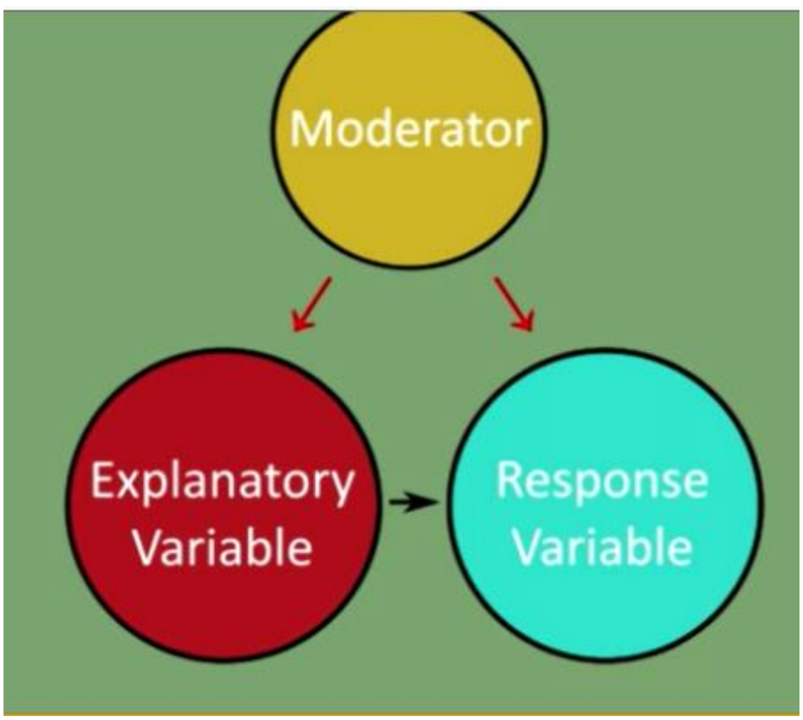
12. Moderation

Section 12.1: Moderation with ANOVA
Section 12.2: Graphing Moderation with ANOVA
Section 12.3: Moderation with Chi-Square
Section 12.4: Graphing Moderation with Chi-Square
Section 12.5: Moderation with Correlation
Section 12.6: Graphing Moderation with Correlation

Section 12.1: Moderation with ANOVA

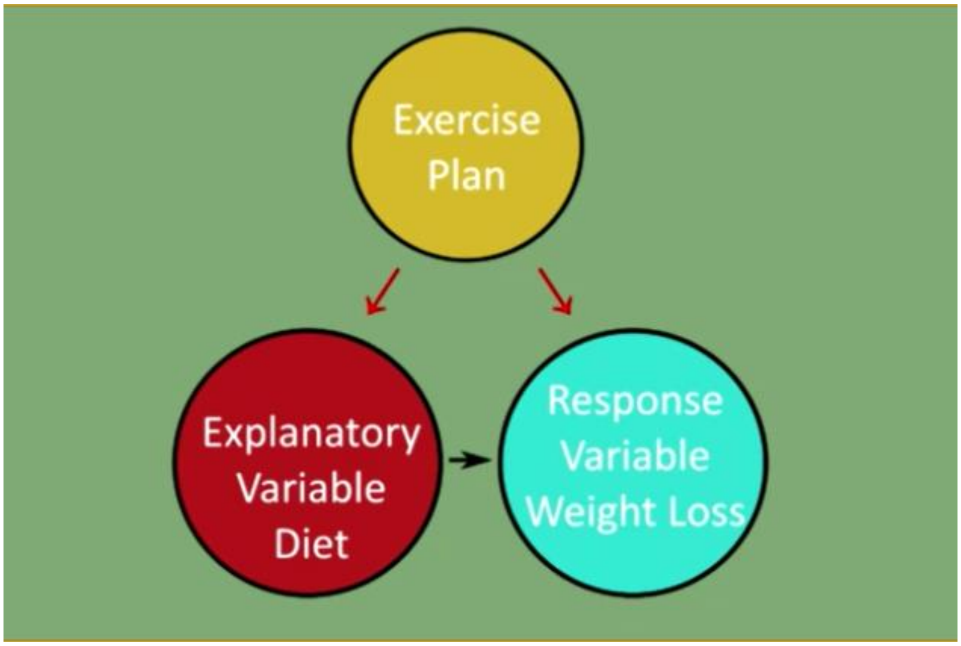
Statistical interaction describes a relationship between two variables that is dependent upon, or moderated by, a third variable. For instance do you prefer ketchup or soy sauce? Obviously, your answer depends on what food you're eating. If you're eating sushi then your preference may be soy sauce. If you're having a burger and fries, you may prefer ketchup.

In this case, the third variable is referred to as the moderating variable or simply the moderator. The effect of a moderating variable is often characterized statistically as an interaction; that is, a third variable that affects the direction and/or strength of the relation between your explanatory, or X variable, and your response, or Y variable.



What if the population we are studying has different subgroups? Could it be that, like the soy sauce ketchup example, different subgroups could have a moderating effect on our association of interest?

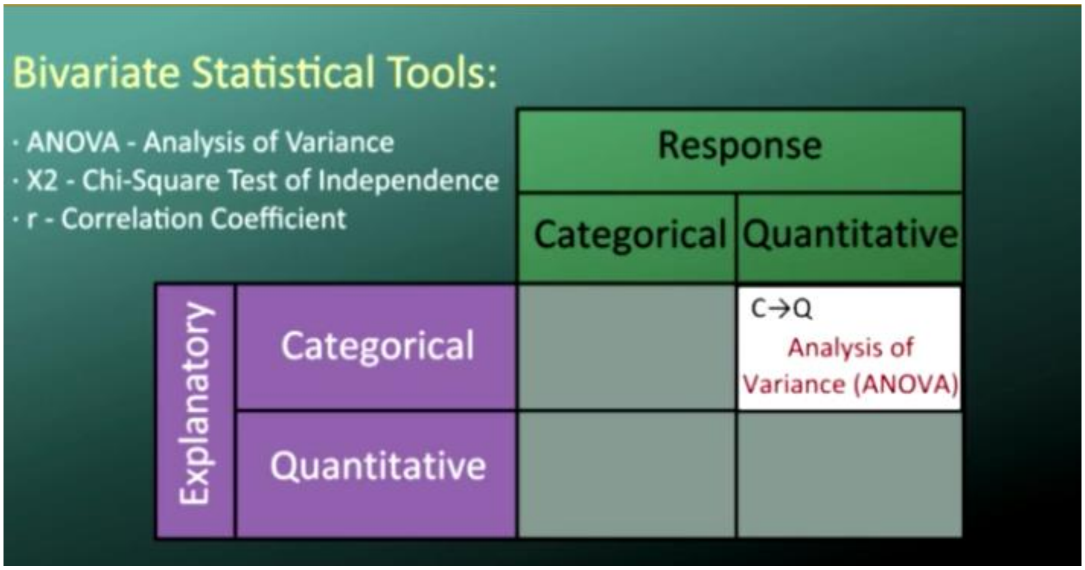
To explore this idea we're going to use a hypothetical study and some made up data. In our imaginary study we're looking at two diets and their effect on weight loss. Diet A is a low-carbohydrate plan. Diet B is a low fat plan. Our hypothetical study also recorded data on which exercise program participants chose; cardiovascular exercise or weight training. Our variables of interest are Diet and Weight Loss. We've added this third variable, Exercise Plan to help us understand moderation, or statistical interaction.



So what is the association between diet plan A and B, our explanatory variable, and weight loss, our quantitative response variable? This table shows our hypothetical data showing diet, weight loss, and exercise plan.

Cardio	B	7.5
Cardio	B	5
Cardio	B	7.2
Cardio	B	9.1
Weights	A	9.7
Weights	A	7.1
Weights	A	7.2
Weights	A	10.3
Weights	A	10.9
Weights	A	4.2
Weights	A	9.4
Weights	A	10

Since we have a categorical explanatory variable, diet plan A or B, and a quantitative response variable, that is weight loss, we will of course need to use analysis of variance to evaluate the association.



The resulting output for this analysis is shown here.

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Diet	2	A B

Number of Observations Read	40
Number of Observations Used	40

The ANOVA Procedure
Dependent Variable: WeightLoss

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	284.622250	284.622250	12.00	0.0013
Error	38	901.381500	23.720566		
Corrected Total	39	1186.003750			

R-Square	Coeff Var	Root MSE	WeightLoss Mean
0.239984	40.62879	4.870376	11.98750

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Diet	1	284.6222500	284.6222500	12.00	0.0013

As you can see, we're testing the association between diet A and B and weight loss. There are 40 observations in the data set. The F value is 12 and it's associated with a significant p-value. That is, a p-value less than 0.05. While this tells us that there is a significant association between diet type and weight loss, to understand that association we need to look at the output generated by the mean statement.

Here we see that the average one month weight loss for diet A is about 14.7 pounds and that the average one month weight loss for diet B is about 9.3 pounds.

Level of Diet	N	WeightLoss	
		Mean	Std Dev
A	20	14.6550000	6.30208612
B	20	9.3200000	2.77936002

So, in conjunction with the significant p-value, we can say that diet plan A is associated with significantly greater weight loss than diet plan B.

Here we show the finding graphically as a bar chart with diet, the explanatory variable, on the x-axis and the mean weight loss, our response variable, on the y-axis.

But what about our third variable, exercise program? Would we get the same results in terms of the association between diet and weight loss for those participants using cardio and those participants using weight training?

In statistics, moderation occurs when the relationship between two variables depends on a third variable. In this case, the third variable is referred to as the moderating variable, or simply the moderator. The effect of the moderating variable is often characterized statistically as an interaction. That is, a third variable that effects the direction and or strength of the relation between your explanatory and response variables. So, does type of exercise program affect the direction or strength of the relationship between diet and weight loss?

The standard way of asking this question in the context of analysis of variance is to move to the use of a two way or two factor analysis of variance, rather than the one way or one factor ANOVA that we've been using.

Instead, we're going to take a less-standard approach that can be consistently used across each of the inferential tools, that is ANOVA, Chi-Square, and Pearson Correlation. In each of these contexts, we're actually going to be asking the question, is our explanatory variable associated with our response variable, for each population sub-group or each level of our third variable?

That is, are diet type and weight loss associated for those doing the cardio exercise program? And are diet and weight loss associated for those using the weight-training program?

Here are the results of the example analysis.

The ANOVA Procedure

Exercise=Cardio

Class Level Information		
Class	Levels	Values
Diet	2	A B

Number of Observations Read	20
Number of Observations Used	20

The ANOVA Procedure
Dependent Variable: WeightLoss

Exercise=Cardio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	896.4605000	896.4605000	280.68	<.0001
Error	18	57.4890000	3.1938333		
Corrected Total	19	953.9495000			

The ANOVA table examining the relationship between diet and weight loss, for those in the cardio exercise group, shows a large f-value and a significant associated p-value.

When examining the means table, we see that for those involved in the cardio exercise program, diet A is associated with greater weight loss, 20.5 pounds on average, than diet B, which is 7.1 pounds on average.

The ANOVA Procedure
Dependent Variable: WeightLoss

Exercise=Cardio

Level of Diet	N	WeightLoss	
		Mean	Std Dev
A	10	20.5000000	1.93218357
B	10	7.1100000	1.62921249

The association between diet and weight loss for those involved in the weight training exercise program is also significant. It has a large f-value.

Exercise=Weights

Class Level Information		
Class	Levels	Values
Diet	2	A B

Number of Observations Read	20
Number of Observations Used	20

The ANOVA Procedure
Dependent Variable: WeightLoss

Exercise=Weights

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	36.99200000	36.99200000	10.58	0.0044
Error	18	62.93000000	3.49611111		
Corrected Total	19	99.92200000			

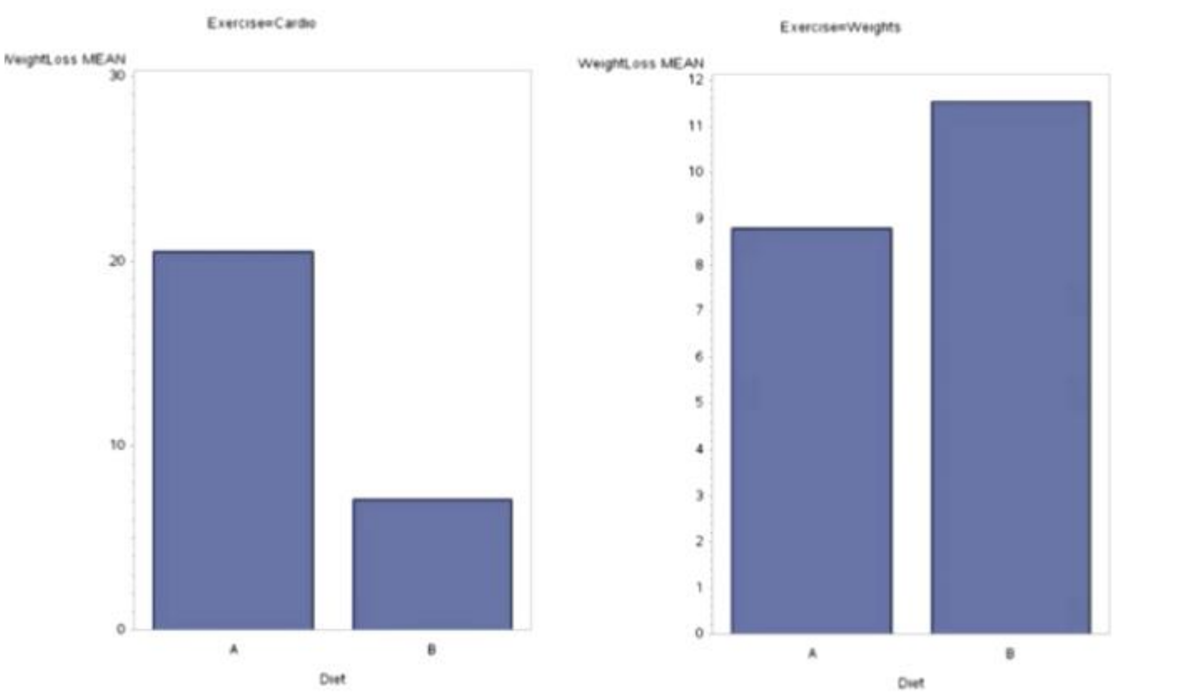
However, the means show that the association is in the opposite direction. For those involved in weight training, diet B is associated with greater weight loss, at 11.5 pounds, compared to diet A, which is only 8.8 pounds.

The ANOVA Procedure
Dependent Variable: WeightLoss

Exercise=Weights

Level of Diet	N	WeightLoss Mean	WeightLoss Std
A	10	8.8100000	2.0474
B	10	11.5300000	1.6733

Here, these results are shown graphically.



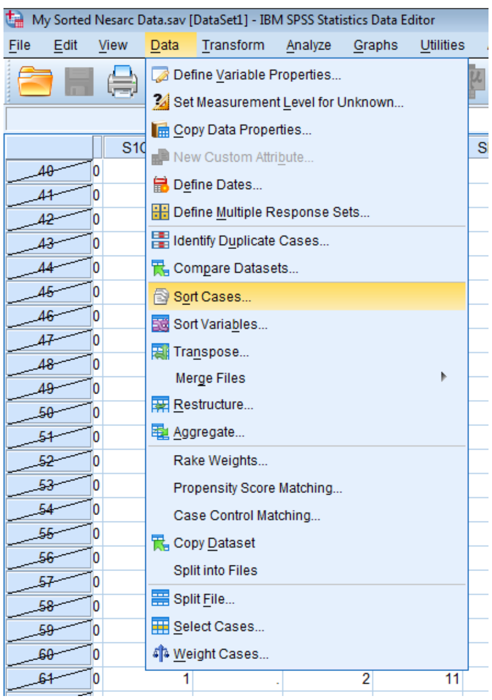
As you can see, the relationship between diet and weight loss depends on which exercise program is being used. When using cardio, diet A is significantly better for weight loss than diet B. When using weights, diet B is significantly better for weight loss than diet A. Thus, we can say there is a significant statistical interaction between the variables diet and weight loss, and the type of exercise (our third variable) moderates the association between diet and weight loss.

Suppose we did not evaluate exercise as a possible moderator and instead focused only on the association between diet and weight loss for the entire population. Based on this graph, obviously, we would've incorrectly concluded that diet A is better than diet B. As we now know, that is true only if we're looking at the cardio group.

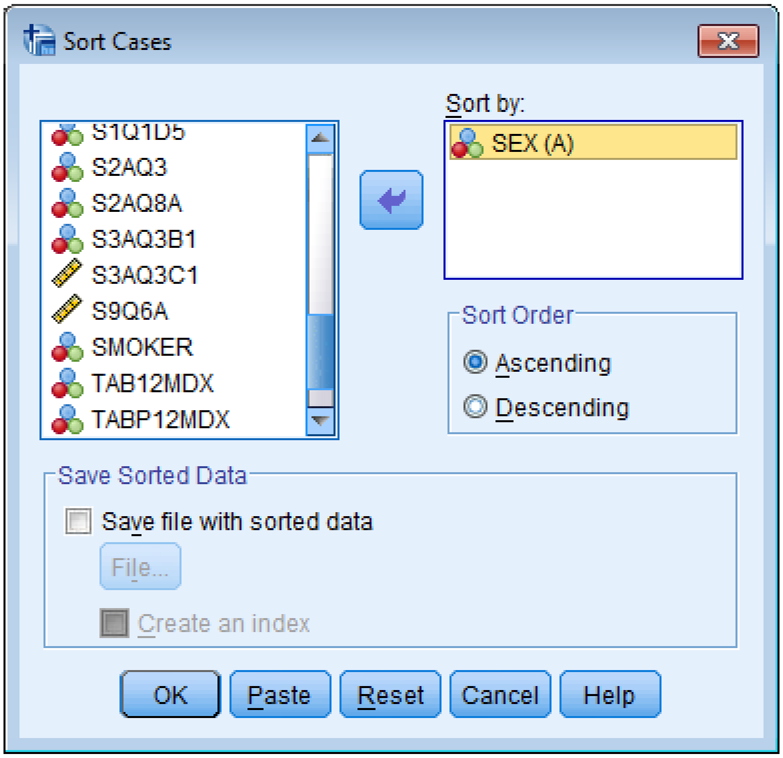
Let's return to our ANOVA example in which we asked 'Is major depression associated with a smoking quantity among current young adult smokers'? Since we are testing for moderation we need to include a third variable of Sex to ask the question for each population sub-group or each level of our third variable. In hypothesis testing terms, "Are the mean number of cigarettes smoked per month equal or not equal for those individuals with and without major depression?" and "Does this association differ by sex? That is, does it differ for males vs. females?"

The explanatory variable here is categorical with two levels, MAJORDEPLIFE, that is, the presence or absence of major depression. The response variable of NUMCIGMO_EST, smoking quantity, measured by the number of cigarettes smoked per month ranges from 1 to 2940. The third variable SEX of the participant where 1=Male and 2=Female.

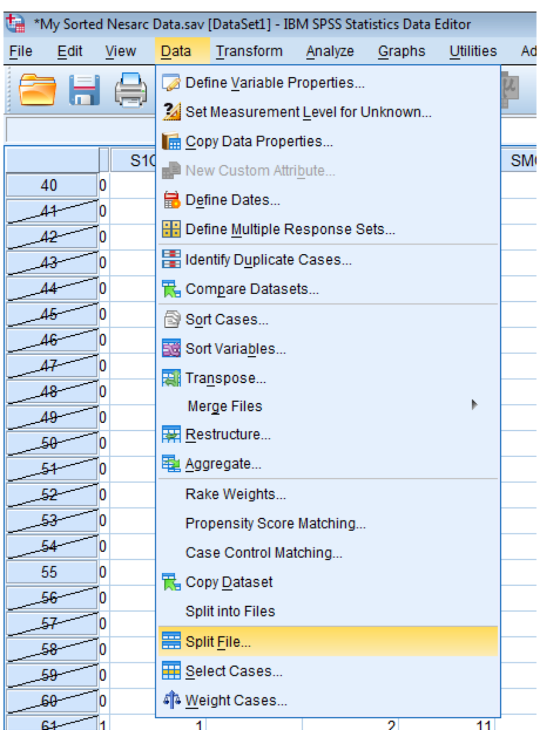
1. Go to **Data > Sort Cases**.



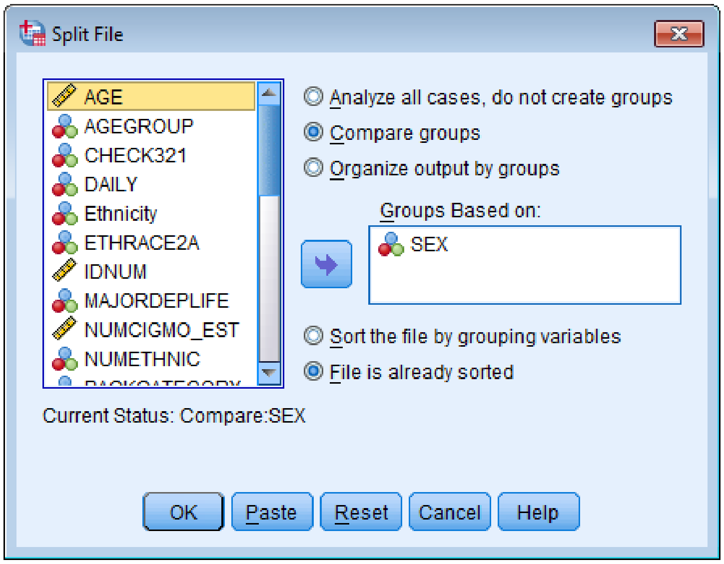
2. Select the name of the third variable (i.e., SEX) from the window on the left and move it to the **Sort by:** window on the right using the arrow. Click **OK**.



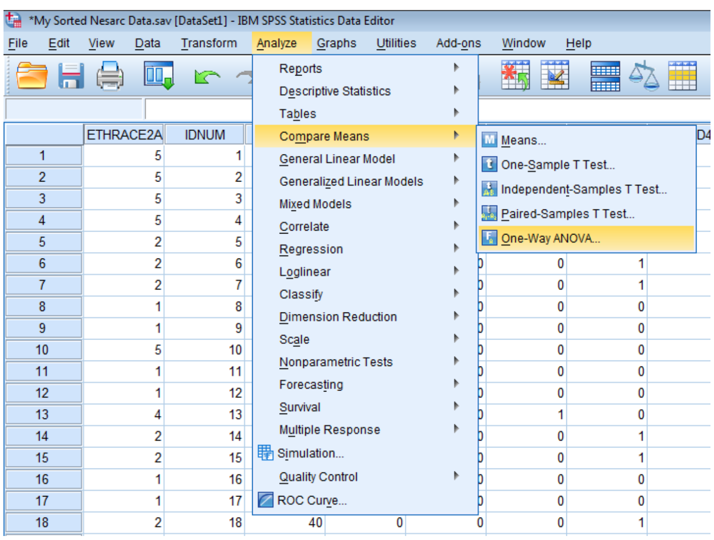
3. Go to **Data > Split File**.



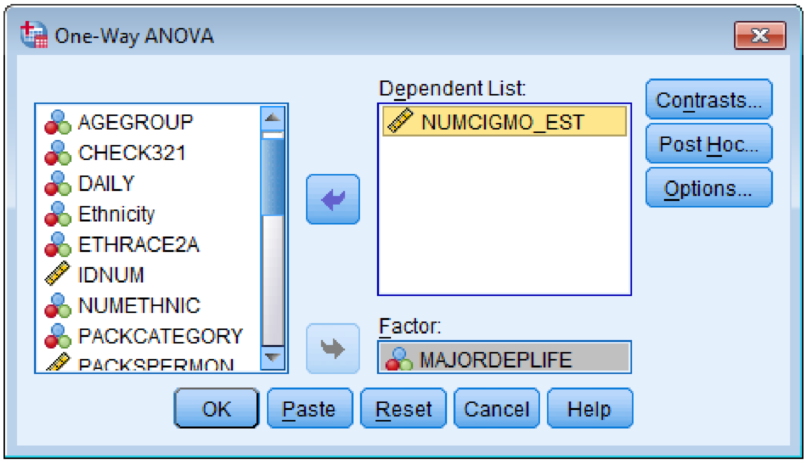
4. Using the arrow move the third variable from the window on the left to the **Groups Based on:** window on the right. Click **Compare groups** and **File is already sorted**, then **OK**.



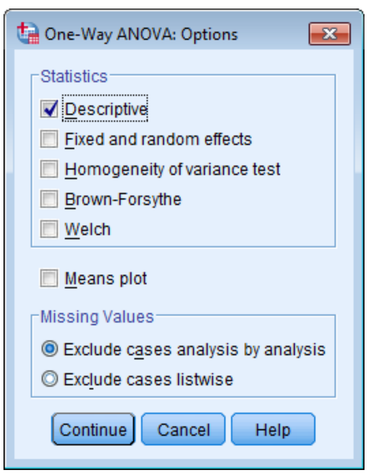
5. Go to **Analyze > Compare Means > One-Way ANOVA…**



6. From the variables list on the left, use the arrows to put your categorical explanatory variable, **MAJORDEPLIFE,** in the **Factor:** window and your quantitative response variable, **NUMCIGMO_EST,** in the **Dependent List:** window.



7. Click the **Options...** button and check the box by **Descriptive**. Click **Continue** then **OK**.



The first table in our output is the **Descriptives** table that shows us different values for each level of the categorical explanatory variable. It is slightly different from what we saw when we ran ANOVA in that now we have a set of descriptives for Male and for Female. That is, one set for each level of our third variable of Sex.

Descriptives

number of cigarettes smoked per month

Sex of Participant		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Male	Absence of Major Depression	689	336.0377	300.61346	11.45246	313.5518	358.5237	1.00	2940.00
	Presence of Major Depression	163	373.8405	321.65062	25.19362	324.0903	423.5907	1.00	2400.00
	Total	852	343.2700	304.91902	10.44635	322.7663	363.7736	1.00	2940.00
Female	Absence of Major Depression	564	284.4965	221.41235	9.32314	266.1841	302.8088	1.00	1200.00
	Presence of Major Depression	281	322.5427	266.18200	15.87909	291.2852	353.8003	1.00	1800.00
	Total	845	297.1485	237.75829	8.17913	281.0947	313.2024	1.00	1800.00

The first column, **N**, is the number of participants for each level and then total number of participants data that was analyzed with ANOVA. So here we see our categorical explanatory variable MAJORDEPLIFE has two levels with each categories label. We can see that 852 male and 845 female observations were included in the analysis.

The **Mean** column shows that male young adult smokers without major depression, as indicated by a value of zero, smoke an average of 336 cigarettes per month, and that males with major depression, indicated by a value of 1, smoke on average 374 cigarettes per month. Female young adult smokers without major depression, as indicated by a value of zero, smoke an average of 285 cigarettes per month, and that females with major depression, indicated by a value of 1, smoke on average 323 cigarettes per month. Overall we see that males whether they have absence or presence of major depression smoking on average more than females with or without major depression.

The second table in our output is the **ANOVA** table. Directly below the table title you will see the variable label for your quantitative response variable **NUMCIGMO_EST**. It is slightly different from what we saw when we ran ANOVA in that now we have an F statistics and p-value for Male and for Female. That is, one set for each level of our third variable of Sex.

ANOVA

number of cigarettes smoked per month

Sex of Participant		Sum of Squares	df	Mean Square	F	Sig.
Male	Between Groups	188371.039	1	188371.039	2.028	.155
	Within Groups	78933870.87	850	92863.377		
	Total	79122241.91	851			
Female	Between Groups	271489.130	1	271489.130	4.824	.028
	Within Groups	47438989.48	843	56274.009		
	Total	47710478.61	844			

You should recall in our original ANOVA the calculated F statistic found in the **F** column in this output, is 3.55. The significance, probability, or p value, associated with this F statistic, is labeled **Sig.**, and as you can see, the p value is .060, just over our p .05 cut point.

Our calculated F statistic for males found in the top row of the **F** column in this output, is 2.028. The significance, probability, or p-value, associated with this F statistic, is labeled **Sig.**, and as you can see, the p value is .155, over our .05 cut point.

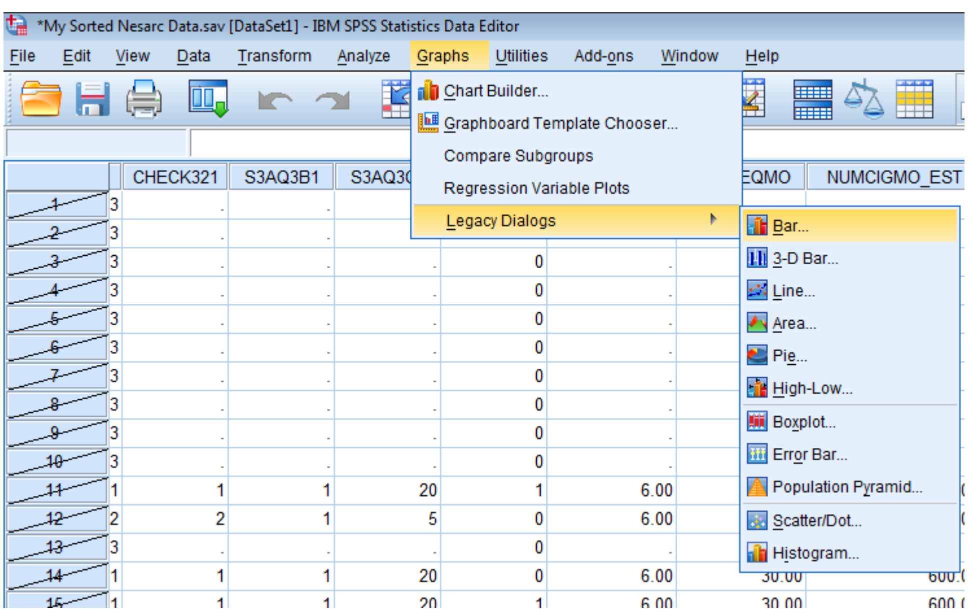
Our calculated F statistic for females found in the bottom row of the **F** column in this output, is 4.824. The significance, probability, or p-value, associated with this F statistic, is labeled **Sig.**, and as you can see, the p value is .028, below the p .05 cut point.

As you can see, the relationship between major life depression and smoking quantity depends on your sex. For males we did not find a statistically significant relationship, that is, no significant difference in amount of cigarettes smoked when comparing those with and without major depression. For females we did find a statistically significant relationship, that is, a significant difference in amount of cigarettes smoked when comparing those with and without major depression. More specifically, females with major depression smoked significantly more cigarettes than those without major depression. Thus, we can say there is a significant statistical interaction between the sex and major depression when predicting smoking quantity. In other words, sex (our third variable) moderates the association between major depression and smoking quantity.

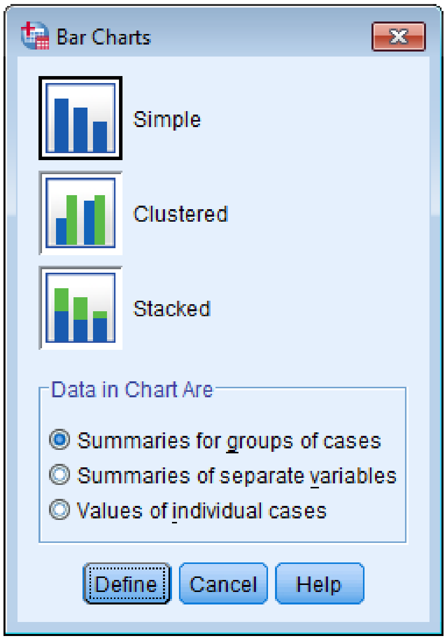
Section 12.2: Graphing Moderation with ANOVA

We have already sorted the data and split the data, therefore, if we complete the appropriate steps for a bivariate graph for a categorical explanatory variable and a quantitative response variable we will get two graphs, one for males and one for females.

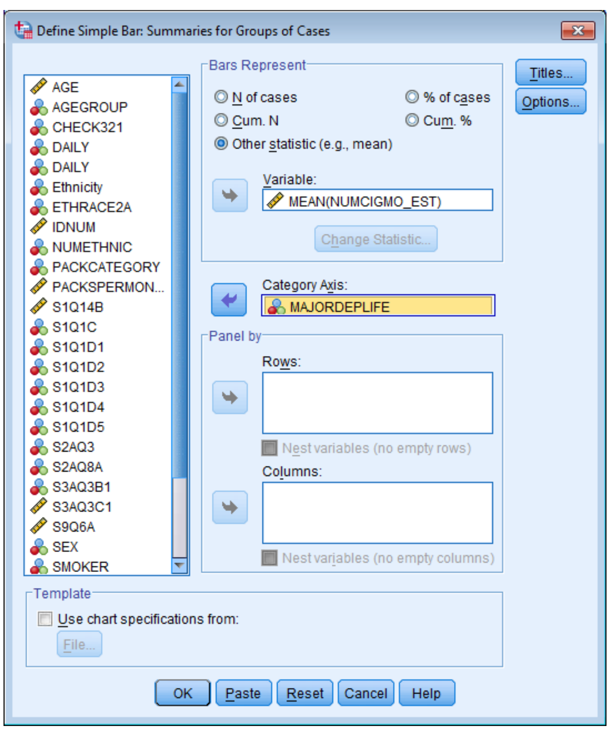
1. Click **Graphs > Legacy Dialogs**.



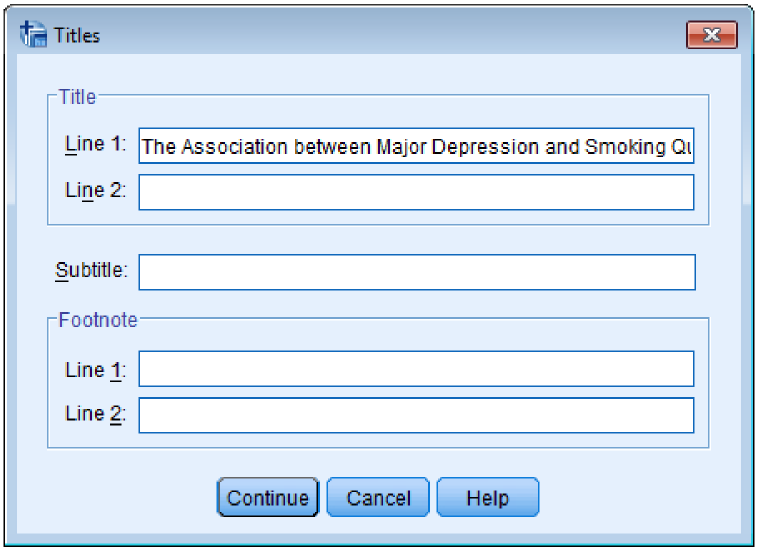
2. Click on the graph left of **Simple > Define.**



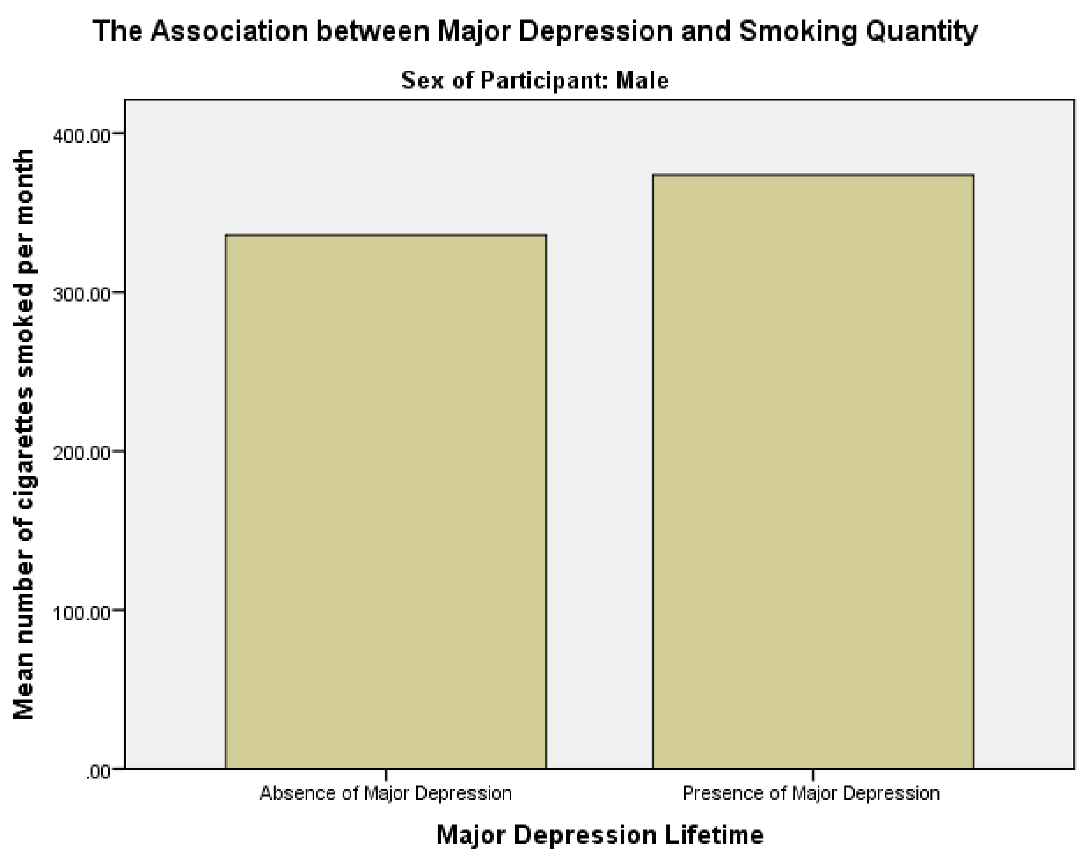
3. In the top middle **Bars Represent** box Click **Other statistic (e.g., mean)**. Using the arrow directly below, move the **Quantitative Response Variable** from the left window to below **Variable:**. Use the next arrow down to move the Categorical Explanatory Variable to the **Category Axis:** window. Click **Titles…** in the upper right corner.

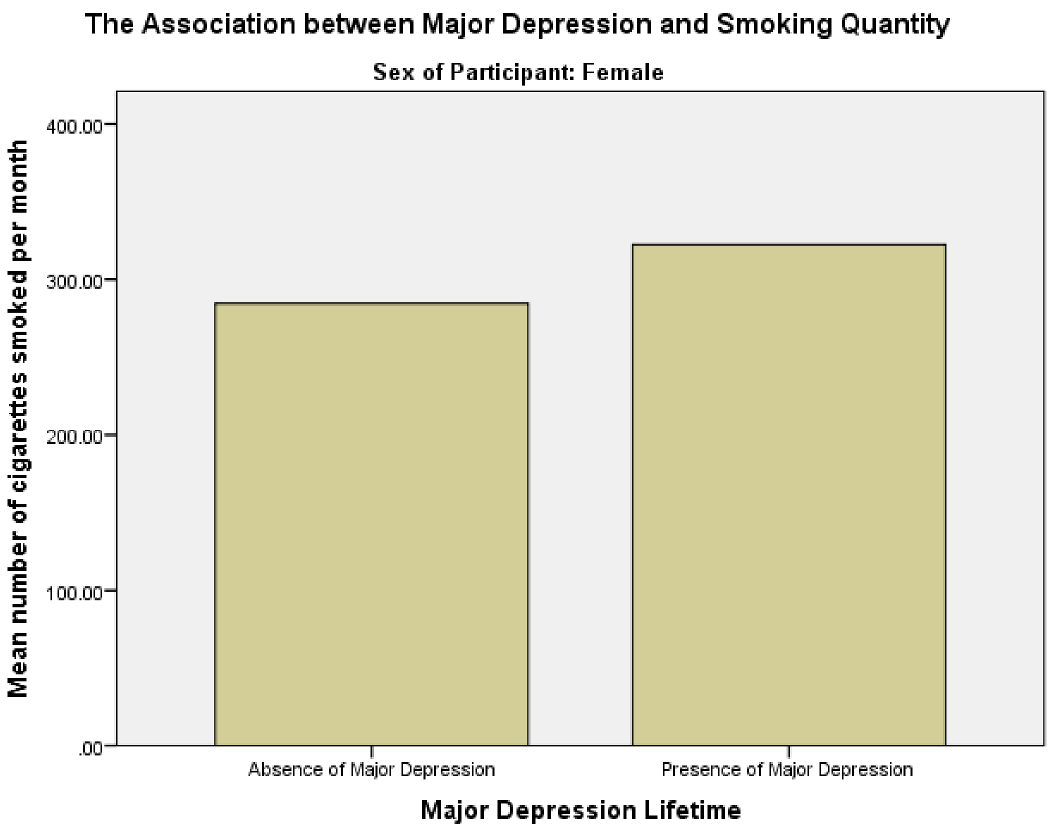


4. In the top window, **Line 1:**, appropriately title your graph. Click **Continue > OK**.



The output below shows two graphs, one for Male and one for Female.





Using the Chart Editor previously explained make the appropriate changes to your graphs.

Suppose we did not evaluate sex as a possible moderator and instead focused only on the association between major depression and smoking quantity for the entire population. Based on this p-value's and graphs, obviously, we would've incorrectly concluded that participants with major depression smoke significantly more than those that do not have major depression. As we now know, that is true only if we're looking at females.

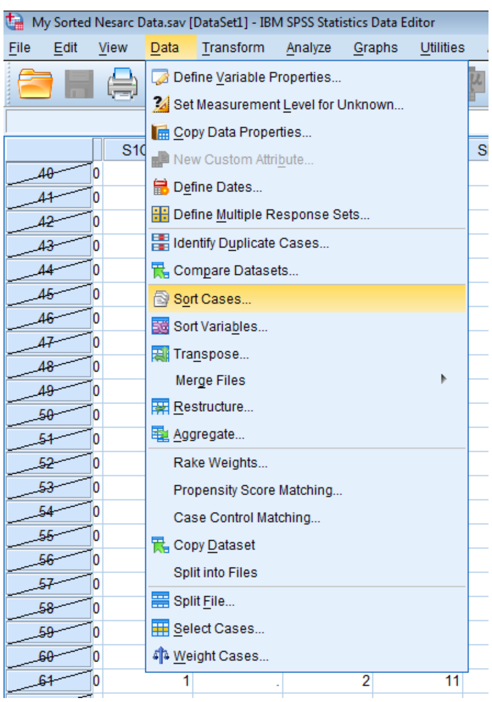
Section 12.3: Moderation with Chi-Square

Now let's evaluate third variables as potential moderators in the context of chi-square tests of independence. Using the NESARC data and asking the question, is smoking associated with nicotine dependence?

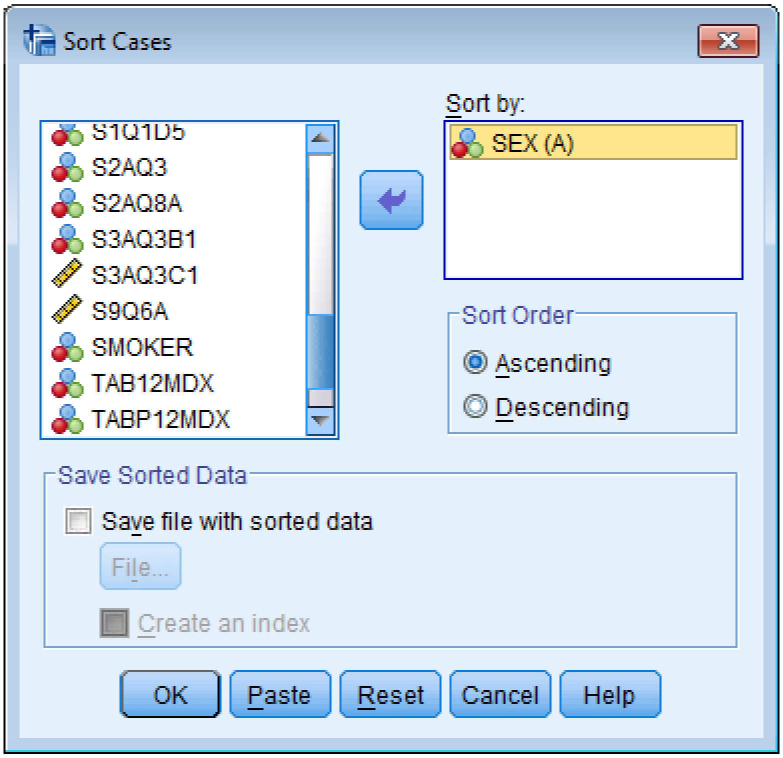
Let's return to our Chi-Square example using the NESARC data in which we asked 'Is smoking frequency associated with nicotine dependence among current young adult smokers'? Since we are testing for moderation we need to include a third variable of Sex to ask the question for each population sub-group or each level of our third variable. In hypothesis testing terms, Or in hypothesis testing terms, "Is smoking frequency and nicotine dependence independent or dependent?; That is, are the rates of nicotine dependence equal or not equal among individuals from my different smoking frequency categories. And is the answer to this question similar or different for males and females?

For this analysis, we're going to use the same categorical explanatory variable and categorical response variable we used when learning about Chi-Square in SPSS tutorial 10. If you recall the categorical explanatory variable has 6 levels, the number of days smoked per month, which we called USFREQMO, with the following categorical values: smoking approximately 1 day/month, 2.5 days/ month, 5 days/month, 14 days/month, 22 days/month and 30 days/month. The response variable, called TAB12MDX, is categorical with 2 levels--the presence or absence of nicotine dependence in the past 12 months. The third variable SEX of the participant where 1=Male and 2=Female.

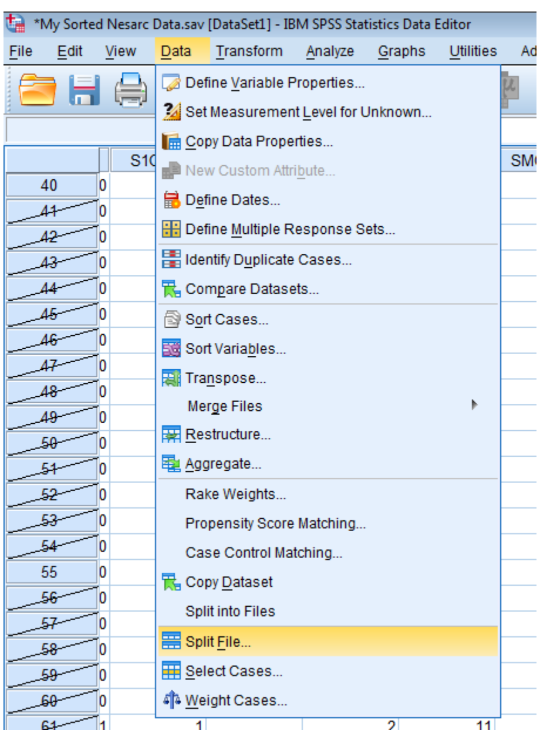
1. Go to **Data > Sort Cases**.



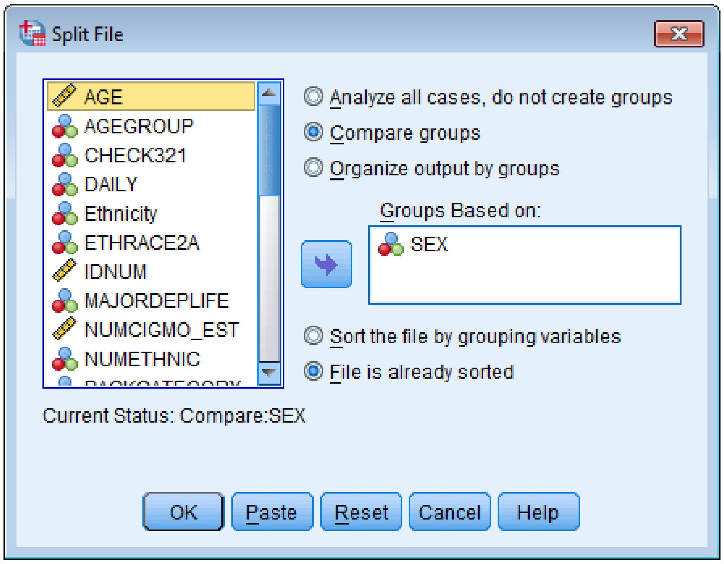
2. Select the name of the third variable (i.e., SEX) from the window on the left and move it to the **Sort by:** window on the right using the arrow. Click **OK**.



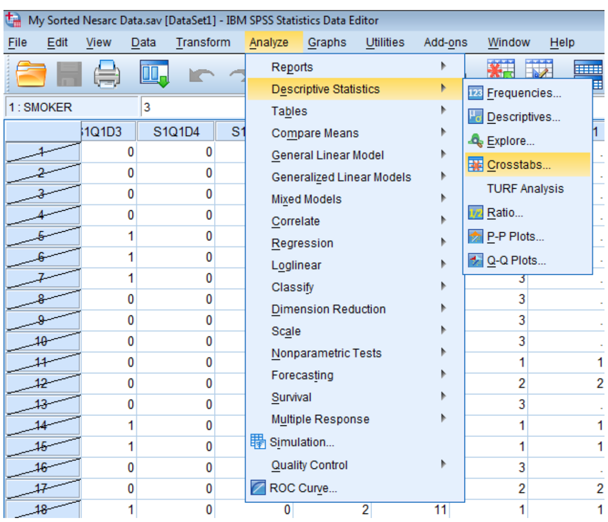
3. Go to **Data > Split File**.



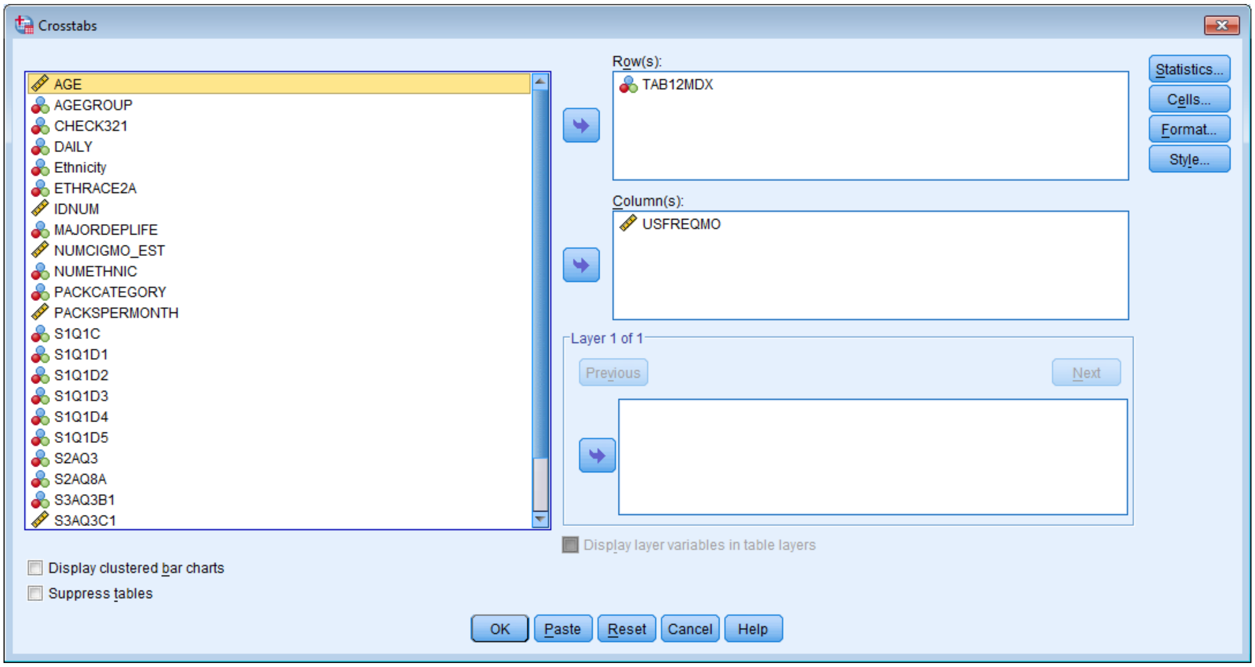
4. Using the arrow move the third variable from the window on the left to the **Groups Based on:** window on the right. Click **Compare groups** and **File is already sorted**, then **OK**.



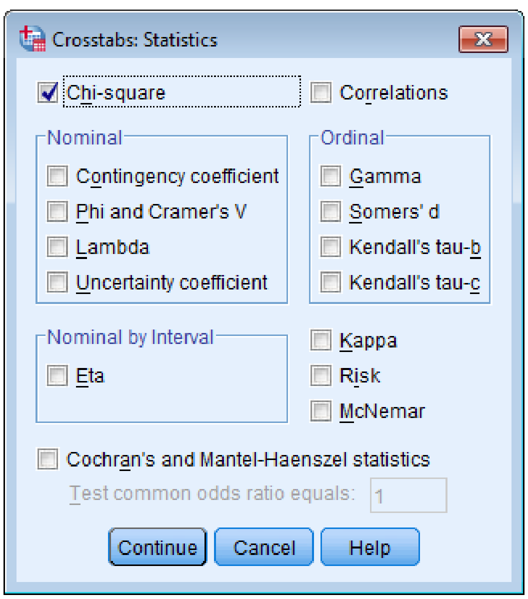
5. Click **Analyze > Descriptive Statistics > Crosstabs**.



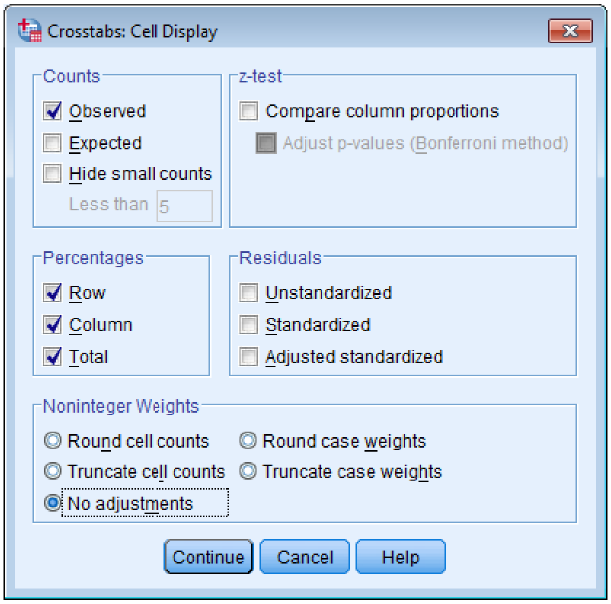
6. Using the arrows move your categorical explanatory variable to the window labeled **Column(s):** and your categorical response variable to the window labeled **Row(s):** then click **Statistics**.



7. Check **Chi-square** then click **Continue**.



8. Click **Cells…** Ensure that **Observed** is checked. Check **Row, Column**, and **Total** in the Percentages box. In the bottom box click **No adjustments**, then click **Continue > OK**.



The first table, **Case Processing Summary** table, is similar to what we saw when we previously used Chi-Square in the SPSS output showing you the number of participants used in the analysis. The difference is that it is divided by each level of the third variable (i.e., Sex). There is the **Valid** column showing the number of participants, the **Missing** data column of those not used in the analysis, and the **Total** number of participants per level of the third variable.

Case Processing Summary

		Cases					
		Valid		Missing		Total	
Sex of Participant		N	Percent	N	Percent	N	Percent
Male	Nicotine Dependence Past 12 Months * Number of Days Smoked in a Usual Month	854	99.8%	2	0.2%	856	100.0%
Female	Nicotine Dependence Past 12 Months * Number of Days Smoked in a Usual Month	849	99.9%	1	0.1%	850	100.0%

The table below should look familiar to the one we saw when running a Chi-Square. It shows the response variable by the explanatory variable with an additional breakdown by our third variable that is Male and Female. Remember that this table is known as the cross tabs or cross tabulation table where you can see a myriad of numbers and percentages with such labels as **Count** (i.e., frequency), **% within Nicotine Dependence** (i.e., row percentage for response variable), **% within Number of Days Smoked** (i.e., column percentage for explanatory variable)., and **% of Total** for each level of the third variable.

Remember we are not interested in the column percentages for those observations without nicotine dependence, indicated with a dummy code of 0 (i.e., Absence of). Instead, we're interested in describing the <u>presence of</u> nicotine dependence within the smoking frequency groups that is these column percentages (i.e., % within Number of Days Smoked) circled with blue.

Nicotine Dependence Past 12 Months * Number of Days Smoked in a Usual Month Crosstabulation

Sex of Participant				Number of Days Smoked in a Usual Month						Total
				1.00	2.50	6.00	14.00	22.00	30.00	
Male	Nicotine Dependence Past 12 Months	Absence of	Count	28_{a}	25_{a}	39_{a}	35_{a}	$24_{a, b}$	264_{b}	415
			% within Nicotine Dependence Past 12 Months	6.7%	6.0%	9.4%	8.4%	5.8%	63.6%	100.0%
			% within Number of Days Smoked in a Usual Month	90.3%	75.8%	75.0%	63.6%	64.9%	40.9%	48.6%
			% of Total	3.3%	2.9%	4.6%	4.1%	2.8%	30.9%	48.6%
		Presence of	Count	3_{a}	8_{a}	13_{a}	20_{a}	$13_{a, b}$	382_{b}	439
			% within Nicotine Dependence Past 12 Months	0.7%	1.8%	3.0%	4.6%	3.0%	87.0%	100.0%
			% within Number of Days Smoked in a Usual Month	9.7%	24.2%	25.0%	36.4%	35.1%	59.1%	51.4%
			% of Total	0.4%	0.9%	1.5%	2.3%	1.5%	44.7%	51.4%
	Total		Count	31	33	52	55	37	646	854
			% within Nicotine Dependence Past 12 Months	3.6%	3.9%	6.1%	6.4%	4.3%	75.6%	100.0%
			% within Number of Days Smoked in a Usual Month	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
			% of Total	3.6%	3.9%	6.1%	6.4%	4.3%	75.6%	100.0%
Female	Nicotine Dependence Past 12 Months	Absence of	Count	36_{a}	$28_{a, b}$	$30_{a, b}$	$24_{a, b}$	$17_{b, c}$	257_{c}	392
			% within Nicotine Dependence Past 12 Months	9.2%	7.1%	7.7%	6.1%	4.3%	65.6%	100.0%
			% within Number of Days Smoked in a Usual Month	90.0%	87.5%	83.3%	66.7%	54.8%	38.1%	46.2%
			% of Total	4.2%	3.3%	3.5%	2.8%	2.0%	30.3%	46.2%
		Presence of	Count	4_{a}	$4_{a, b}$	$6_{a, b}$	$12_{a, b}$	$14_{b, c}$	417_{c}	457
			% within Nicotine Dependence Past 12 Months	0.9%	0.9%	1.3%	2.6%	3.1%	91.2%	100.0%
			% within Number of Days Smoked in a Usual Month	10.0%	12.5%	16.7%	33.3%	45.2%	61.9%	53.8%
			% of Total	0.5%	0.5%	0.7%	1.4%	1.6%	49.1%	53.8%
	Total		Count	40	32	36	36	31	674	849
			% within Nicotine Dependence Past 12 Months	4.7%	3.8%	4.2%	4.2%	3.7%	79.4%	100.0%
			% within Number of Days Smoked in a Usual Month	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
			% of Total	4.7%	3.8%	4.2%	4.2%	3.7%	79.4%	100.0%

Each subscript letter denotes a subset of Number of Days Smoked in a Usual Month categories whose column proportions do not differ significantly from each other at the .05 level.

The **Chi-Square Tests** table below shows the calculation of the chi square statistic along with the associated p-value for each level of our third variable. You will only use the **Pearson Chi-Square** row in this table. Our p-value of .0001 clearly tells us that smoking and nicotine dependence are associated.

Chi-Square Tests

Sex of Participant		Value	df	Asymp. Sig. (2-sided)
Male	Pearson Chi-Square	70.215[a]	5	.000
	Likelihood Ratio	74.525	5	.000
	Linear-by-Linear Association	66.967	1	.000
	N of Valid Cases	854		
Female	Pearson Chi-Square	97.467[b]	5	.000
	Likelihood Ratio	104.894	5	.000
	Linear-by-Linear Association	97.239	1	.000
	N of Valid Cases	849		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 15.06.

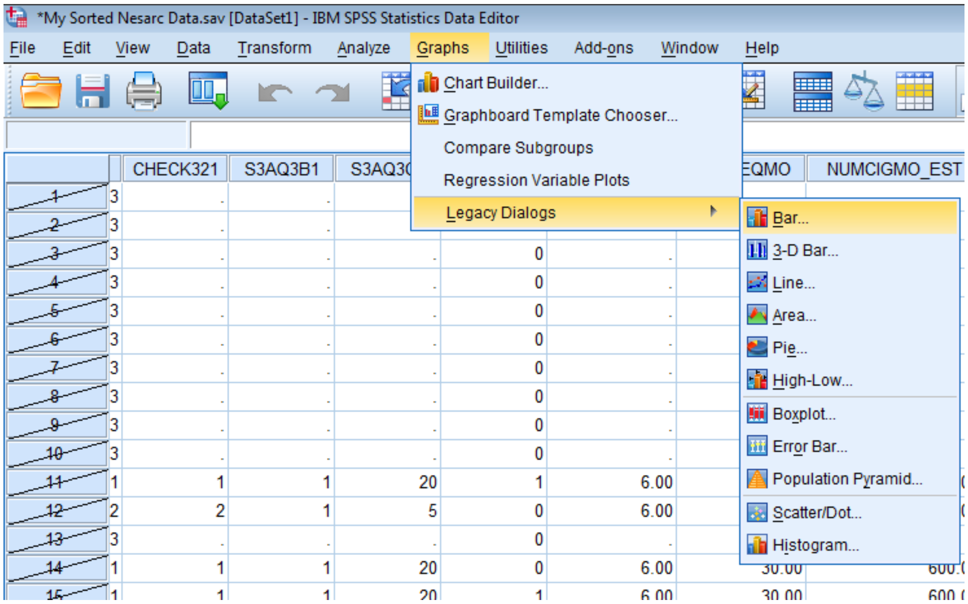
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.31.

We can see that for both male and female young adult smokers a large Pearson Chi-Square value and p-value below .05. In this case we would say one's sex does not moderate the relationship, or significant statistical interaction, between number of days smoked in a month and nicotine dependence.

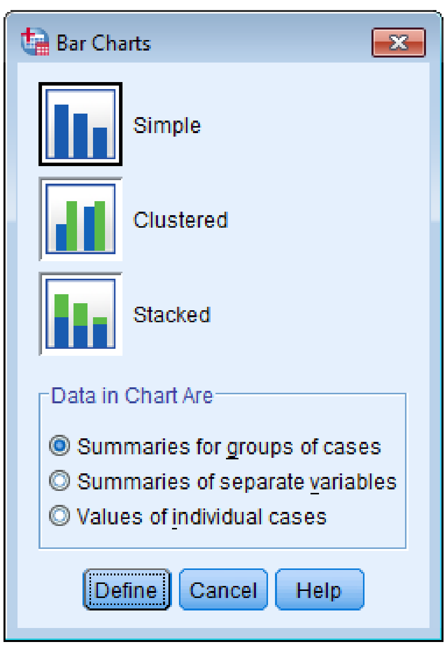
Section 12.4: Graphing Moderation with Chi-Square

We have already sorted the data and split the data, therefore, if we complete the appropriate steps for a bivariate graph for a categorical explanatory variable and a categorical response variable we will get two graphs, one for males and one for females.

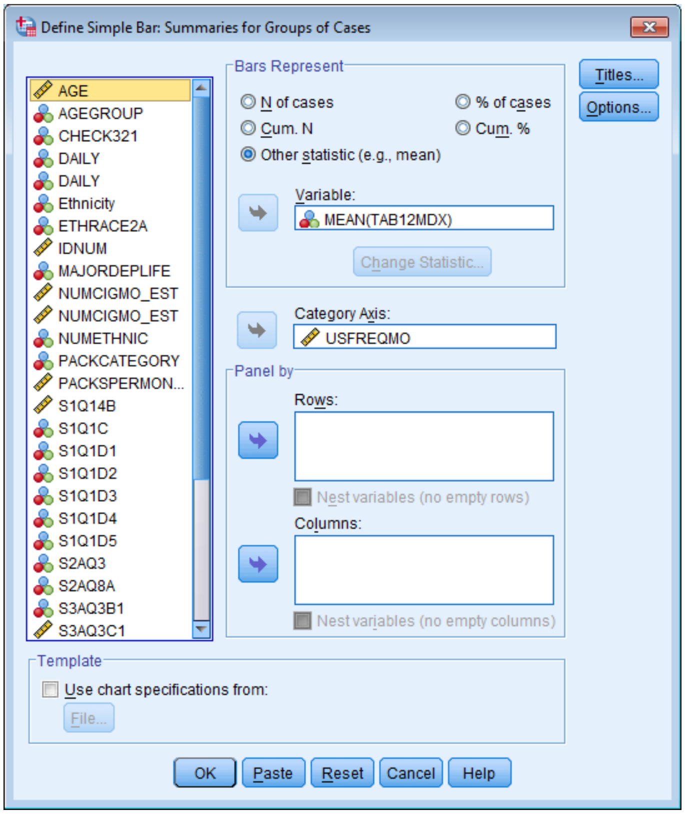
1. Click **Graphs > Legacy Dialogs**.



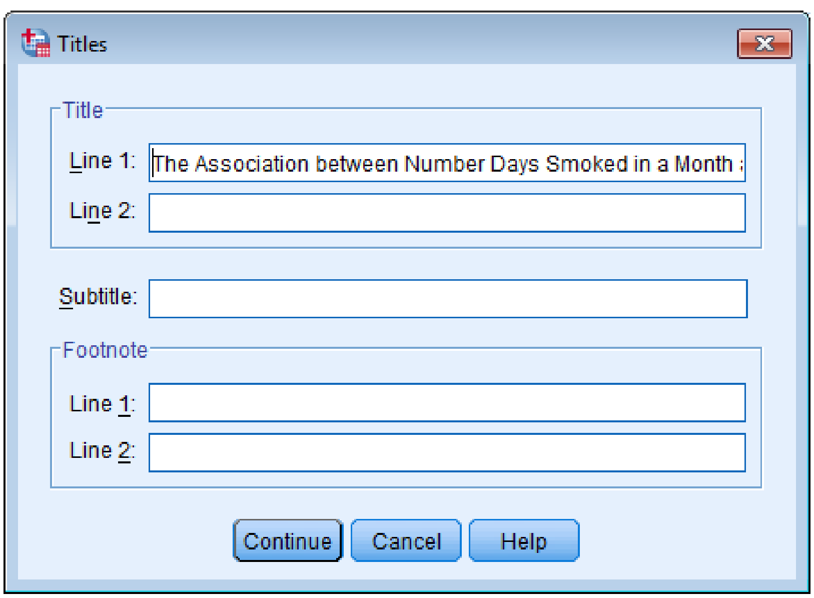
2. Click on the graph left of **Simple > Define.**



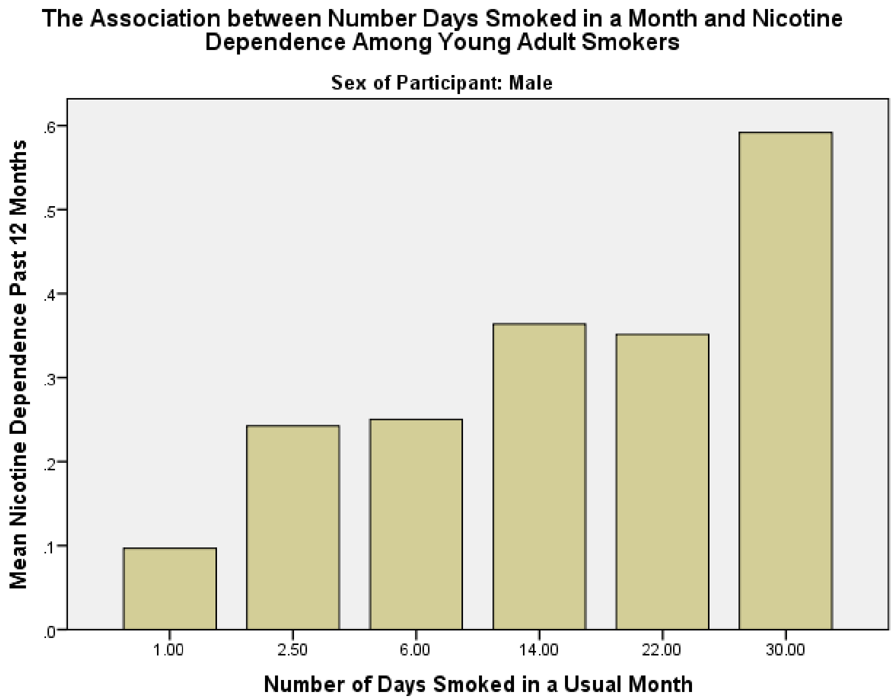
3. In the top middle **Bars Represent** box Click **Other statistic (e.g., mean)**. Using the arrow directly below, move the **Categorical Response Variable** from the left window to below **Variable:**. Use the next arrow down to move the Categorical Explanatory Variable to the **Category Axis:** window. Click **Titles…** in the upper right corner.

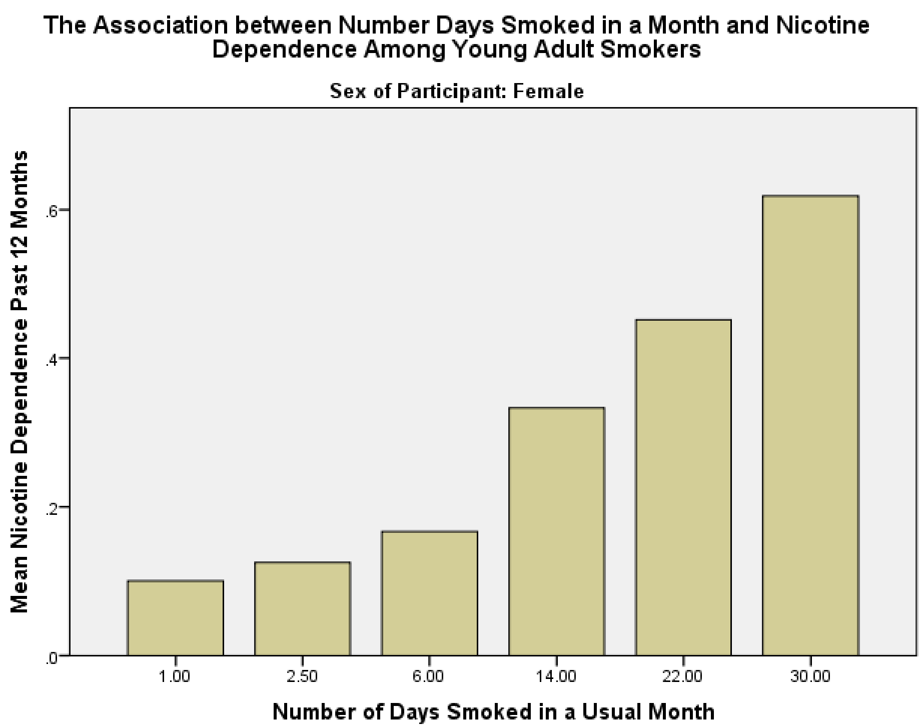


4. In the top window, **Line 1:**, appropriately title your graph. Click **Continue > OK**.



The output below shows two graphs, one for Male and one for Female.



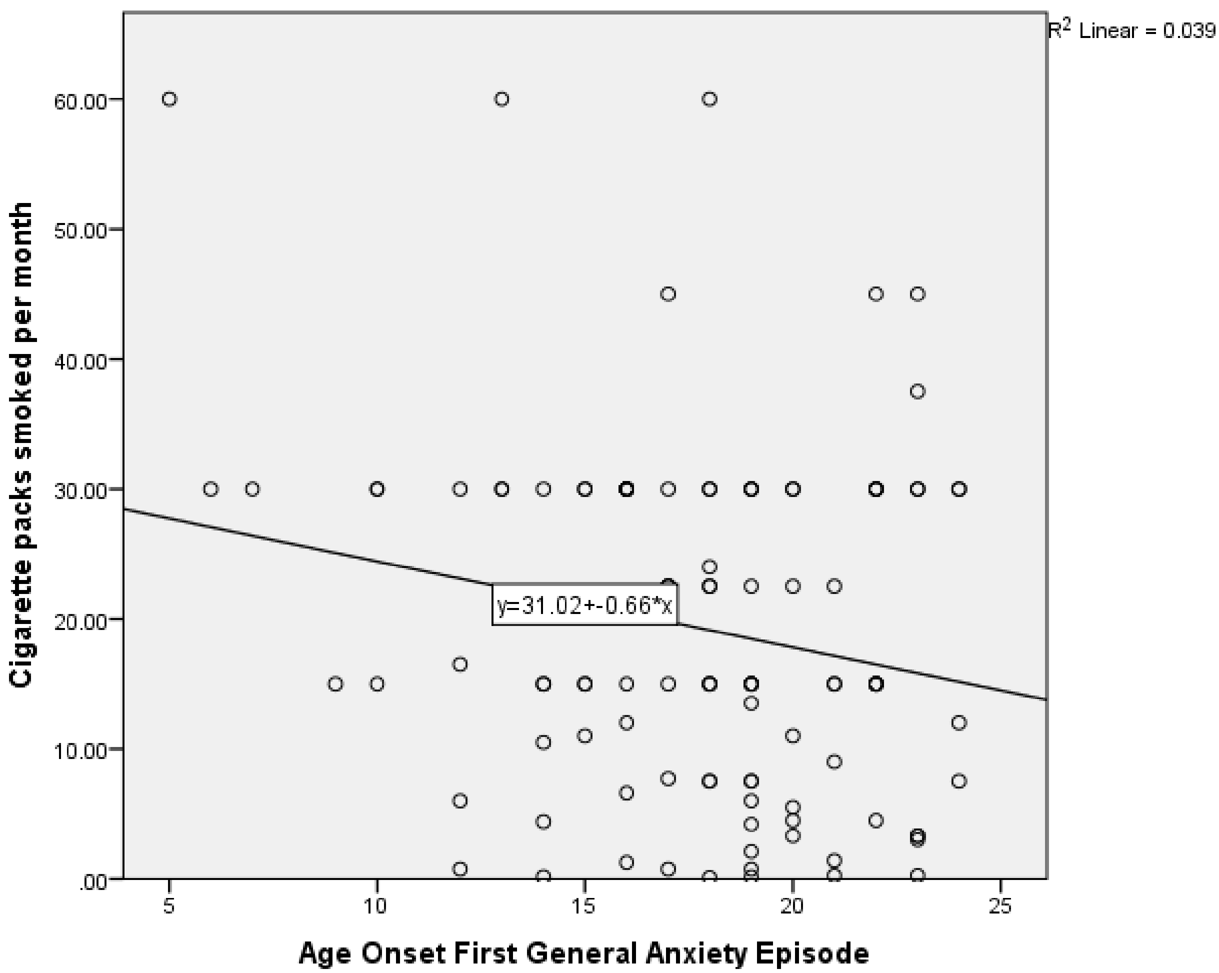


Using the Chart Editor previously explained make the appropriate changes to your graphs.

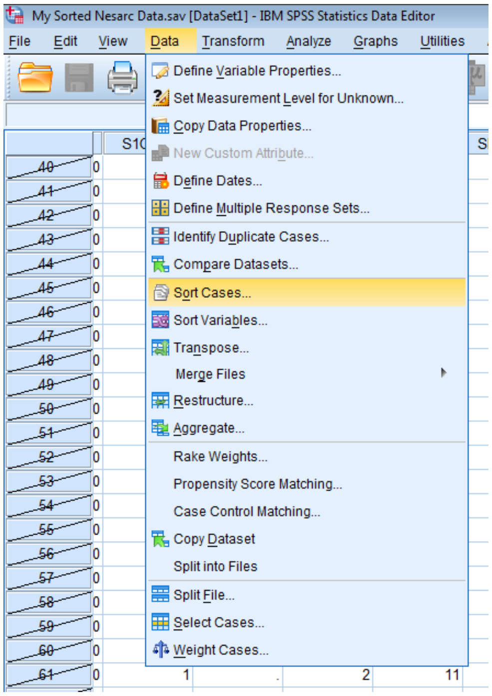
Based on this p-value's and graphs, we can conclude that sex does not moderate the relationship between number of days smoked in a month and nicotine dependence. For both male and female young adult smokers, a higher level of smoking behavior is associated with higher rates of nicotine dependence.

Section 12.5: Moderation with Correlation

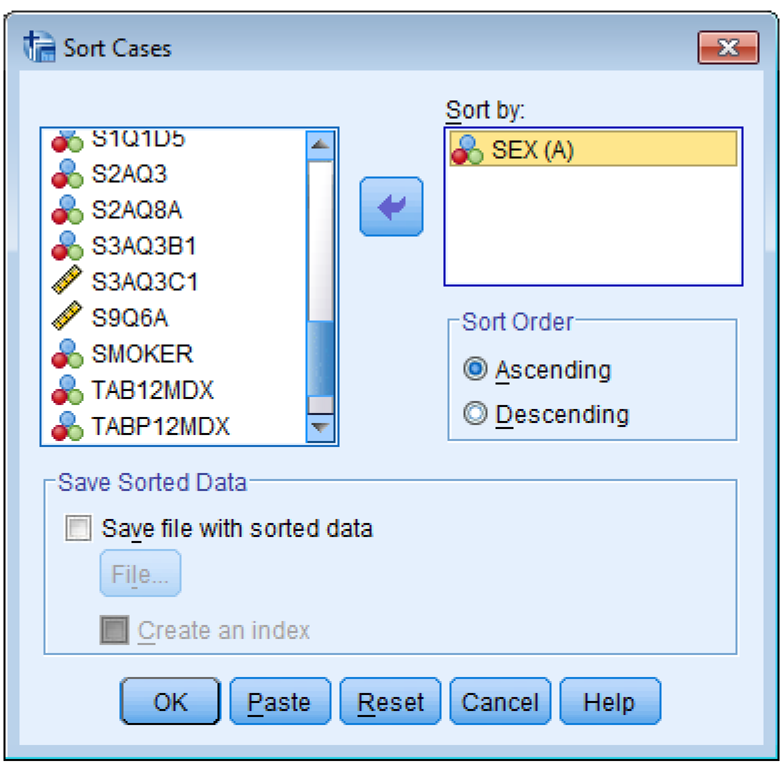
Now let's evaluate third variables as potential moderators in the context of correlation. Using the NESARC data and asking the question 'Is there an Association between Age Onset First General Anxiety Episode (S9Q6A) by Number of Packs Smoked Per Month (PACKSPERMONTH) across Sex in Young Adult Smokers. If you will remember we found a correlation coefficient age of onset of first anxiety disorder and number of packs of cigarettes smoked per month of approximately -.198 with a p-value of .040, which told us that the relationship is statistically significant. But might this relationship--this correlation--differ based on sex of the participant?



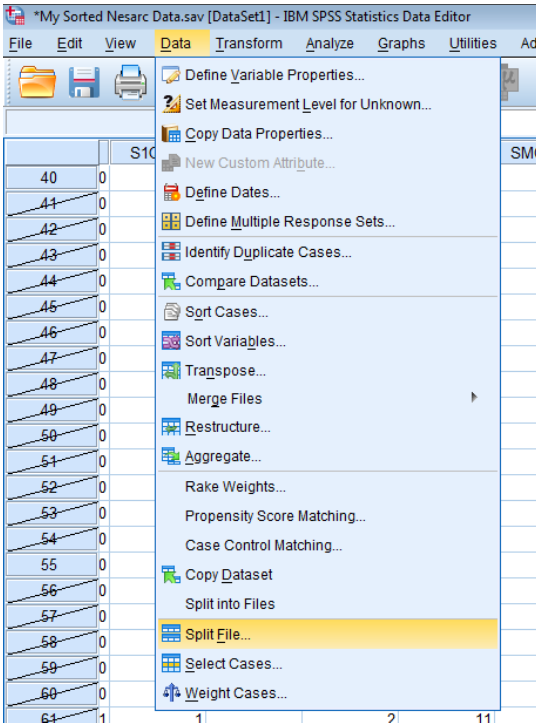
1. Go to **Data > Sort Cases**.



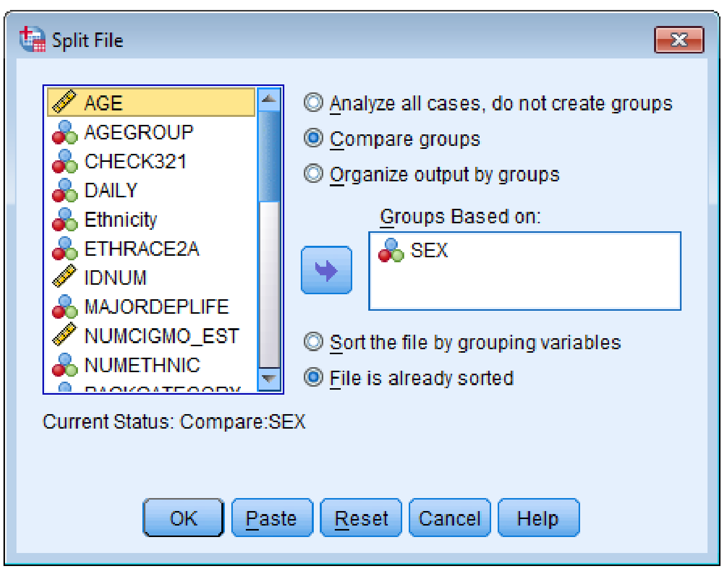
2. Select the name of the third variable (i.e., SEX) from the window on the left and move it to the **Sort by:** window on the right using the arrow. Click **OK**.



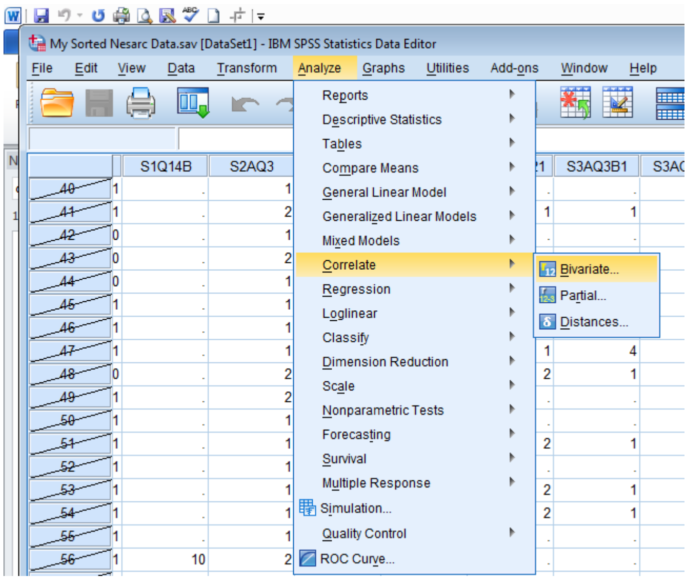
3. Go to **Data > Split File**.



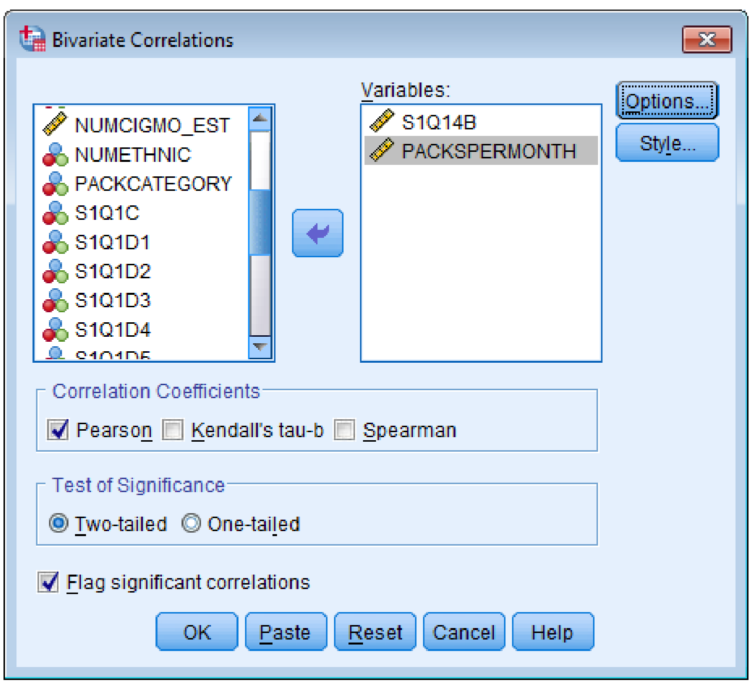
4. Using the arrow move the third variable from the window on the left to the **Groups Based on:** window on the right. Click **Compare groups** and **File is already sorted**, then **OK**.



5. Click **Analyze > Correlate > Bivariate**.



6. Select your Quantitative Explanatory Variable and Quantitative Response Variable from the left hand side using the arrow to move to the **Variables:** window. Click **OK**.



The SPSS output table below should look familiar to what we saw when we ran correlation. The difference is that it is broken down by each level of our third variable. To locate the correlation coefficients of interest and the associated p-values, we need to examine the Pearson Correlation Coefficient table here, and find the row and column where our two variables of interest intersect.

Correlations

Sex of Participant			Age Onset First General Anxiety Episode	Cigarette packs smoked per month
Male	Age Onset First General Anxiety Episode	Pearson Correlation	1	-.336*
		Sig. (2-tailed)		.028
		N	43	43
	Cigarette packs smoked per month	Pearson Correlation	-.336*	1
		Sig. (2-tailed)	.028	
		N	43	852
Female	Age Onset First General Anxiety Episode	Pearson Correlation	1	-.126
		Sig. (2-tailed)		.317
		N	65	65
	Cigarette packs smoked per month	Pearson Correlation	-.126	1
		Sig. (2-tailed)	.317	
		N	65	845

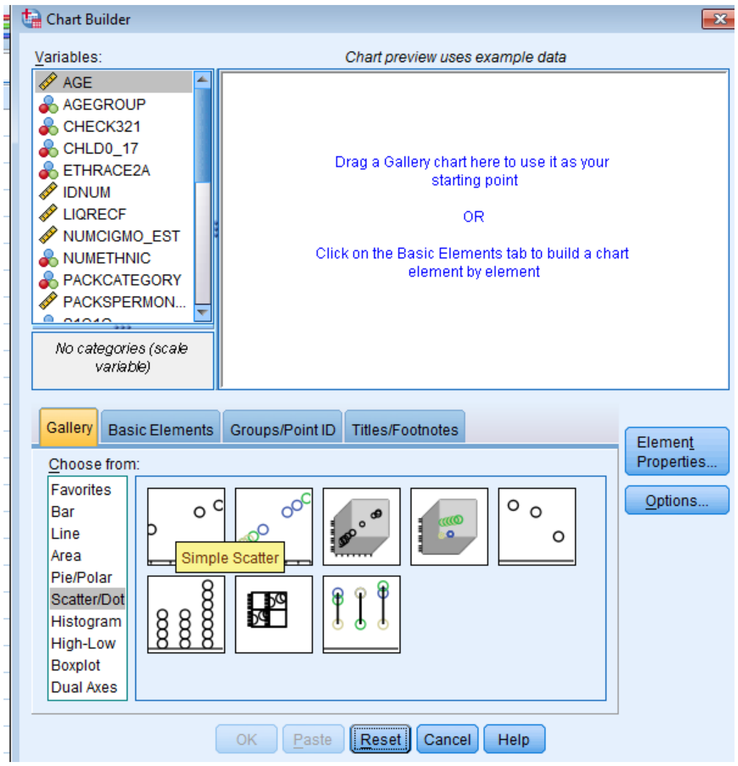
*. Correlation is significant at the 0.05 level (2-tailed).

If you will remember we found a correlation coefficient of approximately -.198 with a p-value of .040, which told us that the relationship is statistically significant. When we examine the correlation coefficient between age onset first general anxiety episode and cigarette packs smoked per month we see p-value of .028 and a correlation coefficient of -.336 in males. This is a statistically significant negative moderate relationship. For females we see a p-value of .317 and a correlation coefficient of -.126 which is a non-significant relationship.

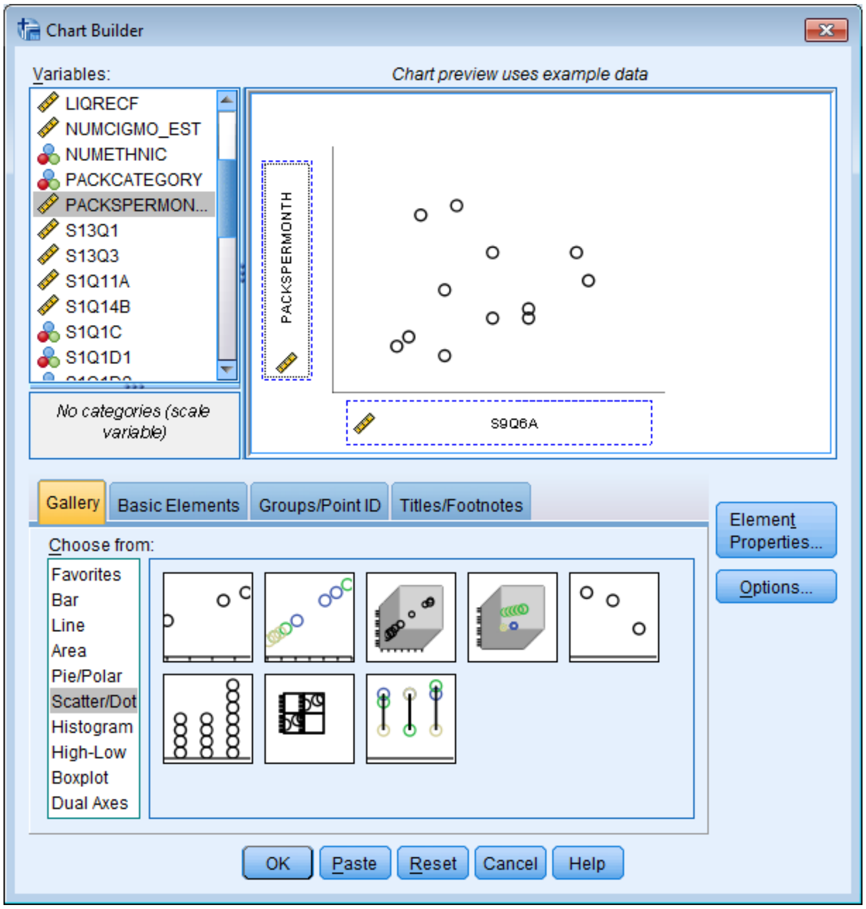
Section 12.6: Graphing Moderation with Correlation

When we graph these findings onto the associated scatterplots we are able to better visualize the significant and non-significant relationship.

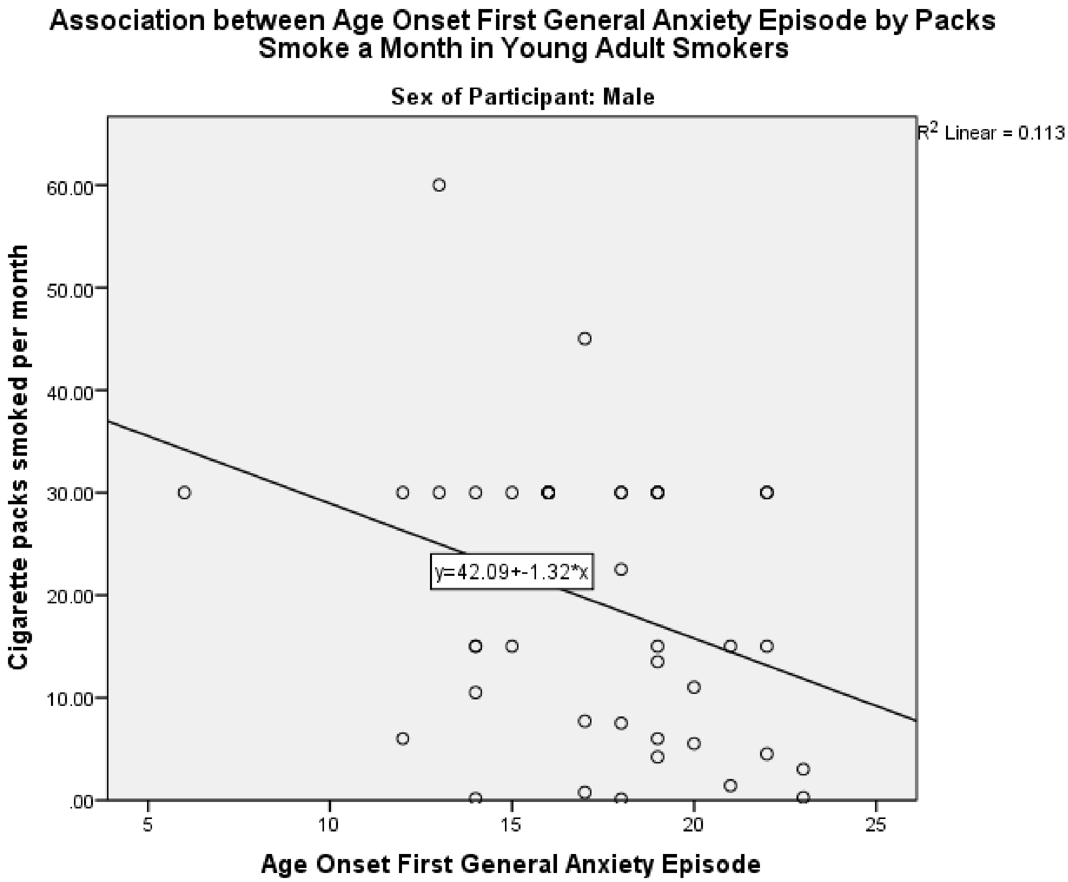
1. Click **Graphs > Chart Builder**. In the lower window under **Choose from:**, Click **Scatter/Dot**, drag the top left Simple Scatter up to the Chart preview window.

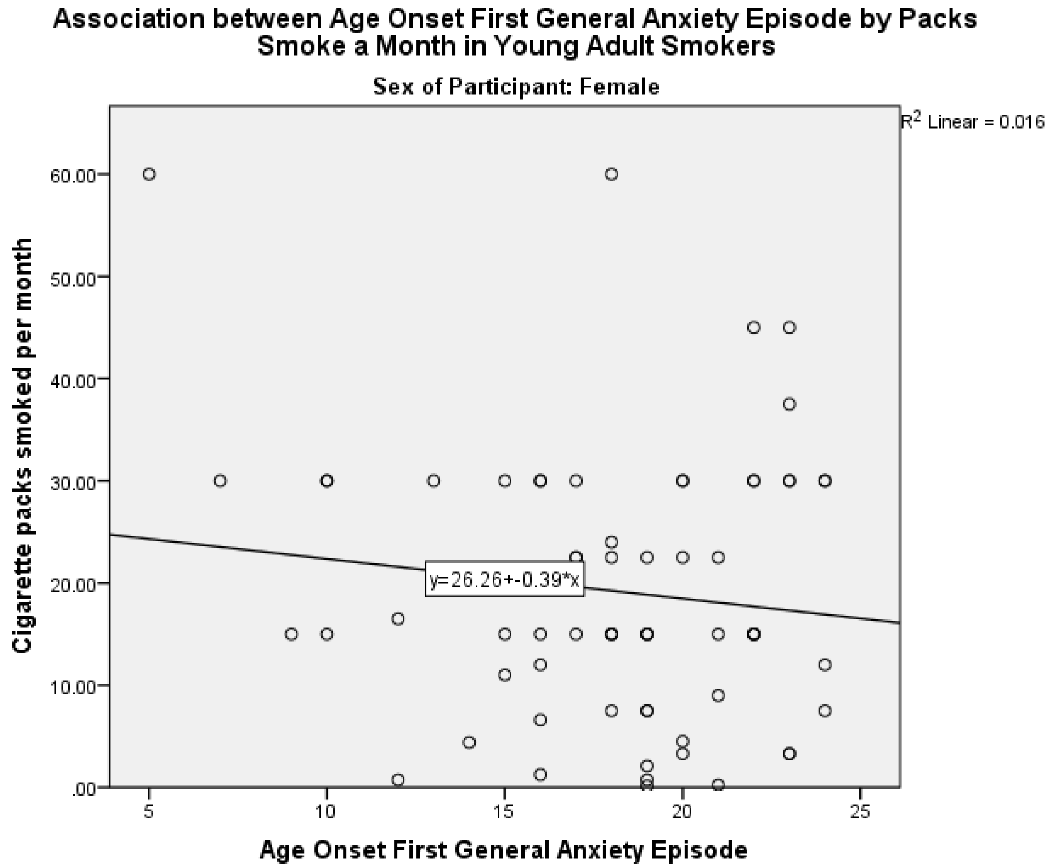


2. From the upper left **Variables:** window, drag your Quantitative Explanatory variable, S9Q6A, to the X-Axis and your Quantitative Response variable, PACKSPERMONTH, to the Y-Axis.



3. In the output window double click on the graph then use **Options** and the **Element Properties** to set a graph title and customize the graph. Estimating a line of best fit within each scatterplot shows the direction of each association.





Asking questions about statistical interactions can be an interesting way to explore your data and your associations of interest. This is not difficult to do using the skills you've acquired thus far. There are more advanced topics than we can cover here, such as multi-variance technique that can be very powerful. But even without these techniques we can still use bivariate inferential tools of ANOVA, Chi-Square, and Correlation to describe our sample, make inferences about the larger population, and really begin to understand what relationships these associations hold under what conditions or at what levels of our third variable.

Now that we've found associations can we assume that association implies causation? We'll answer that question next.