SPSS Workshop Step-by-Step Guide

Note on dataset: this guide uses examples from the teachingratings_spss.xlsx dataset, which contains **modified** data from a published study (see documentation tab in excel data spreadsheet).
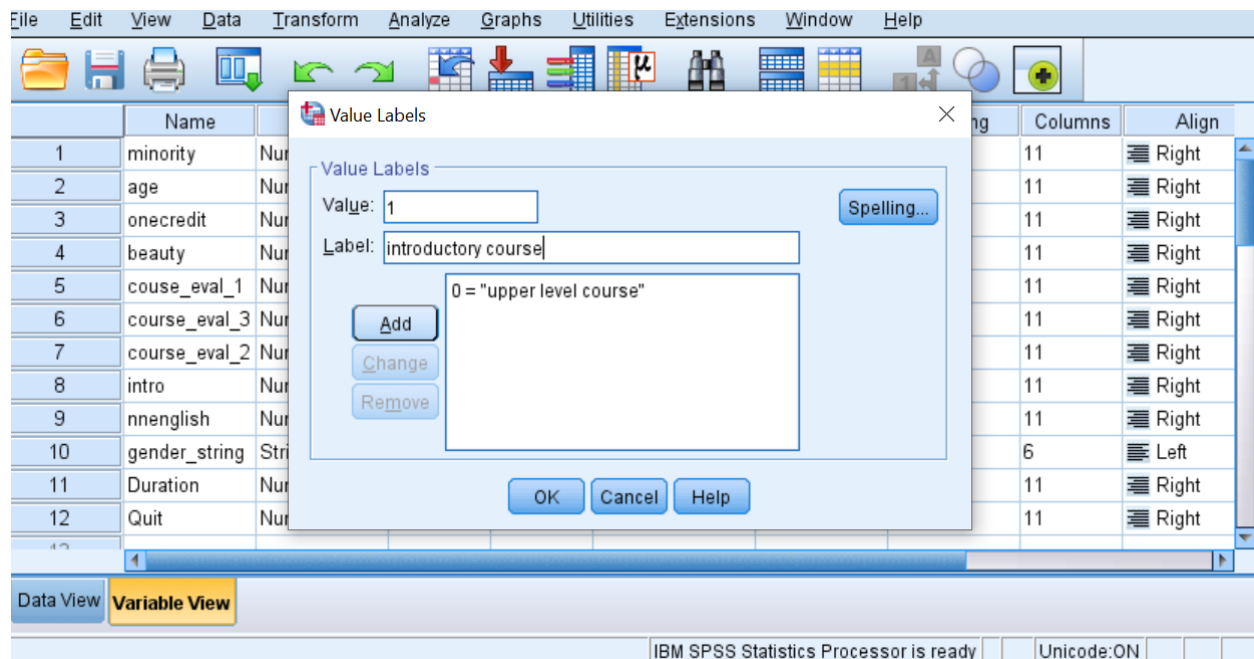
Please note that several variables are completely fabricated for the purpose of this workshop, including "course_eval_1"; "course_eval_2"; "Duration"; and "Quit"
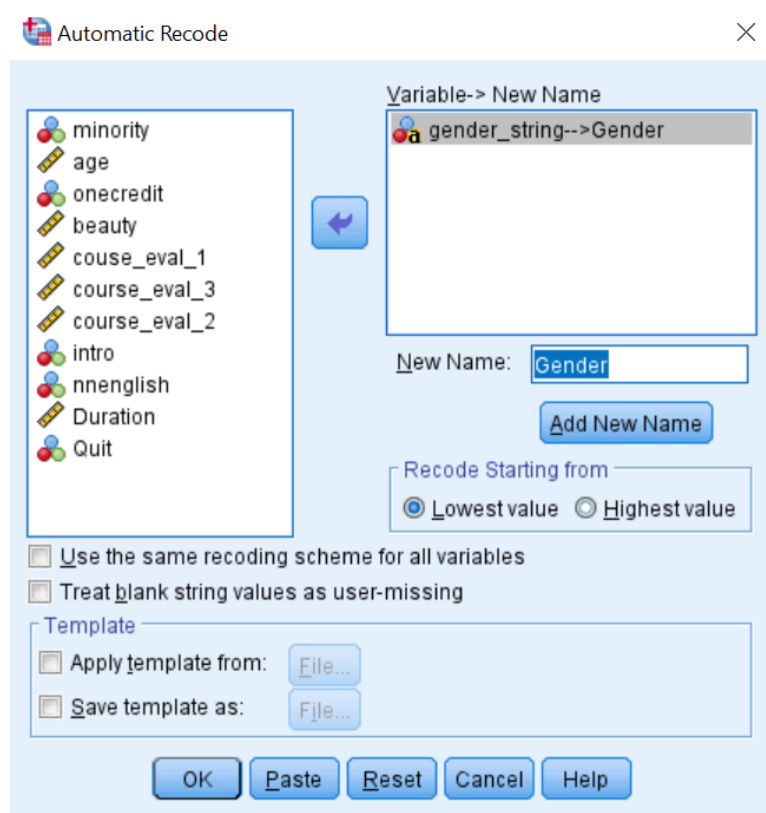
Importing Data

File > Import Data > [select appropriate file type, e.g. Excel] > make sure "Read variable names from first row" is checked

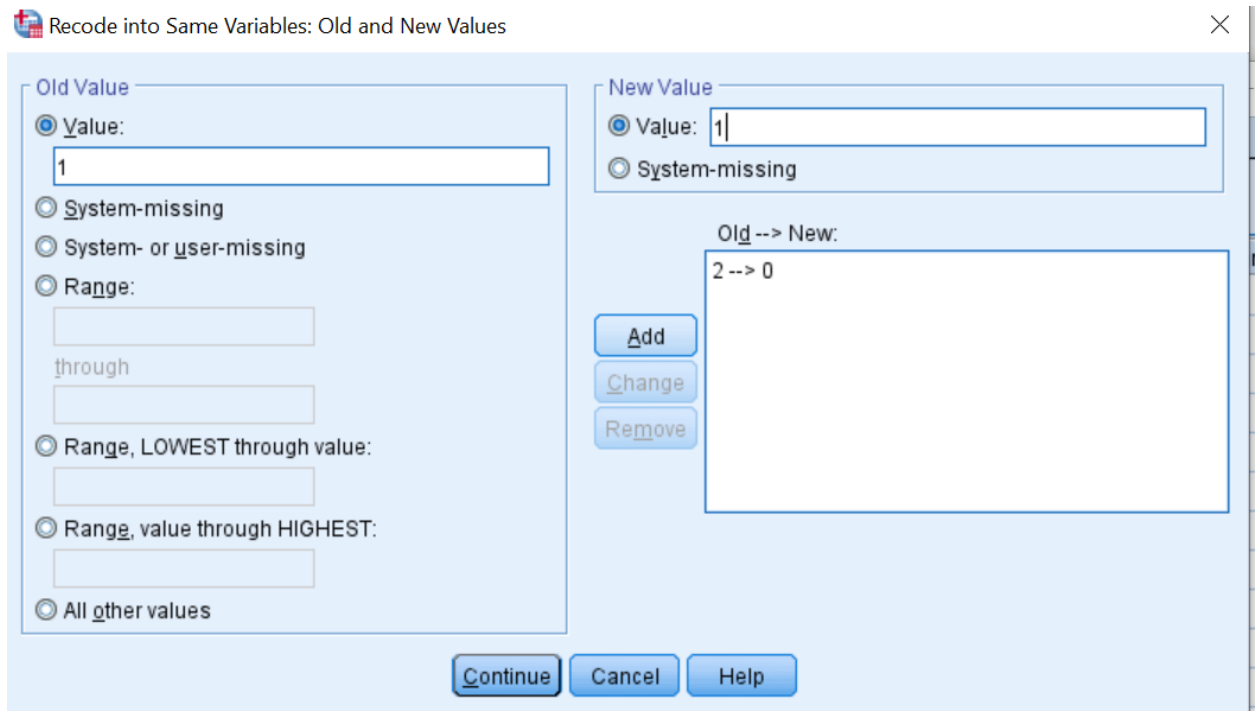## Exploring and Modifying/Cleaning Data

- Note, if you see labels in a dialog box instead of a list of variable names, right click the list and select "Display Variable Names"
- Reorder variables manually in Variable View tab:
    - Click and hold the row number (e.g. 6 for "course_eval_3" variable) and drag up or down (you will see a thin red guideline)
- Add value and variable labels
    - Click into empty cells for "Labels" column and type description of variable. Example: for the "beauty" variable, type "rating of instructor's physical appearance" in the corresponding Label cell
    - To add labels to values for a variable, click the blank cell in the Values column and then the small button with three dots (…). Example: click the cell for Values for the intro variable, for value enter 1 and for label enter "introductory level course" then click add. Repeat for value of 0 and label of "upper level course", click add, then OK. Use the original Excel dataset ("Documentation" tab) as a guide for entering the remaining variable and value labels.

The SPSS Variable View with a "Value Labels" dialog box open:

	Name					ng	Columns	Align
1	minority	Nur					11	Right
2	age	Nur					11	Right
3	onecredit	Nur					11	Right
4	beauty	Nur					11	Right
5	couse_eval_1	Nur					11	Right
6	course_eval_3	Nur					11	Right
7	course_eval_2	Nur					11	Right
8	intro	Nur					11	Right
9	nnenglish	Nur					11	Right
10	gender_string	Stri					6	Left
11	Duration	Nur					11	Right
12	Quit	Nur					11	Right

Value Labels dialog:
Value: 1
Label: introductory course
0 = "upper level course"
Add / Change / Remove
OK / Cancel / Help
Spelling...

Data View | Variable View

IBM SPSS Statistics Processor is ready   Unicode:ON

- Change a string variable to numeric
  - Transform > Automatic recode > gender_string (click arrow to put this variable in the "Variable → New Name" box) > enter "Gender"as New Name > click "Add New Name" > click "OK"

Automatic Recode dialog:

minority
age
onecredit
beauty
couse_eval_1
course_eval_3
course_eval_2
intro
nnenglish
Duration
Quit

Variable-> New Name
gender_string-->Gender

New Name: Gender
Add New Name

Recode Starting from
◉ Lowest value ○ Highest value

☐ Use the same recoding scheme for all variables
☐ Treat blank string values as user-missing

Template
☐ Apply template from: File...
☐ Save template as: File...

OK Paste Reset Cancel Help

- If the new recode (1 and 2 as dummy codes for the gender variable) is fine, leave as is. If you need to recode again (change 2 to 0), follow these instructions: Transform > Recode into Same Variable > move Gender to empty box using arrow > click "Old and New Values" > enter 2 in Value under "Old Value" menu and 0 in Value under "New Value" menu > click Add. Repeat for 1 to 1 and click add, then Continue. Click OK.
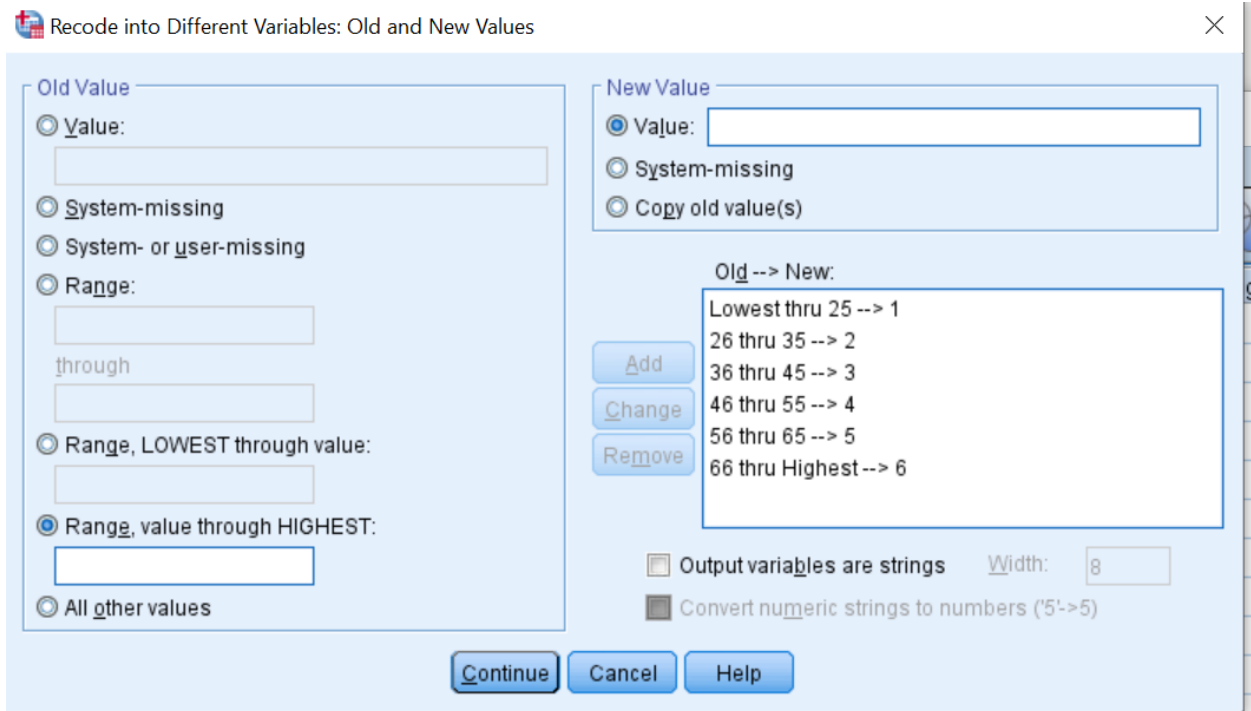

- Don't forget to update the value labels!
- Computing a new variable:
  - Transform > Compute Variable > type a name for the new variable in "Target Variable" box (e.g. Total_CourseEvals) > build a math operation or function to compute the new variable (e.g. course_eval_1 + course_eval_2 + course_eval_3). Click OK. Add label and value labels (if applicable) to new variable in Variable View.
- Recoding existing variable using computation
  - Oops, we wanted total average for course evals, not sum total!. Transform > Compute Variable > [keep same name for Target Variable, e.g. Total_CourseEvals] > find "Statistical" in Function Group menu and click on it > double click on "Mean" in "Functions and Special Variables" list. Replace ? characters with course_eval_1, course_eval_2, course_eval_3. Expression should look like: MEAN(course_eval_1, course_eval_2, course_eval_3). Click OK and then Yes to replace existing variable. Rename variable (click on the name in Variable View and type over previous name) to Mean_AllCourseEvals
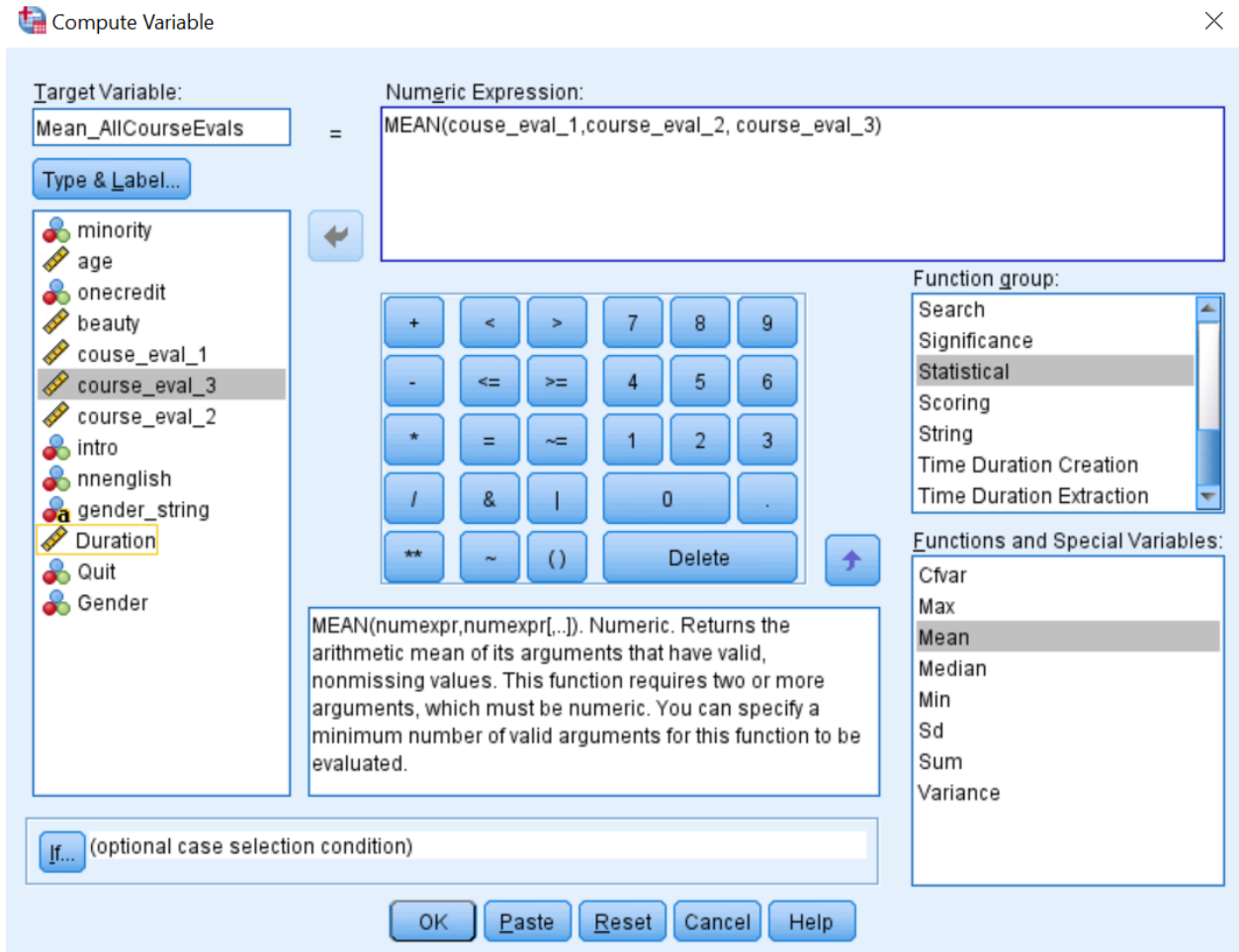
- Creating a new variable using an existing variable:
  - Transform > Recode into Different Variable > move "age" into the "Input Variable →
    Output Variable" empty box using the arrow > put "age_categories" as the Name and
    "age variable divided into buckets of 10 years" as the Label in the Output Variable box.
    Click "Change." Click "Old and New Values"
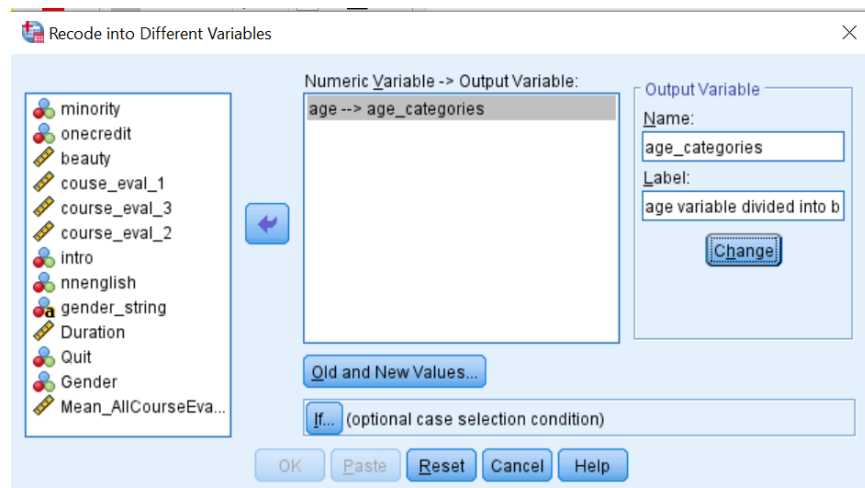
- Click "Range, LOWEST through value:" and enter 25. For new value, enter 1. Click Add. Click "Range:" and enter 26 as the lower bound and 35 as the upper, then enter 2 for new value and click Add. Repeat until 66 years (following the 56 – 65 year old bucket). Click "Range, value through HIGHEST:" and enter 66 and then 6 for new value. Click Add. Click Continue. Click OK.

Recode into Different Variables: Old and New Values

Old Value
- Value:
- System-missing
- System- or user-missing
- Range:
  through
- Range, LOWEST through value:
- Range, value through HIGHEST:
- All other values

New Value
- Value:
- System-missing
- Copy old value(s)

Old --> New:
Lowest thru 25 --> 1
26 thru 35 --> 2
36 thru 45 --> 3
46 thru 55 --> 4
56 thru 65 --> 5
66 thru Highest --> 6

Add
Change
Remove

☐ Output variables are strings   Width: 8
☐ Convert numeric strings to numbers ('5'->5)

Continue   Cancel   Help

- Create a new variable using a cut point (e.g. mean or median usually):
  - Transform > Recode into Different Variable > move "beauty" into the "Input Variable → Output Variable" empty box using the arrow > type "beauty_binary" as the Name and "Beauty as a binary with average and under average beauty in one category and above average beauty in the other category" as the Label in the Output Variable box. Click "Change". Click "Old and New Values"
  - click "Range, LOWEST through value:" and enter 0. When working with numeric variables, you would have to first find the mean or median in order to enter a cut point (by selecting descriptive statistics for that variable, see next task). For new value, enter 0. Click Add. Click "All other values" for old value and enter 1 for new value. Click Add. Click Continue. Click OK.

- Basic descriptive stats:
  - Analyze > Descriptive Statistics > Descriptives > [move variable of interest to empty "Variable(s)" box, for example Mean_AllCourseEvals]. Click Options to see if any apply. Click OK.
- Basic data analysis by group:
  - Data > Sort Cases > move "gender" to Sort By empty box > click OK. Data > Split File > click "Organize Output by Groups" > move gender into "Groups Based on" empty box > Click OK.

- o Analyze > Descriptive Statistics > Descriptives > Mean_AllCourseEvals.
- o Data > Split File > click "Analyze all cases"
- Basic data analysis using a filter (could also be filtered by group):
 - o Data > Select Cases > click "if condition is satisfied" > click "If" > type a math operation or function (example: beauty <= 0).
  - o Analyze > Descriptives > Mean_AllCourseEvals (see output in figure below)

→ **Descriptives**

Gender = male

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
Mean_AllCourseEvals	268	2.07	4.97	4.0343	.55446
Valid N (listwise)	268				

a. Gender = male

Gender = female

Descriptive Statistics[a]

|	N	Minimum	Maximum	Mean	Std. Deviation
Mean_AllCourseEvals	195	2.27	4.87	3.8677	.53880
Valid N (listwise)	195				

a. Gender = female
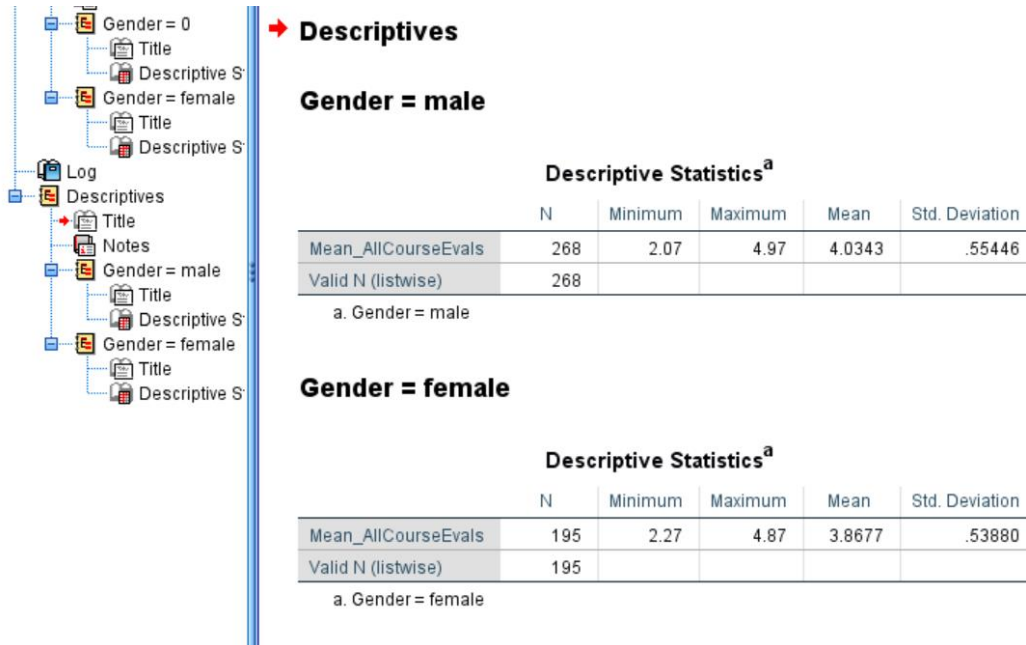
- o Remove filter!!! Data > Select Cases > All Cases > OK
- Comparing means across 2 or more variables (no statistical analysis):
  - o Analyze > Compare Means > Means > move variables to dependent and independent list as appropriate for your analysis (e.g. Mean_AllCourseEvals to dependent and Gender to independent) > OK
- Basic PLOTS/CHARTS:
  - o Boxplot: Graphs > Boxplot > Simple > Check "Summaries for Groups of Cases" > Click Define > beauty as variable and gender as category axis
  - o Histogram: Analyze > Descriptive Statistics > Frequencies > move Mean_AllCourseEvals to Variable(s) empty box > click Charts > click Histogram and also the box to "show normal curve on histogram", if desired > Continue > OK
  - o Scatter plot: Graphs > Scatter/Dot > Simple Scatter > Define > Mean_AllCourseEvals as dependent variable (Y-Axis) and Beauty as independent variable (X-Axis) > OK
- Cross Tabs/Chi-Square Test
  - o Analyze > Descriptive Statistics > Crosstabs > beauty_binary as row and gender as column > click Statistics > check Chi Square > Continue > OK
- Bivariate correlations (e.g. correlation matrix with Pearson or Spearman coefficient):
  - o Analyze > Correlate > Bivariate > move Mean_AllCourseEvals, age, and beauty to Variables empty box > OK
- Independent Samples T-Test:

- o   Analyze > Compare Means > Independent Samples T Test > move Mean_AllCourseEvals
      to Test Variable and gender to Grouping Variable > Define Groups > Use Specified
   Values (0 and 1 since that's how gender is coded) > Continue > OK
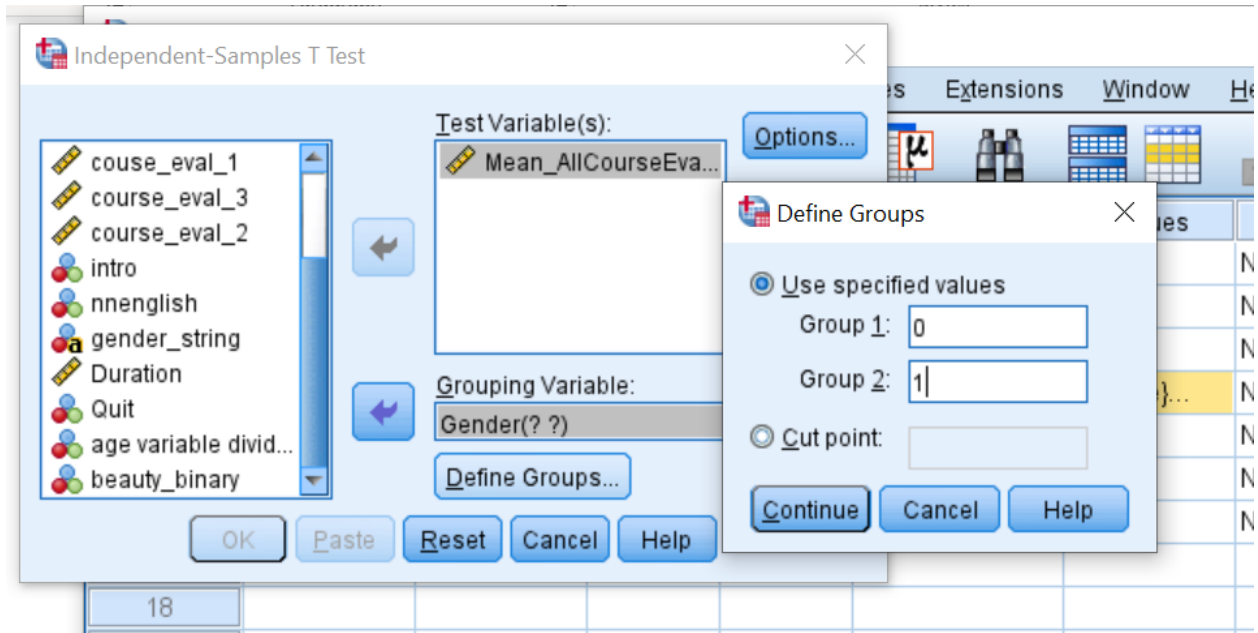
- **Paired T-Test**
  - o   Analyze > Compare Means > Paired Samples T Test > move course_eval_1 and
        course_eval_3 as the first pair (Pair 1) > OK
- **General Linear Models**
  - o   One-Way ANOVA: Analyze > General Linear Model > Univariate > Mean_AllCourseEvals
        as dependent and gender as fixed factor
  - o   Two-way ANOVA (factorial ANOVA): Analyze > General Linear Model > Univariate >
        Mean_AllCourseEvals as dependent and gender and beauty_binary as fixed factors. Click
        Plots > gender on x-axis and beauty_binary as separate line > click Add > Ok
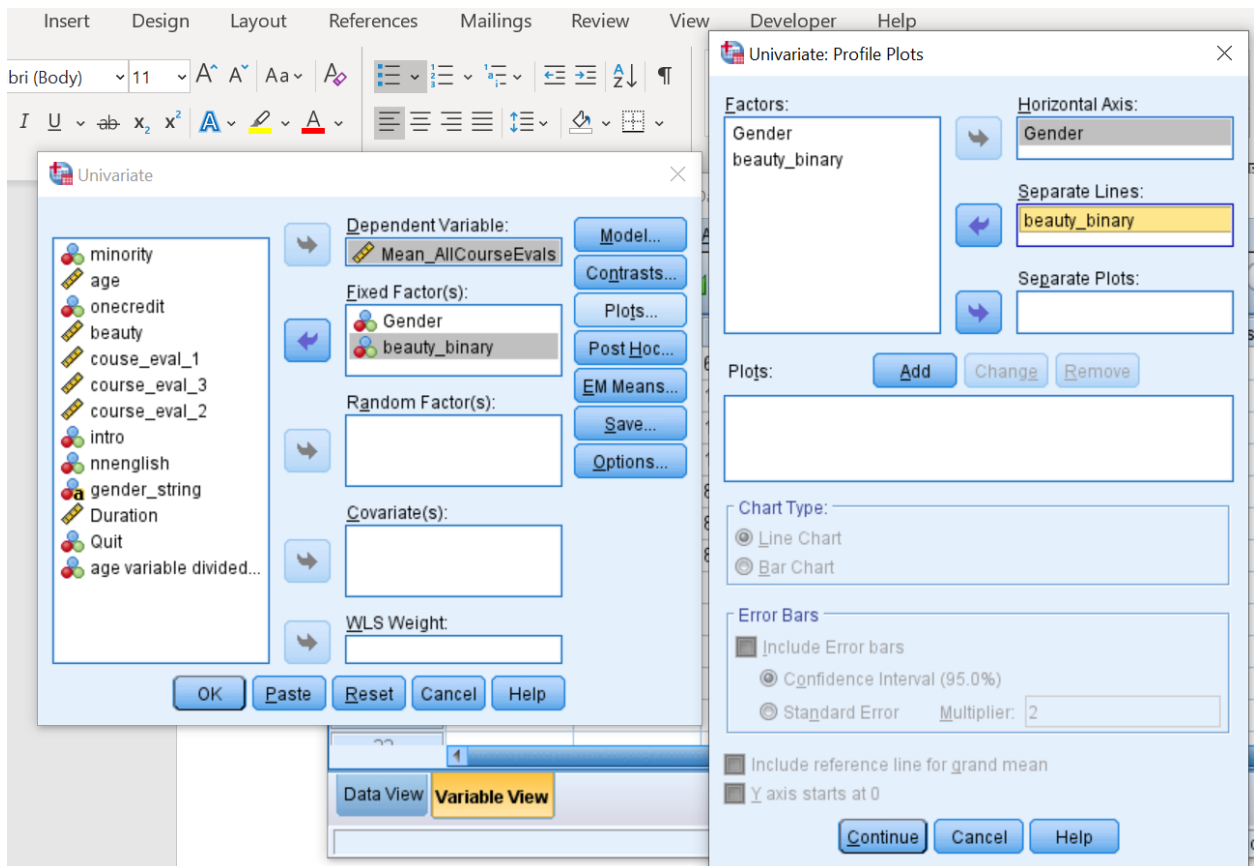
- OLS regression
  - Analyze > Regression > Linear > Mean_AllCourseEvals as dependent and beauty and age as independent
- Survival Analysis
  - Kaplan-Meier Plot: Analyze > Survival > Kaplan-Meier > move Duration to Time and Quit to Status > Define event > single value (put 1 which means the individual quit before the end of their contract) > move gender to factor > click Compare Factor > check off "Log rank" > Continue > click Options > Plots > Survival (or Hazard) > Continue > OK
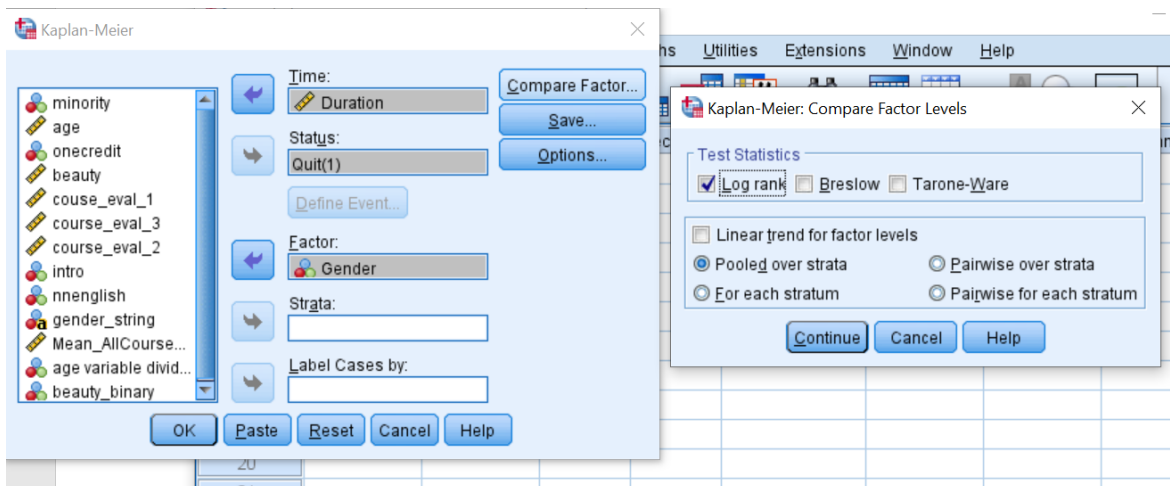
- o Cox Regression: Analyze > Survival > Cox regression > move Duration to Time and Quit to Status > Define event (type 1) > Continue > put gender, Mean_AllCourseEvals, and age_categories as covariates > Click Categorical > move gender and age_categories to Categorical Covariates empty box > click on one at a time and change contrast to Simple and reference category to First (click change each time) > Continue > Options > Model Statistics > Check box for CI for exp(B) > Continue > click Plots > click Survival and/or Hazard then move gender to "Separate Lines for" > Continue > OK
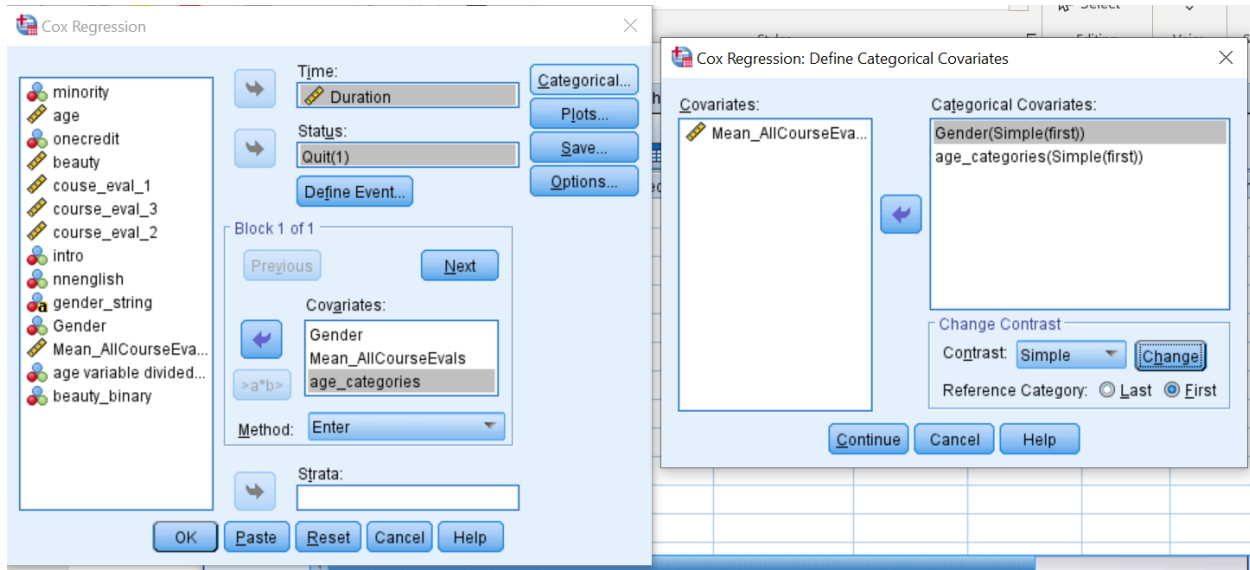
- o Note on cox regression: to add an interaction term, select two or more variables from the variable list (e.g. by holding CTRL on a pc or Command on a Mac) and then the button >a*b> to add to covariates list.