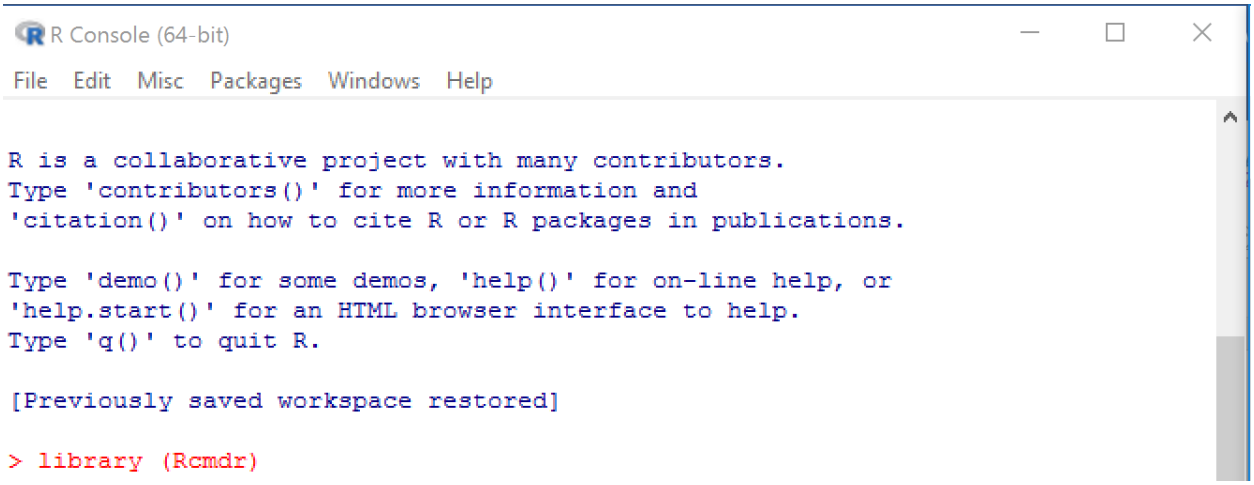
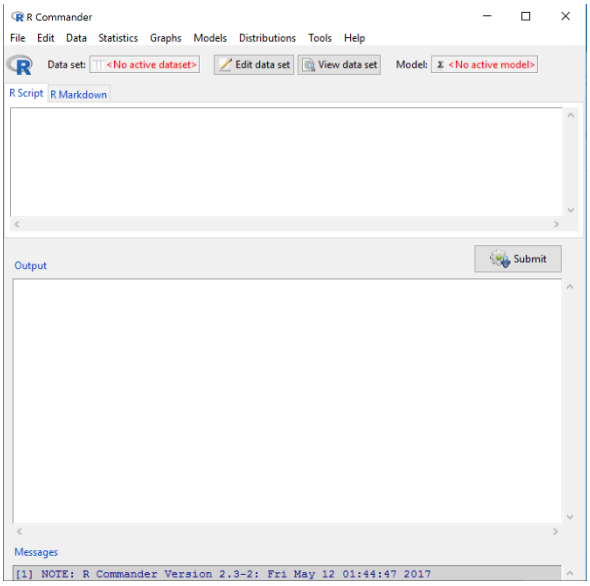
Create Your Own Adventure! R Commander (Rcmdr) & Statistics Worksheet

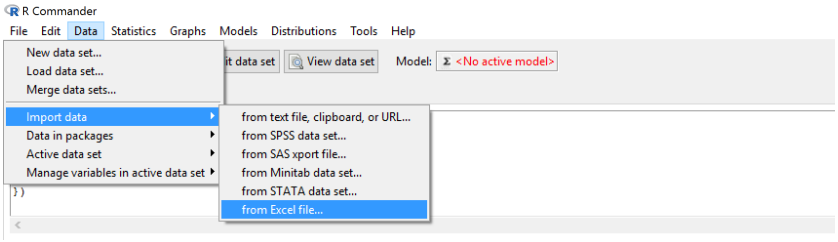
1. Open R (i.e. R x64 3.3.2) in your computer
2. Load R Commander by typing: library(Rcmdr) on your R console



3. By now you should be active in R Commander window



4. Load your dataset into R, for example: **satisfaction.xlsx** by clicking on menu Data > Import data > From Excel file…



5. After loading your data, you can check your dataset by clicking 

Start here…

Do you have more than one question assessing the same construct (e.g., Satisfaction with Life Scale has 5 questions, all intended to assess life satisfaction) → Go to 1. Data Processing

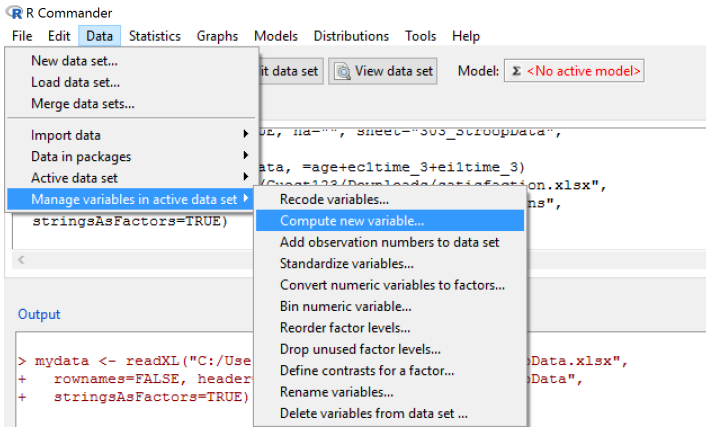
Are you trying to describe the characteristics of your sample? → Go to 2. Descriptives

Are you trying to test hypotheses? → Go to 3. Hypothesis Testing

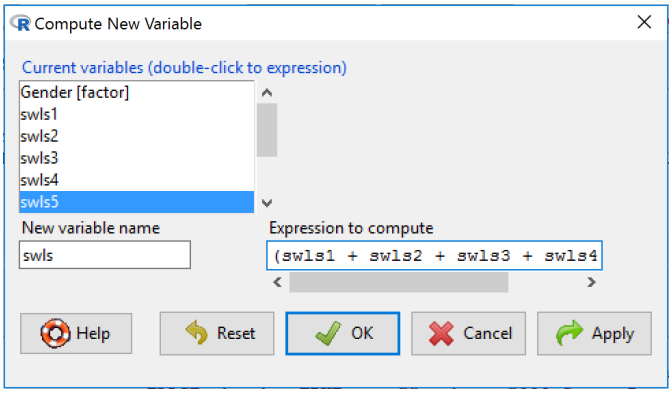
1. **Data Processing**

Are any of your items supposed to be reverse coded? If NO → Go to 1.1. If YES → Go to 1.2
Are you trying to calculate internal reliability for your survey measure? → Go to 1.3
Do you need to create grouping variables out of a continuous variable? → Go to 1.4

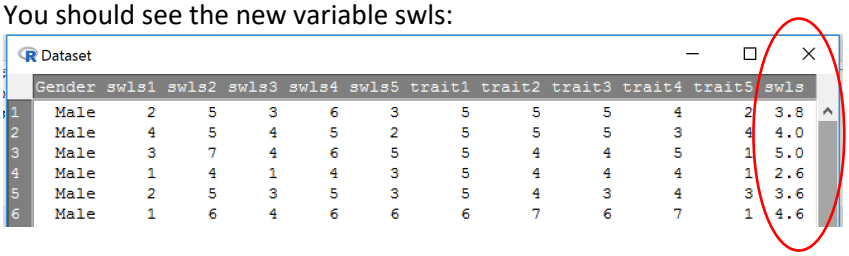
1.1. Create a new variable by making the average of all the items in your scale.
Example: swls = (swls1 + swls2 + swls3 + swls4 + swls5)/5
Data → Manage variables in active data set → Compute new variable…



In "New variable name" box, type your new variable name.
<u>* R is case sensitive, be careful in typing the variable name or any command</u>
In "Expression to compute" box, type (or double click) variables you want to average → in between each variable name type + → at the start and end of your variables in the box, type ( ) → type / → type the total number of variables you have in this scale → press OK



You should see the new variable swls:

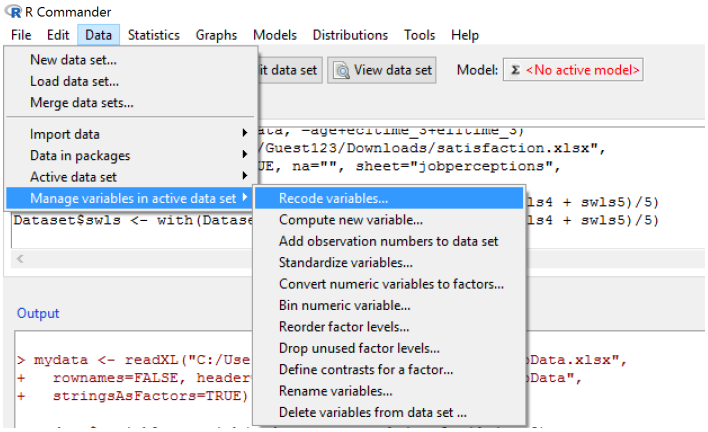


	Gender	swls1	swls2	swls3	swls4	swls5	trait1	trait2	trait3	trait4	trait5	swls
1	Male	2	5	3	6	3	5	5	5	4	2	3.8
2	Male	4	5	4	5	2	5	5	5	3	4	4.0
3	Male	3	7	4	6	5	5	4	4	5	1	5.0
4	Male	1	4	1	4	3	5	4	4	4	1	2.6
5	Male	2	5	3	5	3	5	4	3	4	3	3.6
6	Male	1	6	4	6	6	6	7	6	7	1	4.6

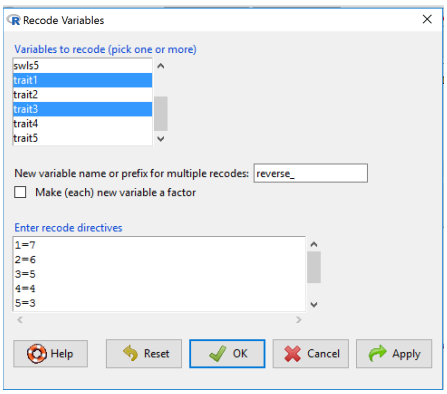
IMPORTANT: This new variable you created will be the one you use in your hypothesis testing, not the individual items.

1.2. Reverse code the items you need to have reverse coded
Example: for a 1-7 scale in variables trait1 and trait3, recode: 1=7, 2=6, 3=5, 4=4, 5=3, 6=2, 7=1

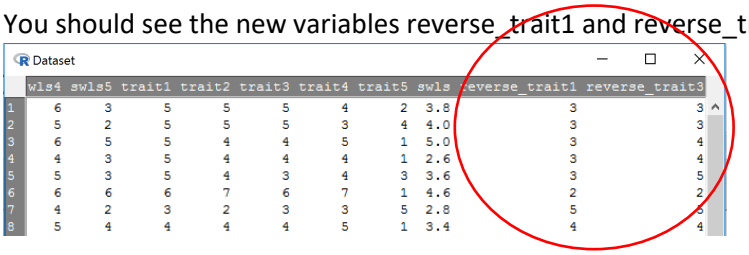
Data → Manage variables in active data set → Recode variables…



In "Variables to recode" box, you can choose multiple variables e.g. click variable trait1 and Ctrl + click trait3 →
In "New variable name" box, you can type: reverse_ →
Because we want these variables to be numeric, uncheck the box "Make (each) new variable a factor" →
In "Enter recode directives" type:
1=7 click ENTER
2=6 click ENTER
3=5 click ENTER
4=4 click ENTER
5=3 click ENTER
6=2 click ENTER
7=1
➔ Click OK



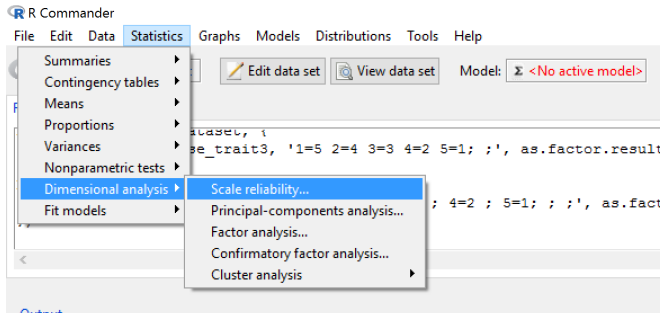
You should see the new variables reverse_trait1 and reverse_trait3:



	wls4	swls5	trait1	trait2	trait3	trait4	trait5	swls	reverse_trait1	reverse_trait3
1	6	3	5	5	5	4	2	3.8	3	3
2	5	2	5	5	5	3	4	4.0	3	3
3	6	5	5	4	4	5	1	5.0	3	4
4	4	3	5	4	4	4	1	2.6	3	4
5	5	3	5	4	3	4	3	3.6	3	5
6	6	6	6	7	6	7	1	4.6	2	2
7	4	2	3	2	3	3	5	2.8	5	5
8	5	4	4	4	4	5	1	3.4	4	4

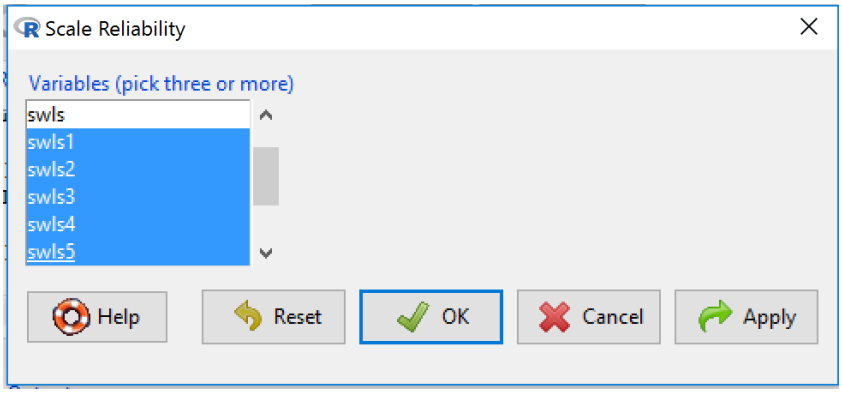
Now, go to → 1.1. to create your new variables.

1.3. To calculate internal reliability/Cronbach's Alpha
Statistics→ Dimensional analysis → Scale reliability



Pick variables you want to test:

Example: Click swls1 → Ctrl + click swls2 → Ctrl + click swls3 → Ctrl + click swls4 → Ctrl + click swls5 → OK

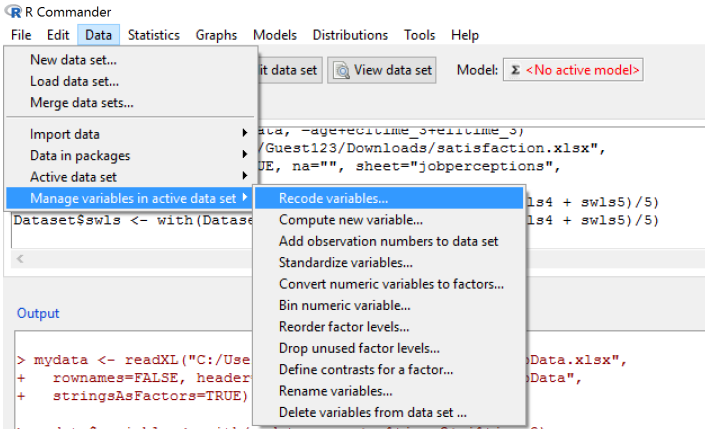


Important values to note: Cronbach's Alpha

1.4. Creating a 2-level grouping variable (dummy coding)

Example: Group scale 1-7 into: Low (1-4), High (5-7) would be (1="low", 2="low", 3="low" , 4 =" low"); (5="high", 6="high", 7="high").

Data → Manage variables in active data set → Recode variables…

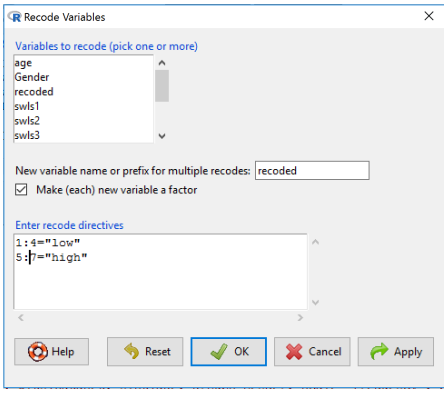


In “Variables to recode” box, click variable trait1 →
In “New variable name” box, type: recoded_ →
Because we want these variables to be factor, check the box “Make (each) new variable a factor” →
In “Enter recode directives” box, there are 2 ways to type, first:
1 , 2, 3, 4 = ”low” click ENTER
5, 6, 7 = ”high”

OR you can type:

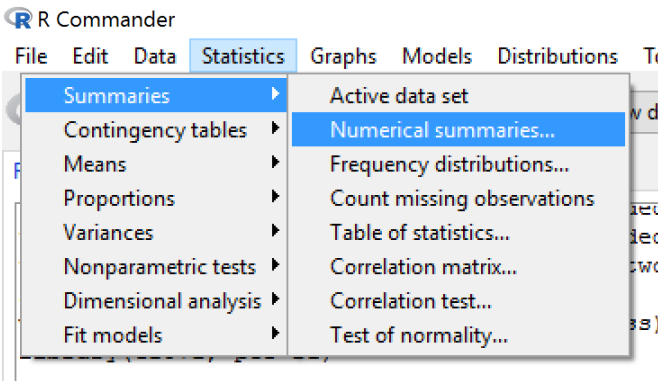
1 : 4 = ”low” click ENTER
5 : 7 = ”high”

Click OK

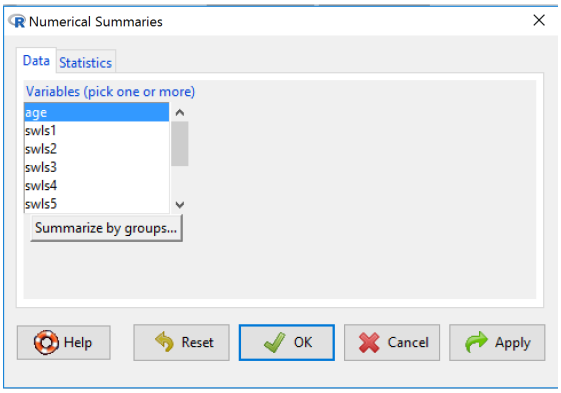


2. **Descriptives**
   Is your variable continuous? If YES → Go to 2.1 Means and Standard Deviations. If NO → Go to 2.2 Frequencies

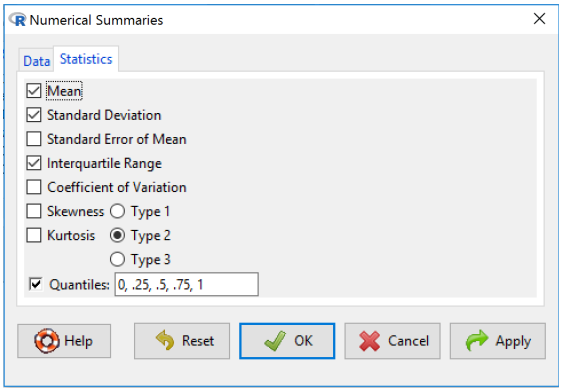
   2.1. Means and Standard Deviations: This is for continuous variables ONLY (e.g., age)
   Statistics → Summaries → Numerical summaries…



➔ Pick the variable you want to choose, e.g. age



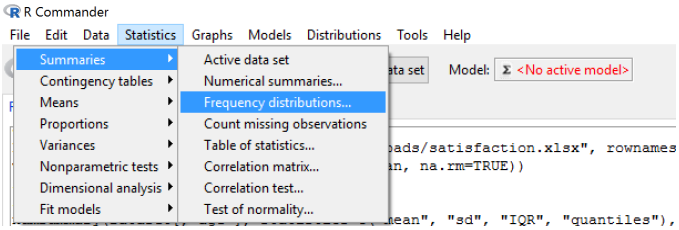
➔ Select what statistics you want to compute in "Statistics" tab → click OK



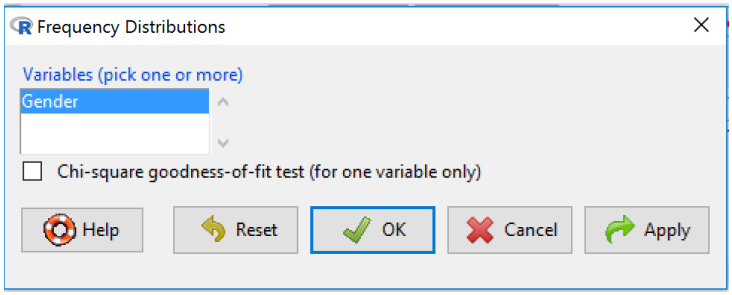
Important values to note: Mean, Standard Deviation {ADD DISTRIBUTION PLOT}

To add distribution plot: Graphs… → Histogram

2.2. Frequencies: These are for categorical variables ONLY (e.g., gender, class standing)
How to run frequencies (counts/percentages):
Statistics → Summaries → Frequency distributions…



➔ Pick the variable you want to choose, e.g. gender → click OK



Important Values to note: Frequency, Percent

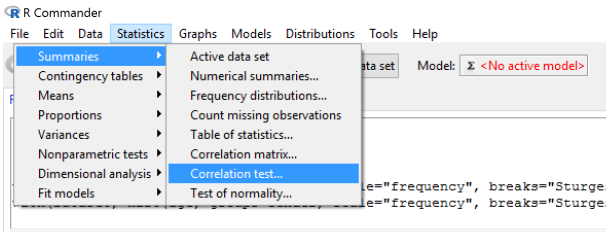
3. Hypothesis Testing

Do you have more than one predictor variable that you want to account for in one single analysis? If Yes → Go to 4. Multiple Regression. If No, keep reading below.
Is your independent variable (or predictor) categorical? (e.g., condition, gender, class year). If YES → Go to 3.2 Categorical Predictors. If NO, it's continuous → go to 3.1 Continuous Predictors

3.1. So your predictor/independent variable is continuous. Is your dependent (or outcome) variable also continuous? If YES → Go to 3.1.1 Correlation. If No → Go to 3.1.2 Logistic Regression

3.1.1. Correlation (both predictor and outcome = continuous)

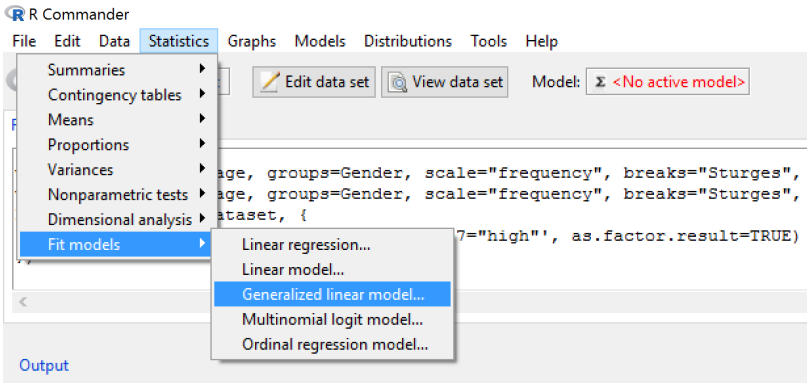
Statistics → Summaries → Correlation test..



Important values to note: Pearson Correlation (r) and Sig (p value).

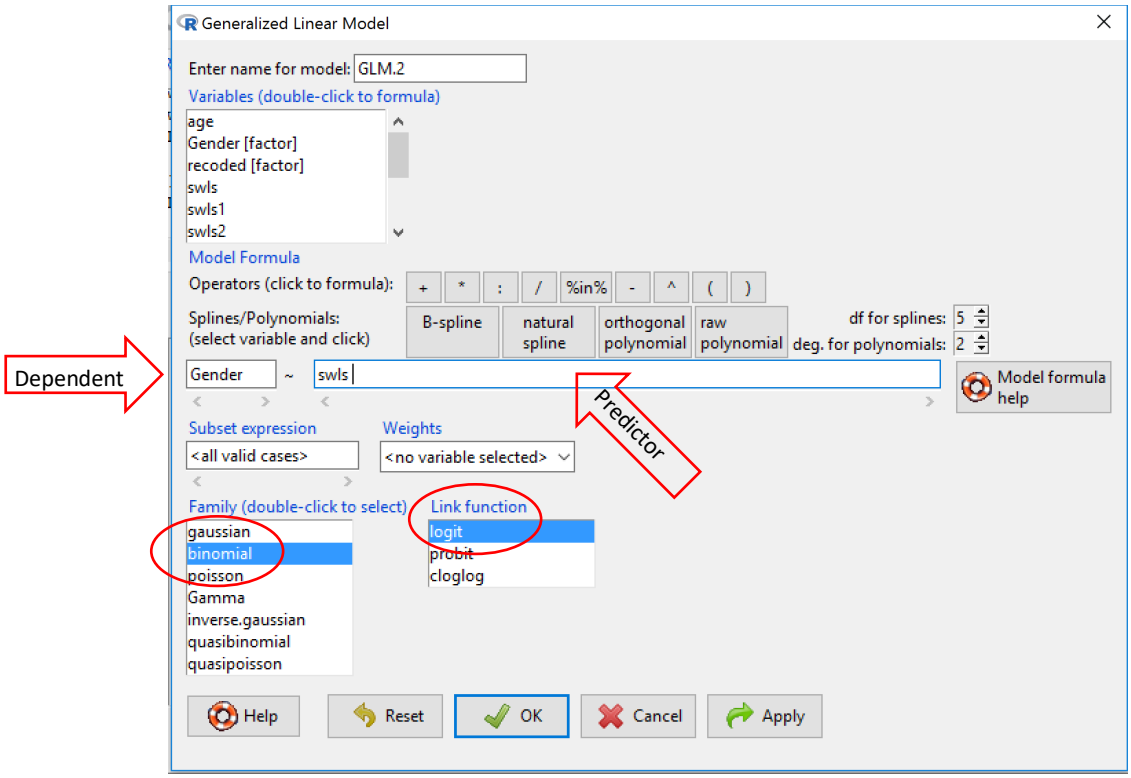
3.1.2. Logistic Regression (predictor continuous, outcome = categorical, with two levels/choices)

Does your outcome variable have more than 2 levels? If YES → go to 1.4 Creating a two-level grouping variable. If NO, keep reading below.
Statistics → Fit models → Generalized linear model..



Enter your categorical dependent variable in the dependent box → Enter your continuous predictor in the predictor box → Select Family "binomial" → Link function "logit"

Example: dependent variable: Gender , predictor: swls



Important values to note: Beta, Std Error, df, Sig (p value).

3.2. Categorical Predictors. So your predictor/independent variable is categorical. Is your dependent (or outcome) variable continuous? If YES → Go to 3.2.1 T-Test/ANOVA. If NO → Go to 3.2.2 Chi-Square

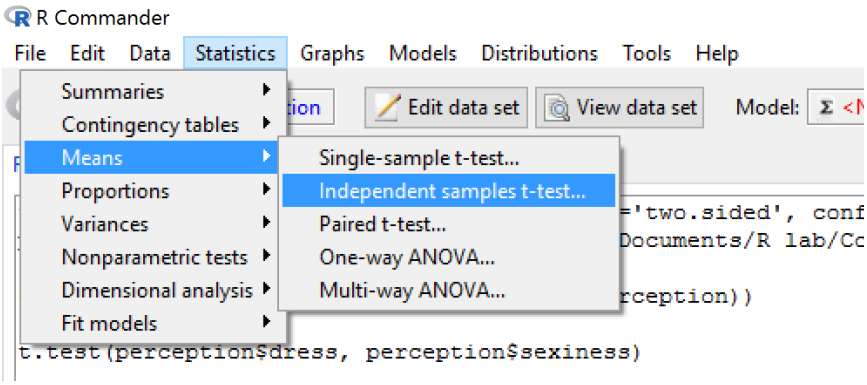
3.2.1. T-Test/ANOVA (Predictor is categorical, outcome is continuous)

Does your predictor variable have more than two conditions or groups (e.g., freshmen, sophomores, juniors, seniors)? If NO → Go to 3.2.1.1 T-Test. If Yes → Go to 3.2.1.2 ANOVA

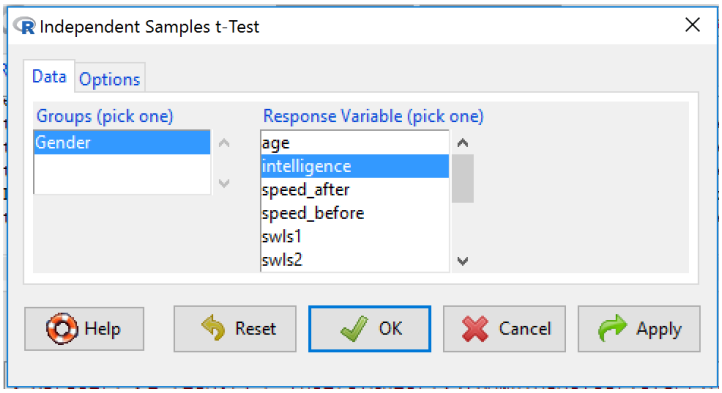
3.2.1.1. T-Test (Two conditions/groups/sets of variables). Are you trying to compare two different groups of people? If YES → Go to 3.2.1.1.1 Independent Samples T-Test. If NO, I'm trying to compare one person's responses on two different sets of questions → Go to 3.2.1.1.2 Paired Samples T-Test

3.2.1.1.1. Independent Samples T-Test (are the means between two groups different – between groups design)

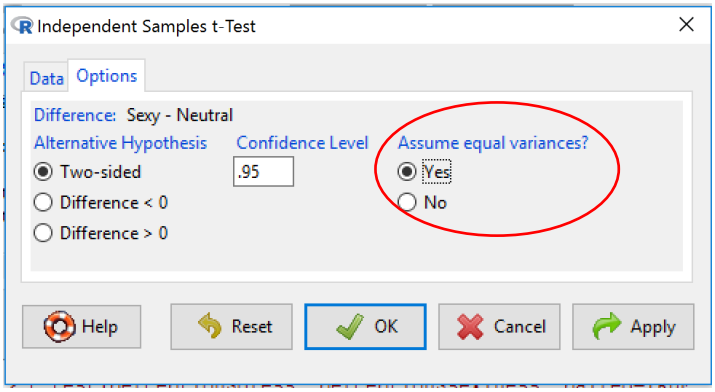
Statistics → Means → Independent Samples T-test →



Define Groups (categorical) → Response Variable (continuous)



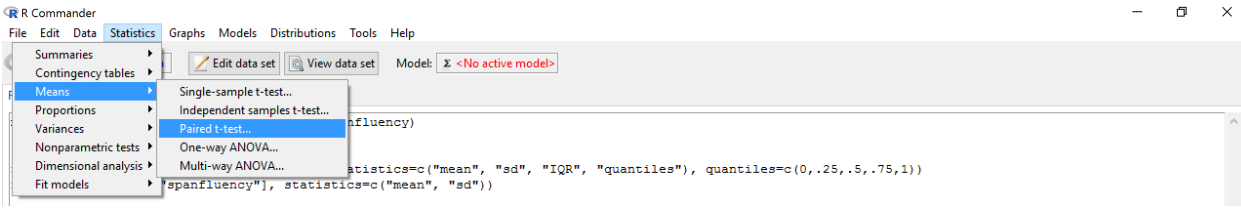
On "Options" tab, set the equal variance assumption (or not) → OK



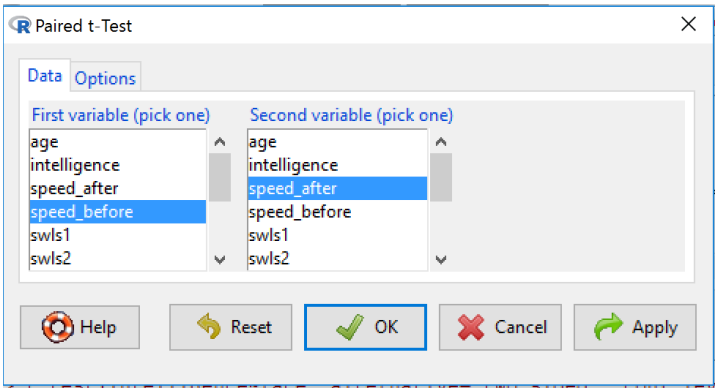
Important values to note: t, df, sig (p value), mean, std. deviation

3.2.1.1.2. Paired Samples T-Test (within-groups design)

Statistics → Means → Paired t-test

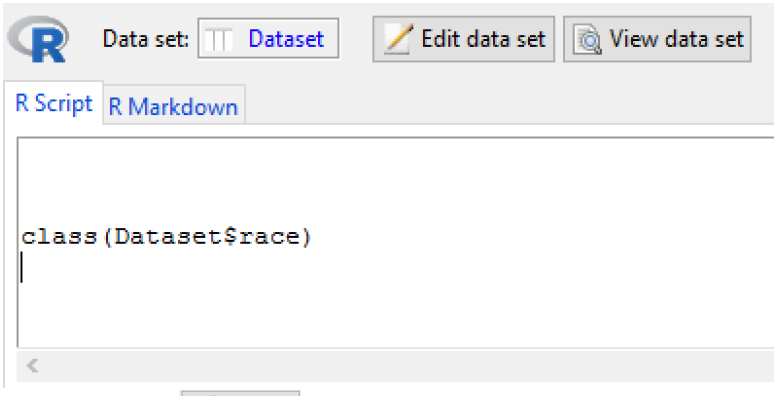


Pick first variable (before) and second variable (after) → click OK



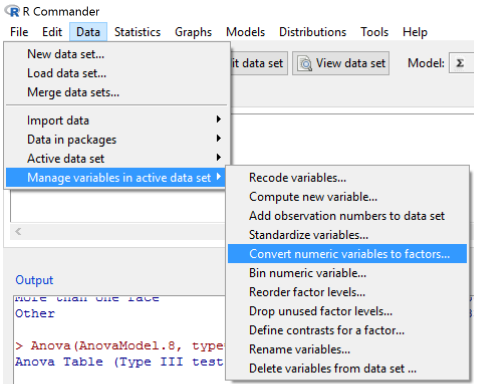
Important values to note: t, df, sig (p value), mean, std. deviation

3.2.1.2. ANOVA. So your predictor variable has more than one level/group.
Check the categorical variable (e.g. race) has been set as factor, by typing:

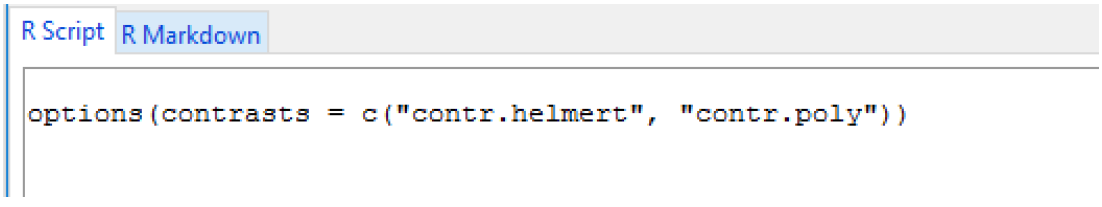


Click "Submit"  button

If the variable hasn't been set as factor, change it by clicking Data → Manage variables in active data set → Convert numeric variables to factors:

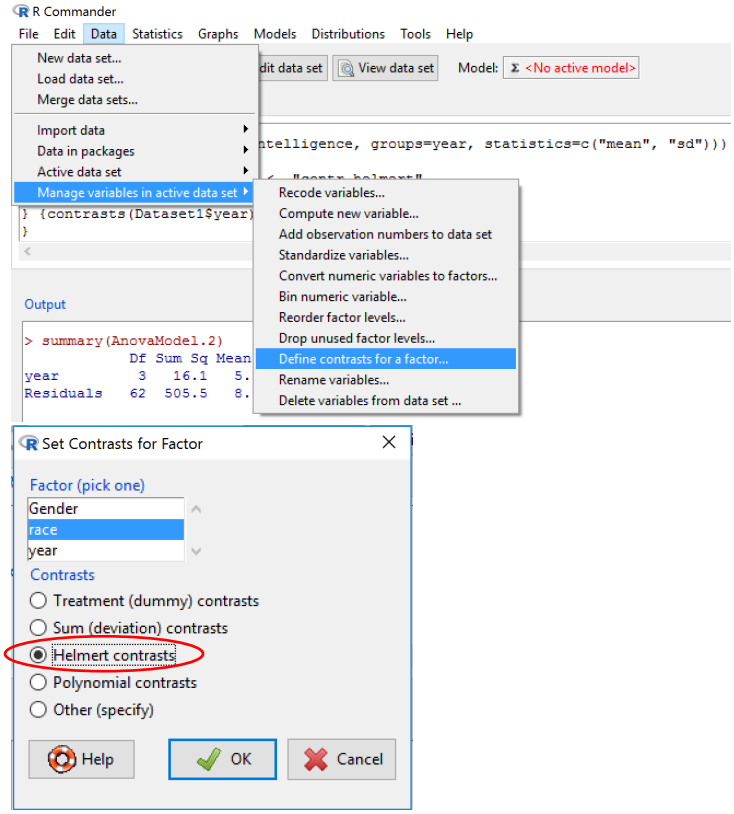


Next, set the default contrast in R to helmert, by typing in the command box →

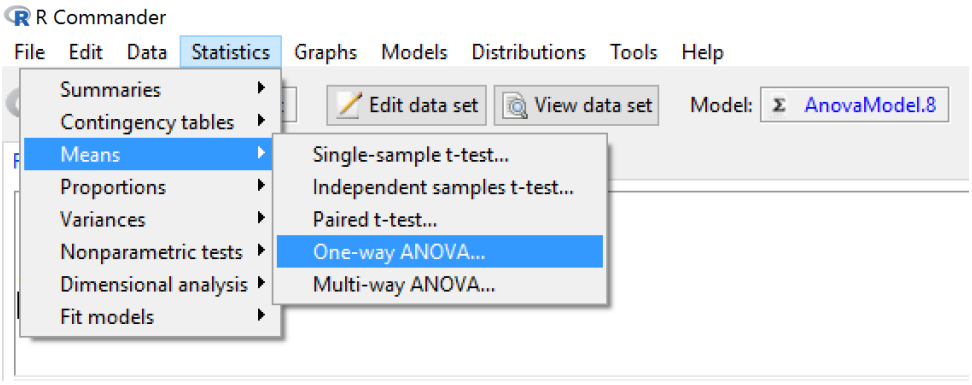


OR

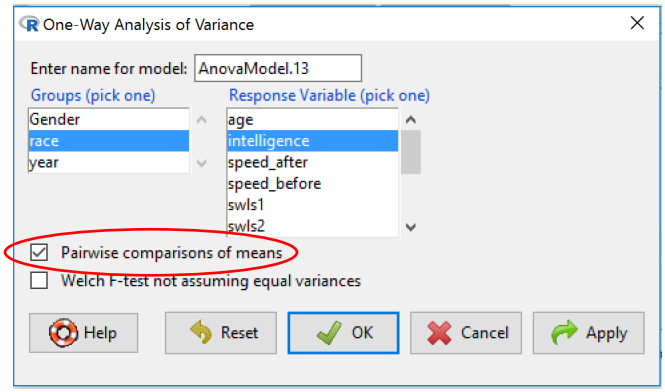
Data → Manage variables in active data set → Define contrasts for a factor…



Next, conduct ANOVA: Statistics → Means → One-way ANOVA…



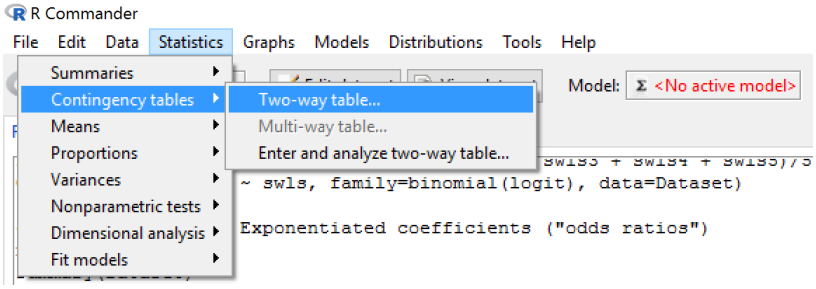
Don't forget to tick the "Pairwise comparisons of means" box for Tukey Comparison



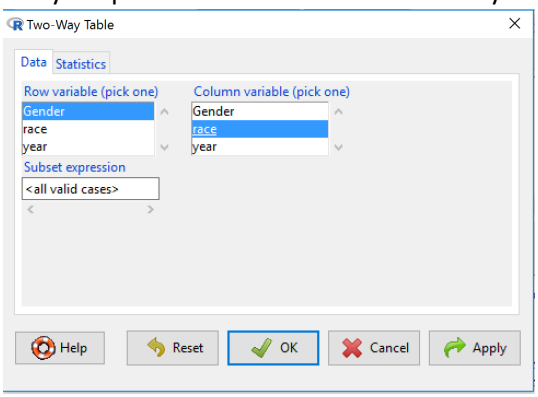
Important values to note: F, df, Sig (p value). If p < .05, check Tukey multiple comparisons for which groups are significantly different.

3.2.2. Chi-Square (Both predictor and outcome are categorical)

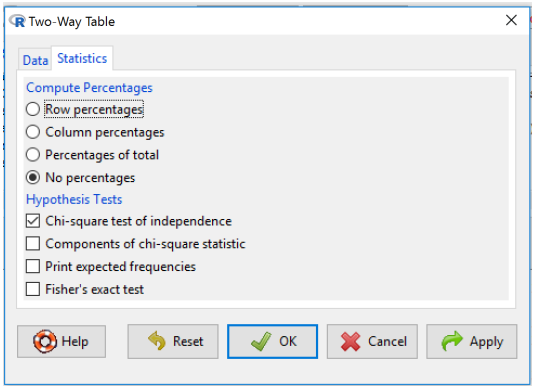
Statistics → Contingency tables → Two-way table…



Put your predictor in the row box → Put your outcome in the column box →



You can set the output from Chi Square Hypothesis Tests on Statistics tab

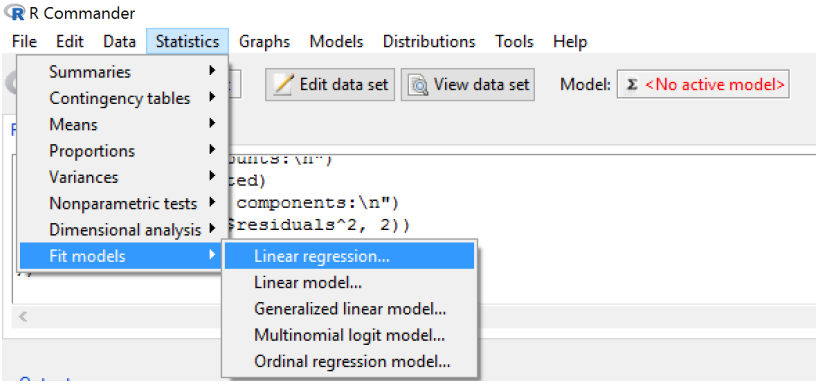


Important values to note: Pearson Chi-Square Value, df, Asymp Sig (p value).

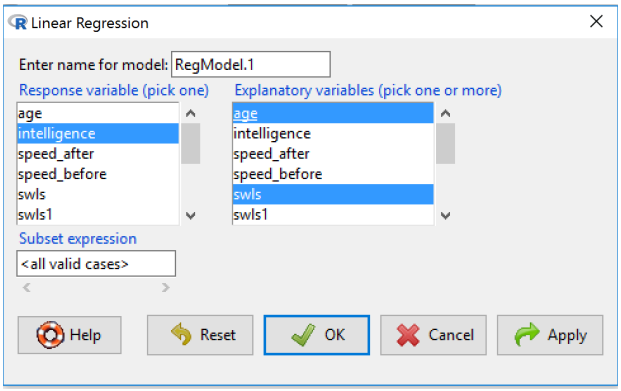
4. **Multiple Regression:** (IMPORTANT: Outcome variable is continuous in this analysis. If your outcome is categorical, you will need to go to 3.1.2 Logistic Regression).
   Are all of your predictors continuous? IF YES→ go to 4.1 Analyzing Multiple Regression. If NO → go to 1.4 Creating a 2-level grouping variable. When done with this 1.4, you can come back to 4.1.

4.1. Analyzing Multiple Regression

Statistics → Fit models → Linear regression →



[enter dependent variable into response box] → [enter predictor variables into explanatory variable(s) box] → click OK



Important values to note in Model Summary: R, R Square, Adjusted R Square; in ANOVA box: F, df, Sig (p value). in Coefficients box: Beta, Std. Error, Sig for each of your variables (ignore constant).