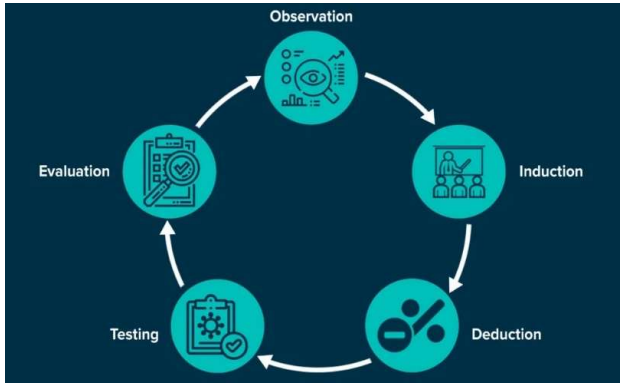
Empirical research in management and economics

Exercise

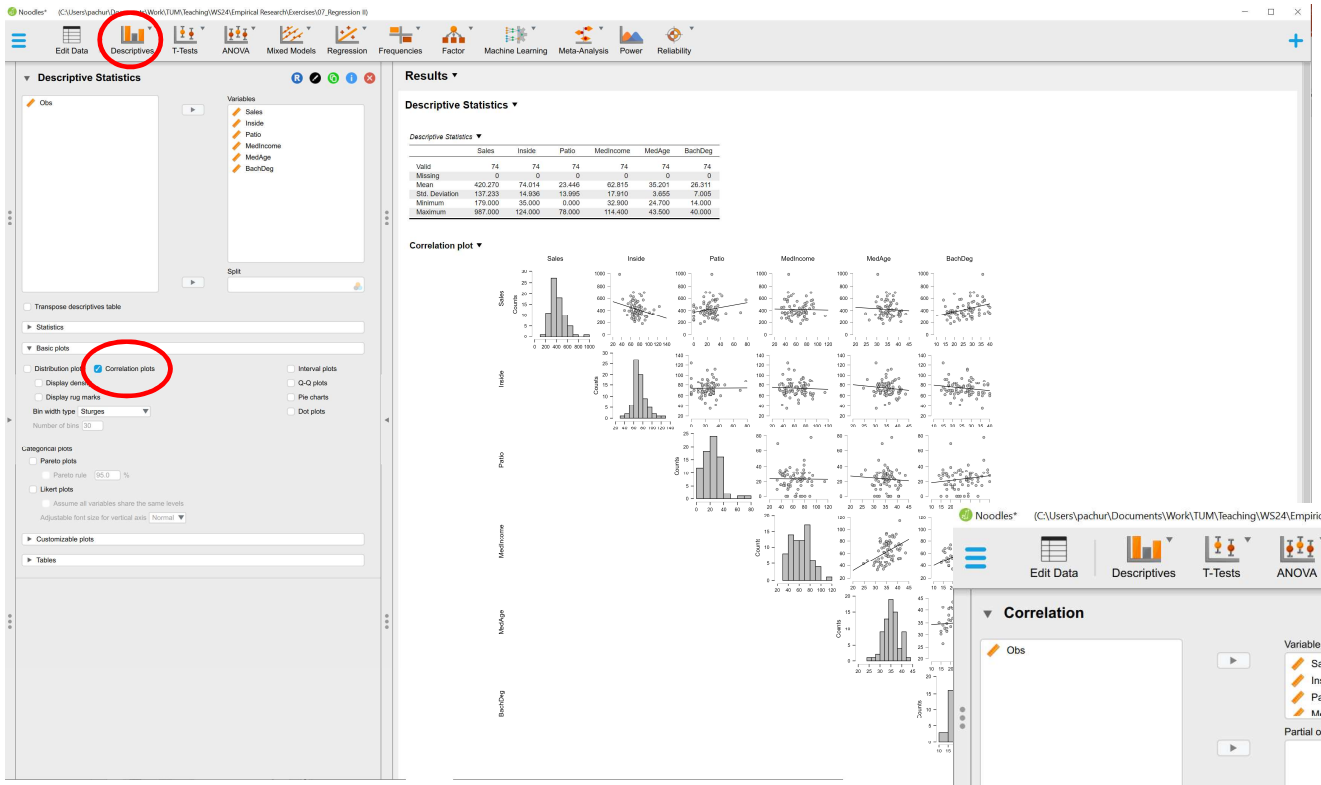
Thorsten Pachur, Linus Hof, Rebecca West, Sebastian Hellmann, Nuno Busch

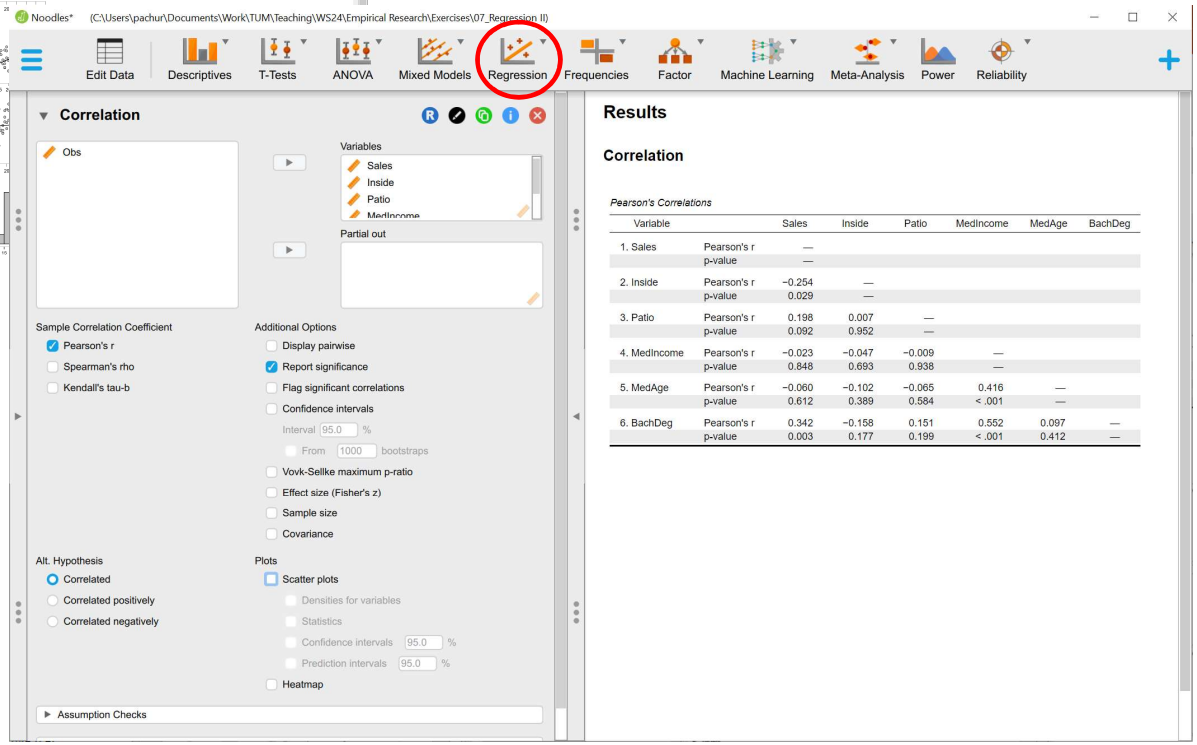
Technical University of Munich
School of Management
Chair of Behavioral Research Methods

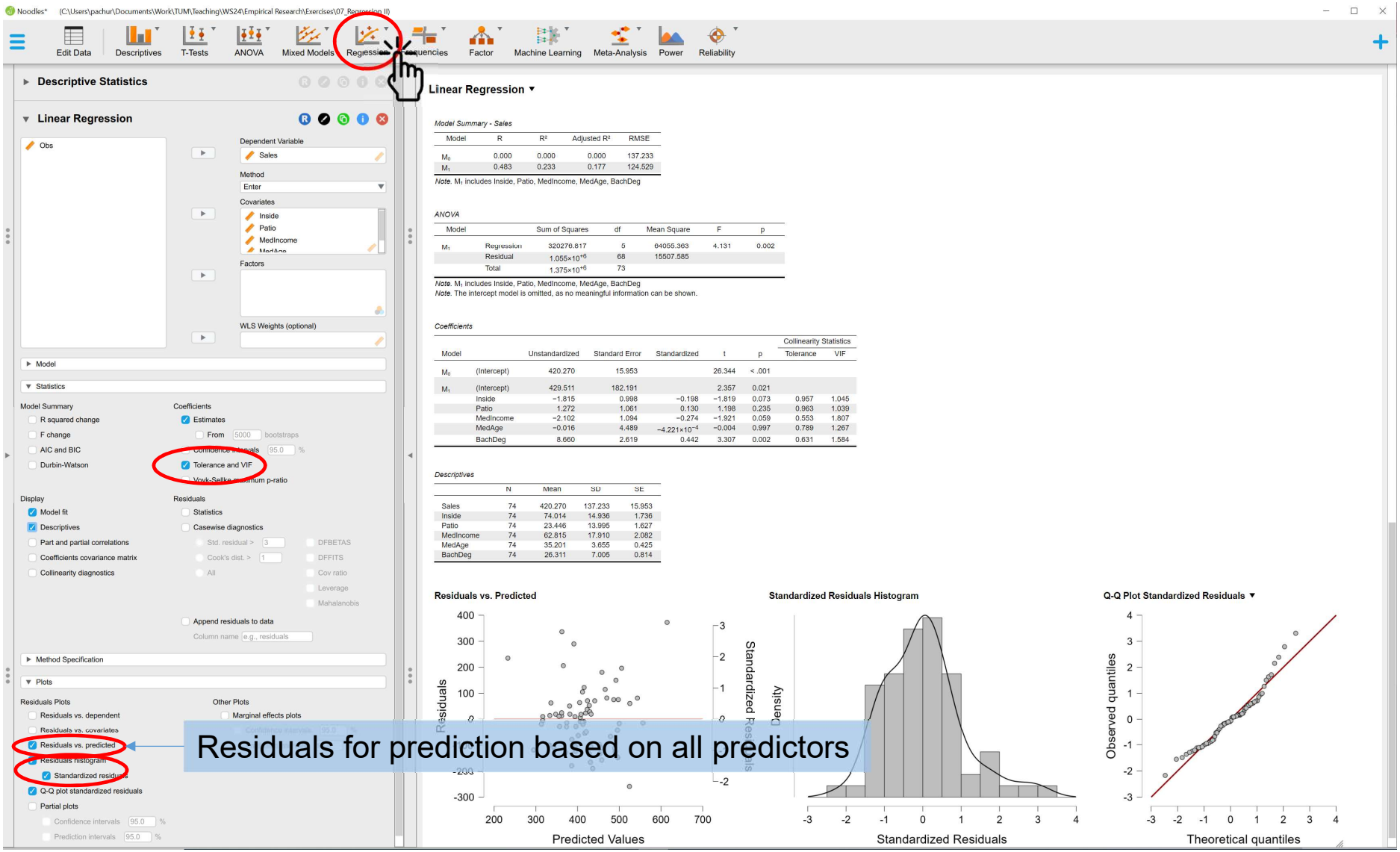


Exercise I

- Open dataset “Noodles.csv“
  →Annual sales (in $) per square foot of floor space as function of
  - Number of interior seats (Seats-Inside)
  - Number of outside seats (Seats-Patio)
  - Median family income in neighborhood (in thousand $) (MedIncome)
  - Median age in neighborhood (in years) (MedAge)
  - Percentage of population with at least a Bachelor‘s degree (BachDeg%)
- Which of the restaurant characteristics predict sales?
  - Run a multiple regression analysis
  - Check the assumptions (linearity, homoscedasticity, normality, multicollinearity)
  - Calculate **by hand** the standardized regression coefficients for the predictors “Seats-Inside“ and “Seats-Patio“


Edit Data
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Machine Learning
Meta-Analysis
Power
Reliability
Descriptive Statistics
Obs
Variables
Sales
Inside
Patio
MedIncome
MedAge
BachDeg
Split
Transpose descriptives table
Statistics
Basic plots
Distribution plots
Correlation plots
Display density
Display rug marks
Bin width type
Sturges
Number of bins
Interval plots
Q-Q plots
Pie charts
Dot plots
Categorical plots
Pareto plots
Pareto rule
Likert plots
Assume all variables share the same levels
Adjustable font size for vertical axis
Normal
Customizable plots
Tables
Results
Descriptive Statistics
Descriptive Statistics
Sales Inside Patio MedIncome MedAge BachDeg
Valid 74 74 74 74 74 74
Missing 0 0 0 0 0 0
Mean 420.270 74.014 23.446 62.815 35.201 26.311
Std. Deviation 137.233 14.936 13.995 17.910 3.655 7.005
Minimum 179.000 35.000 0.000 32.900 24.700 14.000
Maximum 987.000 124.000 78.000 114.400 43.500 40.000
Correlation plot
Sales
Inside
Patio
MedIncome
MedAge
BachDeg
Counts


Noodles* (C:\Users\pachur\Documents\Work\TUM\Teaching\WS24\Empirical Research\Exercises\07_Regression II)
Edit Data
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Machine Learning
Meta-Analysis
Power
Reliability
Correlation
Obs
Variables
Sales
Inside
Patio
MedIncome
Partial out
Sample Correlation Coefficient
Pearson's r
Spearman's rho
Kendall's tau-b
Additional Options
Display pairwise
Report significance
Flag significant correlations
Confidence intervals
Interval 95.0 %
From 1000 bootstraps
Vovk-Sellke maximum p-ratio
Effect size (Fisher's z)
Sample size
Covariance
Alt. Hypothesis
Correlated
Correlated positively
Correlated negatively
Plots
Scatter plots
Densities for variables
Statistics
Confidence intervals 95.0 %
Prediction intervals 95.0 %
Heatmap
Assumption Checks
Results
Correlation
Pearson's Correlations
Variable Sales Inside Patio MedIncome MedAge BachDeg
1. Sales Pearson's r — — — — — —
p-value — — — — — —
2. Inside Pearson's r −0.254 — — — — —
p-value 0.029 — — — — —
3. Patio Pearson's r 0.198 0.007 — — — —
p-value 0.092 0.952 — — — —
4. MedIncome Pearson's r −0.023 −0.047 −0.009 — — —
p-value 0.848 0.693 0.938 — — —
5. MedAge Pearson's r −0.060 −0.102 −0.065 0.416 — —
p-value 0.612 0.389 0.584 < .001 — —
6. BachDeg Pearson's r 0.342 −0.158 0.151 0.552 0.097 —
p-value 0.003 0.177 0.199 < .001 0.412 —


Edit Data
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Machine Learning
Meta-Analysis
Power
Reliability
Descriptive Statistics
Linear Regression
Obs
Dependent Variable
Sales
Method
Enter
Covariates
Inside
Patio
MedIncome
Factors
WLS Weights (optional)
Model
Statistics
Model Summary
R squared change
F change
AIC and BIC
Durbin-Watson
Coefficients
Estimates
From 5000 bootstraps
Tolerance and VIF
Display
Model fit
Descriptives
Part and partial correlations
Coefficients covariance matrix
Collinearity diagnostics
Residuals
Statistics
Casewise diagnostics
Append residuals to data
Method Specification
Plots
Residuals Plots
Residuals vs. dependent
Residuals vs. covariates
Residuals vs. predicted
Residuals histogram
Standardized residuals
Q-Q plot standardized residuals
Partial plots
Other Plots
Marginal effects plots
Linear Regression
Model Summary - Sales
Model R R² Adjusted R² RMSE
M0 0.000 0.000 0.000 137.233
M1 0.483 0.233 0.177 124.529
Note. M1 includes Inside, Patio, MedIncome, MedAge, BachDeg
ANOVA
Model Sum of Squares df Mean Square F p
M1 Regression 320276.817 5 64055.363 4.131 0.002
Residual 1.055×10^+6 68 15507.585
Total 1.375×10^+6 73
Note. M1 includes Inside, Patio, MedIncome, MedAge, BachDeg
Note. The intercept model is omitted, as no meaningful information can be shown.
Coefficients
Model Unstandardized Standard Error Standardized t p Collinearity Statistics Tolerance VIF
M0 (Intercept) 420.270 15.953 26.344 < .001
M1 (Intercept) 429.511 182.191 2.357 0.021
Inside −1.815 0.998 −0.198 −1.819 0.073 0.957 1.045
Patio 1.272 1.061 0.130 1.198 0.235 0.963 1.039
MedIncome −2.102 1.094 −0.274 −1.921 0.059 0.553 1.807
MedAge −0.016 4.489 −4.221×10^−4 −0.004 0.997 0.789 1.267
BachDeg 8.660 2.619 0.442 3.307 0.002 0.631 1.584
Descriptives
N Mean SD SE
Sales 74 420.270 137.233 15.953
Inside 74 74.014 14.936 1.736
Patio 74 23.446 13.995 1.627
MedIncome 74 62.815 17.910 2.082
MedAge 74 35.201 3.655 0.425
BachDeg 74 26.311 7.005 0.814
Residuals vs. Predicted
Residuals
Predicted Values
Standardized Residuals Histogram
Density
Standardized Residuals
Q-Q Plot Standardized Residuals
Observed quantiles
Theoretical quantiles
Residuals for prediction based on all predictors

Standardized regression coefficients

→ Allows for a comparison of regression coefficients between predictors

$$\beta_i = b_i \times \frac{SD_{X_i}}{SD_Y}$$

$$\widehat{Sales} = 429.511 - 1.815 \times Inside + 1.272 \times Patio$$

$$\beta_{inside} = -1.815 \times \frac{14.936}{137.233} = -.198$$

$$\beta_{patio} = 1.272 \times \frac{13.995}{137.233} = .130$$

$$z(\widehat{Sales}) = 0 - 0.198 \times z(Inside) + .130 \times z(Patio)$$

Intercept!

Interpretation of standardized regression coefficients: changes in terms of **standard deviations in z-transformed data**

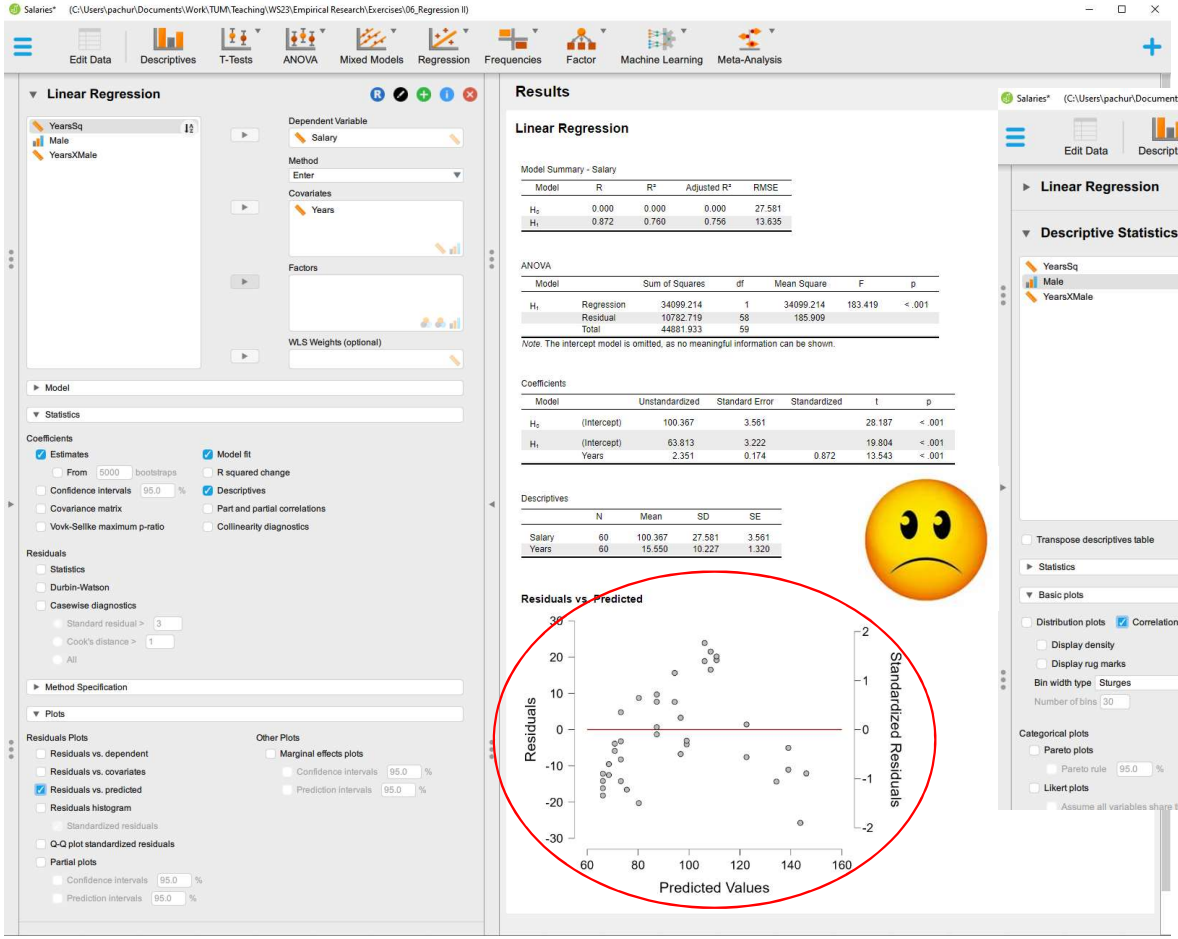
SD_{sales} = 137.233
SD_{inside} = 14.936
SD_{patio} = 13.99

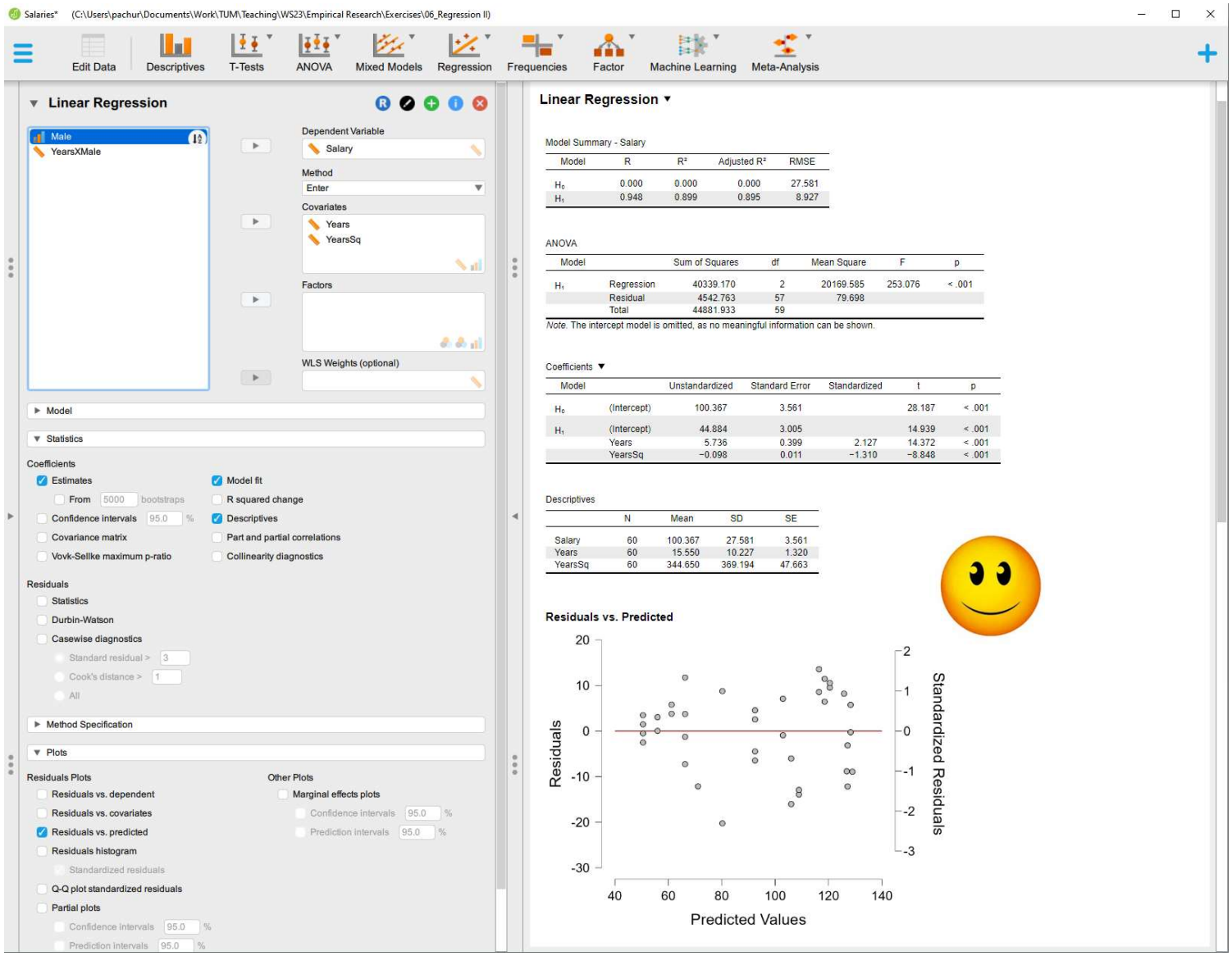
Exercise II: Dealing with nonlinearity

- Open dataset “Salaries.csv“ in JASP

  → Salaries of 60 engineers along with their years of experience
- Build a regression model to test whether and how salary is associated with years of experience
- Check for linearity
- Add squared years of experience (“YearsSq“) as predictor

  → Years of experience is only associated with higher salary up to a point and the effect then flattens


Edit Data
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Machine Learning
Meta-Analysis
Linear Regression
YearsSq
Male
YearsXMale
Dependent Variable
Salary
Method
Enter
Covariates
Years
Factors
WLS Weights (optional)
Model
Statistics
Coefficients
Estimates
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Residuals
Durbin-Watson
Casewise diagnostics
Method Specification
Plots
Residuals Plots
Residuals vs. dependent
Residuals vs. covariates
Residuals vs. predicted
Residuals histogram
Q-Q plot standardized residuals
Partial plots
Other Plots
Marginal effects plots
Results
Linear Regression
Model Summary - Salary
Model R R² Adjusted R² RMSE
H₀ 0.000 0.000 0.000 27.581
H₁ 0.872 0.760 0.756 13.635
ANOVA
Model Sum of Squares df Mean Square F p
H₁ Regression 34099.214 1 34099.214 183.419 < .001
Residual 10782.719 58 185.909
Total 44881.933 59
Note. The intercept model is omitted, as no meaningful information can be shown.
Coefficients
Model Unstandardized Standard Error Standardized t p
H₀ (Intercept) 100.367 3.561 28.187 < .001
H₁ (Intercept) 63.813 3.222 19.804 < .001
Years 2.351 0.174 0.872 13.543 < .001
Descriptives
N Mean SD SE
Salary 60 100.367 27.581 3.561
Years 60 15.550 10.227 1.320
Residuals vs. Predicted
Residuals
Standardized Residuals
Predicted Values
Descriptive Statistics
Variables
Split
Transpose descriptives table
Basic plots
Distribution plots
Correlation plots
Display density
Display rug marks
Bin width type
Sturges
Number of bins
Categorical plots
Pareto plots
Likert plots
Interval plots
Q-Q plots
Pie charts
Dot plots
Descriptive Statistics
Salary Years
Valid 60 60
Missing 0 0
Mean 100.367 15.550
Std. Deviation 27.581 10.227
Minimum 48.000 1.000
Maximum 134.000 35.000
Correlation plot
Salary
Years
Counts

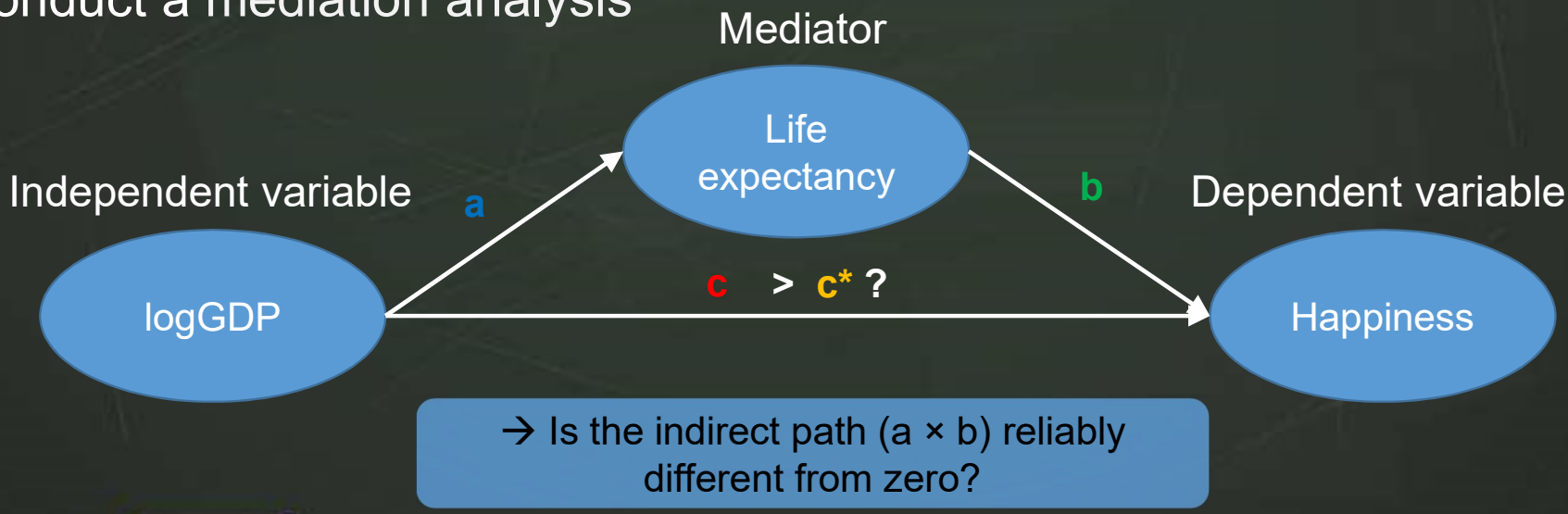

Salaries* (C:\Users\pachur\Documents\Work\TUM\Teaching\WS23\Empirical Research\Exercises\06_Regression II)
Edit Data
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Machine Learning
Meta-Analysis
Linear Regression
Male
YearsXMale
Dependent Variable
Salary
Method
Enter
Covariates
Years
YearsSq
Factors
WLS Weights (optional)
Model
Statistics
Coefficients
Estimates
From 5000 bootstraps
Confidence intervals 95.0 %
Covariance matrix
Vovk-Sellke maximum p-ratio
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Residuals
Statistics
Durbin-Watson
Casewise diagnostics
Standard residual > 3
Cook's distance > 1
All
Method Specification
Plots
Residuals Plots
Residuals vs. dependent
Residuals vs. covariates
Residuals vs. predicted
Residuals histogram
Standardized residuals
Q-Q plot standardized residuals
Partial plots
Confidence intervals 95.0 %
Prediction intervals 95.0 %
Other Plots
Marginal effects plots
Confidence intervals 95.0 %
Prediction intervals 95.0 %
Linear Regression
Model Summary - Salary
Model R R² Adjusted R² RMSE
H0 0.000 0.000 0.000 27.581
H1 0.948 0.899 0.895 8.927
ANOVA
Model Sum of Squares df Mean Square F p
H1 Regression 40339.170 2 20169.585 253.076 < .001
Residual 4542.763 57 79.698
Total 44881.933 59
Note. The intercept model is omitted, as no meaningful information can be shown.
Coefficients
Model Unstandardized Standard Error Standardized t p
H0 (Intercept) 100.367 3.561 28.187 < .001
H1 (Intercept) 44.884 3.005 14.939 < .001
Years 5.736 0.399 2.127 14.372 < .001
YearsSq −0.098 0.011 −1.310 −8.848 < .001
Descriptives
N Mean SD SE
Salary 60 100.367 27.581 3.561
Years 60 15.550 10.227 1.320
YearsSq 60 344.650 369.194 47.663
Residuals vs. Predicted
Residuals
Standardized Residuals
Predicted Values
20
10
0
-10
-20
-30
2
1
0
-1
-2
-3
40
60
80
100
120
140

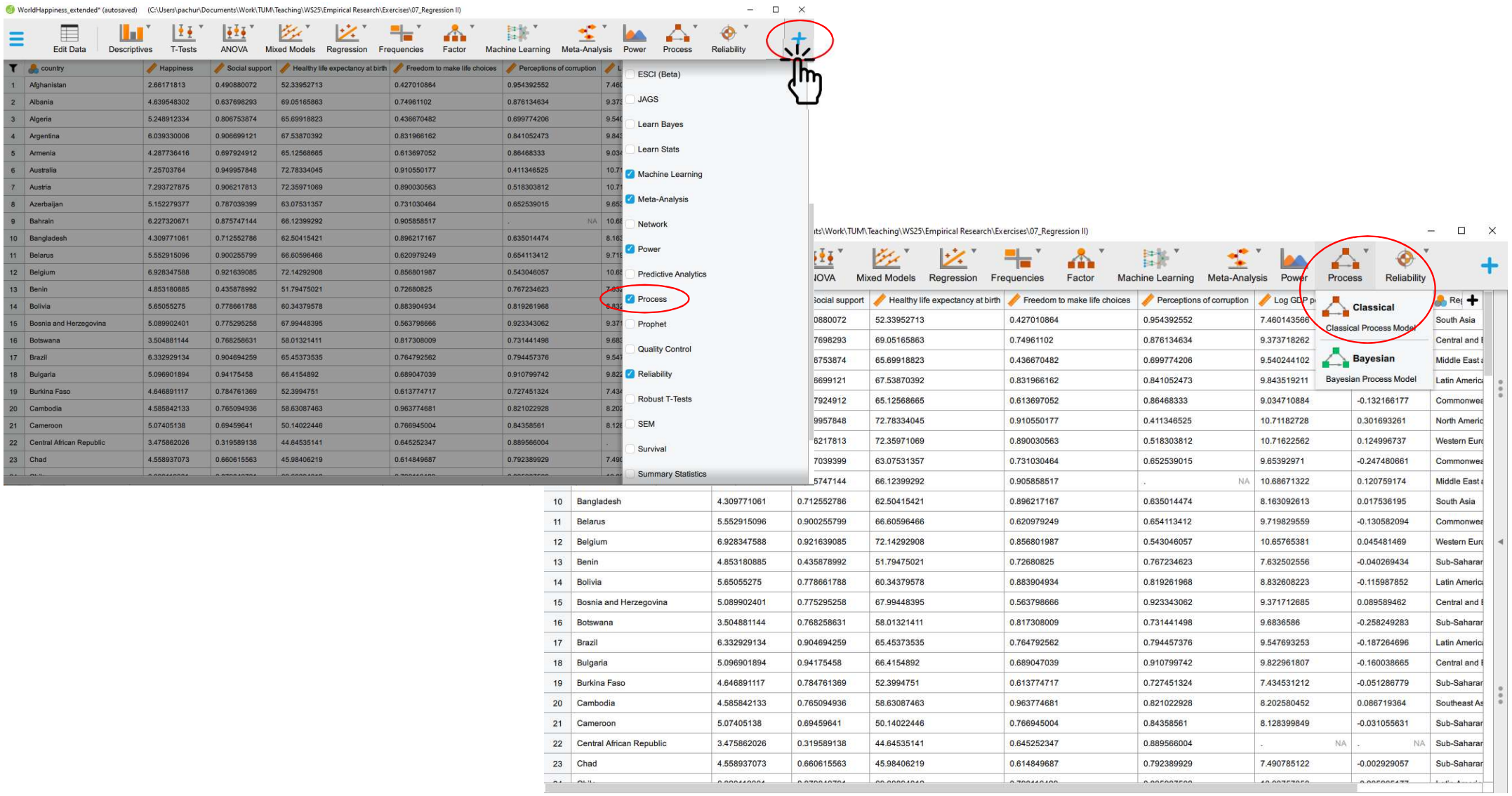
Exercise III: Mediation analysis

Open dataset “WorldHappiness_extended.csv“ in JASP

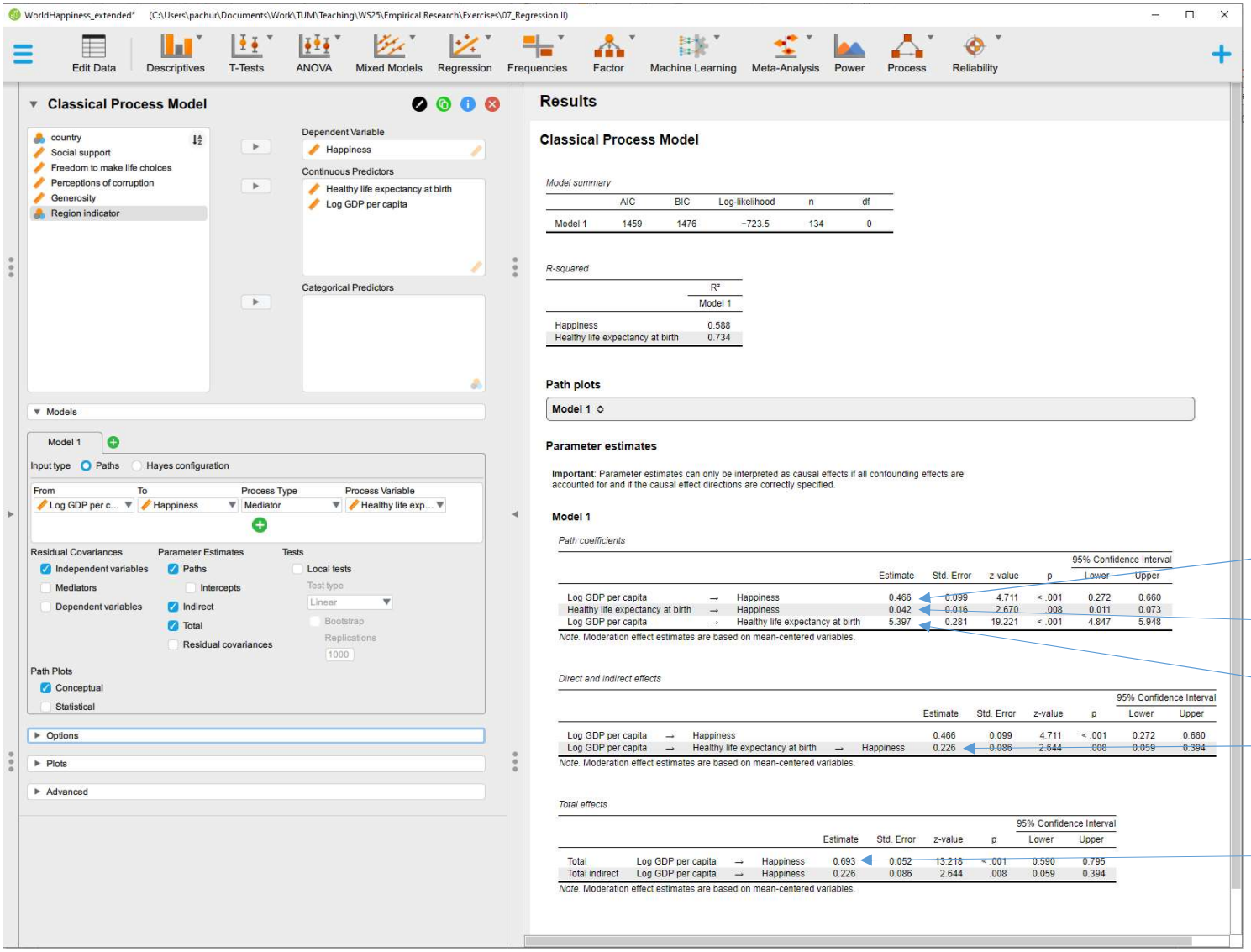
Is the effect of logGDP on happiness mediated by life expectancy (i.e., health conditions) in a country?

→ Conduct a mediation analysis





	country	Happiness	Social support	Healthy life expectancy at birth	Freedom to make life choices	Perceptions of corruption	Log GDP p		
1	Afghanistan	2.66171813	0.490880072	52.33952713	0.427010864	0.954392552	7.460143566		South Asia
2	Albania	4.639548302	0.637698293	69.05165863	0.74961102	0.876134634	9.373718262		Central and E
3	Algeria	5.248912334	0.806753874	65.69918823	0.436670482	0.699774206	9.540244102		Middle East a
4	Argentina	6.039330006	0.906699121	67.53870392	0.831966162	0.841052473	9.843519211		Latin America
5	Armenia	4.287736416	0.697924912	65.12568665	0.613697052	0.86468333	9.034710884	-0.132166177	Commonwea
6	Australia	7.25703764	0.949957848	72.78334045	0.910550177	0.411346525	10.71182728	0.301693261	North Americ
7	Austria	7.293727875	0.906217813	72.35971069	0.890030563	0.518303812	10.71622562	0.124996737	Western Eur
8	Azerbaijan	5.152279377	0.787039399	63.07531357	0.731030464	0.652539015	9.65392971	-0.247480661	Commonwea
9	Bahrain	6.227320671	0.875747144	66.12399292	0.905858517	. NA	10.68671322	0.120759174	Middle East a
10	Bangladesh	4.309771061	0.712552786	62.50415421	0.896217167	0.635014474	8.163092613	0.017536195	South Asia
11	Belarus	5.552915096	0.900255799	66.60596466	0.620979249	0.654113412	9.719829559	-0.130582094	Commonwea
12	Belgium	6.928347588	0.921639085	72.14292908	0.856801987	0.543046057	10.65765381	0.045481469	Western Eur
13	Benin	4.853180885	0.435878992	51.79475021	0.72680825	0.767234623	7.632502556	-0.040269434	Sub-Saharan
14	Bolivia	5.65055275	0.778661788	60.34379578	0.883904934	0.819261968	8.832608223	-0.115987852	Latin Americ
15	Bosnia and Herzegovina	5.089902401	0.775295258	67.99448395	0.563798666	0.923343062	9.371712685	0.089589462	Central and E
16	Botswana	3.504881144	0.768258631	58.01321411	0.817308009	0.731441498	9.6836586	-0.258249283	Sub-Saharan
17	Brazil	6.332929134	0.904694259	65.45373535	0.764792562	0.794457376	9.547693253	-0.187264696	Latin Americ
18	Bulgaria	5.096901894	0.94175458	66.4154892	0.689047039	0.910799742	9.822961807	-0.160038665	Central and E
19	Burkina Faso	4.646891117	0.784761369	52.3994751	0.613774717	0.727451324	7.434531212	-0.051286779	Sub-Saharan
20	Cambodia	4.585842133	0.765094936	58.63087463	0.963774681	0.821022928	8.202580452	0.086719364	Southeast As
21	Cameroon	5.07405138	0.69459641	50.14022446	0.766945004	0.84358561	8.128399849	-0.031055631	Sub-Saharan
22	Central African Republic	3.475862026	0.319589138	44.64535141	0.645252347	0.889566004	. NA	. NA	Sub-Saharan
23	Chad	4.558937073	0.660615563	45.98406219	0.614849687	0.792389929	7.490785122	-0.002929057	Sub-Saharan



Paths

c*

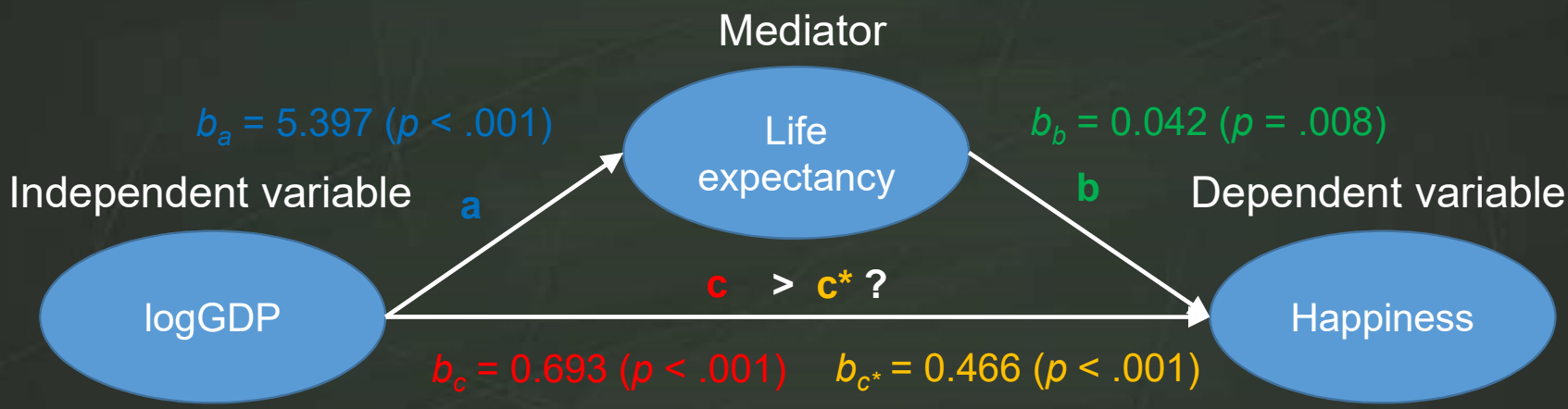
b

a

a × b

c

Exercise III: Mediation analysis



Statistical evaluation of the indirect path (a × b)

a × b = 0.226, $CI_{95\%}$ = [0.059, 0.394], p = .008