

Trellis Plots



Revised: 2/1/2022



Summary.....	1
Sample Data.....	2
Trellis Plots for Numeric Y	3
Trellis Plots for Categorical Y	9
Trellis Plots for Y Versus X	14
Trellis Plots for Z Versus X and Y	19

Summary

Trellis Plots are segmented plots that display data for each combination of one or more conditioning variables. For example, histograms of the distribution of height among individuals might be displayed side-by-side for men and women. The plots are designed to help users visualize how data change across levels of the conditioning variables.

Statgraphics provides 4 types of trellis plots:

1. *Numeric Y*: displays characteristics of a single quantitative variable at different combinations of 1 or 2 conditioning variables using either a histogram, box and whisker plot or normal probability plot.
2. *Categorical Y*: displays characteristics of a single categorical variable at different combinations of 1 or 2 conditioning variables using either a barchart, piechart or donut chart.
3. *Y vs X*: displays the relationship between 2 variables at different combinations of 1 or 2 conditioning variables using either a scatterplot or a regression curve.

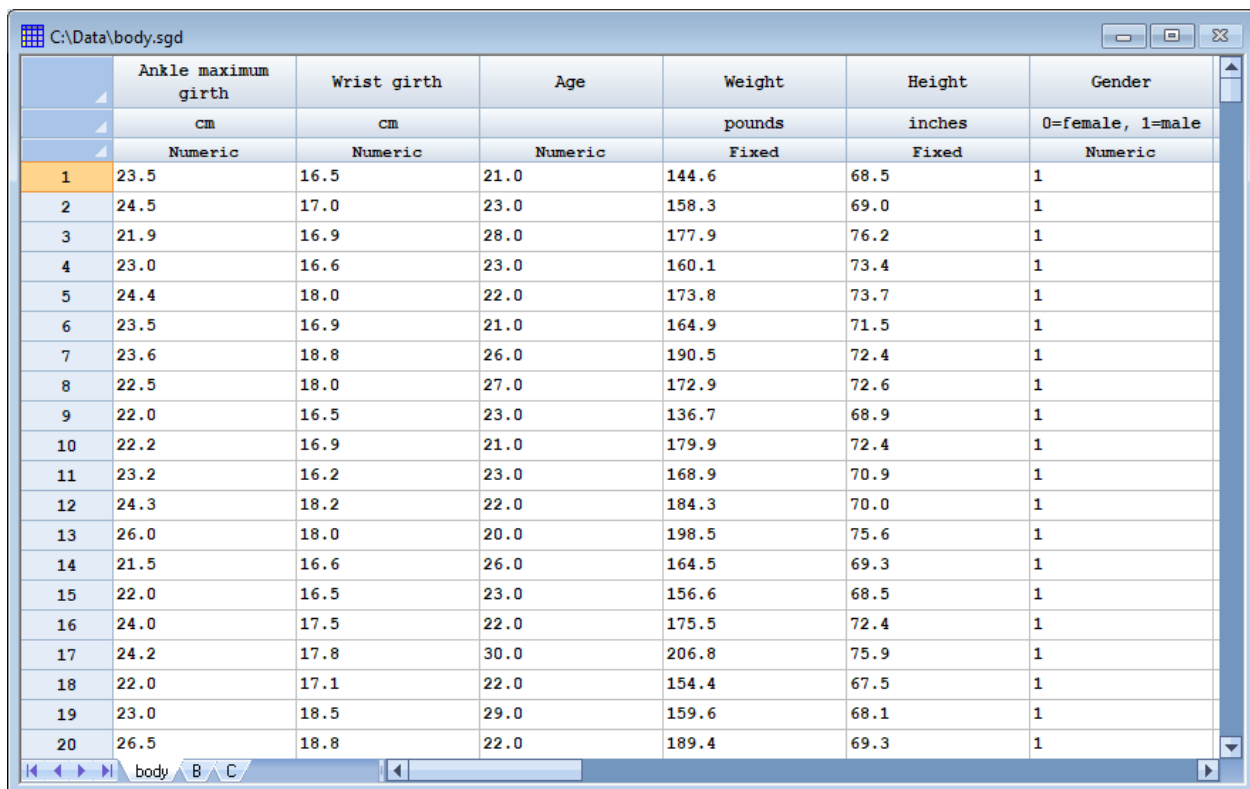
4. *Z vs X and Y*: displays the relationship between 3 variables at different combinations of 1 or 2 conditioning variables using a bubble chart, a regression model, or a nonparametric smoother.

Sample StatFolio: *trellis.sgp*

Sample Data File: *body.sgd*

Sample Data

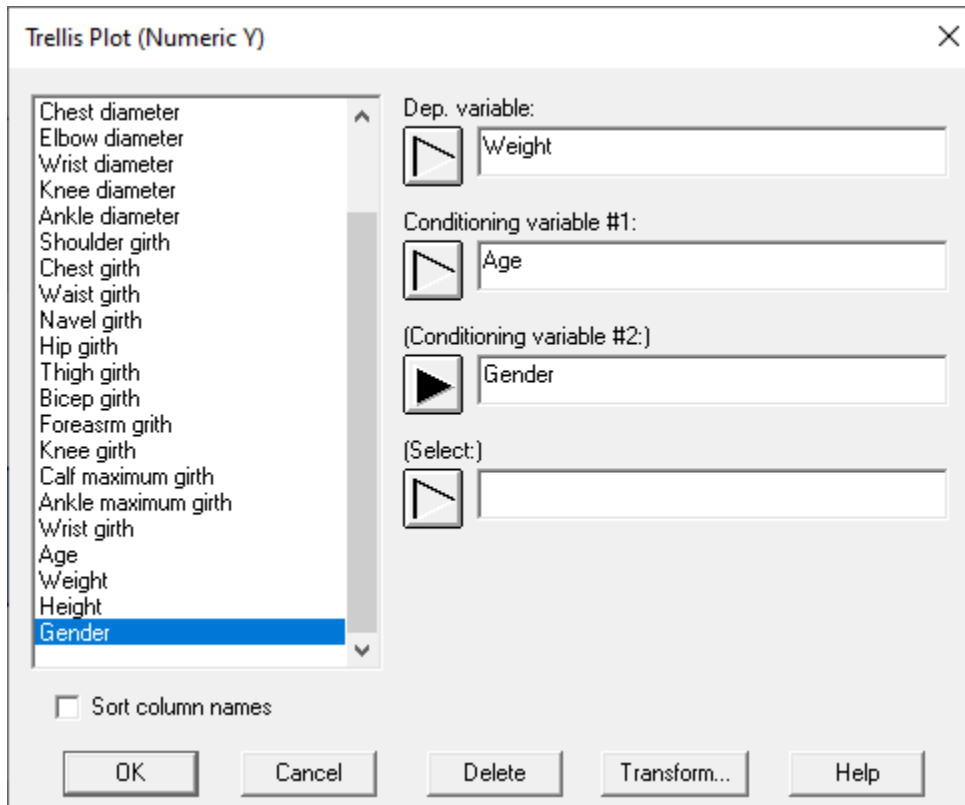
The file named *body.sgd* contains data describing physical characteristics of 247 men and 267 women sampled at fitness centers in California. It was downloaded from the data archive of the Journal of Statistics Education (http://jse.amstat.org/jse_data_archive.htm). The file contains measurements of 27 body dimensions for each individual, as well as information on their height, weight, age and gender. A small section of the file is shown below:



	Ankle maximum girth	Wrist girth	Age	Weight	Height	Gender
	cm	cm		pounds	inches	0=female, 1=male
	Numeric	Numeric	Numeric	Fixed	Fixed	Numeric
1	23.5	16.5	21.0	144.6	68.5	1
2	24.5	17.0	23.0	158.3	69.0	1
3	21.9	16.9	28.0	177.9	76.2	1
4	23.0	16.6	23.0	160.1	73.4	1
5	24.4	18.0	22.0	173.8	73.7	1
6	23.5	16.9	21.0	164.9	71.5	1
7	23.6	18.8	26.0	190.5	72.4	1
8	22.5	18.0	27.0	172.9	72.6	1
9	22.0	16.5	23.0	136.7	68.9	1
10	22.2	16.9	21.0	179.9	72.4	1
11	23.2	16.2	23.0	168.9	70.9	1
12	24.3	18.2	22.0	184.3	70.0	1
13	26.0	18.0	20.0	198.5	75.6	1
14	21.5	16.6	26.0	164.5	69.3	1
15	22.0	16.5	23.0	156.6	68.5	1
16	24.0	17.5	22.0	175.5	72.4	1
17	24.2	17.8	30.0	206.8	75.9	1
18	22.0	17.1	22.0	154.4	67.5	1
19	23.0	18.5	29.0	159.6	68.1	1
20	26.5	18.8	22.0	189.4	69.3	1

Trellis Plots for Numeric Y

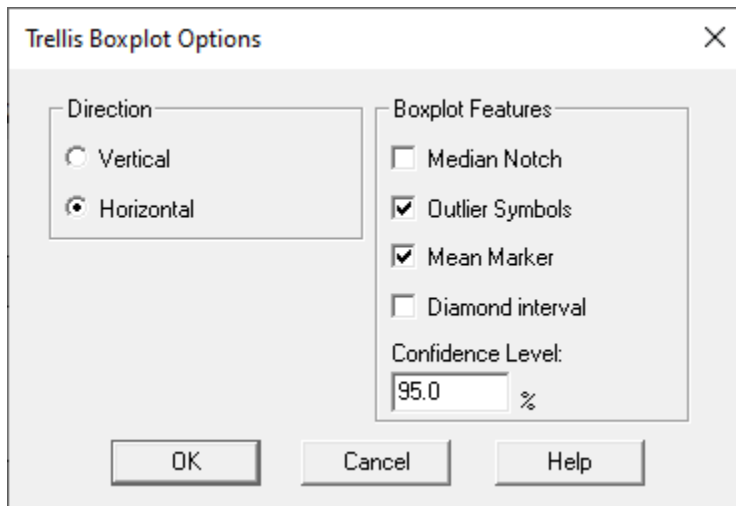
To display characteristics of a single quantitative variable, the *Numeric Y* trellis plot may be used. It begins by displaying the data input dialog box shown below:



- **Dep. variable:** numeric dependent variable for which statistics will be displayed.
- **Conditioning variable #1:** first variable used to segment the data. May be numeric or categorical.
- **Conditioning variable #2:** optional second variable used to segment the data. May be numeric or categorical.
- **Select:** optional subset selection.

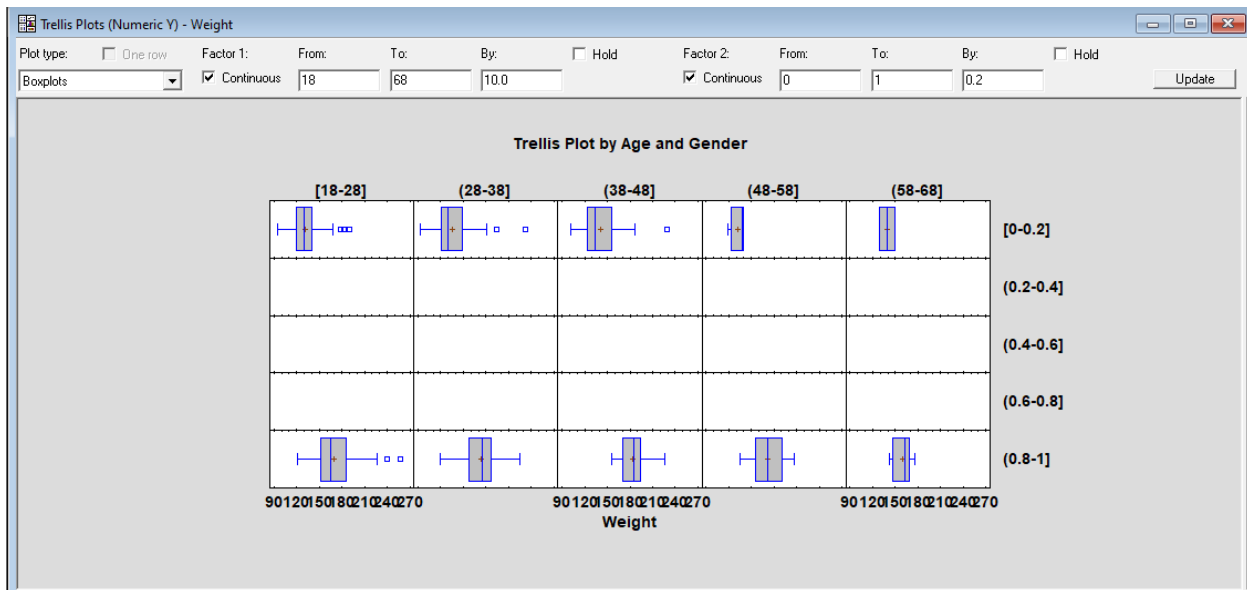
Graphs will be drawn displaying characteristics of the dependent variable at various combinations of the conditioning variables.

A second dialog box will then be displayed containing options for the *trellis boxplot*, which is the default graph for displaying numeric data:



- **Direction:** the orientation of the plot, corresponding to the direction of the whiskers.
- **Median Notch:** if selected, a notch will be added to the plot showing an approximate $100(1-\alpha)\%$ confidence interval for the median at the default system confidence level (set on the *General* tab of the *Preferences* dialog box on the *Edit* menu).
- **Outlier Symbols:** if selected, indicates the location of outside points.
- **Mean Marker:** if selected, shows the location of the sample mean as well as the median.
- **Diamond interval:** if selected, a diamond will be added to the plot showing a $100(1-\alpha)\%$ confidence interval for the mean.
- **Confidence level:** specifies the confidence level $100(1-\alpha)\%$ for the diamond interval.

After completing the second dialog box, a Statlet window is displayed showing multiple box and whisker plots:

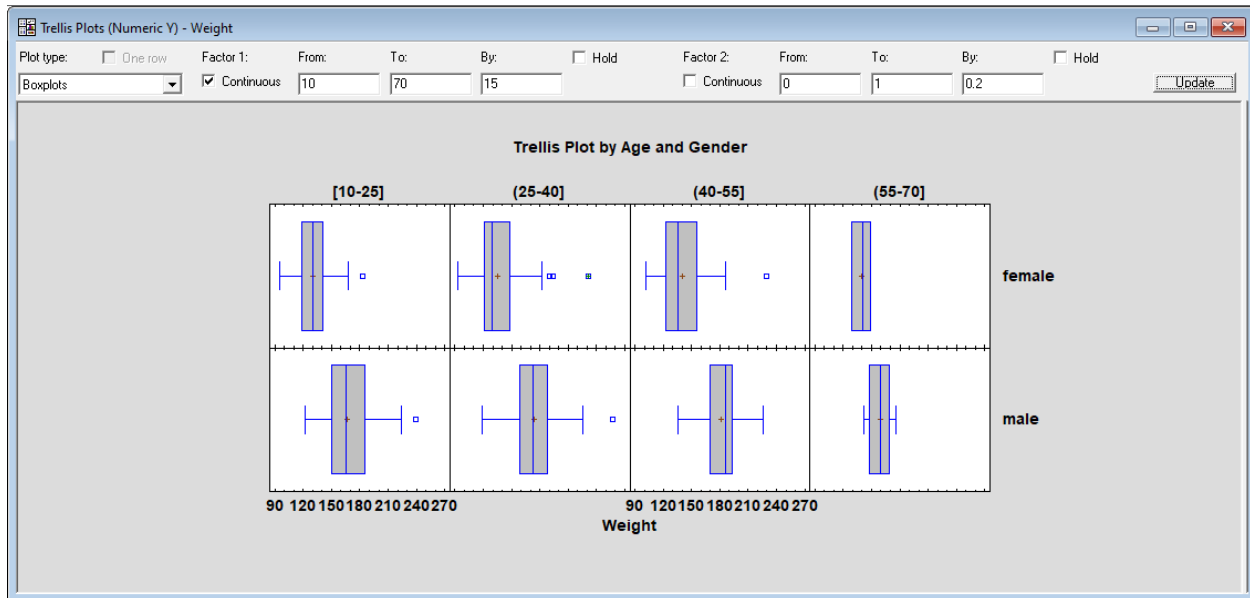


Each column of the plot shows data for different combinations of the first conditioning variable, while each row shows data for different combinations of the second conditioning variable.

The controls on the Statlet toolbar let you make changes to the display:

- **Plot type:** you have the choice of displaying boxplots, frequency plots, or normal probability plots.
- **Continuous:** specify whether each conditioning factor is continuous or discrete. If continuous, the range of the factor will be divided into intervals based on the *from*, *to* and *by* settings. If not continuous, a plot will be drawn for each unique level of the factor.
- **Hold:** check this box to prevent the division of continuous factors from reverting to the default values if the data is reread.

Changing factor 1 to range from 10 to 70 by 15 and factor 2 to be non-continuous results in the following display:

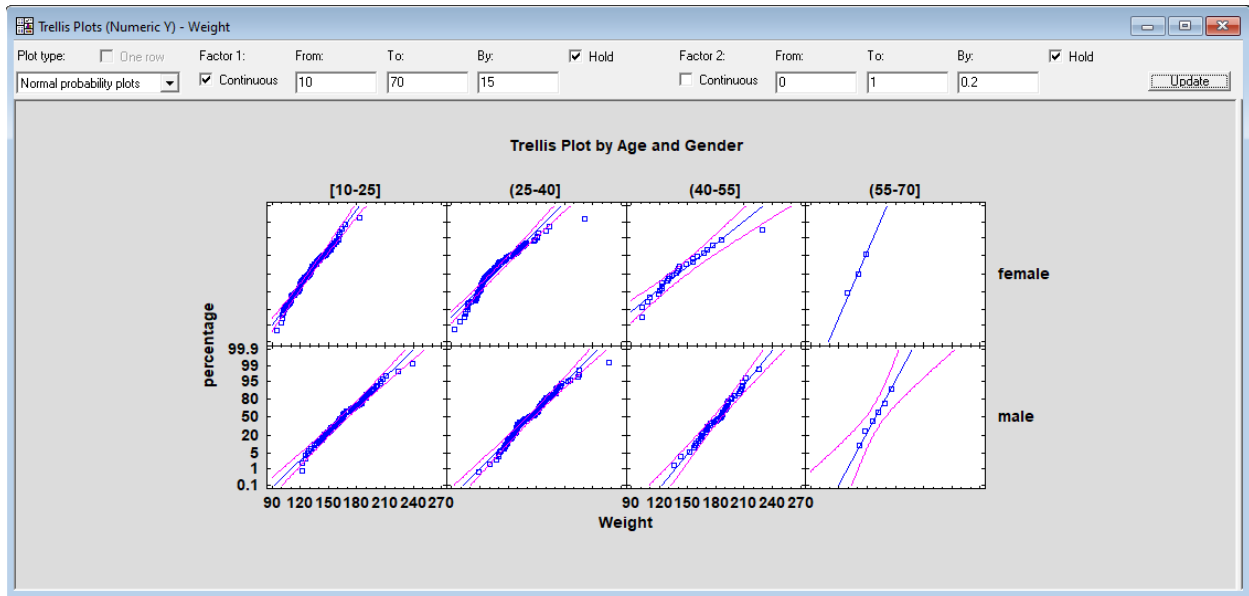


Note the labels for gender. These were automatically obtained from the value labels of the *gender* column in the Databook.

Changing the *plot type* to *Frequency plots* replaces the boxplots with frequency histograms:



Changing the *plot type* to *Normal probability plots* results in the following display:



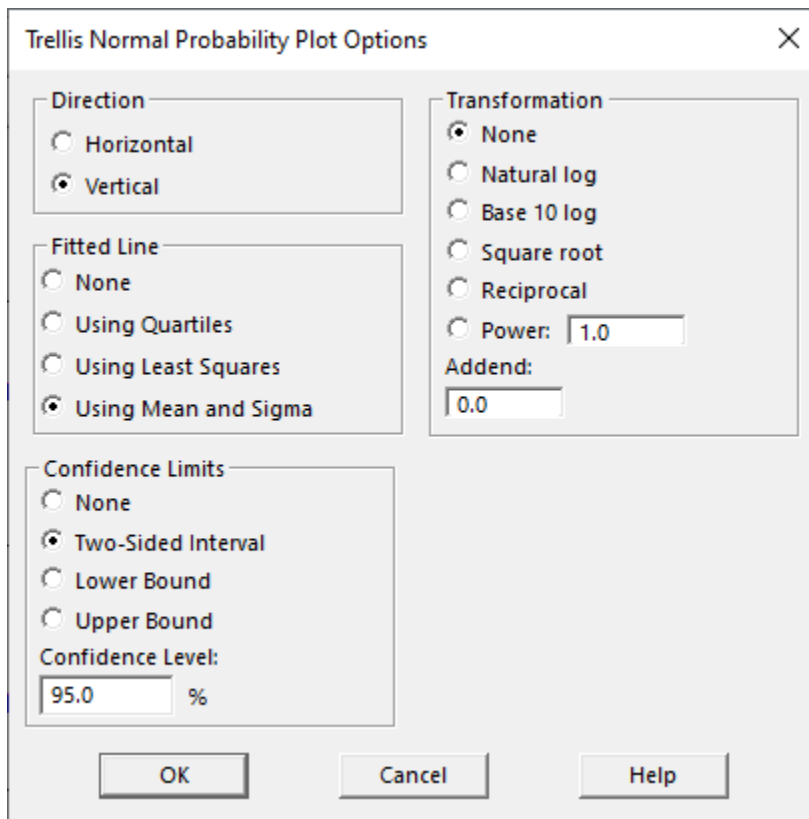
Analysis Options

The dialog box that appears if you right-click on the plot and select *Analysis Options* depends on the plot type. The *Trellis Boxplot Options* dialog box was shown earlier. There is also a dialog box that controls frequency histograms:

- **Number of classes:** the number of intervals into which the data will be divided. Intervals are adjacent to each other and of equal width.
- **Lower Limit:** lower limit of the first interval.

- **Upper Limit:** upper limit of the last interval.
- **Hold scaling:** maintains the selected number of intervals and limits even if the source data change. By default, the number of classes and the limits are recalculated whenever the data change. This is necessary so that all observations are displayed even if some of the updated data fall beyond the original limits.
- **Counts:** if *Relative*, the height of the bars represent the number of observations in a single interval. If *Cumulative*, the height represents the number of observations in the indicated interval and all intervals to its left.
- **Plot Type:** if *Histogram*, the class frequencies are displayed as a barchart. If *Polygon*, the frequencies are displayed using a connected line chart.
- **Add normal curve:** if checked, a normal density function with the same mean and standard deviation as the data is added to the plot.

A separate *Analysis Options* dialog box controls the normal probability plots:

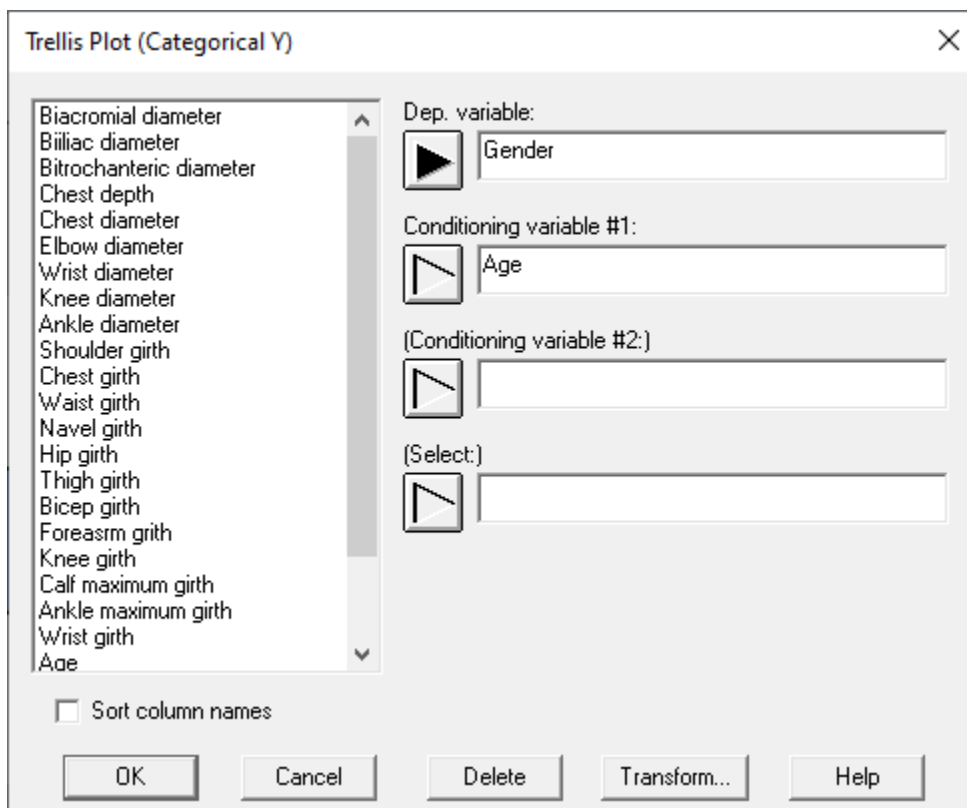


- **Direction:** the orientation of the plot. If vertical, the *Percentage* is displayed on the vertical axis. If horizontal, *Percentage* is displayed on the horizontal axis.

- **Fitted Line:** the method used to fit the reference line to the data. If *Using Quartiles*, the line passes through the median when *Percentage* equals 50 with a slope determined from the interquartile range. If *Using Least Squares*, the line is fit by least squares regression of the normal quantiles on the observed order statistics. If *Using Mean and Sigma*, the line is determined from the mean and standard deviation of the n observations. The method based on quartiles puts more weight on the shape of the data near the center and is often able to show deviations from normality in the tails that would not be evident using the other methods.
- **Confidence limits:** displays confidence limits around the fitted line (only available when estimating the line *Using mean and sigma*).
- **Transformation:** the selected transformation is applied to the data before being plotted. The assumption of a normal distribution therefore applies in the transformed metric.

Trellis Plots for Categorical Y

To display characteristics of a single categorical variable, the *Categorical Y* trellis plot may be used. It begins by displaying the data input dialog box shown below:



Trellis Plot (Categorical Y)

Dep. variable: Gender

Conditioning variable #1: Age

(Conditioning variable #2:)

(Select:)

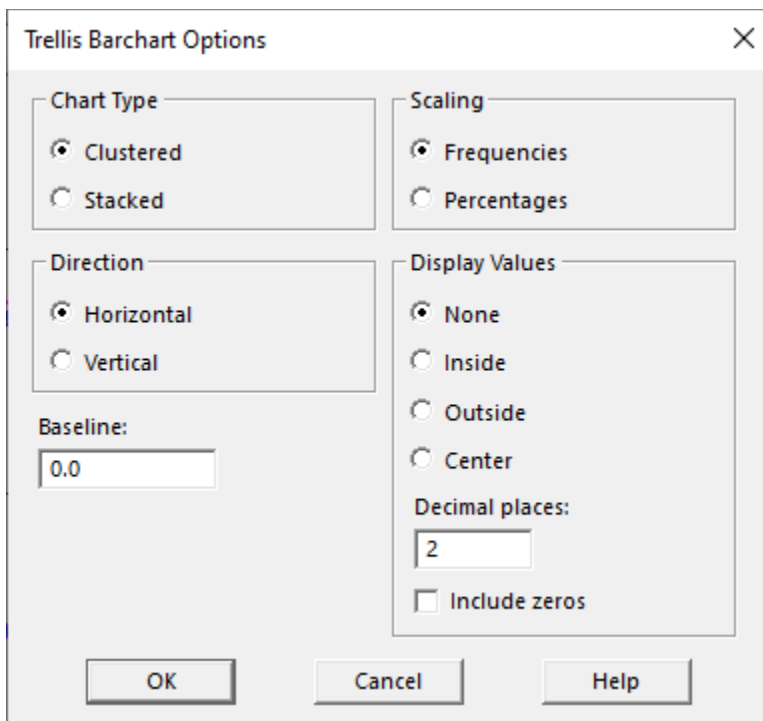
Sort column names

OK Cancel Delete Transform... Help

- **Dep. variable:** categorical dependent variable for which statistics will be displayed.
- **Conditioning variable #1:** first variable used to segment the data. May be numeric or categorical.
- **Conditioning variable #2:** optional second variable used to segment the data. May be numeric or categorical.
- **Select:** optional subset selection.

Graphs will be drawn displaying characteristics of the dependent variable at various combinations of the conditioning variables.

A second dialog box will then be displayed containing options for the *trellis barchart*, which is the default graph for displaying categorical data:



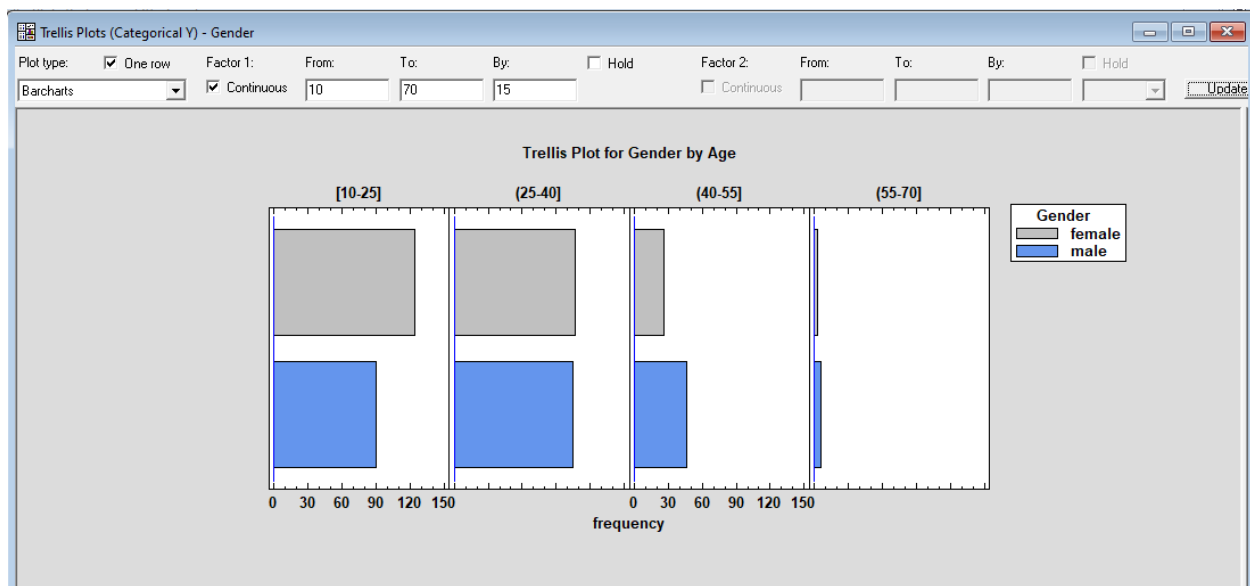
- **Chart Type:** The bars may be clustered side by side or stacked one upon the other.
- **Scaling:** whether the axis scale shows the frequencies in each class or the percentages in each class.
- **Direction:** whether the bars extend horizontally or vertically.
- **Baseline:** the value from which the bars extend.

- **Display Values:** whether to display plotted values as labels. If *inside*, labels will be plotted near the ends of the bars. If *outside*, labels will be plotted next to the bars. If *center*, labels will be centered in the bars.
- **Decimal places:** number of decimal places to be shown for each displayed value.
- **Include zeroes:** whether to display values for bars with zero frequency.

After completing the second dialog box, a Statlet window is displayed showing

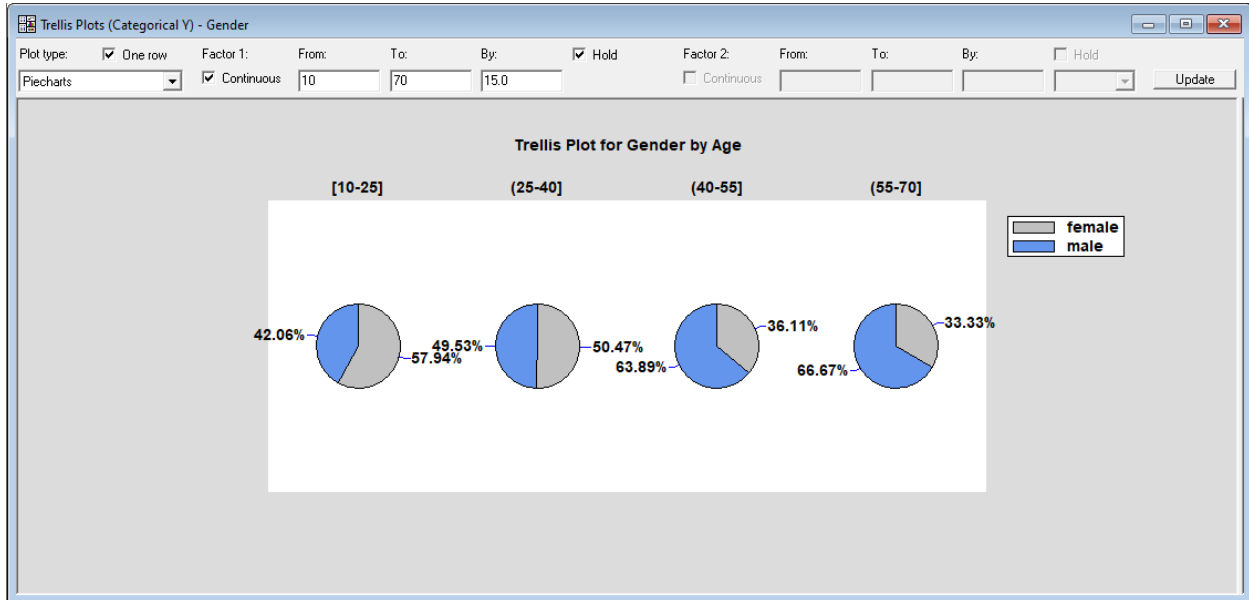


Each plot shows data for different levels of the first conditioning variable. Checking the *One row* box and changing the settings for Factor 1 displays the following:

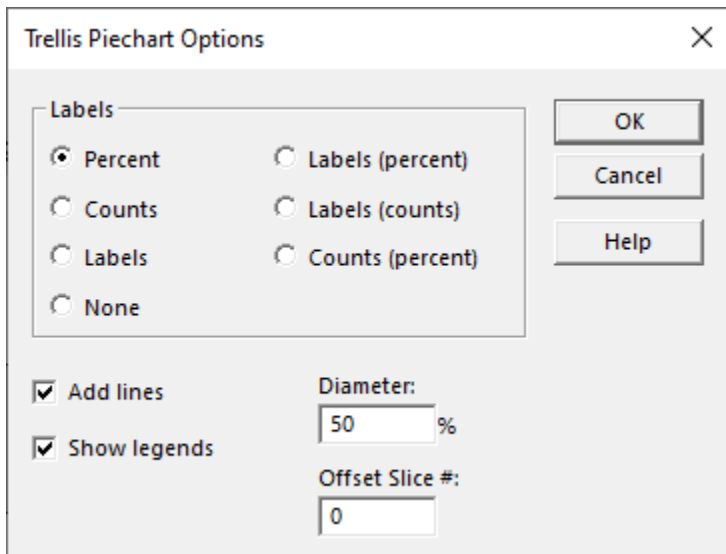


Clearly, females outnumber males only in the younger age brackets.

The *Plot type* field may be used to display other types of graphs. For example, changing it to *Piechart* creates the following:



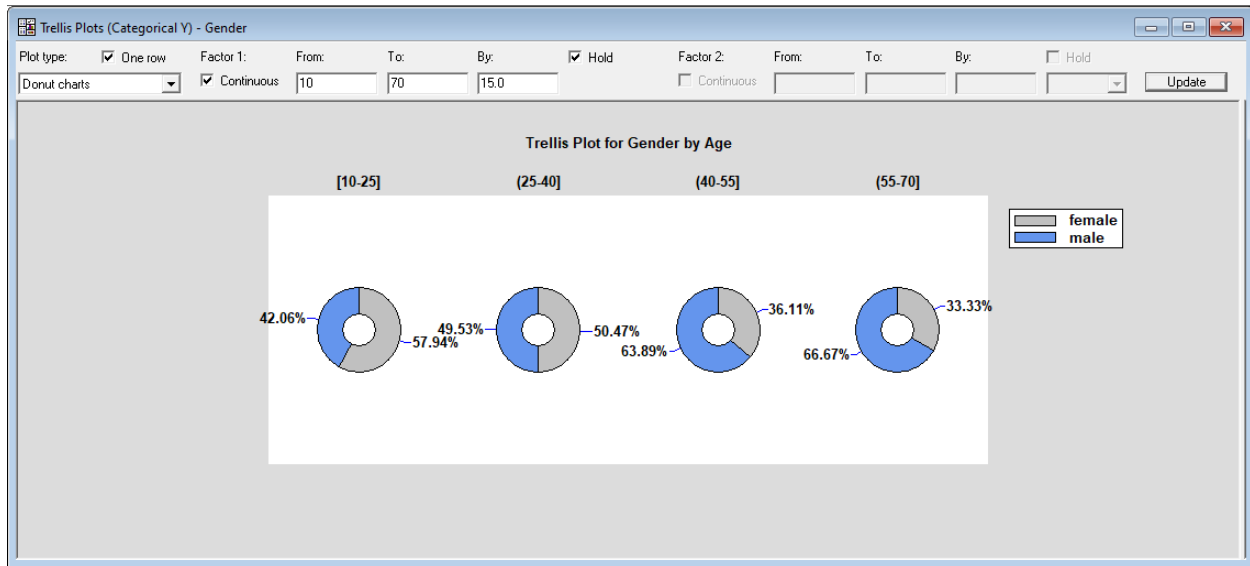
The *Analysis Options* dialog box displays options for changing the appearance of the plot:



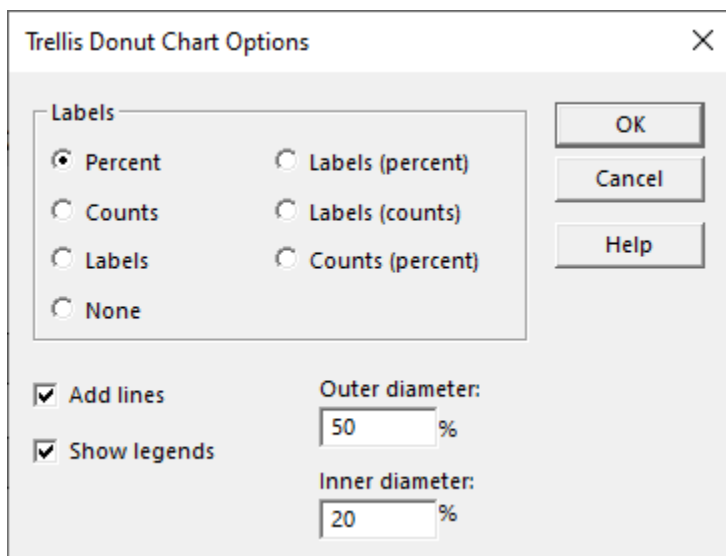
- **Labels:** the desired information to be placed next to each slice of the pie.
- **Add lines:** whether to extend lines from each slice to its label.
- **Show legends:** whether to display a legend block to the right of the plot.

- **Diameter:** the size of the pie relative to the size of the plotting area. Reducing this value makes the pie smaller.
- **Offset Slice #:** number of a slice between 1 and k to offset from the rest of the pie.

A third type of plot is the *Donut chart*:



The *Analysis Options* dialog box is shown below:

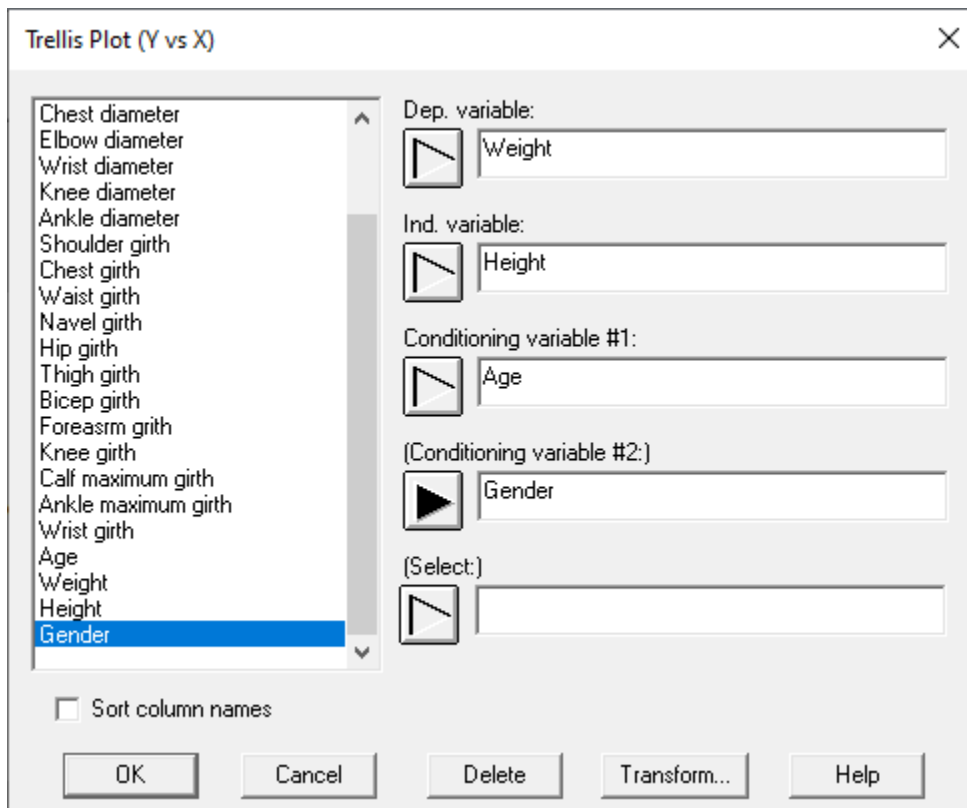


- **Labels:** the desired information to be placed next to each section of the donut.
- **Add lines:** whether to extend lines from each section to its label.

- **Show legends:** whether to display a legend block to the right of the plot.
- **Outer diameter:** the size of the donut relative to the size of the plotting area. Reducing this value makes the donut smaller.
- **Inner diameter:** the size of the donut hole relative to the size of the plotting area.

Trellis Plots for Y Versus X

Trellis plots may also be created to display the relationship between a numeric dependent variable Y and a numeric independent variable X. Selecting *Trellis Plots – Y vs X* begins by displaying the data input dialog box shown below:

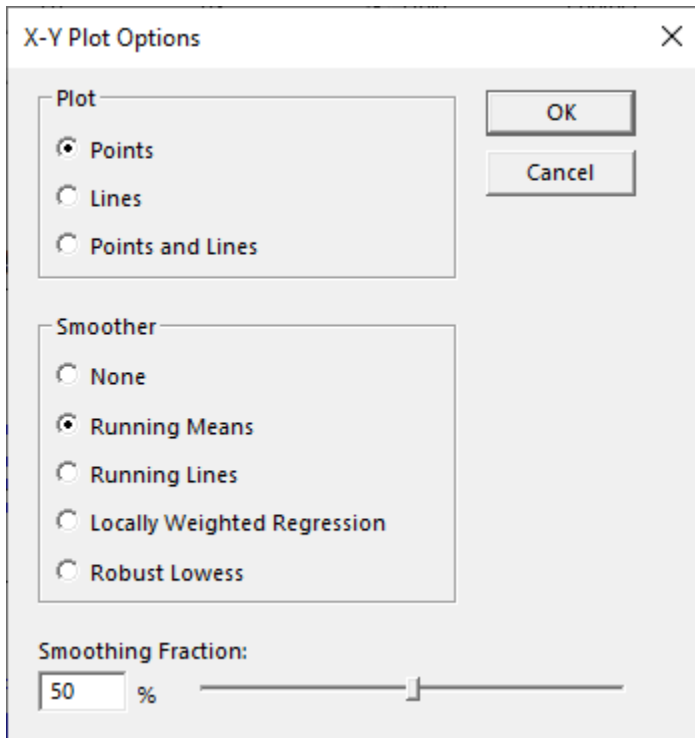


It then displays a Statlet window displaying scatterplots for various combinations of the conditioning variables:



Note that the scaling of the conditioning factors has been changed as in earlier trellis plots.

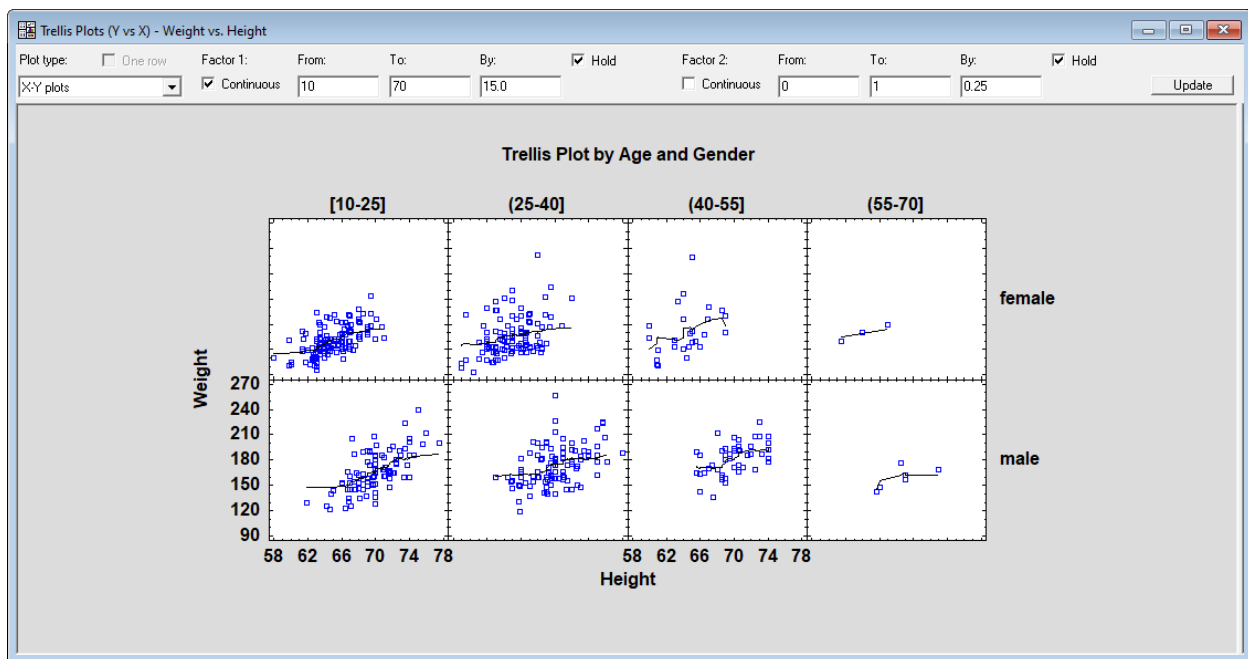
Analysis Options may be used to add features to the plot:



- **Plot:** changes the display from a scatterplot to a connected line plot.

- **Smoother:** specifies the type of smoothing to be performed on each plot, if any. The smoothed values are created by sorting the data according to the value of X, estimating the height of the function at each value of X using the q nearest neighbors to that value, and then connecting the estimated heights with a line. Details of the various smoothing procedures are contained in the PDF document titled *Graphics Options*.
- **Fraction:** the percentage of observations in each plot used to select the q nearest neighbors.

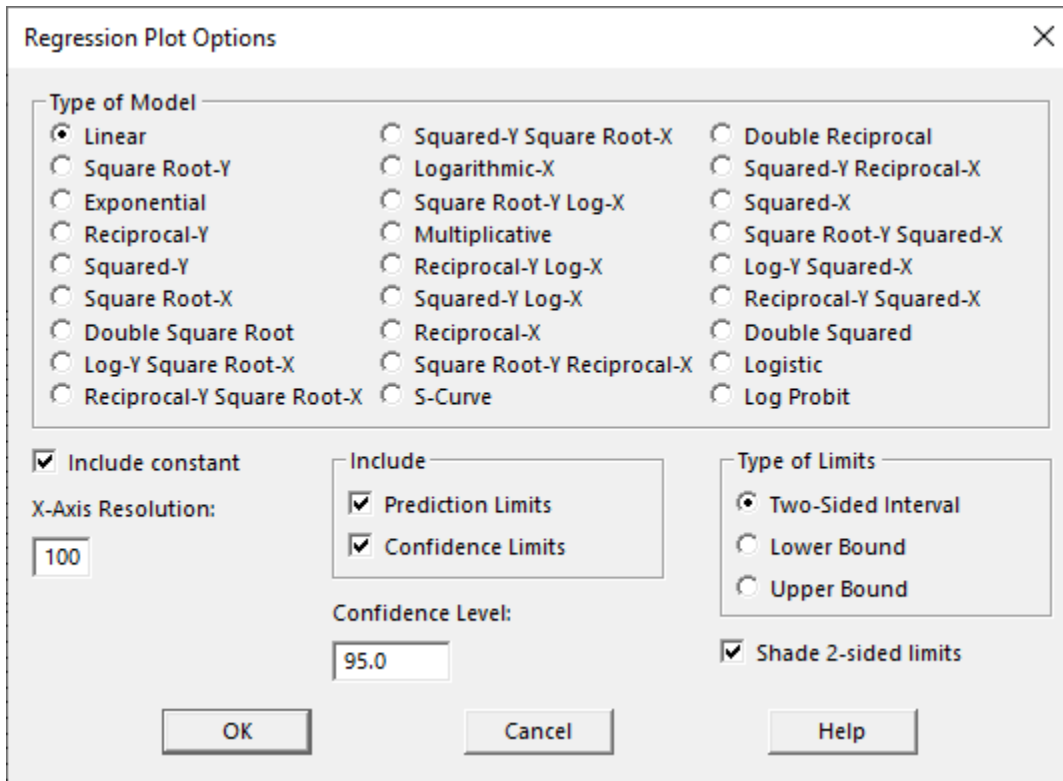
The plot below shows the result of adding running means based on a 50% fraction.



Changing the *Plot type* field to *Regression plots* adds a least squares regression line to each segment of the plot:



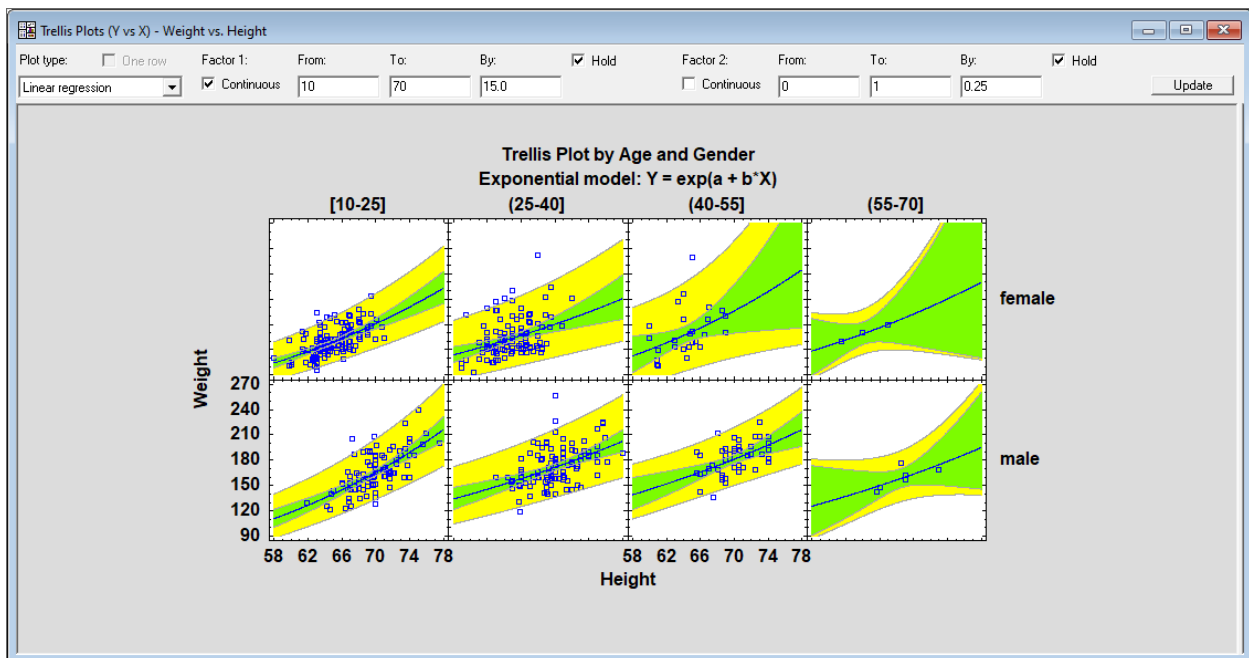
The *Analysis Options* dialog box when plotting regression lines is shown below:



- **Type of Model:** the type model to fit to the data.

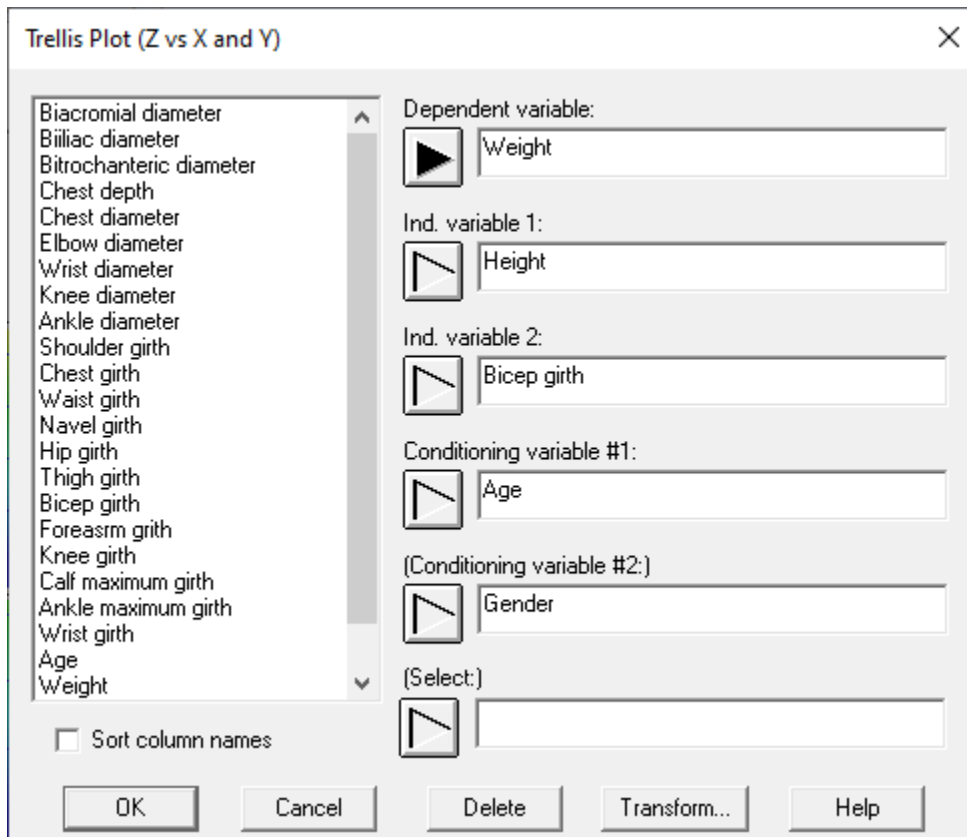
- **Include constant:** whether to include a constant term in the model. If not, the model will be constrained to pass through the origin.
- **X-Axis Resolution:** the number of values of X at which the line is determined when plotting. Higher resolutions result in smoother plots.
- **Include:** the limits to include on the plot.
- **Confidence Level:** the confidence percentage for the limits.
- **Type of Limits:** whether to plot two-sided intervals or one-sided bounds.
- **Shade 2-sided limits:** If checked, the area between the limits will be shaded using a fill color.

Plotting both prediction and confidence intervals creates the following plot:



Trellis Plots for Z Versus X and Y

Trellis plots may also be created to display the relationship between a numeric dependent variable Z and 2 numeric independent variables X and Y. Selecting *Trellis Plots – Z vs X and Y* begins by displaying the data input dialog box shown below:



Trellis Plot (Z vs X and Y)

Dependent variable:

Ind. variable 1:

Ind. variable 2:

Conditioning variable #1:

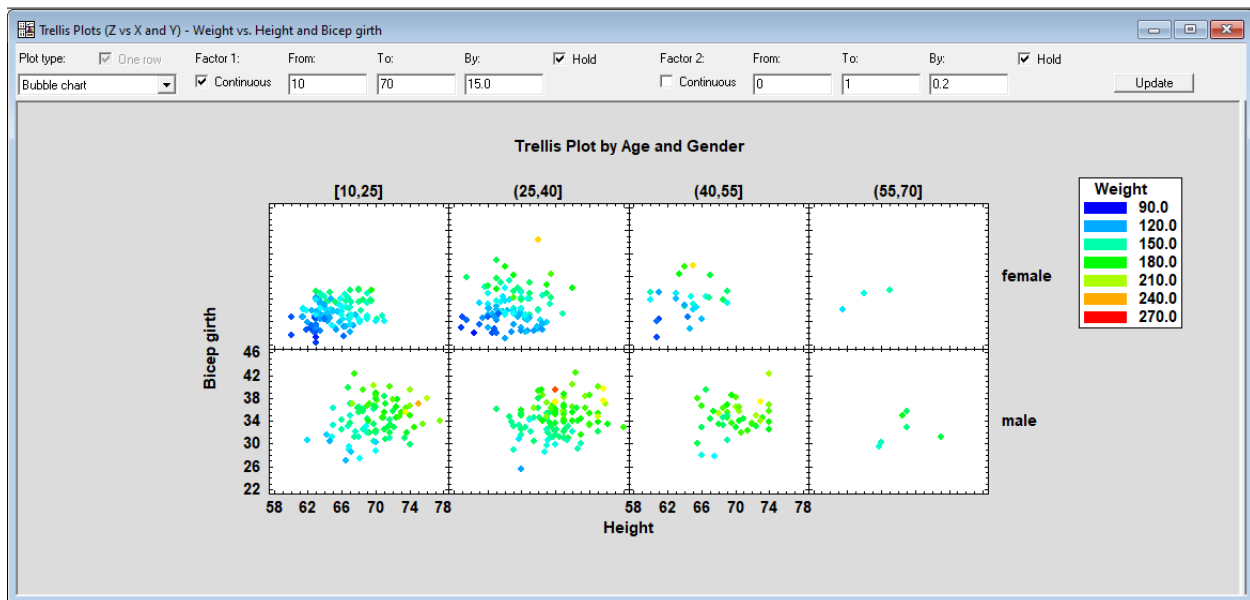
(Conditioning variable #2:)

(Select:)

Sort column names

OK Cancel Delete Transform... Help

The default plot that is created is a bubble chart in which each observation is colored according to the value of *weight*:

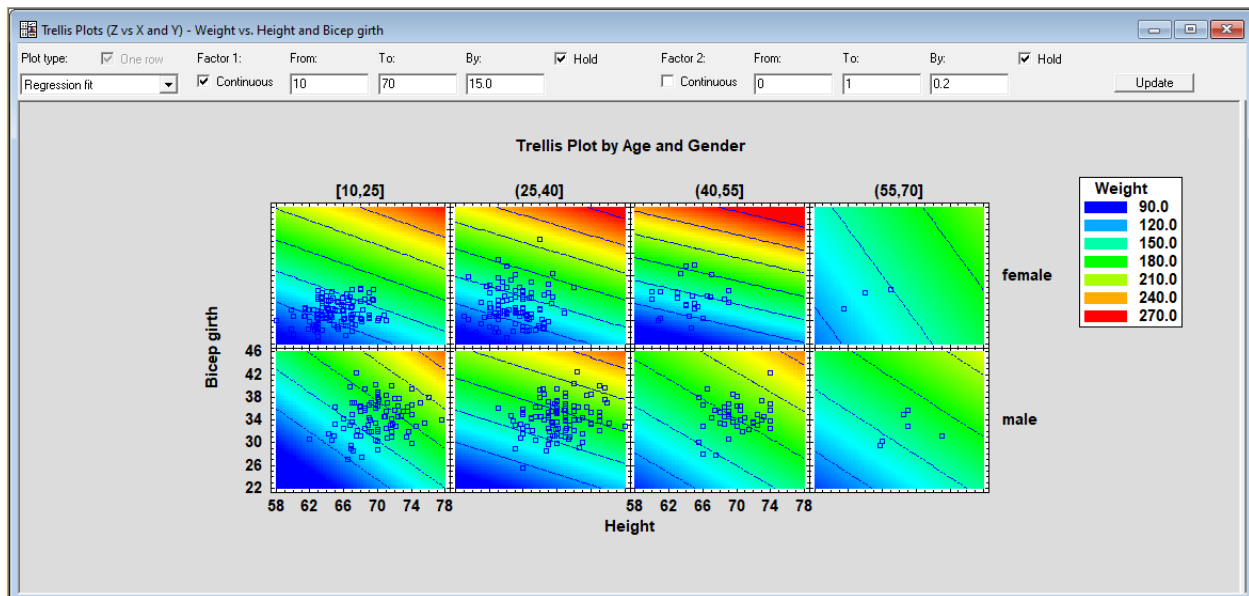


Note that the scaling of the conditioning factors has been changed as in earlier trellis plots.

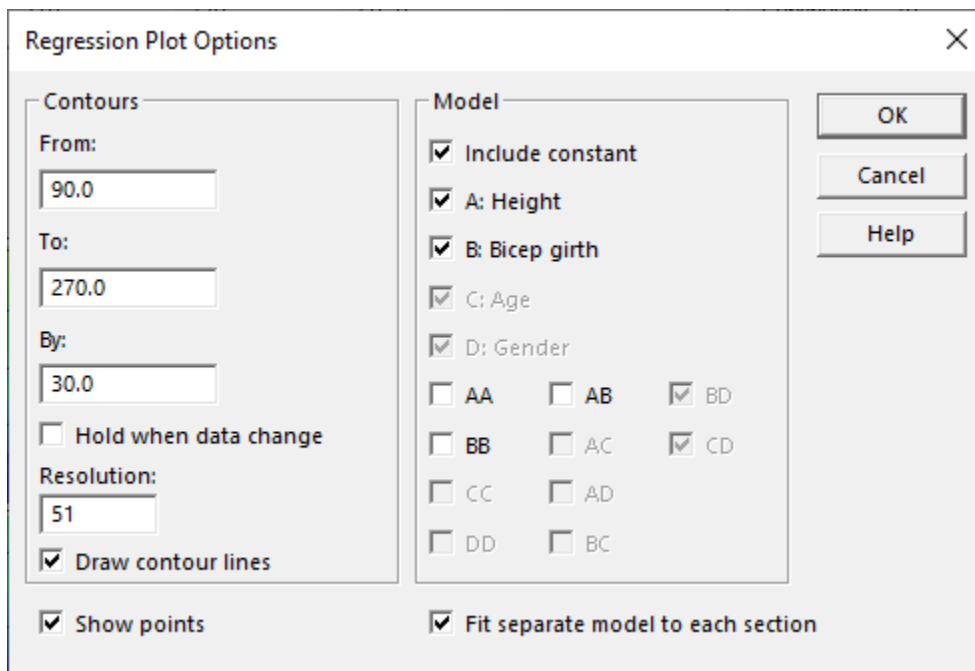
The contour levels and size of the bubbles may be changed using *Analysis Options*:

- **Contours:** *From* and *To* specify the range of values that define the coloring of the plot. If contour lines are included, the *By* field defines the distance between the contour lines. Check *Hold when data change* to keep these settings from changing when the data are reread.
- **Bubbles:** defines the range of sizes for the bubbles. If the minimum and maximum values are different, each bubble's size will be proportional to the value of the dependent variable.

If a regression model is selected, the trellis plot displays the fitted model with respect to each of the independent variables at various combinations of the conditioning variables:



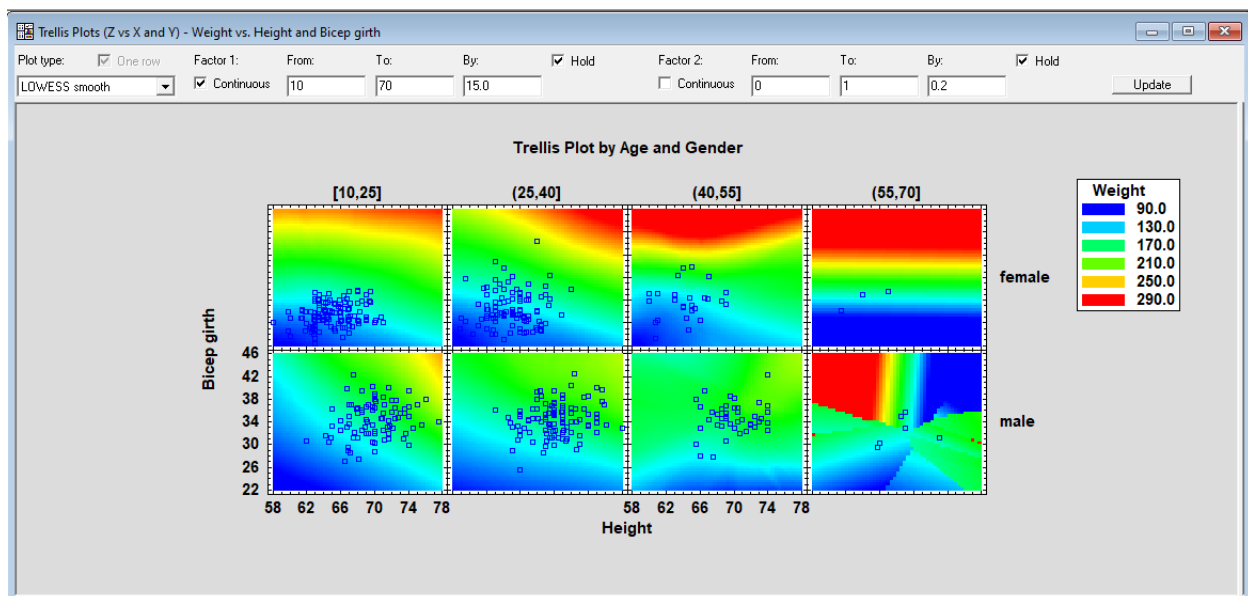
Analysis Options may be used to add features to the plot:



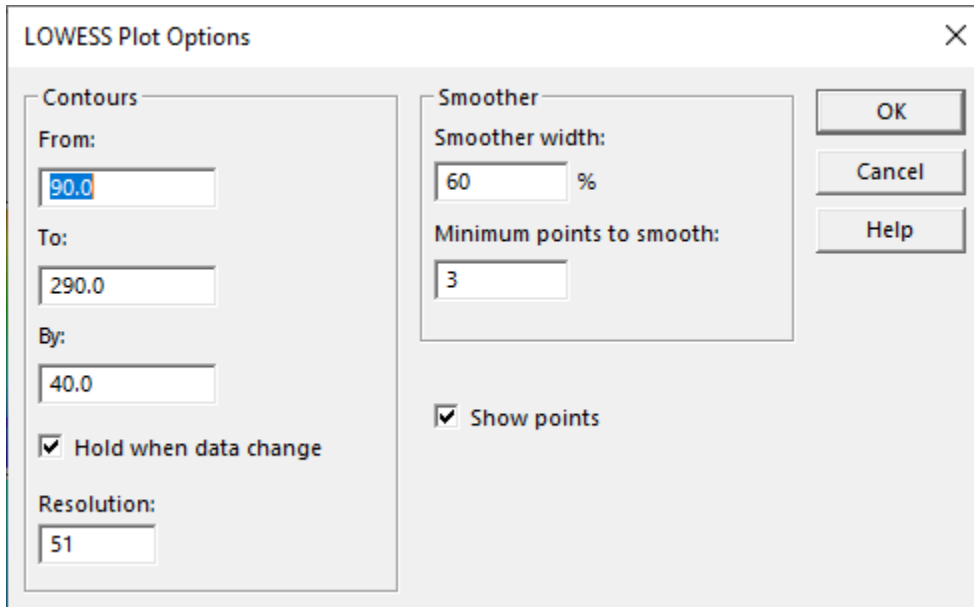
- **Contours:** *From* and *To* specify the range of values that define the coloring of the plot. If contour lines are included, the *By* field defines the distance between the contour lines. Check *Hold when data change* to keep these settings from changing when the data are reread.
- **Draw contour lines:** draws lines in addition to colored regions.
- **Show points:** displays the original data used to fit the regression model.

- **Include constant:** whether to include a constant term in the model. If not, the model will be constrained to pass through the origin.
- **Model:** select the effects to be included in the fitted regression model. A , B , C and D are the first-order main effects. AA , BB , CC and DD are the second-order quadratic terms (for continuous variables only). The other terms represent cross-products or two-factor interactions.
- **Fit separate model to each section:** if selected, a separate regression model is fit in each section using only the data values in that section. In such cases, all terms involving either C or D are ignored. If not selected, a single model is fit using all of the data including terms for C and D .
- **Resolution:** the number of values of X and Y at which the model is evaluated when plotting. Higher resolutions result in smoother plots.

If a *LOWESS smooth* is selected, a smoother is applied to the data values in each section of the plot. At many locations throughout each section, a weighted regression is fit to the data values that are closest to that location and the estimated value at that location is plotted. This provides a model-independent estimate of the dependent variable throughout the region. For sections with a reasonable amount of data, the result is usually quite good. For sections with few data values, the smoother may not work well.



The *Analysis Options* dialog box specifies certain aspects of the smoother and how the contour levels are determined:



- **Contours:** *From* and *To* specify the range of values that define the coloring of the plot. If contour lines are included, the *By* field defines the distance between the contour lines. Check *Hold when data change* to keep these settings from changing when the data are reread.
- **Resolution:** the number of values of *X* and *Y* at which the model is evaluated when plotting. Higher resolutions result in smoother plots.
- **Smoother width:** percentage of nearest points at each location that are included in the smooth.
- **Minimum points to smooth:** if the number of data values in a section is less than this number, smoothing will not be performed.
- **Show points:** displays the original data used to fit the regression model.