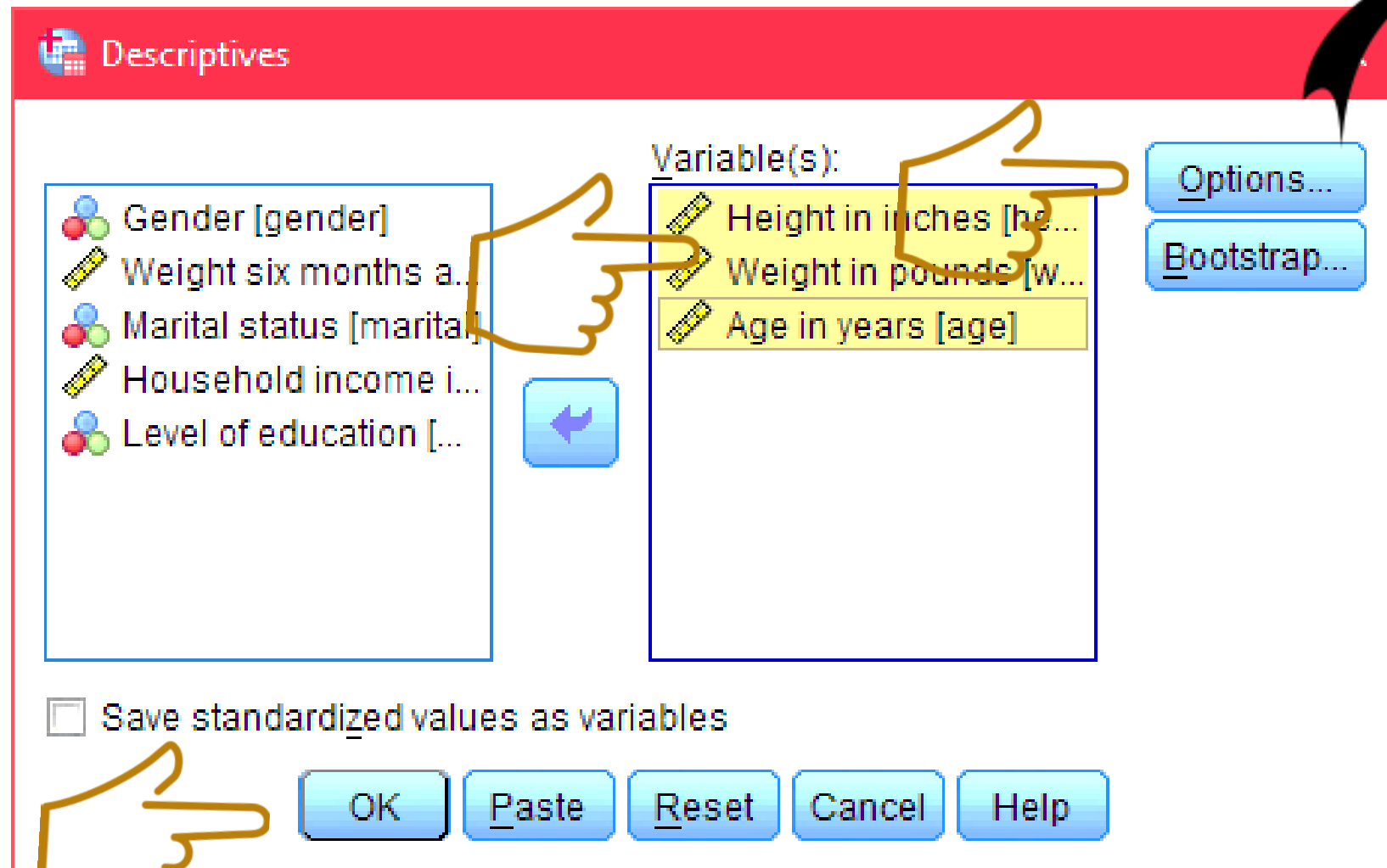
Descriptives → scale variables

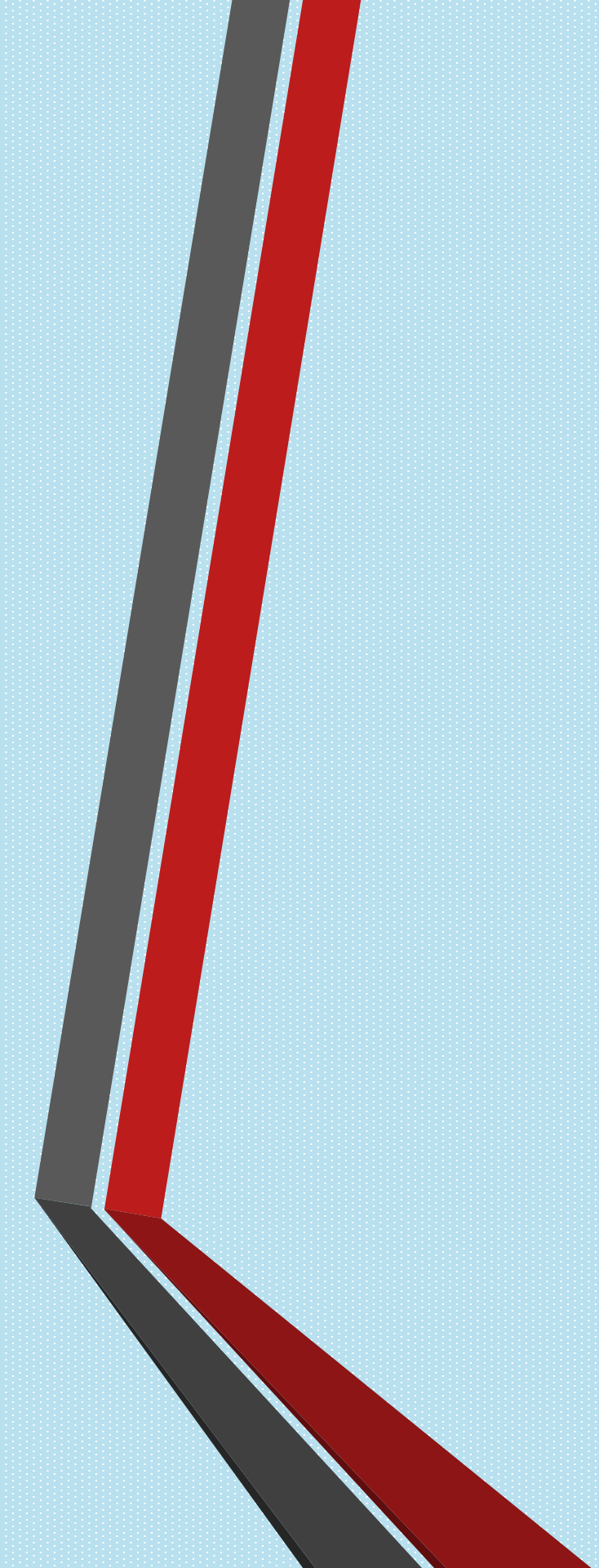
Practice

Use the **descriptives.sav** file to obtain quick summaries for the **Descriptives** of the following variables: *height*, *weight* and *age*.



- In “**Descriptives**” window, move the variables to be analyzed into “**Variable(s)**” box
- Click “**Option**” button and select the appropriate options in “**Descriptives: Options**” window → Click the **Continue** button
- Click **OK** button.





Explore

Explore

- To deeply investigate scale variables, with or without grouping variables (sample subsets).
- To produce confidence intervals, normality tests, and plots

The image shows the SPSS Explore dialog box and its Options sub-dialog. The main dialog has a list of variables on the left, including Gender, Height, Weight, Age, Marital status, Household income, and Level of education. The 'Dependent List' contains 'Scale variables' and the 'Factor List' contains 'Categorical variables'. The 'Options...' button is highlighted with a red circle and an arrow pointing to a callout box. The callout box contains the text 'Options to handle missing data'. The Options sub-dialog is open, showing the 'Missing values' section with three radio buttons: 'Exclude cases listwise', 'Exclude cases pairwise' (which is selected), and 'Report values'. The 'Continue', 'Cancel', and 'Help' buttons are visible at the bottom of the sub-dialog.

1

2

Options to handle missing data

Explore: Options

Missing values

Exclude cases listwise

Exclude cases pairwise

Report values

Continue Cancel Help

① Explore: Statistics

Explore: Statistics

Descriptives
Confidence Interval for Mean: 95 %

M-estimators
 Outliers
 Percentiles

Continue Cancel Help

② Explore: Plots

Explore: Plots

Boxplots
 Factor levels together
 Dependents together
 None

Descriptive
 Stem-and-leaf
 Histogram

Normality plots with tests

Spread vs Level with Levene Test
 None
 Power estimation
 Transformed Power: Natural log
 Untransformed

Continue Cancel Help

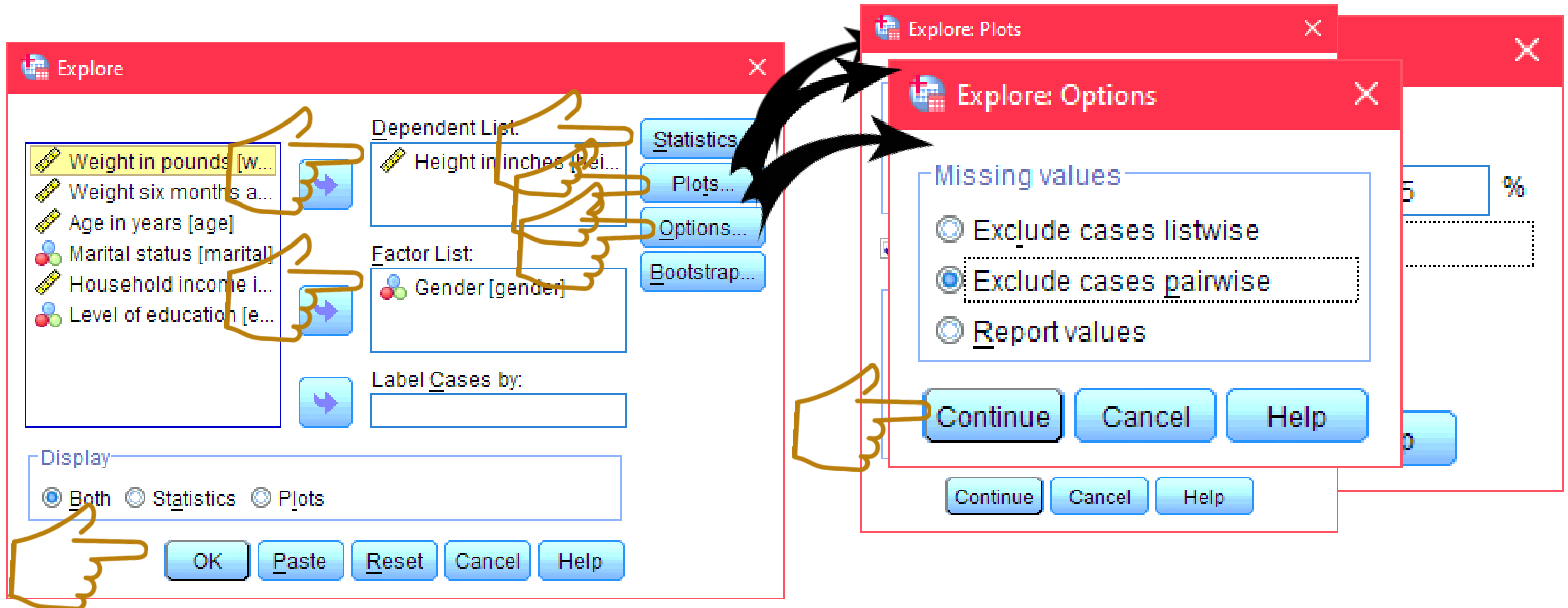
Spread vs Level with Levene Test → for testing the homogeneity of variance between the groups.

Practice

Use the **descriptives.sav** file to describe the *height* of study participants by their *gender*.



- In “**Explore**” window, move “*Height*” variable to the “**Dependent List**” box and “*Gender*” to the “**Factor List**” box.
- Click “**Statistics**” button → tick Descriptives → Click the **Continue** button
- Click “**Plots**” button → tick the options as shown below → Click the **Continue** button
- Click “**Options**” button → tick Exclude cases pairwise → Click the **Continue** button
- Click **OK** button.



Output



→ Explore

Gender

Case Processing Summary

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Height in inches	Male	207	100.0%	0	0.0%	207	100.0%
	Female	193	100.0%	0	0.0%	193	100.0%

5% Trimmed mean → The mean calculated after excluding the lowest 5% and highest 5% of the data

Descriptives

Gender			Statistic	Std. Error	
Height in inches	Male	Mean	66.7887	.22977	
		95% Confidence Interval for Mean	Lower Bound	66.3357	
			Upper Bound	67.2417	
		5% Trimmed Mean	66.8937		
		Median	67.3000		
		Variance	10.928		
		Std. Deviation	3.30580		
		Minimum	56.92		
		Maximum	72.81		
		Range	15.89		
		Interquartile Range	3.63		
		Skewness	-.548-	.169	
		Kurtosis	.070	.337	
	Female	Mean	66.6453	.29028	
		95% Confidence Interval for Mean	Lower Bound	66.0728	
			Upper Bound	67.2179	
		5% Trimmed Mean	66.8751		
		Median	67.5000		
		Variance	16.263		
		Std. Deviation	4.03268		
		Minimum	55.00		
		Maximum	72.69		
		Range	17.69		
		Interquartile Range	4.84		
		Skewness	-.863-	.175	
		Kurtosis	.188	.348	



Normality and Homogeneity

Tests of Normality

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Height in inches	Male	.091	207	.000	.970	207	.000
	Female	.112	193	.000	.931	193	.000

a. Lilliefors Significance Correction

Significant difference from normal distribution → Non-normal distribution

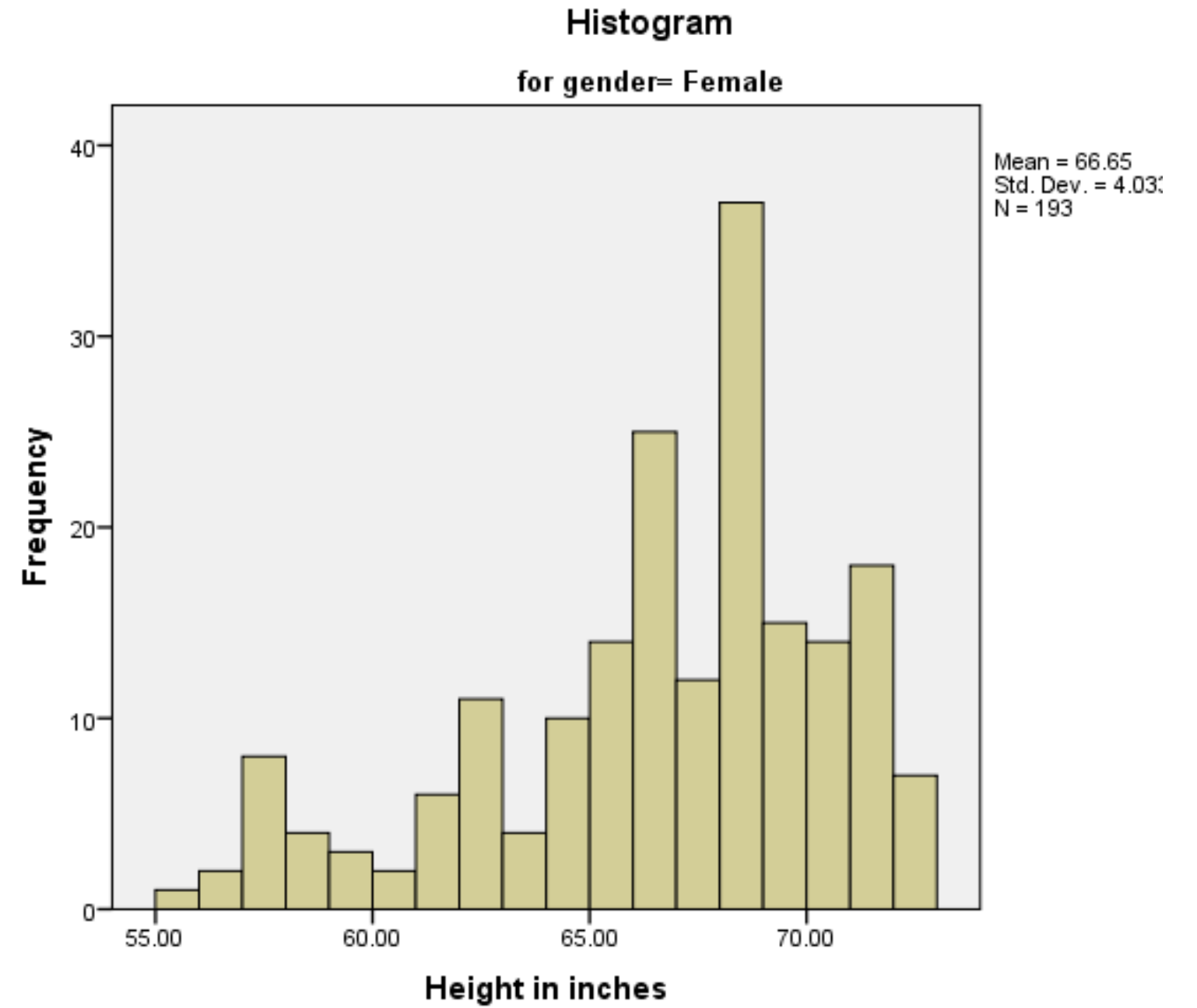
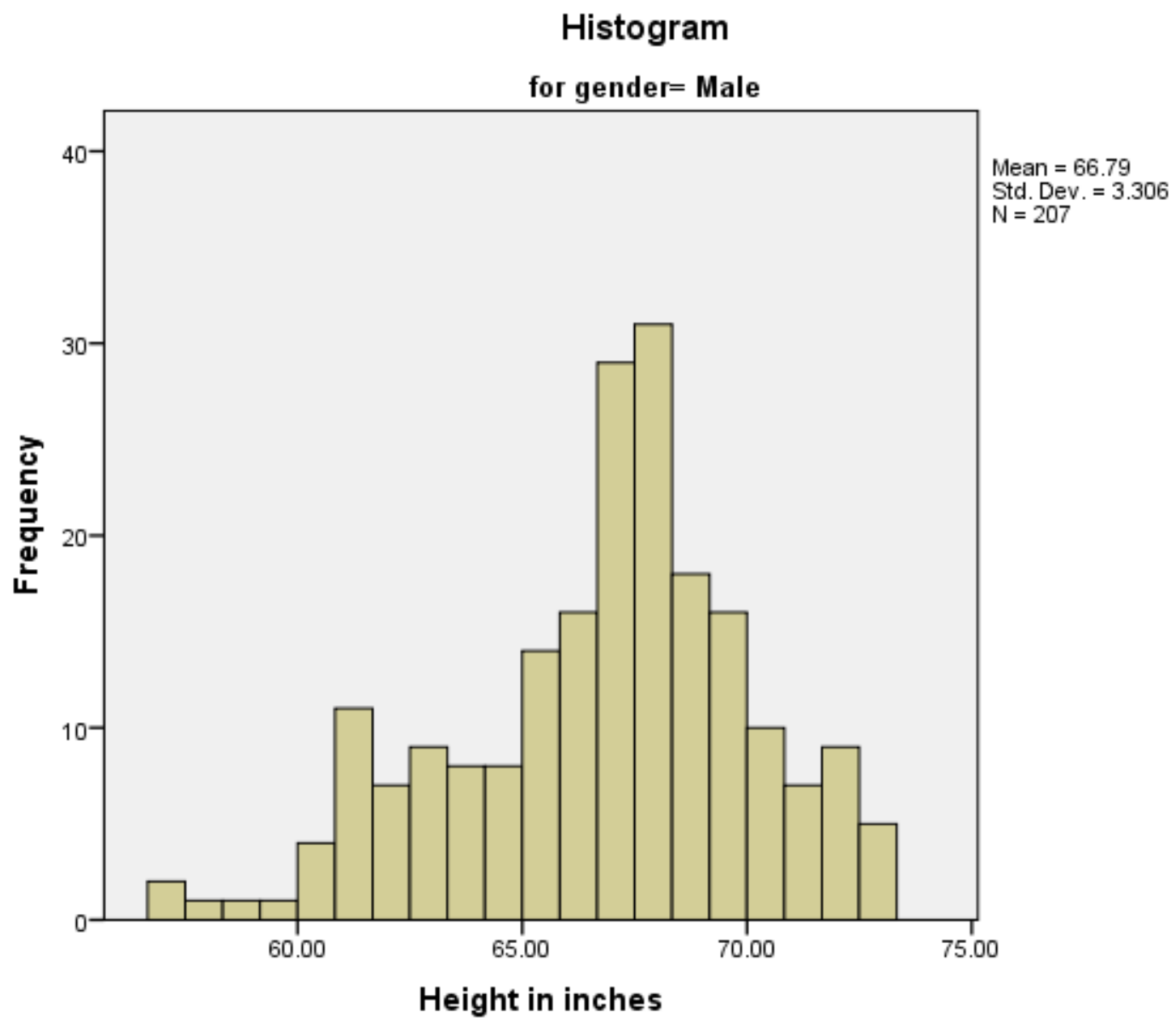
Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Height in inches	Based on Mean	6.092	1	398	.014
	Based on Median	5.011	1	398	.026
	Based on Median and with adjusted df	5.011	1	378.961	.026
	Based on trimmed mean	5.594	1	398	.019

Significant difference in variance between groups

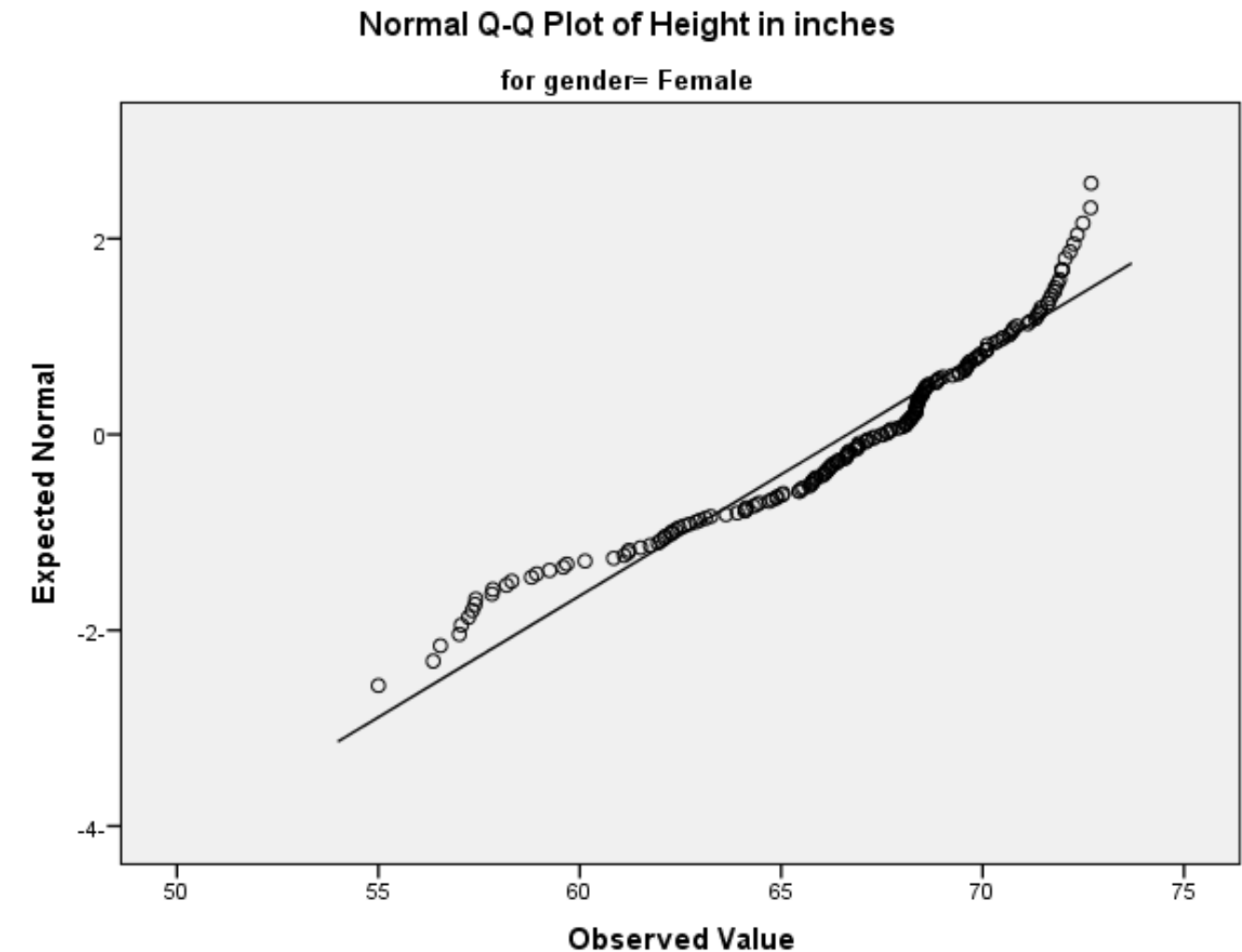
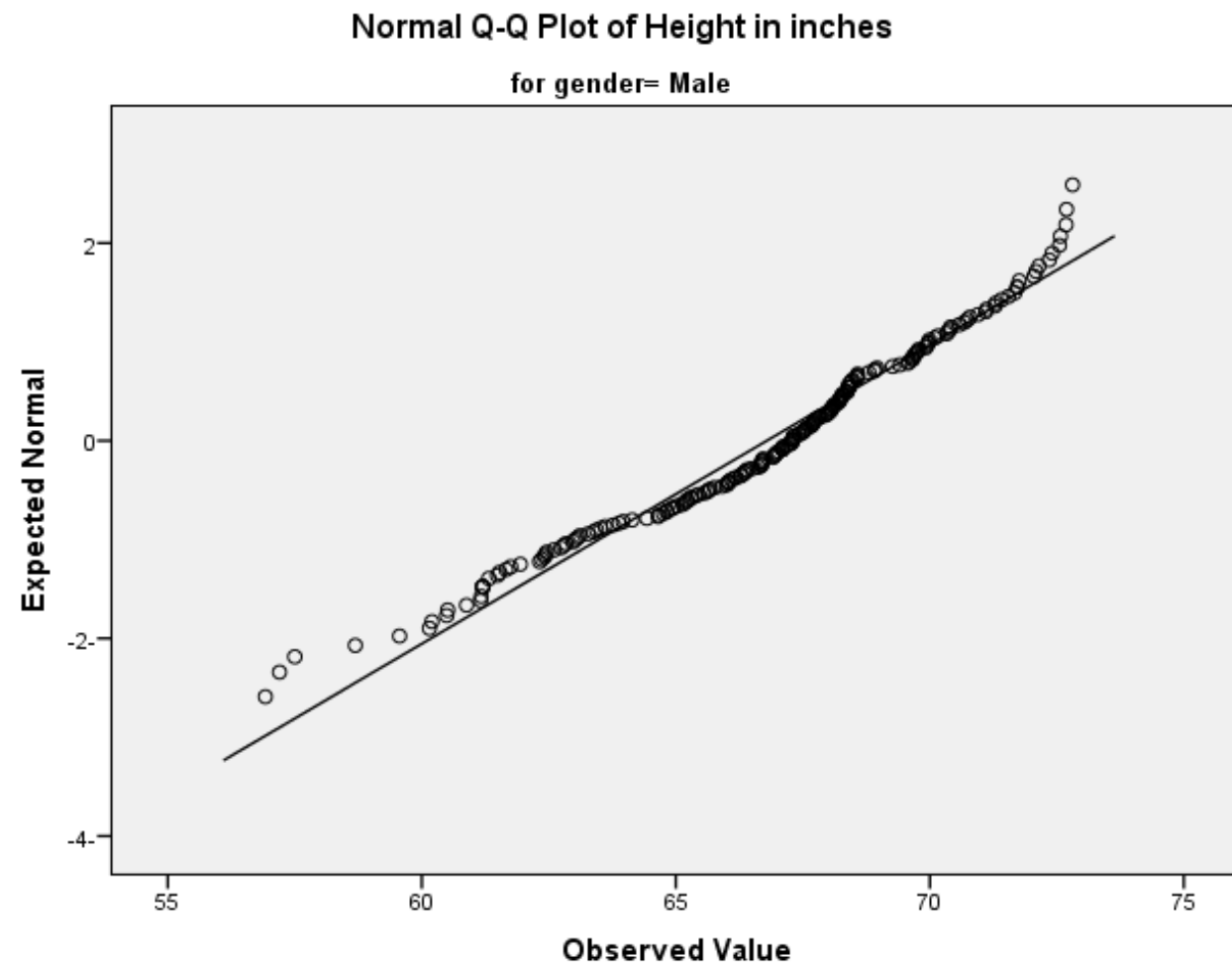
Histograms

Histograms



Normal Q-Q Plots

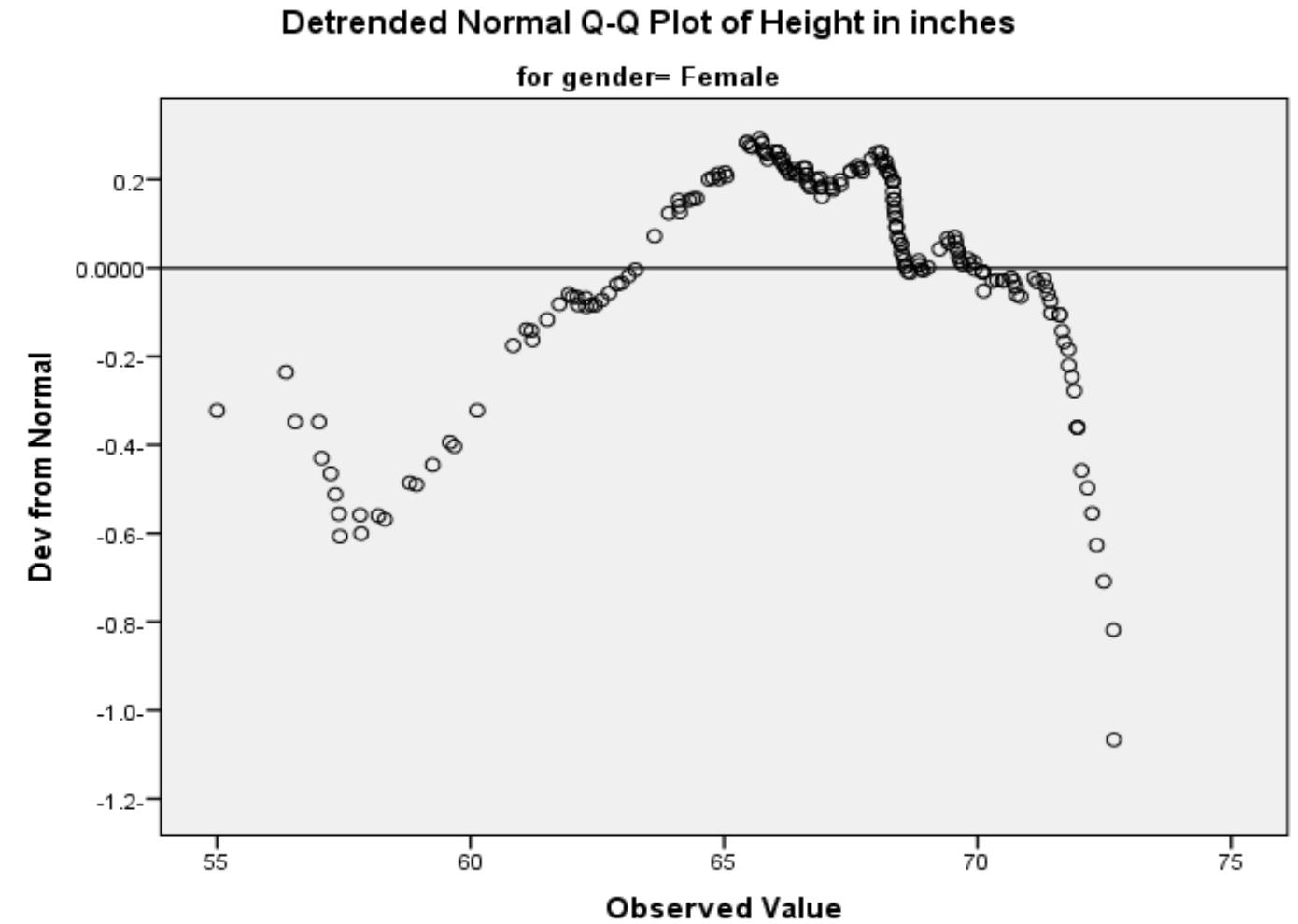
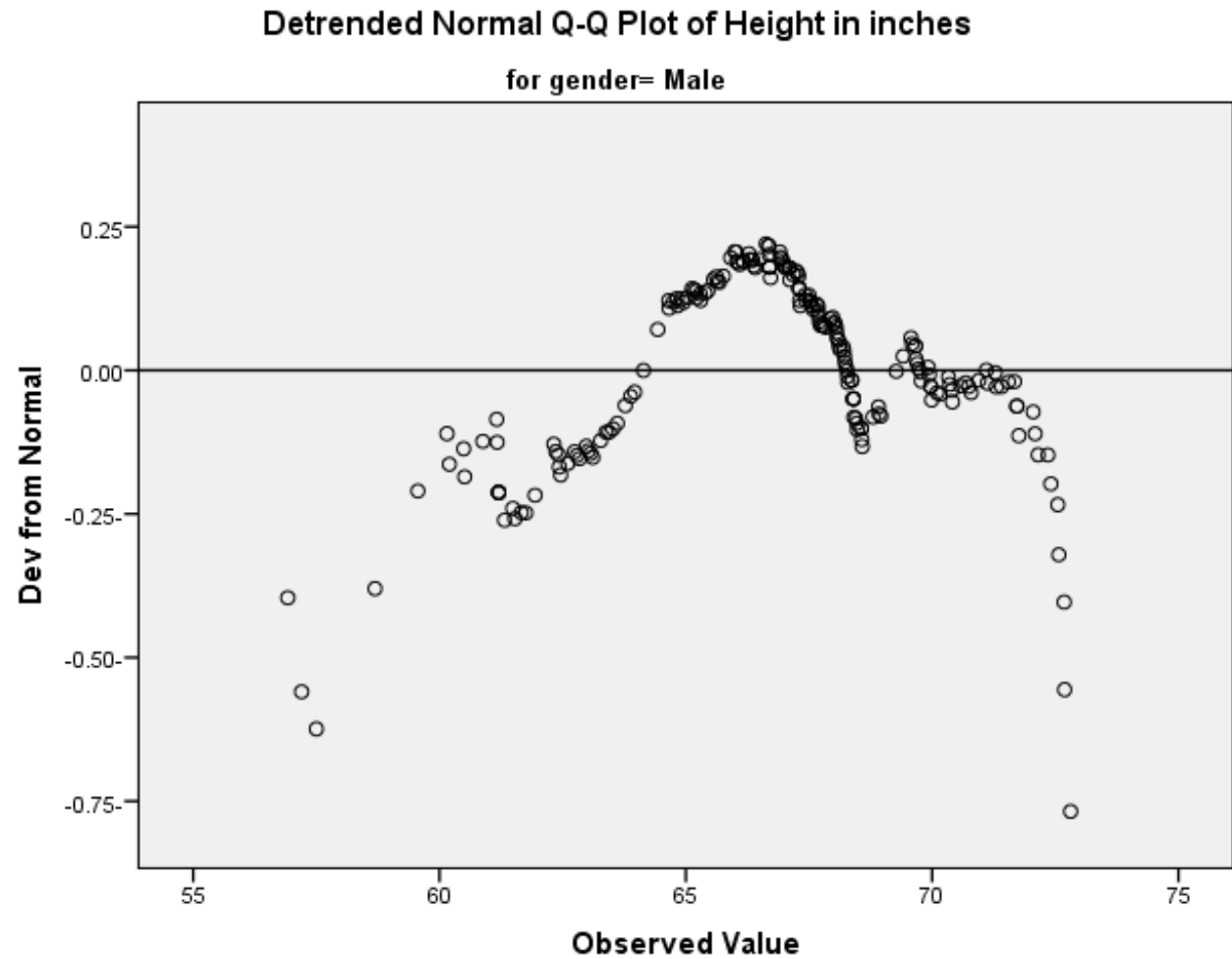
Normal Q-Q Plots



Normal Q-Q plot: a scatterplot of two sets of quantiles against one another. If both sets of quantiles came from the same distribution, we should see the points forming a diagonal line that's roughly straight.

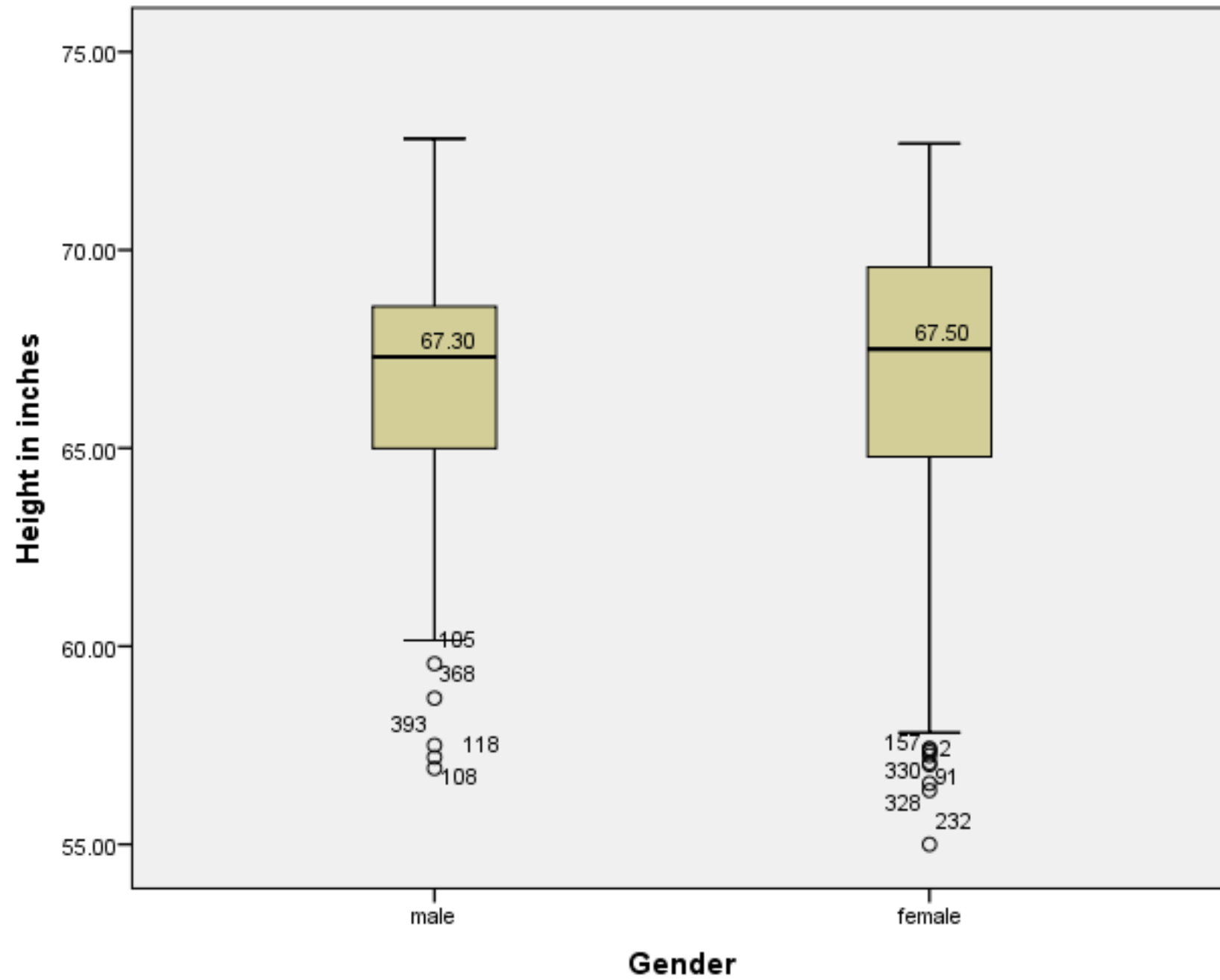
Detrended Normal Q-Q Plots

Detrended Normal Q-Q Plots



Detrended normal Q-Q plot: a scatterplot of deviations from the normal distribution around a straight line that represents the perfectly normal distribution..

Boxplots

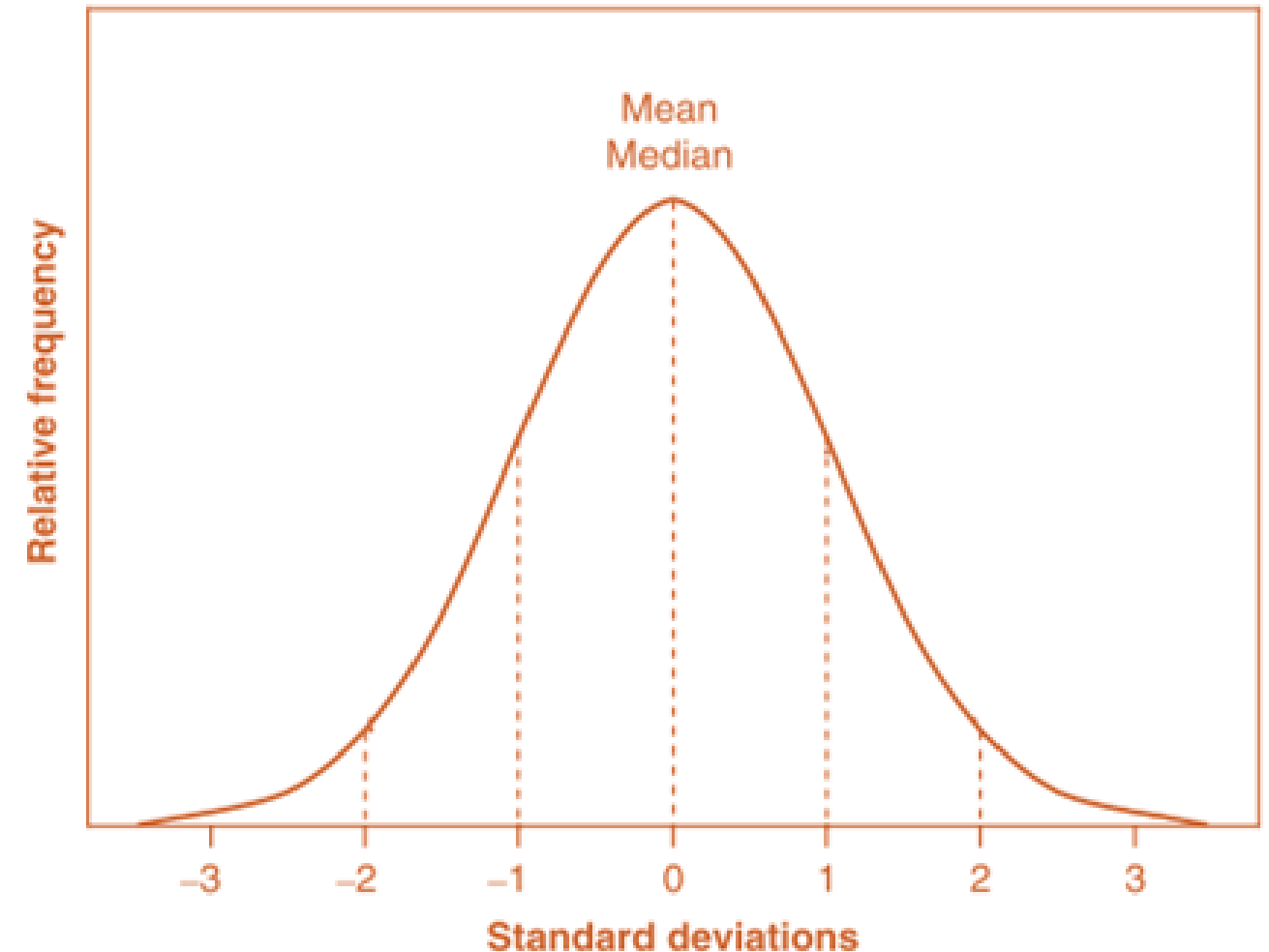




Data Distribution

Normal Distribution

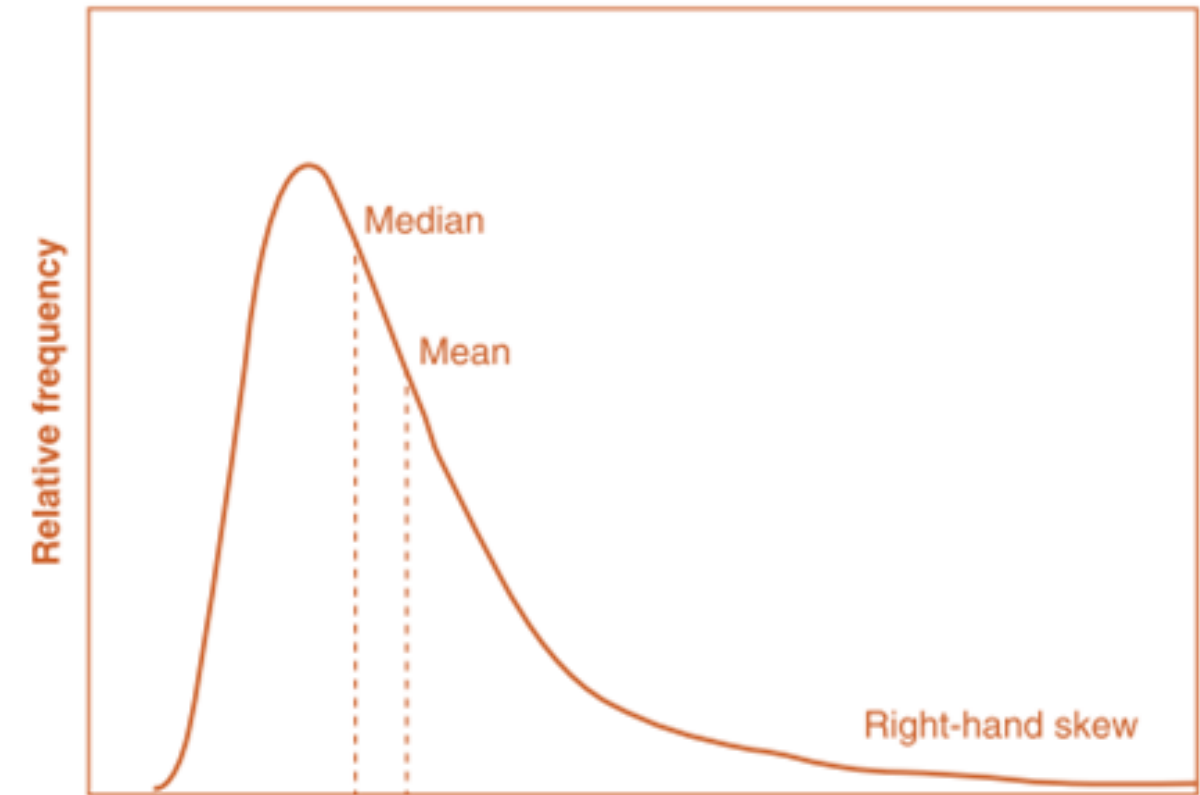
- Bilaterally symmetrical bell-shaped curve.
- Approximately equal mean and median.
- Data description: mean and SD.
- Standard normal distribution (Z distribution): mean = 0 and SD = 1.

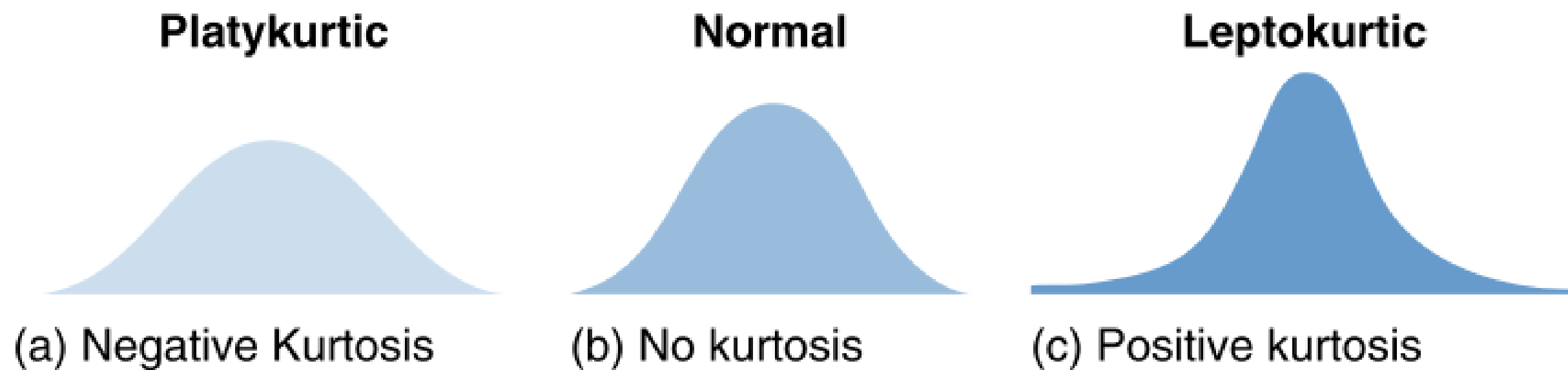
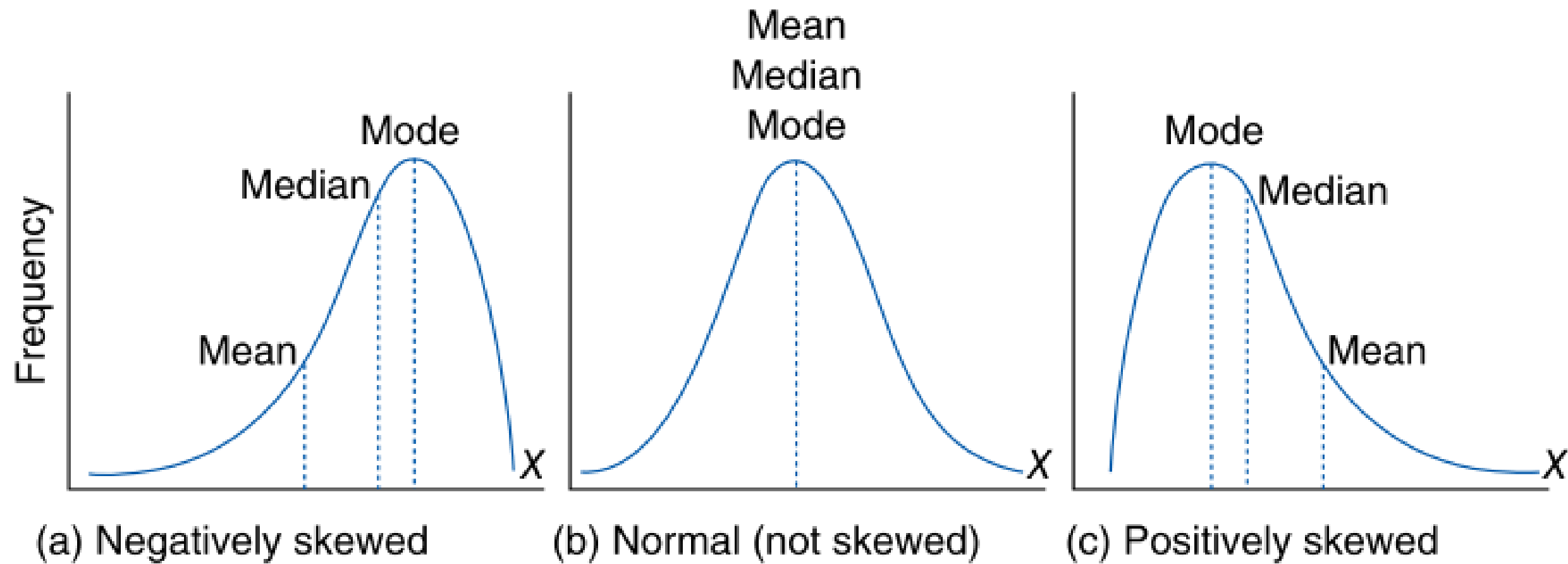



- The following assumptions about the data distribution can be made:
 - 68% of the data falls within 1 SD of the mean.
 - 95% of the data falls within 2 SD of the mean.
 - 99.7% of the data falls within 3 SD of the mean.

Skewed Distribution

- Asymmetrical curve.
- Positive skewness: $\text{mean} > \text{median}$.
- Negative skewness: $\text{mean} < \text{median}$.
- Data description: median and IQR (also used for data from very small sample sizes).







Checking for Data Distribution Normality

Checking for Data Distribution Normality

- A prerequisite for many statistical tests.

Methods

Visual inspection

- Lack objectivity.
- Through graphs and plots (histograms, Q-Q plots, P-P plots, boxplots)

Statistical tests

- Objective judgement of normality.
- Not sensitive enough at low sample sizes
- Overly sensitive to large sample sizes.



Most statisticians prefer to rely on numerical methods and statistical tests for normality

Statistical tests for normality

- In SPSS, there are two tests for the normality of data distribution.
 - ① Kolmogorov–Smirnov (K-S) test
 - ② Shapiro–Wilk (S-W) test
- For both tests, a P -value (Sig.) less than 0.05 provides evidence that the distribution is significantly different from normal → Non-parametric tests should be used for statistical analysis.

Practice

Use the **descriptives.sav** file to check for the distribution normality of *weight* according to *gender* using normality tests.



- In “**Explore**” window, move “*Weight*” variable to the “**Dependent List**” box and “*Gender*” to the “**Factor List**” box.
- Click “**Plots**” button → tick “Normality plots with tests” → Click the **Continue** button
- Click **OK** button.

The image shows two screenshots of the SPSS software interface. The left screenshot is the main 'Explore' dialog box. It features a list of variables on the left, including 'Height in inches', 'Weight six months ago', 'Age in years', 'Marital status', 'Household income', and 'Level of education'. The 'Dependent List' contains 'Weight in pounds', and the 'Factor List' contains 'Gender'. The 'Plots...' button is highlighted with a dashed border and a hand icon. Below the dialog, the 'Display' section has 'Both' selected, and the 'OK' button is also highlighted with a hand icon. The right screenshot is the 'Explore: Plots' sub-dialog box. It shows options for 'Boxplots' (Factor levels together, Dependents together, None) and 'Descriptive' (Stem-and-leaf, Histogram). The 'Normality plots with tests' checkbox is checked and highlighted with a hand icon. Below it, the 'Spread vs Level with Levene Test' section has 'None' selected. The 'Continue' button is highlighted with a hand icon. A large black arrow points from the 'Plots...' button in the first window to the 'Explore: Plots' window.

Output

Tests of Normality

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Weight in pounds	Male	.105	207	.000	.960	207	.000
	Female	.127	193	.000	.947	193	.000

a. Lilliefors Significance Correction

The P-values of both K-S and S-W tests are significant
→ Suggesting non-normal distribution (Null hypothesis is rejected)



One-Sample Kolmogorov– Smirnov Test for Normality

One-Sample K-S Test for Normality

Purpose

- A nonparametric test for normality.
- To test if a continuous variable follows a normal distribution.

Command Path

Menu

Analyze → Nonparametric Tests → One Sample

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads "decriptives.sav [DataSet1] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The Analyze menu is open, showing a list of statistical analysis options. The "Nonparametric Tests" option is highlighted in yellow, and its sub-menu is also open, with "One Sample..." selected. The data editor window shows a table with columns "gender" and "height", and rows numbered 388 to 403. The "age" column is also visible in the background data view.

age	marital	incom
35	Married	9
40	Married	4
50	Married	25
31	Married	4
39	Unmarried	10
50	Married	12
21	Married	2
31	Unmarried	1
39	Unmarried	6
26	Unmarried	2
29	Unmarried	1
		2
		1

Practice

Use the **descriptives.sav** file to check for the distribution normality of *weight* using one-sample K-S test.



- Click the **One Sample** button, and tick “Automatically compare observed data to hypothesized” in the “**Objective**” window → Click the “**Fields**” button.
- Tick the option “Use custom field assignments” in the “**Fields**” window and move “Weight” to “**Test Fields**” box → Click the “**Settings**” button.
- Tick the option “Customize settings”, tick “Kolmogorov-Smirnov test” → click “Options” and select “Normal” > “Use sample data” → Click the **OK** and **Run** buttons.

The screenshot shows the IBM SPSS Statistics Data Editor interface with the following components:

- Data Editor:** A table with columns 'gender' and 'height'. The first few rows are:

gender	height
Male	63.6
Male	69.6
Female	58.8
Female	66.8
Male	67.7
Male	57.5
Male	68.4
Male	70.4
Male	67.4
Male	65.1
Female	70.3
Female	66.1
Female	70.1
- One-Sample Nonparametric Tests Dialog:**
 - Objective:** "Automatically choose tests" is selected.
 - Fields:** "Choose Tests" is selected.
 - Settings:** "Customize tests" is selected. "Test observed data" is checked.
- Kolmogorov-Smirnov Test Options Dialog:**
 - Hypothesized Distributions:** "Normal" is selected.
 - Distribution Parameters:** "Use sample data" is selected. Mean: 0, Std.Dev.: 1.
 - Uniform:** "Use sample data" is selected. Min: 0, Max: 1.
 - Exponential:** "Sample mean" is selected. Mean: 0.
 - Poisson:** "Sample mean" is selected. Mean: 0.

Hand-drawn annotations include yellow boxes around the "One Sample..." menu item, the "Settings" tab, and the "Run" button. A black arrow points from the "Settings" tab to the "Options..." button in the Kolmogorov-Smirnov dialog.

Output

→ Nonparametric Tests

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Weight in pounds is normal with mean 155.415 and standard deviation 15.91.	One-Sample Kolmogorov-Smirnov Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Non-normal distribution of weight in pounds



Standardizing Data

(Z-scores)

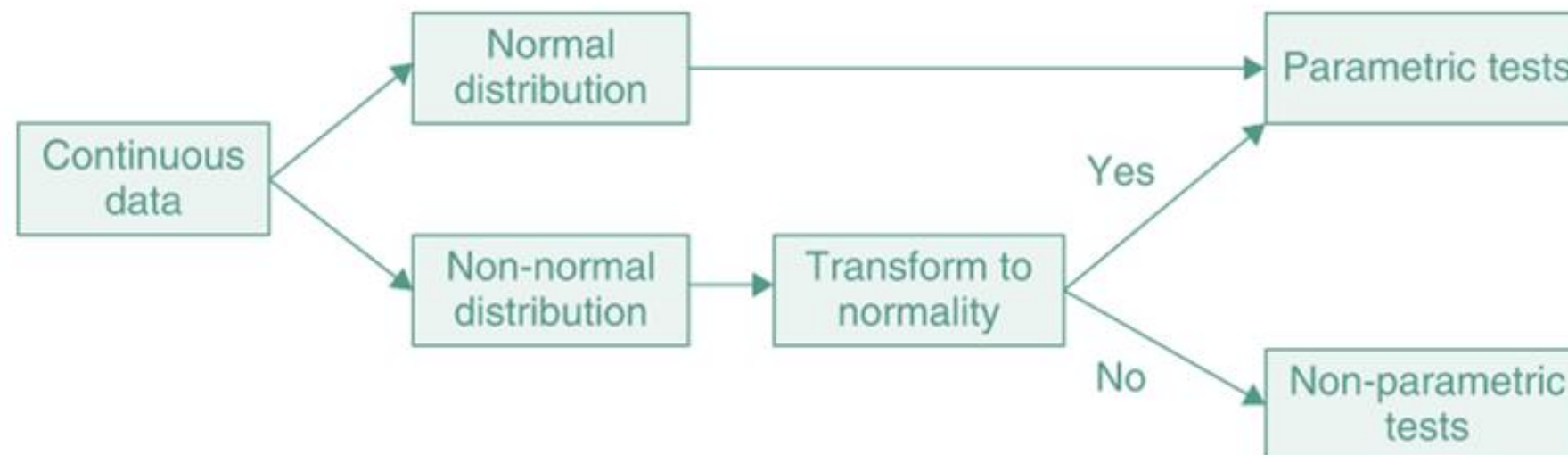
Standardizing data

- **“Descriptives”** can be used to standardize values → To convert values to Z-scores (Z-values or standard normal scores)

Z-score: SD units of each value away from the mean → Positive Z-score (above the average) and negative Z-score (below the average)

Purpose

- To convert normal distribution to standard normal distribution, and to convert non-normal to normal distribution (if possible).
- To allow for comparing different scores from different normal distributions (irrespective of the measurement units).



Command Path

Menu

Analyze → Descriptive Statistics → Descriptives

decriptives.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform **Analyze** Direct Marketing Graphs Utilities Add-ons W

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale

Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...

406 :	gender	height
388	Male	63.6
389	Male	69.6
390	Female	58.8
391	Female	66.8
392	Male	67.7
393	Male	57.5
394	Male	68.4
395	Male	70.4
396	Male	67.4
397	Male	65.1
398	Female	70.3

Descriptives

Variable(s):

- Gender [gender]
- Height in inches [he...]
- Weight in pounds [w...]
- Weight six months a...
- Age in years [age]
- Marital status [marital]
- Household income i...
- Level of education [...]

Save standardized values as variables

Options...
Bootstrap...

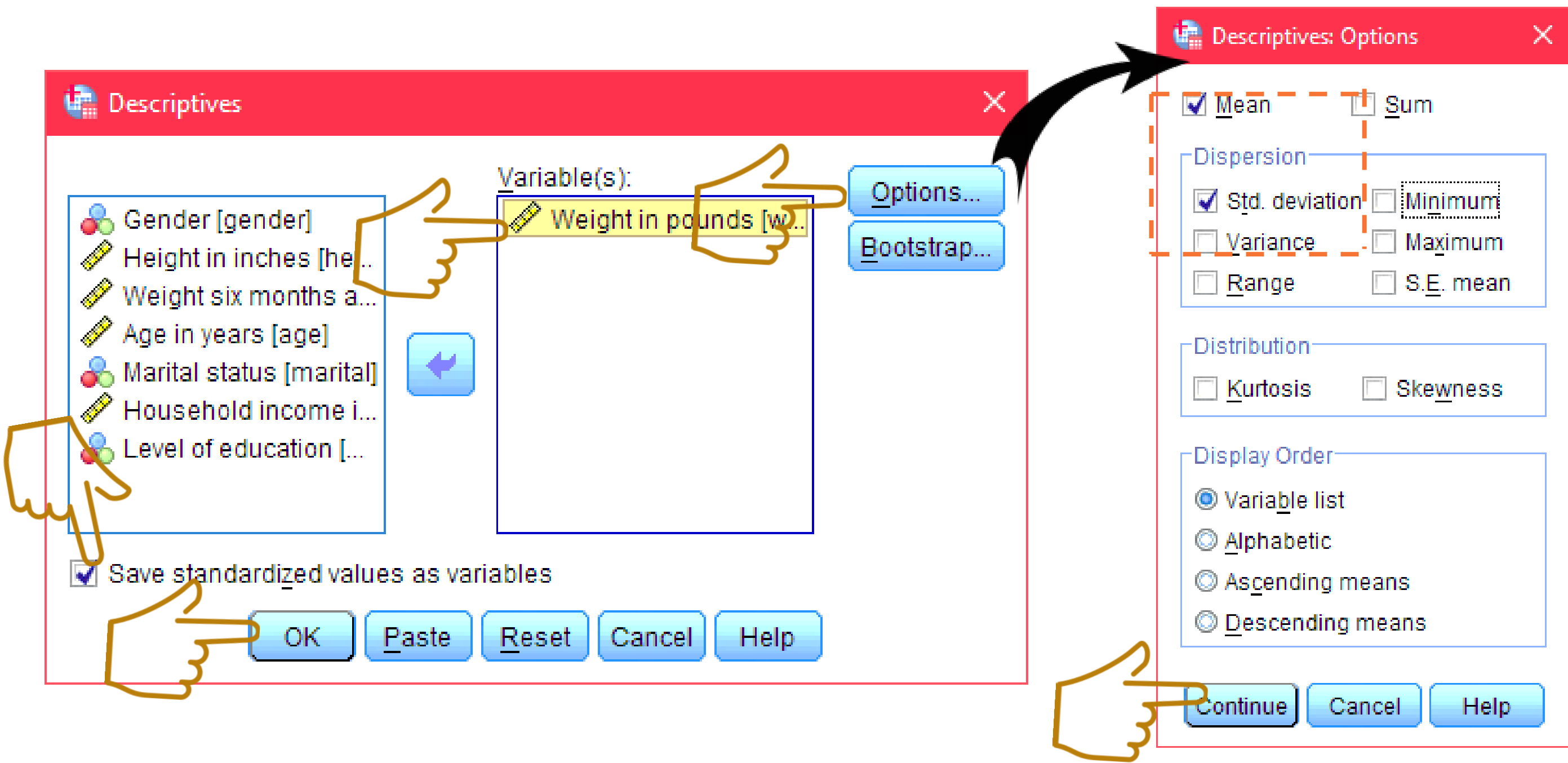
OK Paste Reset Cancel Help

Practice

Use the **descriptives.sav** file to standardize the *Weight* variable.

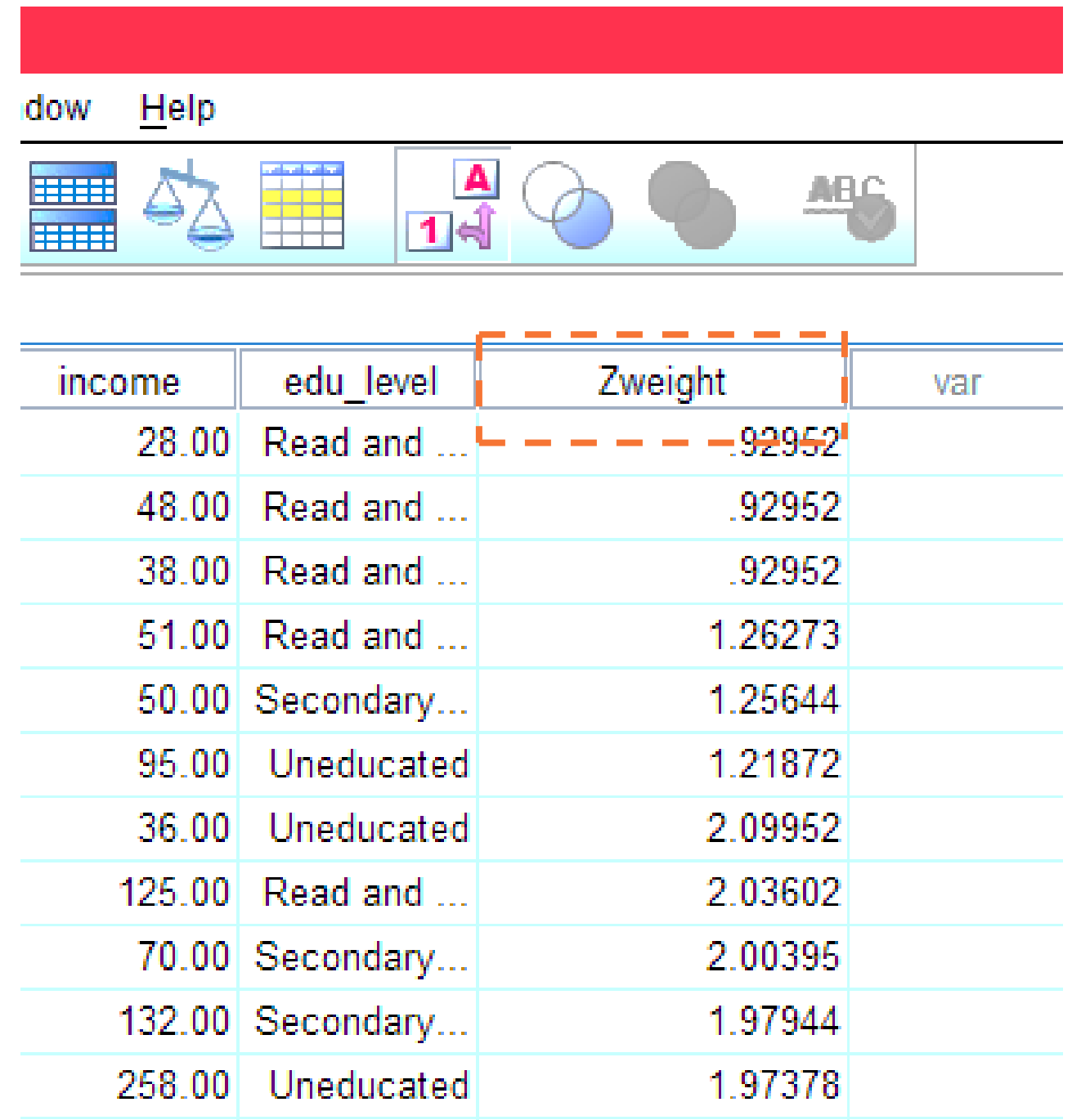


- In “**Descriptives**” window, move “*Weight*” variable to the “**Variable(s)**” box and tick the option “Save standardized values as variables” → Click “**Options**” button.
- Tick “Mean” and “Standard deviation” options (*default*) in the “**Descriptives: Options**” window → Click the **Continue** button
- Click **OK** button.



Output

- A new variable called “Zweight” is added to the dataset → Standard normal distribution.



income	edu_level	Zweight	var
28.00	Read and92952	
48.00	Read and92952	
38.00	Read and92952	
51.00	Read and ...	1.26273	
50.00	Secondary...	1.25644	
95.00	Uneducated	1.21872	
36.00	Uneducated	2.09952	
125.00	Read and ...	2.03602	
70.00	Secondary...	2.00395	
132.00	Secondary...	1.97944	
258.00	Uneducated	1.97378	



**Inferential Statistics:
Choosing Appropriate
Statistical Tests in IBM
SPSS Statistics**

Statistical tests

Parametric


- Used to test statistically significant differences between groups when the study sample has a **normal distribution** regarding the continuous variable of interest.

Nonparametric

- Used to test statistically significant differences between groups when scale data is **non-normally distributed** (distribution-free tests), data of interest are **frequency counts**, or when the **dependent variable is not scaled**.
- Also called “rank tests” because they depend on the ranks of data (ordinal level of measurement).




Parametric tests are more powerful than nonparametric tests.



One-Sample Hypothesis Testing

➤ One-Sample Statistical Tests

- ① One-sample t -test
 - ② Wilcoxon signed-rank test (one sample)
 - ③ Binomial test
 - ④ Chi-square goodness-of-fit test
- 
- A decorative graphic consisting of several parallel white lines of varying lengths, slanted upwards from left to right, located in the bottom right corner of the slide.



One-Sample t -Test

One-Sample (Student's) t-Test

Purpose

To test whether a sample mean is significantly different from a hypothesized mean value in a population (e.g. a national average or a mean from a literature review).

Assumptions

- ① The variable is measured at the scale level.
- ② The data of the variable are normally distributed.
- ③ Existence of a hypothesized mean for comparison.



The t test is robust and can handle violations of the assumption of a normal distribution

Command Path



Menu

Analyze → Compare Means → One Sample T test

The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu path is highlighted: **Analyze** > **Compare Means** > **One-Sample T Test...**

The data editor shows a table with the following data:

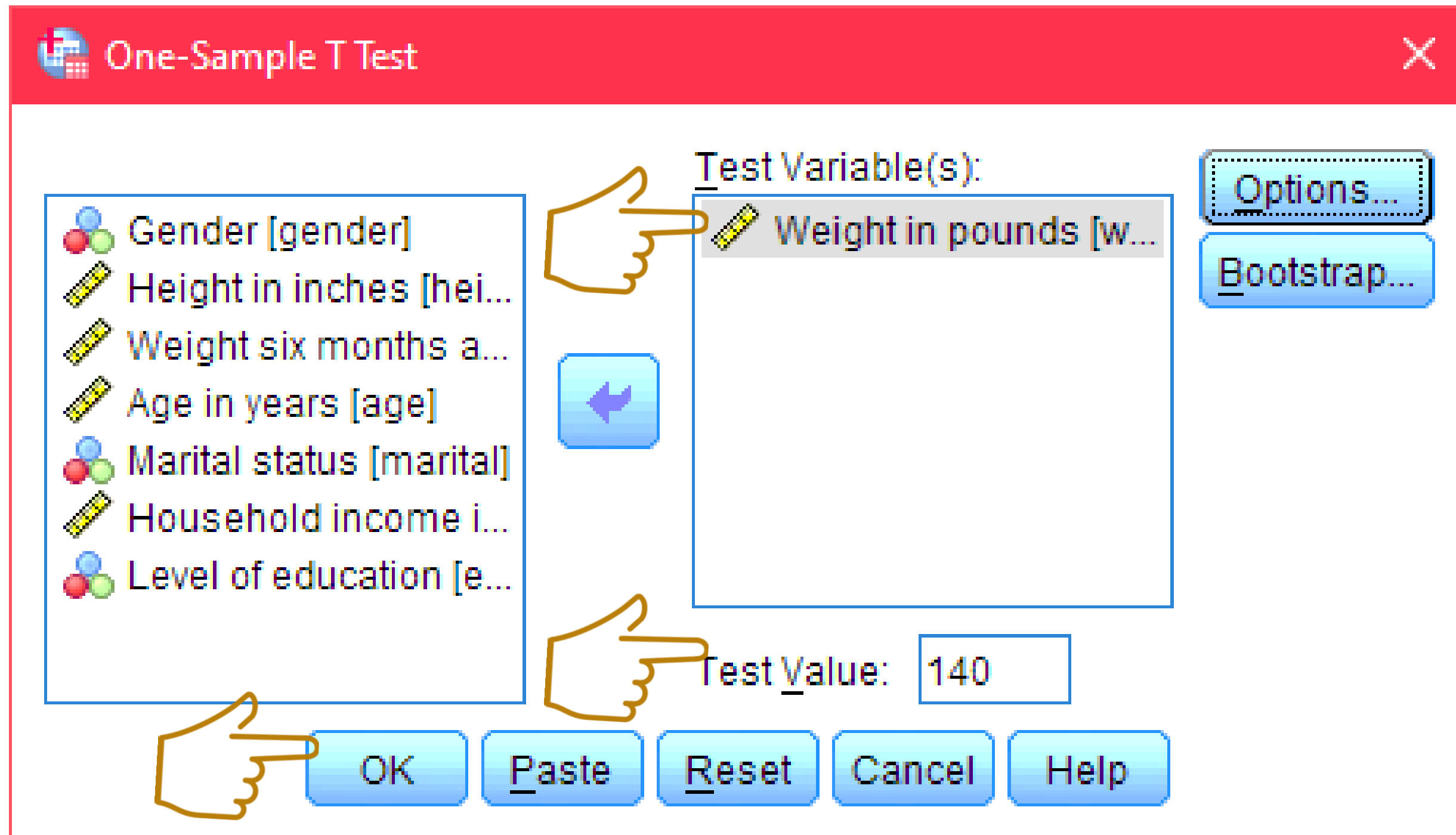
height		
64.9	55	Unmarried
57.0	29	Married
67.3	51	Married
64.4		
58.3		
66.9		
68.6		
68.4		
67.1	45	Unmarried

Practice

Use the **descriptives.sav** file to compare the mean *weight* of the study population (155.42 pounds) with that published in the literature (suppose it to be 140.00 pounds)



- In “**One-Sample T test**” dialog box, move “*Weight*” variable to the “**Test Variable(s)**” panel
- Set the “**Test Value**” at “140” → Click **OK** button.



Output

→ T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Weight in pounds	400	155.4151	15.90601	.79530

One-Sample Test

	Test Value = 140					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Weight in pounds	19.383	399	.000	15.41508	13.8516	16.9786

$P < 0.05 \rightarrow$
statistically significant
difference

The 95% CI does not
include zero \rightarrow
statistically significant
difference

Reporting

The mean weight of the study population was significantly higher than the expected mean of 140.00 pounds reported in literature ($P < 0.001$).





Wilcoxon Signed-Rank Test

(one sample)

One-Sample Wilcoxon Signed-Rank Test

Purpose

To test whether a sample median is significantly different from a hypothesized median value in a population (e.g. a national average or a median from a literature review).

The alternative nonparametric test that may be used when the assumptions required of the one-sample t-test are not met.

It uses the ranks of values of the scale variable measured at the ordinal level.

Assumptions

- ① The variable is measured at the scale level.
- ② The data of the variable are not normally distributed.
- ③ Existence of a hypothesized median for comparison.

Command Path



Menu

Analyze



Nonparametric Tests



One Sample

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads "decriptives.sav [DataSet1] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The Analyze menu is open, showing options like Reports, Descriptive Statistics, Tables, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, Nonparametric Tests, Forecasting, Survival, Multiple Response, and Missing Value Analysis... The Nonparametric Tests menu item is highlighted, and its sub-menu is open, showing One Sample..., Independent Samples..., and Related Samples... The One Sample... option is also highlighted. In the background, a data table is visible with columns for gender and height, and rows numbered 1 through 16.

	gender	height
1	Female	64.9
2	Female	57.0
3	Male	67.3
4	Male	64.4
5	Female	58.3
6	Female	66.9
7	Female	68.6
8	Male	68.4
9	Female	67.1
10	Male	68.2
11	Female	70.1
12	Male	62.3
13	Female	59.5
14	Male	65.5
15	Male	63.9
16	Male	68.2

Practice

Use the **descriptives.sav** file to compare the median *weight* of the study population (153.15 pounds) with that published in the literature (suppose it to be 140.00 pounds)



- Click the **One Sample** button, and tick “Automatically compare observed data to hypothesized” in the “**Objective**” window → Click the “**Fields**” button.
- Tick the option “Use custom field assignments” in the “**Fields**” window and move “Weight” to “**Test Fields**” box → Click the “**Settings**” button.
- Tick the option “Customize settings”, tick “Compare median to hypothesized (Wilcoxon signed-rank test)” → Set the value of “Hypothesized median” to “140.00” → Click **Run** button.

The screenshot shows the IBM SPSS Statistics Data Editor with the 'One-Sample Nonparametric Tests' dialog box open. The dialog is in the 'Settings' tab, and the 'Compare median to hypothesized (Wilcoxon signed-rank test)' option is selected. The 'Hypothesized median' is set to 140. Hand-drawn arrows indicate the steps: selecting 'One Sample...' from the 'Analyze' menu, clicking the 'Settings' tab, checking the 'Compare median to hypothesized' option, setting the 'Hypothesized median' to 140, and finally clicking the 'Run' button.

age	marital	incom
35	Married	9
40	Married	4
50	Married	25
31	Married	4
39	Unmarried	10
50	Married	12
21	Married	2
31	Unmarried	1
39	Unmarried	8
26	Unmarried	2
29	Unmarried	1

Output

$P < 0.05 \rightarrow$
statistically significant
difference

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Weight in pounds equals 140.000.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Reporting

The median weight of the study population was significantly higher than the median weight of 140.00 pounds reported in the literature ($P < 0.001$).





Binomial Test

Binomial Test

Purpose

● To determine whether the sample proportion of one category in a binary (dichotomous) variable equals a hypothesized proportion value.

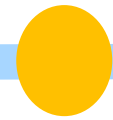
Principally used with categorical data to test the equality of proportions → Whether the proportion of cases in one of only two possible categories is equal to a pre-specified proportion

✦ Also called “**one-sample proportion test**”.

Assumptions

- ① The variable is dichotomous.
- ② The observations are independent.
- ③ Existence of a hypothesized proportion for comparison.

Command Path



Menu

Analyze → Nonparametric Tests → One Sample

The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu path is: **Analyze** > **Nonparametric Tests** > **One Sample...**

The data table contains the following information:

	gender	height
1	Female	64.9
2	Female	57.0
3	Male	67.3
4	Male	64.4
5	Female	58.3
6	Female	66.9
7	Female	68.6
8	Male	68.4
9	Female	67.1
10	Male	68.2
11	Female	70.1
12	Male	62.3
13	Female	59.5
14	Male	65.5
15	Male	63.9
16	Male	68.2

The 'Analyze' menu is open, showing options like Reports, Descriptive Statistics, Tables, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, **Nonparametric Tests** (highlighted), Forecasting, Survival, Multiple Response, and Missing Value Analysis... The 'Nonparametric Tests' submenu is also open, showing **One Sample...** (highlighted), Independent Samples..., Related Samples..., and Legacy Dialogs.

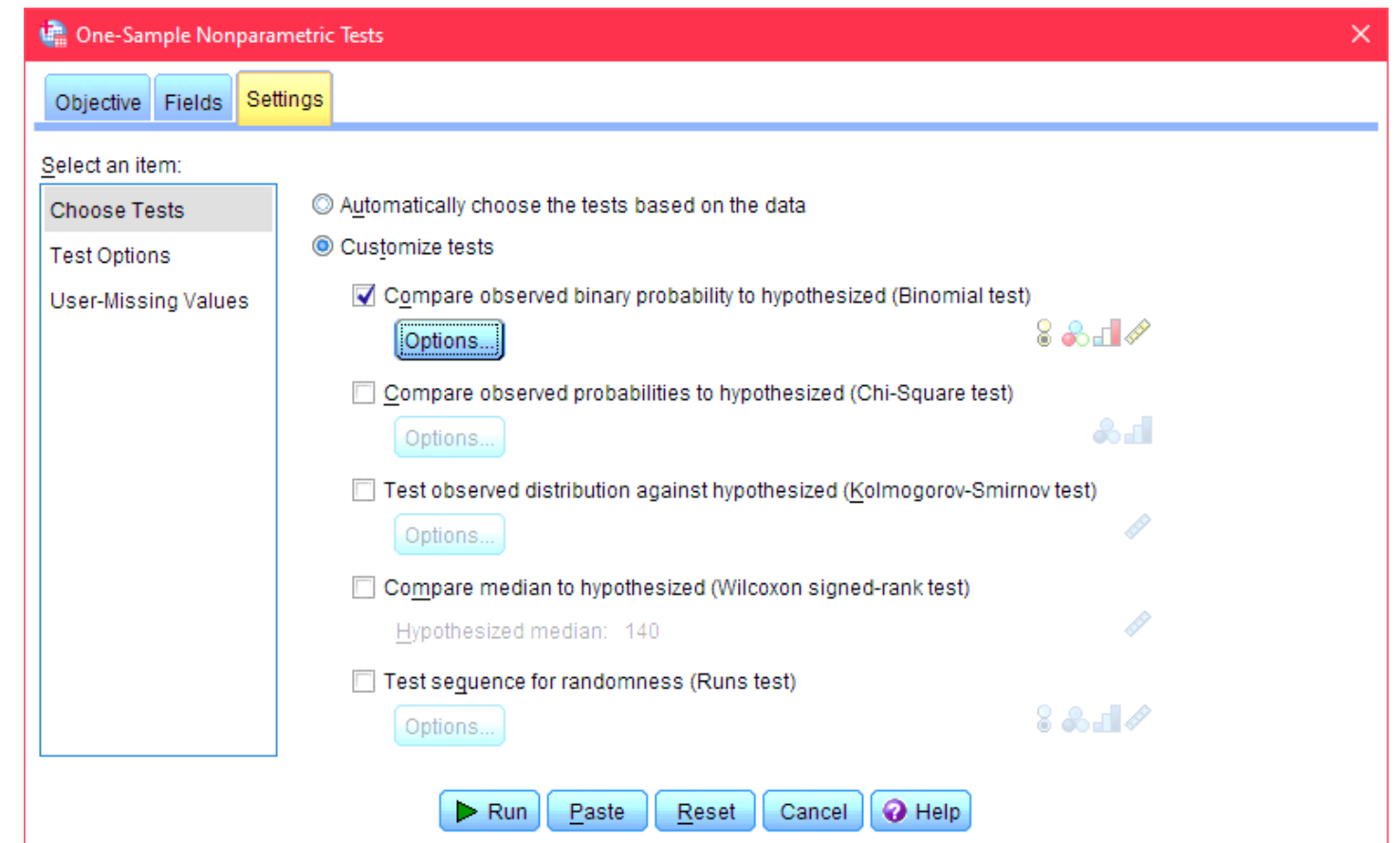
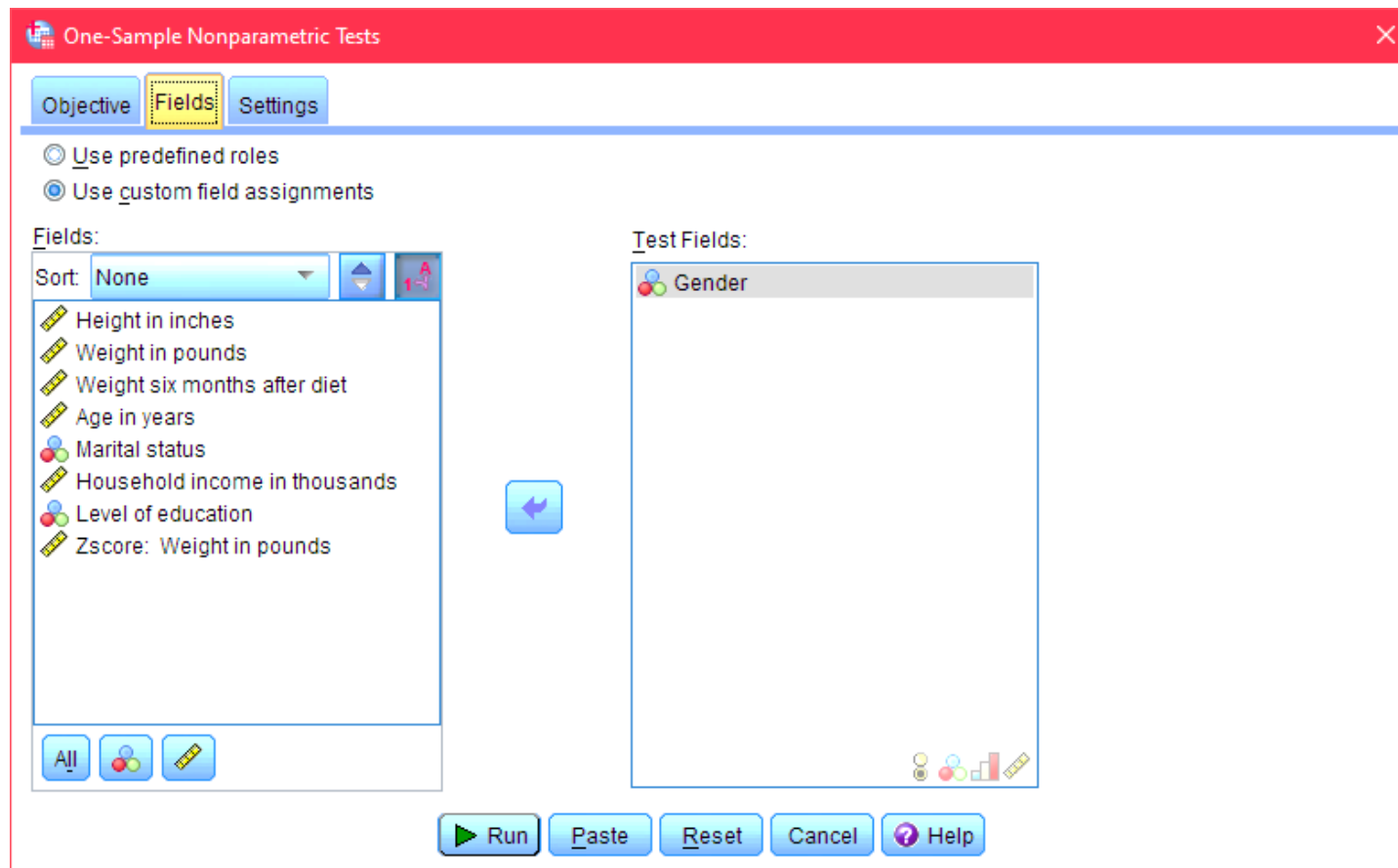
Practice

Use the **descriptives.sav** file to to compare the proportion of males (51.8%) with that published in the literature (suppose it to be 75.0%)

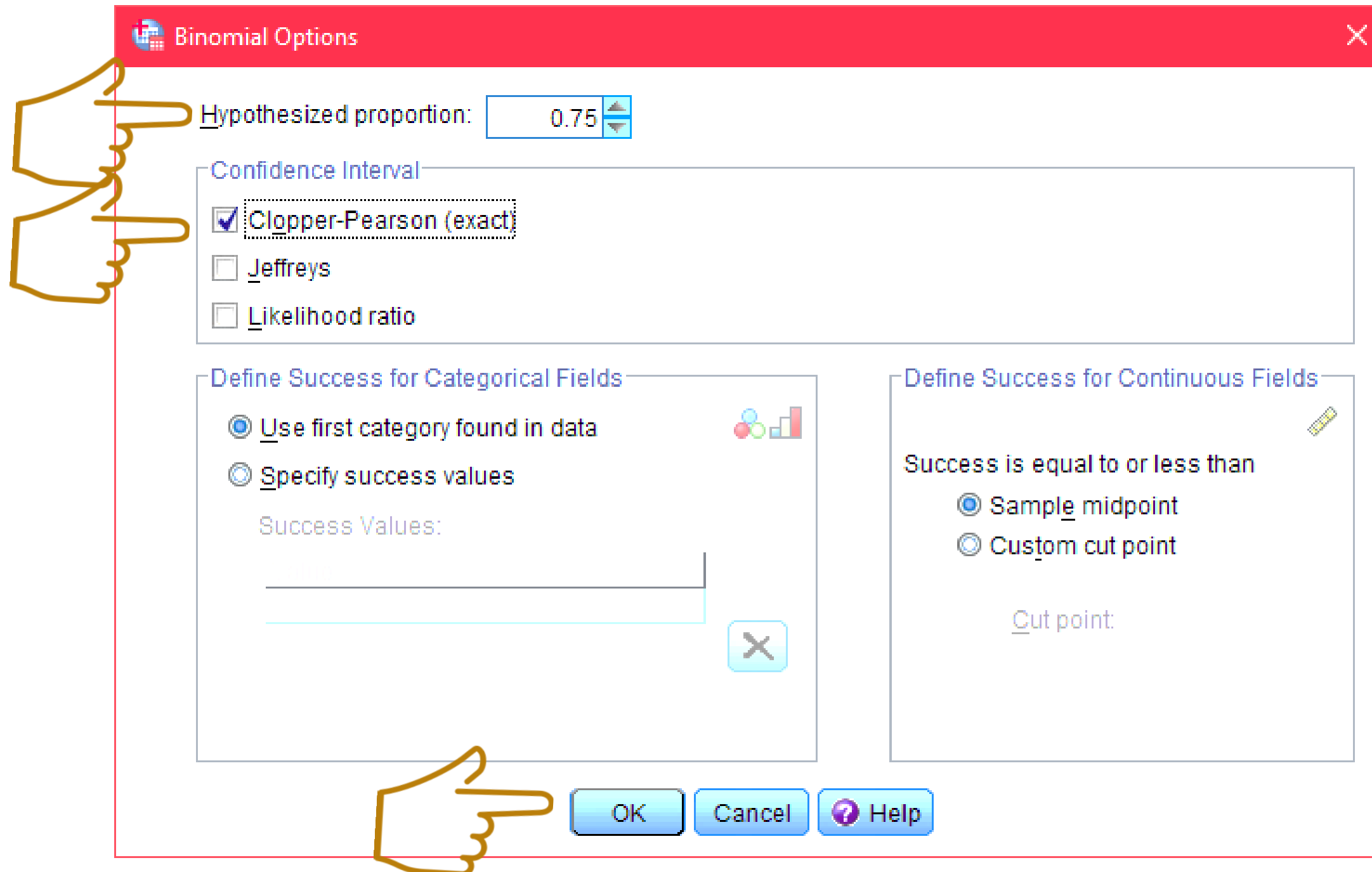


The test compares the first category of the binary variable → One should check the coding of the categories.

- It is performed using the same steps of one-sample Wilcoxon signed-rank test, with the exception that “gender” is moved to the “**Test Fields**” panel and “Compare observed binary probability to hypothesized (Binomial Test)” is chosen in the “**Settings**” window



- Click “**Options**” button, set the “Hypothesized proportion” at “0.75” and tick the option “Clopper-Pearson (exact)” under **Confidence Interval**.
- Click “**OK**”



Output

→ Nonparametric Tests

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories defined by Gender = Female and Male occur with probabilities 0.75 and 0.25.	One-Sample Binomial Test	.000	Reject the null hypothesis.

$P < 0.05 \rightarrow$
statistically significant
difference

Asymptotic significances are displayed. The significance level is .05.

Output

- By double-clicking on the “Hypothesis Test Summary”, we can obtain the 95% CI.

Confidence Interval Summary

Confidence Interval Type	Parameter	Estimate	95% Confidence Interval	
			Lower	Upper
One-Sample Binomial Success Rate (Clopper-Pearson)	Probability (Gender=Male)	.518	.467	.567

Reporting

The proportion of males in the study population (51.8%, 95% CI: 46.7-56.7) was significantly lower than the proportion of 75.0% reported in the literature ($P < 0.001$).



Another Option!

Statistics Data Editor

Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples

age	marital	income	edu_level	Zweight
36	Unmarried	38.00	Read and9
37	Unmarried	51.00	Read and ...	1.2
51	Married	125.00	Read and ...	2.0
42	Married	132.00	Secondary...	1.9
40	Married	26.00	Primary ed...	1.9
58	Unmarried	31.00	Uneducated	1.9
37	Married	62.00	Uneducated	1.8
26	Unmarried	22.00	Uneducated	1.8
54	Unmarried	31.00	Primary ed...	1.8
39	Unmarried	43.00	Read and ...	1.7
39	Married	19.00	Uneducated	1.7
66	Uneducated	66.00	Uneducated	1.6
36	Primary ed...	36.00	Primary ed...	1.5
20	Uneducated	20.00	Uneducated	1.5
41	Unmarried			
49	Married			
23	Unmarried			

Legacy Dialogs
Chi-square...
Binomial...
Runs...

Binomial Test

Test Variable List: Gender [gender]

Height in inches [height]
Weight in pounds [wei...
Weight six months afte...
Age in years [age]
Marital status [marital]
Household income in t...
Level of education [ed...
Zscore: Weight in pou...

Define Dichotomy: Get from data Cut point

Test Proportion: 0.75

Exact...
Options...

OK Paste Reset Cancel Help

Output

→ NPar Tests

Binomial Test

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
Gender	Group 1	Male	207	.52	.000 ^a
	Group 2	Female	193	.48	
	Total	400	1.00	.75	

$P < 0.05 \rightarrow$
statistically significant
difference

a. Alternative hypothesis states that the proportion of cases in the first group $< .75$.




Chi-Square Goodness-of-Fit Test

Chi-Square Goodness-of-Fit Test

Purpose

A nonparametric test used to test whether frequency distribution (proportion) of one categorical variable differs from an expected (theoretical) distribution.



The closer the expected values are to the observed values, the smaller the value of chi-square → Larger chi-square values means larger differences in the observed *vs.* expected values, which may then lead to rejection of the null hypothesis.

✦ Also called **one-sample chi-square test**, or **multinomial test**.

Assumptions

① One categorical variable (nominal or ordinal)

② Independent observations

③ Relatively large sample size

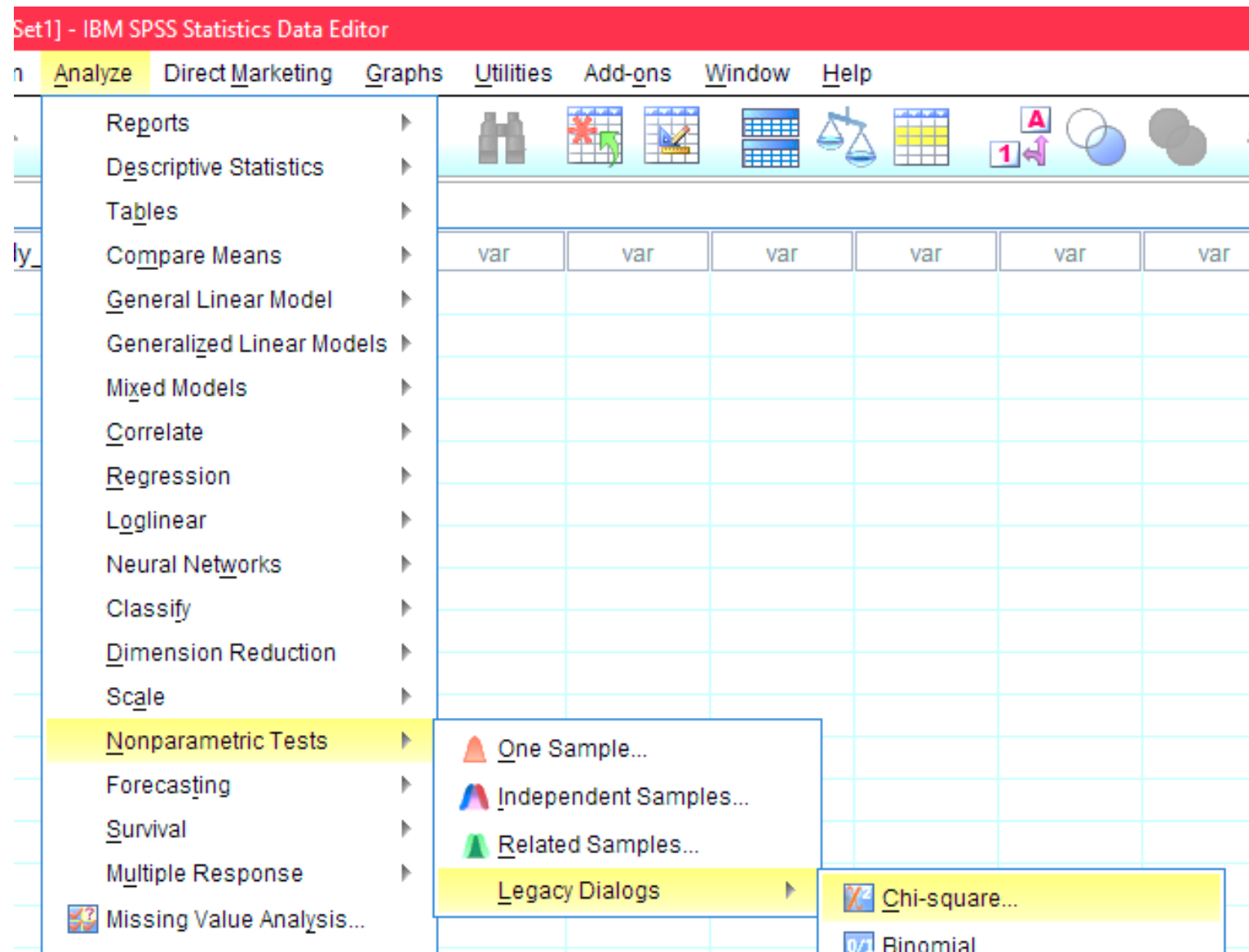
- The expected frequency for each category is at least 1
- The expected frequency is at least 5 for 80% or more of the categories

Command Path



Menu

Analyze → Nonparametric Tests → Legacy Dialogs → Chi-square



Practice

Use the dataset **chi-square_goodness-of-fit.sav** to test if the flavors of candies (orange, lemon and strawberry) are preferred by children equally.



For a **dichotomous variable**, it is better to use a **binomial test** because it gives the exact instead of the approximate significance level.

- In the “**Chi-square Test**” window, move the categorical variable “Preferred flavor of candies” into the “**Test Variable List**” panel → Tick “**All categories equal**” option
- Click “**Options**” button → Tick “**Descriptive**” option → Click “**Continue**” and “**OK**” buttons.

The image displays two screenshots from the SPSS software interface, illustrating the steps to perform a Chi-square test. The first screenshot shows the "Chi-square Test" dialog box. The variable "Preferred flavour of ca" is moved from the left list to the "Test Variable List" panel. The "Expected Values" section has the radio button for "All categories equal" selected. The "Options..." button is highlighted. The second screenshot shows the "Chi-square Test: Options" dialog box. The "Descriptive" checkbox is checked, and the "Continue" button is highlighted. Hand-drawn arrows indicate the flow of actions between the two windows.

Chi-square Test

Test Variable List:

id Preferred flavour of ca

Exact... Options...

Expected Range:

Get from data
 Use specified range

Lower: Upper:

Expected Values:

All categories equal
 Values:

Add Change Remove

OK Paste Reset Cancel Help

Chi-square Test: Options

Statistics:

Descriptive Quartiles

Missing Values:

Exclude cases test-by-test
 Exclude cases listwise

Continue Cancel Help

Output

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Preferred flavour of candies	90	1.79	.786	1	3

No. of cases

Not needed

Categories

Chi-Square Test

Frequencies

Differences between observed and expected frequencies.

Preferred flavour of candies			
	Observed N	Expected N	Residual
Orange	39	30.0	9.0
Lemon	31	30.0	1.0
Strawberry	20	30.0	-10.0
Total	90		

Differs the most from expected

Consistent with the assumption for chi-square test

No. of expected observations are equal as per the hypothesis (no difference)

Test Statistics

	Preferred flavour of candies
Chi-Square	6.067 ^a
df	2
Asymp. Sig.	.048

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 30.0.

$P < 0.05 \rightarrow$
Statistically significant difference from random

Candy flavor were distributed unequally \rightarrow There is a preference for a particular candy flavor, which is "Orange"

Reporting

A chi-square goodness-of-fit test showed that there were statistically significant differences in the preference of the type of candy flavor among children ($\chi^2 = 6.067$, $P = 0.048$), with most children preferring the orange-flavoured candies.



Another Option!

*Chi-square goodness-of-fit.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Hel

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival

	id	candy_
25	25	
26	26	
27	27	
28	28	
29	29	
30	30	
31	31	
32	32	
33	33	
34	43	
35	35	
36	36	
37	37	
38	38	

var var var

One Sample...
Independent Samples...

One-Sample Nonparametric Tests

Objective Fields Settings

Select an item:
Choose Tests
Test Options
User-Missing values

Automatically choose the tests based on the data
 Customize tests

Compare observed binary probability to hypothesized (Binomial test)
Options...

Compare observed probabilities to hypothesized (Chi-Square test)
Options...

Test observed distribution against hypothesized (Kolmogorov-Smirnov test)
Options...

Compare median to hypothesized (Wilcoxon signed-rank test)
Hypothesized median:

Test sequence for randomness (Runs test)
Options...

Run Paste Reset Cancel Help

Output

*Significant difference in
the preference of candy
flavours*

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of Preferred flavour of candies occur with equal probabilities.	One-Sample Chi-Square Test	.048	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.




Remember!

The two data-processing methods for conducting chi-square goodness-of-fit test in SPSS




Nonparametric Tests →
Legacy Dialogs

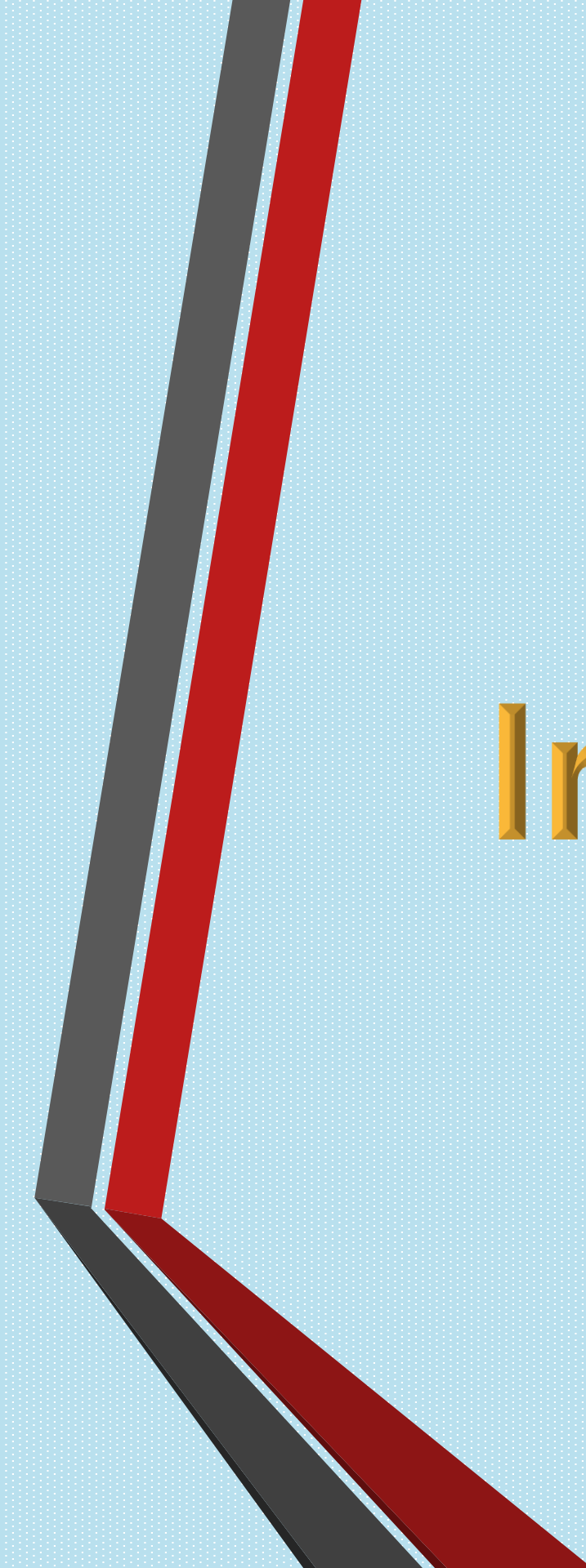
Nonparametric Tests → One
Sample



Hypothesis Testing of Two Samples - **Independent Samples**

➤ Statistical Tests for Two Independent Samples

- ① Independent-samples t -test
 - ② Mann-Whitney U test
 - ③ Chi-square test of independence
- 




Independent-Samples t-Test

Independent-Samples t-Test

Purpose

To compare the means of **two** independent samples.



Significant difference between the sample means → inference that the unknown means of the population are also different.

Also called **Unpaired t-test**

Assumptions

- ① Independent samples (each case must be in one group only).
- ② Independent measurements (only one measurement for each case)
- ③ Dependent variable is scale and normally distributed, while the grouping variable is dichotomous.
- ④ Homogeneity of variances between the two groups.