Research Methods

Regression

Laurens Rook (Delft University of Technology)

TU Delft

Delft University of Technology

Previous lecture

- The correlational method
- We discussed the difference between strong, moderate, and weak correlation coefficients
- We visually inspected and interpreted scatter plots (valence, form, direction and strength of the correlation)

Today

- Regression = linear (and multiple) regression as advanced correlational techniques

- Regression analysis as an advanced correlational technique -> advanced in the sense that it allows for prediction

Learning goals

- You will be able to explain regression analysis as advanced correlational technique
- You will be able to interpret the JASP output of a (linear and multiple) regression analysis

Linear regression

1.

Regression analysis

- **Regression analysis** = a procedure that allows us to predict an individual's score on one variable, based on knowing one (or more) other variable(s)

- Regression analysis involves determining the equation for the best-fitting line for a data set

  - **the regression line** --> a straight line that best represents the relationship between your variables

Regression line

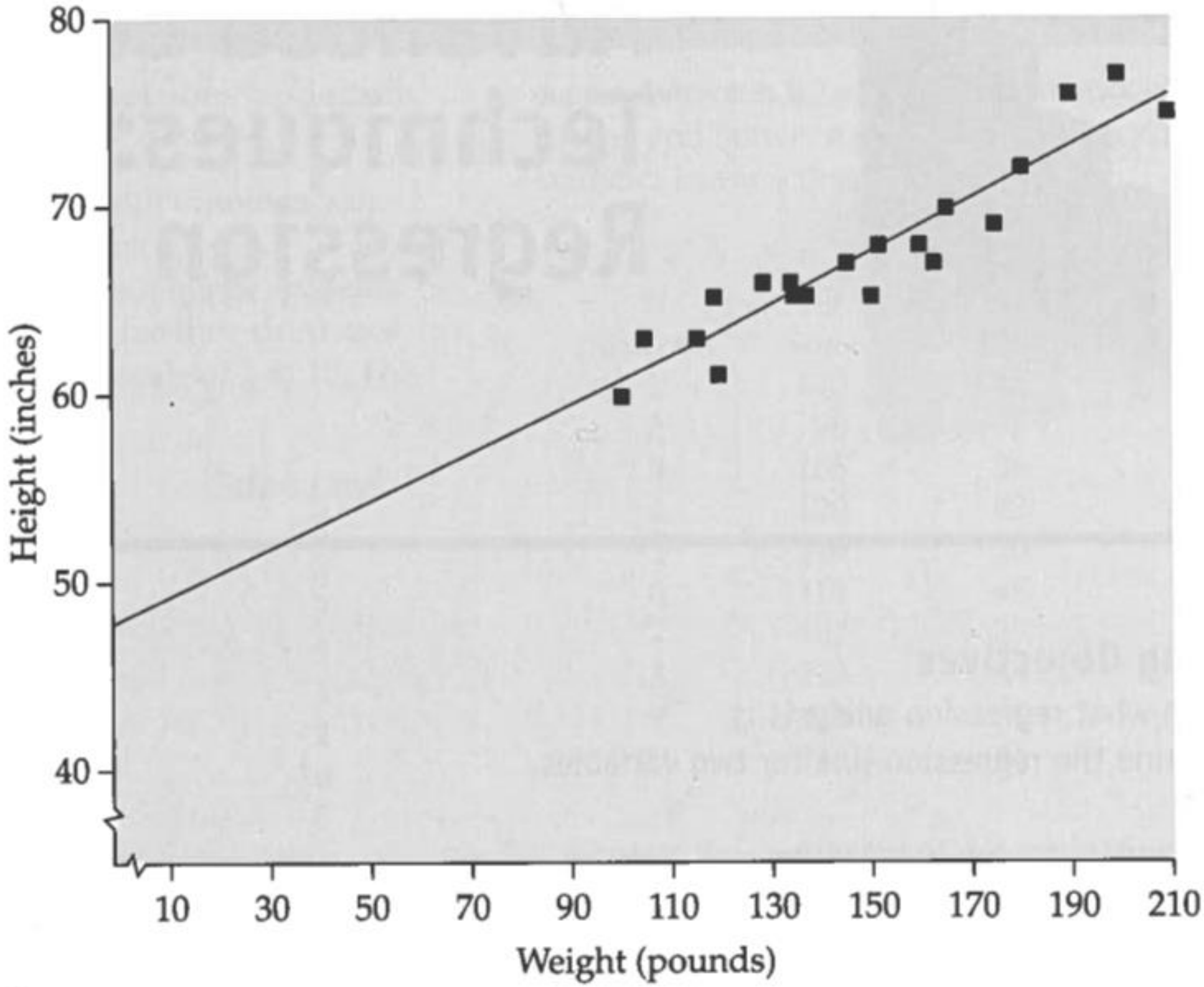
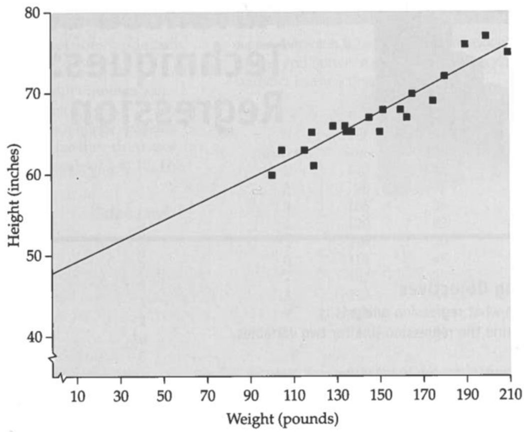


TABLE 20.1 Height and weight data for 20 individuals

WEIGHT (IN POUNDS)	HEIGHT (IN INCHES)
100	60
120	61
105	63
115	63
119	65
134	65
129	66
143	67
151	65
163	67
160	68
176	69
165	70
181	72
192	76
208	75
200	77
152	68
134	66
138	65
$\mu = 149.25$	$\mu = 67.4$
$\sigma = 30.42$	$\sigma = 4.57$

TU Delft

Simple regression model



- $Y' = a + bX + \varepsilon$

Y' = dependent (or outcome / response) variable (DV)

X = independent (or predictor) variable (IV)

a = y-intercept of the line (the point at which the line cuts through the y-axis)

b = slope of the line

ε = random error component (aka. white noise)

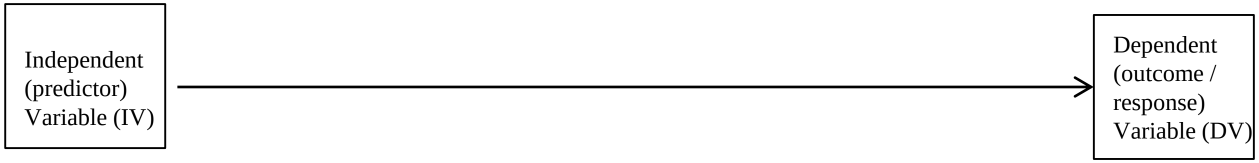
Prediction and regression

- If you are capable of calculating the equation for the regression line, you can use this line to predict from (outcome / response) variable (DV) to (predictor) variable (IV) -> **“to estimate the model”**

- The stronger the relationship between the two variables the more accurate the prediction

  - *that is, the stronger the correlation coefficient* the stronger the **prediction**

In a conceptual model



Linear regression analysis (JASP)

2.

The practice of linear regression in JASP

- In the following part, you will see how to conduct and interpret the output of a linear regression analysis in JASP

- As before, we will use the Chatbot data

Example

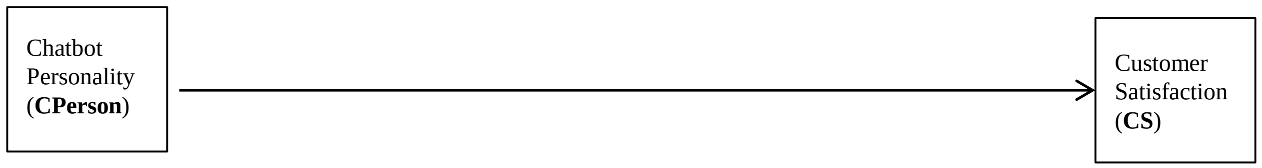
- We begin by formulating a working hypothesis:

A chatbot's personality (CPerson) positively impacts the user's customer satisfaction (CS) in human-chatbot interactions

H_0 and H_a

- The H_0 version of this working hypothesis states that the positive impact of CPerson on CS does not exist
- The H_a version of this working hypothesis states that this positive impact does exist (as properly formulated on the former slide)
- We now run a linear regression to test the H_0 against the H_a (where we, as always, seek to reject the H_0)

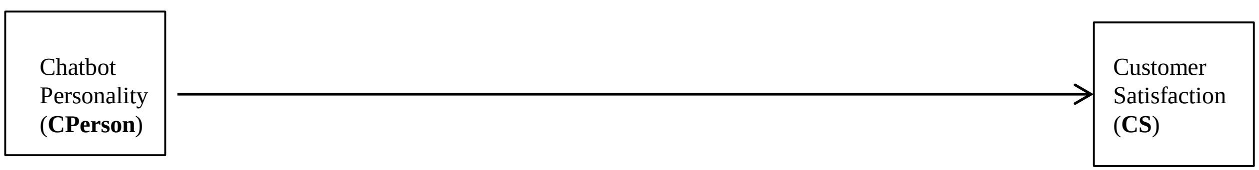
In a conceptual model



Predictor

Outcome / response

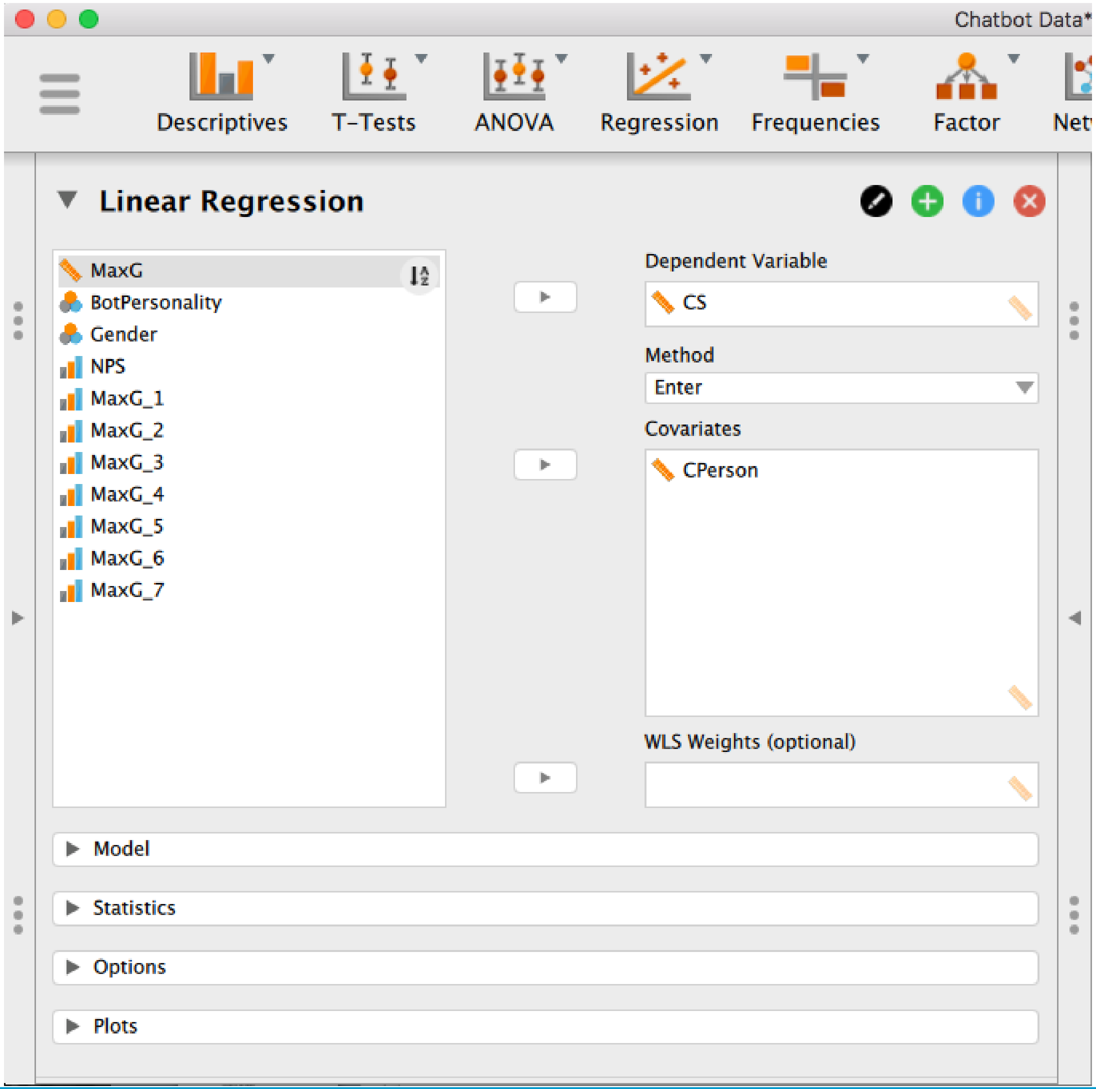
From conceptual to statistical model in JASP (how to put the variables in the right boxes)



The predictor is put in the "Coveriates" box

The response is put in the "Dependent Variable" box

In JASP


Chatbot Data*
Descriptives
T-Tests
ANOVA
Regression
Frequencies
Factor
Linear Regression
MaxG
BotPersonality
Gender
NPS
MaxG_1
MaxG_2
MaxG_3
MaxG_4
MaxG_5
MaxG_6
MaxG_7
Dependent Variable
CS
Method
Enter
Covariates
CPerson
WLS Weights (optional)
Model
Statistics
Options
Plots

Linear regression: interpreting the 3 output tables

- The linear regression command gives you 3 different output tables that allow you to interpret various aspects of the output:

    1. Model Summary
    2. ANOVA (Analysis of Variance)
    3. Coefficients

- The following slides show you how each of those look like & should be interpreted

Regression output

Results

Linear Regression

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.089	0.008	−0.000	0.761

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	0.550	1	0.550	0.950	0.332
	Residual	68.294	118	0.579		
	Total	68.843	119			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	(Intercept)	5.328	0.249		21.426	< .001
	CPerson	0.069	0.071	0.089	0.974	0.332

1. Interpreting the model summary table

- The value of R: (0.089) when you have only one predictor, this value is the correlation between IV and DV

- The value of R^2: (0.008) tells you that the IV accounts for 0.8% of the variation in the DV

  (This means that 99.2% of the variation in our DV is unaccounted for -> there may be better and more IVs to include in the model)

- Adjusted R^2: (-0.000) states how much of the variance in the DV would be accounted for had the model been taken from a different sample in the population (the loss of predictive power when the regression model was tested in a different sample)

2. Interpreting the ANOVA table

- The most important part of the ANOVA table is the *F*-statistic

- The *F*-statistic of 0.950 is associated with significance value of $p < 0.332$

- This tells you that there is less than a 33.2% chance that an *F*-statistic at least this large would happen if the null hypothesis were true -> our model thus does NOT predict well

3. Interpreting the coefficient table

- This table provides the model parameters (the beta values) and their significance
- The value of y-intercept *a*: (Intercept; 5.328) indicates that when the IV is unaccounted for, the model predicts 5.328 of CS
- The value of slope *b*: (0.069) if our IV is increased by 1 unit, our model predicts 0.069 more CS
    - --> so, the model is more predictive when our IV is unaccounted for. Not good!

TU Delft

3. Interpreting the coefficient table (II)

- This table also provides the model parameters (the beta values) and their significance

- The *t*-value (0.974) and the associated *p*-value (0.332) for CPerson: indicate whether the β-value from the previous slide is significantly different from zero. If the p-value is less than 0.05, they are significantly different from zero.

- In our case, they are not. **H_0 is accepted, and H_a is rejected.**

TU Delft

Conclusion of the linear regression analysis

- Our linear regression model does not predict an association between the IV and the DV

- The R, R^2 and adjusted R^2 indicate the lack of association, the non-significant F-statistic and t-statistic indicate that **the H_0 is true** (there is no effect)

- Thus: the positive impact of chatbot personality (Cperson) on customer satisfaction (CS) does not exist

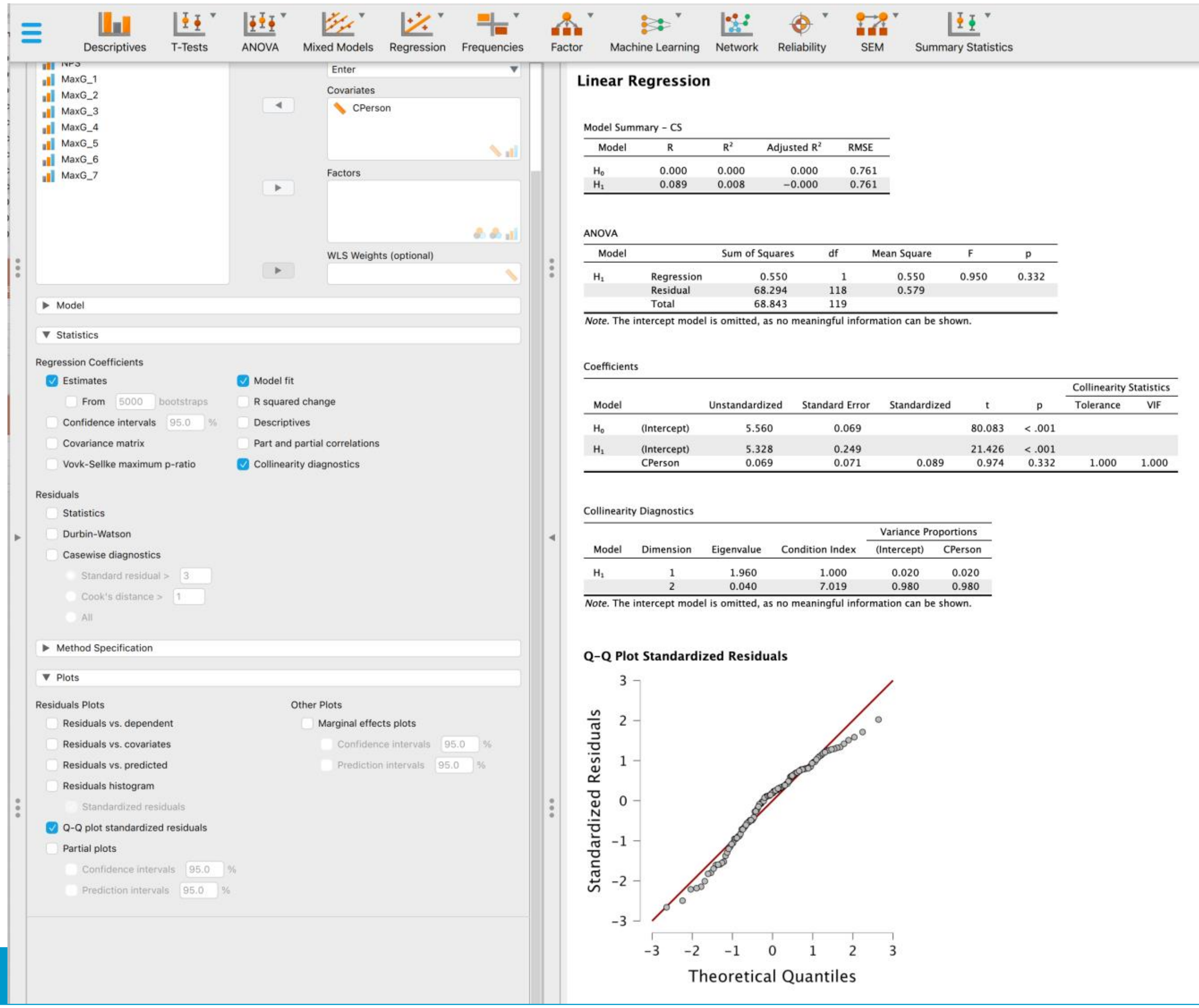
Prediction and regression (II)

- Differences between observed data and the predicted model -> **residuals**

- If observed data values and predicted values from the model have perfect fit: all values will be plotted on the same regression line

- Residuals can be found in regression output in numerical form and (visually) in residuals plots

In JASP



Model Summary – CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.089	0.008	−0.000	0.761

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	0.550	1	0.550	0.950	0.332
	Residual	68.294	118	0.579		
	Total	68.843	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

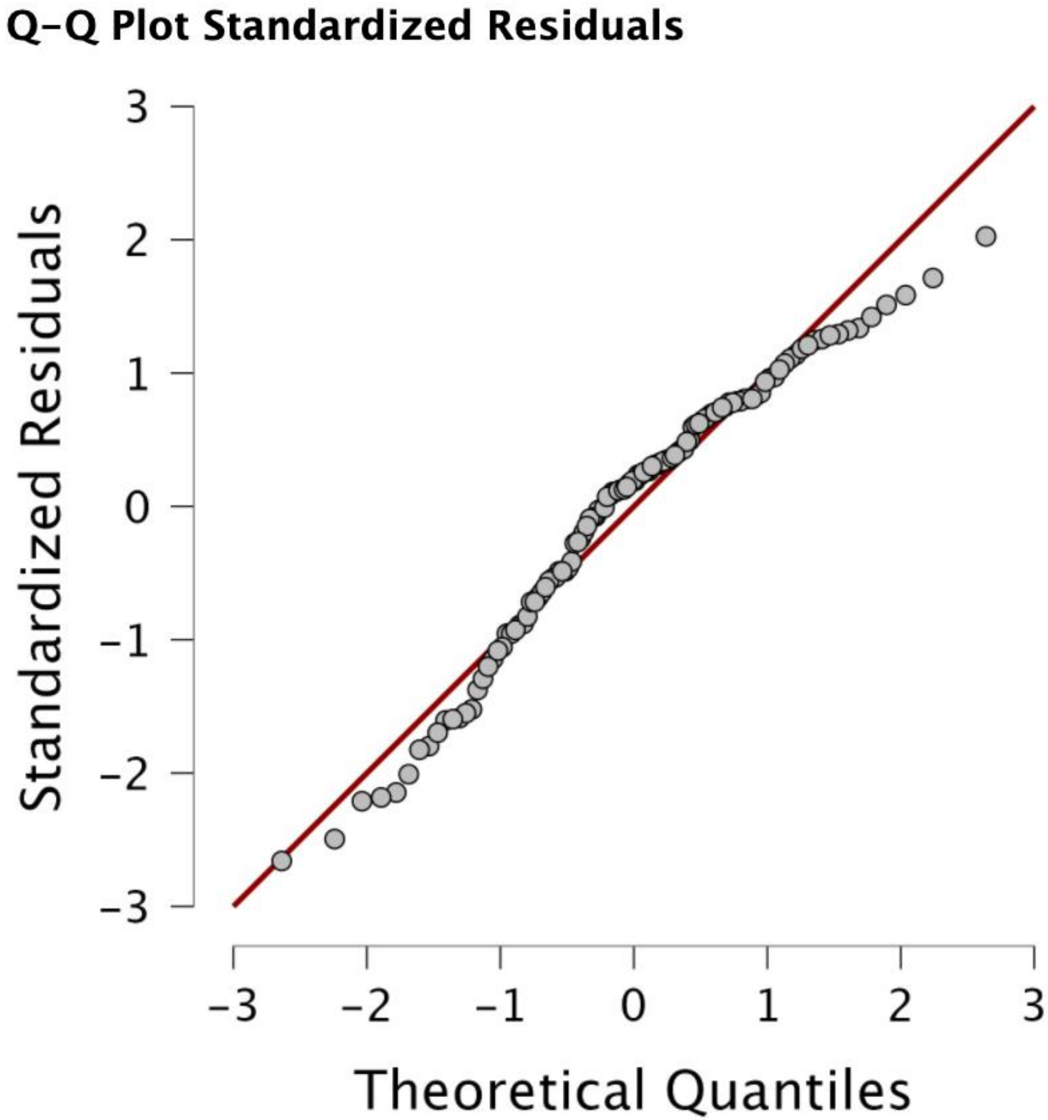
Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	5.328	0.249		21.426	< .001		
	CPerson	0.069	0.071	0.089	0.974	0.332	1.000	1.000

Collinearity Diagnostics

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Intercept)	CPerson
H_1	1	1.960	1.000	0.020	0.020
	2	0.040	7.019	0.980	0.980

Note. The intercept model is omitted, as no meaningful information can be shown.


Q-Q Plot Standardized Residuals
Standardized Residuals
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
Theoretical Quantiles

Interpreting the residuals plot

- The data values clearly are not nicely grouped on the regression line (they deviate from the line -> there's lots of residuals)

- What you see is that the data values for Cperson and the predicted values (visualized in the regression line) do not fit

Conclusion of the visual inspection of the linear regression

- Visual inspection of residuals plot confirms that the residuals are too large to fit our regression line
- Quite likely, our linear regression model does not predict an association between the IV (predictor) and the DV (output variable)

In exam form

- You could be asked to interpret a linear regression output as illustrated in Example 1:
    - the 3 regression output tables
    - and be asked whether H_0 and H_a must be accepted / rejected based on this

Hypothesis (H_a): NPS leads to a higher level of Customer Satisfaction (CS)

Linear Regression

Model Summary – CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.663	0.440	0.435	0.572

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	30.29	1	30.294	92.73	< .001
	Residual	38.55	118	0.327		
	Total	68.84	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	3.154	0.255		12.354	< .001		
	NPS	0.334	0.035	0.663	9.630	< .001	1.000	1.000

Linear Regression

Model Summary - CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.663	0.440	0.435	0.572

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	30.29	1	30.294	92.73	< .001
	Residual	38.55	118	0.327		
	Total	68.84	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	3.154	0.255		12.354	< .001		
	NPS	0.334	0.035	0.663	9.630	< .001	1.000	1.000

Hypothesis (H_a): NPS leads to a higher level of Customer Satisfaction (CS)

Linear Regression

Model Summary – CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.663	0.440	0.435	0.572

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	30.29	1	30.294	92.73	< .001
	Residual	38.55	118	0.327		
	Total	68.84	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	3.154	0.255		12.354	< .001		
	NPS	0.334	0.035	0.663	9.630	< .001	1.000	1.000

$F(1, 118) = 92.73$, $p < .001$ --> less than a 0.1% chance that an F-statistic at least this large would happen if the null hypothesis were true -> **our model thus predicts very well**

TU Delft

Hypothesis (H_a): NPS leads to a higher level of Customer Satisfaction (CS)

Linear Regression

Model Summary – CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.663	0.440	0.435	0.572

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	30.29	1	30.294	92.73	< .001
	Residual	38.55	118	0.327		
	Total	68.84	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	3.154	0.255		12.354	< .001		
	NPS	0.334	0.035	0.663	9.630	< .001	1.000	1.000

- b = 0.334 --> if the predictor increases by 1 unit, our model predicts 0.334 more CS
- $t(118) = 9.630$, $p < .001$ --> less than a 0.1% chance that a t-statistic at least this large would happen if the null hypothesis were true -> **our model is significant and predicts very well (H_a is accepted)**

Multiple regression

3.

What does multiple regression do?

- It is a type of linear regression that involves combining several (i.e., more than one) predictor variables into a single regression equation

- Why important?

- There may be more and other variables in our dataset we could add to our equation for more accurate prediction -> To better predict real life phenomena (which would lead to a larger R^2)

TU Delft

Multiple regression

- What if we have multiple predictors?
  - For example: age, income, education level, …
  - Multiple regression model

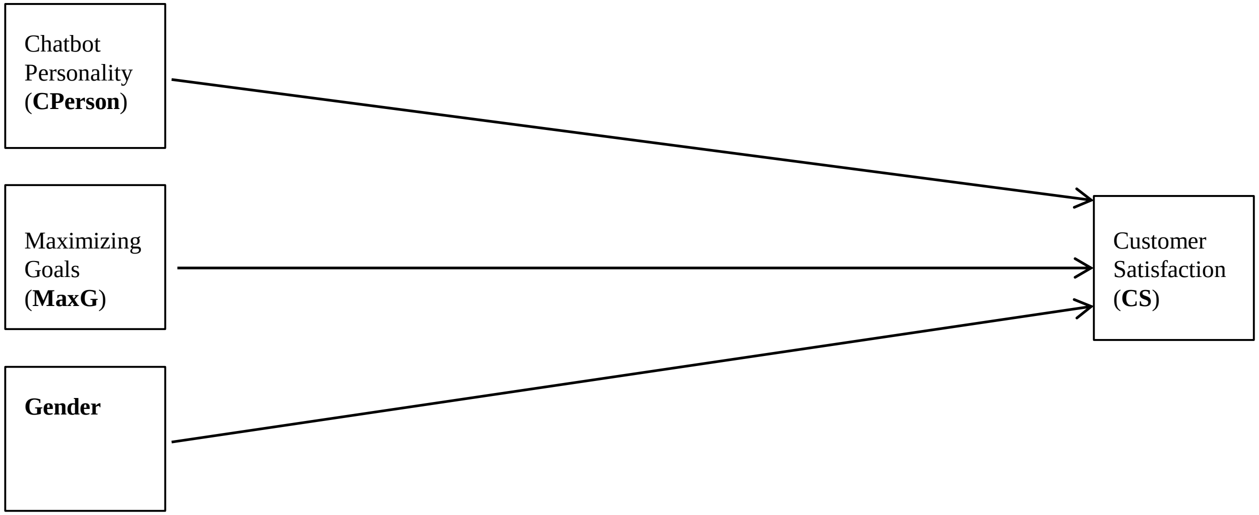
$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

  - $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are the parameters
  - ε is a random variable called the error term
  - Similar procedures as simple regression

What is the added value?

- Taking more predictors into account often leads to better (i.e., more accurate) prediction models
- This leads to a better understanding of the relationships among variables under study
- This is especially relevant in social science (& management) research, where perfect, single, predictors usually are hard to find

In a conceptual model



Example

- We begin by formulating a working hypothesis:

A chatbot's personality (CPerson), the user's gender (Gender), and maximizing personality (MaxG) each positively impact the user's customer satisfaction (CS) in human-chatbot interactions

In JASP

Results

Linear Regression

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.113	0.013	−0.013	0.765

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	0.883	3	0.294	0.502	0.681
	Residual	67.960	116	0.586		
	Total	68.843	119			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	(Intercept)	5.503	0.426		12.904	< .001
	CPerson	0.070	0.072	0.090	0.977	0.331
	MaxG	−0.042	0.095	−0.041	−0.443	0.659
	Gender	−0.090	0.148	−0.056	−0.611	0.543

1. Interpreting the model summary table

- The value of *R*: (0.113) when you have only one predictor, this value is the correlation between IV and DV

- The value of *R^2*: (0.013) tells you that the IV accounts for 1.3% of the variation in the DV

  (This means that 98.7% of the variation in our DV is unaccounted for -> there may be better and more IVs to include in the model)

- Adjusted *R^2*: (-0.013) indicates that we would not loose predictive power when the regression model was tested in a different sample

2. Interpreting the ANOVA table

- In the ANOVA table, we check out the *F*-statistic, and the related *p* value
- The *F*-statistic of 0.502 is associated with significance value of $p < 0.681$
- There is less than a 68.1% chance that an *F*-statistic at least this large would happen if the null hypothesis were true -> our model thus does NOT predict well

TU Delft

3. Interpreting the coefficient table

- The value of y-intercept *a*: (Intercept; 5.503) indicates that when the IVs are unaccounted for, the model predicts 5.503 of CS (again, y-intercept *a* predicts better than the *b* values below)

- The value of slopes b_1, b_2, b_3: (all 0.09 or smaller) if each IV is increased by 1 unit, our model predicts 0.07 more CS for b_1 (note that, for b_2, b_3 our model predicts 0.04 or 0.09 less CS)

3. Interpreting the coefficient table (II)

- This table also provides the model parameters (the beta values) and their significance

- The *t*-value (0.977) and the associated *p*-value (0.331) for Cperson is not significantly different from zero.

- The *t*-value (-0.443) and the associated *p*-value (0.659) for MaxG is not significantly different from zero.

- The *t*-value (-0.611) and the associated *p*-value (0.543) for Gender is not significantly different from zero.

Conclusion of the linear regression analysis

- Our multiple regression model does not predict an association between the IVs and the DV

- The *R*, R^2 and adjusted R^2 indicate the lack of association, the non-significant *F*-statistic and *t*-statistic indicate that the H_0 is true (there is no effect)

- --> The predicted positive impact of chatbot personality (Cperson), Gender, and maximizing personality (MaxG) on customer satisfaction (CS) does not exist

TU Delft

Example

<u>Question 1 (total score 3 points)</u>

Researchers investigated the impact of two personality meta-traits (Stability and Plasticity) on a person's susceptibility to authoritative persuasive messages (Authority). Two (alternative) hypotheses were tested:

> *Hypothesis 1*: People high (vs. low) in Stability are more susceptible to authoritative persuasive messages (Authority).
>
> *Hypothesis 2*: People high (vs. low) in Plasticity are more susceptible to authoritative persuasive messages (Authority).

a) On the next page, you find the JASP output of a regression analysis for this study. Do we accept or reject (alternative) *Hypothesis 1*? [open question; **0.5 point**]. Use the JASP regression output to motivate your answer [open question; **1 point**].

Hypothesis 1: People high (vs. low) in Stability are more susceptible to authoritative persuasive messages (Authority).

Linear Regression ▼

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.257	0.066	0.062	4.100

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	613.2	2	306.59	18.24	< .001
	Residual	8674.4	516	16.81		
	Total	9287.6	518			

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
1	(Intercept)	0.060	3.101		0.019	0.984		
	Stability	0.068	0.015	0.214	4.573	< .001	0.830	1.205
	Plasticity	0.039	0.023	0.080	1.706	0.089	0.830	1.205

TU Delft

Linear Regression ▼

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.257	0.066	0.062	4.100

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	613.2	2	306.59	18.24	< .001
	Residual	8674.4	516	16.81		
	Total	9287.6	518			

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
1	(Intercept)	0.060	3.101		0.019	0.984		
	Stability	0.068	0.015	0.214	4.573	< .001	0.830	1.205
	Plasticity	0.039	0.023	0.080	1.706	0.089	0.830	1.205

Hypothesis 2: People high (vs. low) in Plasticity are more susceptible to authoritative persuasive messages (Authority).

Linear Regression ▼

Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.257	0.066	0.062	4.100

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	613.2	2	306.59	18.24	< .001
	Residual	8674.4	516	16.81		
	Total	9287.6	518			

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
1	(Intercept)	0.060	3.101		0.019	0.984		
	Stability	0.068	0.015	0.214	4.573	< .001	0.830	1.205
	Plasticity	0.039	0.023	0.080	1.706	0.089	0.830	1.205

A small observation

- Researchers often turn to multiple regression rather than linear regression to increase the predictive accuracy of their model
- If you compare the stats for Example 1 (linear) and Example 2 (multiple), you see this principle illustrated (but very modestly)
- Admittedly, the increase is not strong enough for us to accept our H_a

Multicollinearity in multiple regression models

- Multicollinearity is an undesirable situation when one independent variable is a linear function of another independent variable

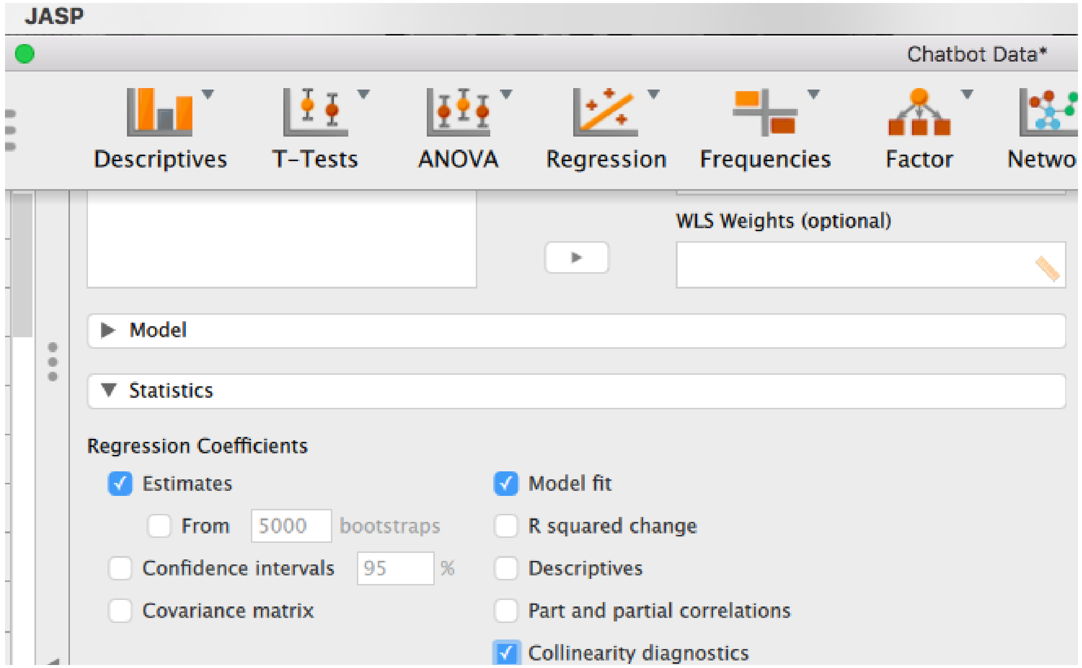
- This occurs when there is (too) strong correlation between the two variables. **Perfect multicollinearity** would be a correlation coefficient of ± 1.000

- it would be impossible to assign unique estimates of the regression coefficients in your prediction model --> i.e., you cannot assess the individual contribution of each predictor in your regression model

Identifying multicollinearity via the Variance Inflation Factor (VIF)

- The better way to identify multicollinearity is to look into the **Variance Inflation Factor** (VIF)
- The VIF indicates whether a predictor has a strong linear relationship with the other predictor(s). The VIF should be lower than 10 (VIF ≥ 10 is a problem)
- The connected **tolerance statistic** also can be used. If tolerance is ≤ 0.2 you have a problem

In JASP



Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	5.503	0.426		12.904	< .001		
	CPerson	0.070	0.072	0.090	0.977	0.331	0.995	1.005
	MaxG	−0.042	0.095	−0.041	−0.443	0.659	0.998	1.002
	Gender	−0.090	0.148	−0.056	−0.611	0.543	0.997	1.003

TU Delft

Interpretation

- In our case, all VIF values are <10; this is good
- In our case, all Tolerance values > 0.2; this is good
- Hence, no multicollinearity problems in our case

TUDelft

Extra: Linear regression with interaction terms

ABSTRACT

Exposure to a creative exemplar generally undermines creativity, because this invites part-set cuing inhibition. We tested and found that this effect occurs in difficult (vs. simple) tasks among actors with high (vs. low) Personal Need for Structure, and especially for individuals (vs. groups).

Creativity and Imitation in Individuals and Groups: Effects of Task Difficulty and Personal Need for Structure

Laurens Rook [1] & Daan van Knippenberg [2]

[1] Delft University of Technology; [2] Rotterdam School of Management, the Netherlands

Contact Details
T +31 (0)15 27 86330
F +31 (0)15 27 83177
E l.rook@tudelft.nl

ABSTRACT

Exposure to a creative exemplar generally undermines creativity, because this invites part-set cuing inhibition. We tested and found that this effect occurs in difficult (vs. simple) tasks among actors with high (vs. low) Personal Need for Structure, and especially for individuals (vs. groups).

Building on prior research on part-set cuing inhibition and creativity [1, 2] we reasoned that the presence of a creative exemplar product generally undermines people's creativity, because this invites imitation. We hypothesized that this mediating effect of imitation especially occurs in difficult (vs. simple) task settings that are more demanding, among actors with a high (vs. low) Personal Need for Structure, who are more averse to ill-defined problems [3], and thus more likely to rely on a creative exemplar product (Figure 1). We further expected this effect to be more pronounced for individuals than for groups, in which members also can back each other up.

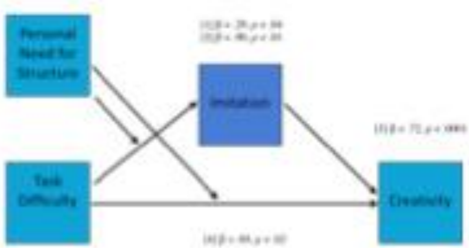
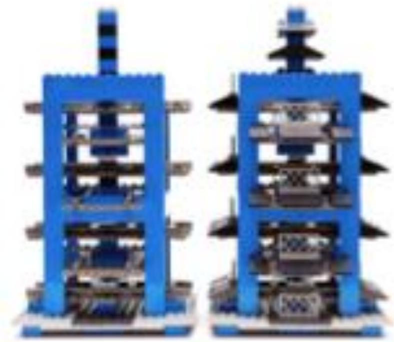


Figure 1. Partial Mediation Model [illegible]



METHODOLOGY

Two hundred and eighty seven students (N_{male} = 156, N_{female} = 131; M_{age} = 20.71, SD = 2.20) were randomly assigned either to groups of three or to the individual condition. Groups (N = 61) and individuals (N = 103) were randomly assigned to a (high versus low) task difficulty factorial design on creativity and imitation, to which Personal Need for Structure (NFS) was added.

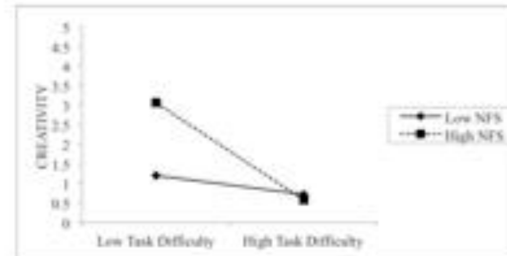
REFERENCES

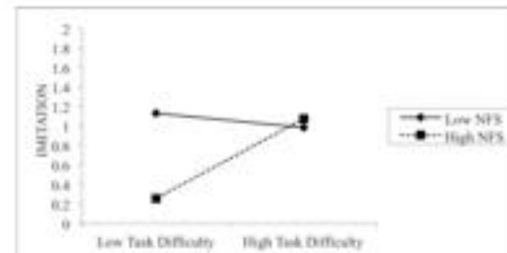
[1] Förster, J., Friedman, R. S., Özelsel, A., & Denzler, M. (2006). Enactment of approach and avoidance behavior influences the scope of perceptual and conceptual attention. *Journal of Experimental Social Psychology, 42*, 133-146.

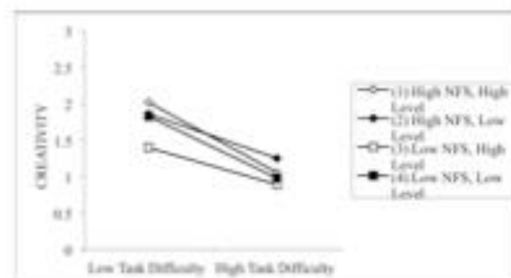
[2] Rook, L. & Van Knippenberg, D. (2011). Creativity and imitation: Effects of regulatory focus and creative exemplar quality. *Creativity Research Journal, 23*, 346-356.

[3] Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. B. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 19-39). Mahwah, NJ: Lawrence Erlbaum.

RESULTS







TU Delft

Why interaction terms?

- The two previous slides illustrate what you can do with multiple regression, if you also compute cross-products of your predictor variables (just an appetizer)

- How would you determine to compute such cross-products in your own data?

  - -->If prior study (theory) has shown the existence of such cross-products, you may decide to add them to your regression model

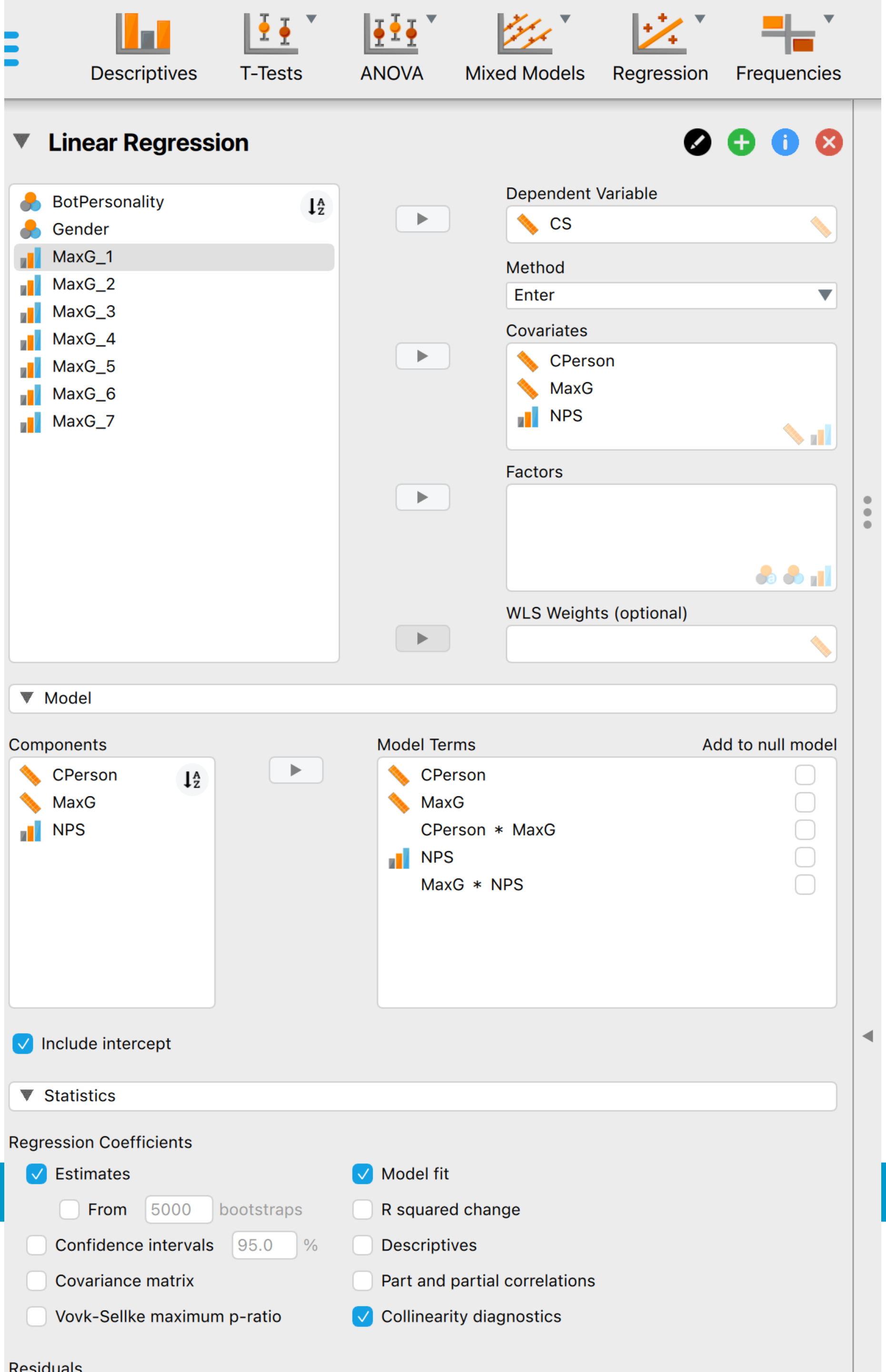
TU Delft

Interaction terms and multicollinearity

- When you include interaction terms in your regression model (indicated with * in a JASP output):
- --> you increase the likelihood that you will have multicollinearity in your data!
- So, it becomes the more important to assess this

In JASP

You build your interaction terms for your regression model here -->



Results

Linear Regression

Model Summary - CS

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.761
H_1	0.670	0.448	0.424	0.577

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	30.87	5	6.175	18.54	< .001
	Residual	37.97	114	0.333		
	Total	68.84	119			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
H_0	(Intercept)	5.560	0.069		80.083	< .001		
H_1	(Intercept)	1.346	1.416		0.950	0.344		
	CPerson	0.176	0.240	0.228	0.733	0.465	0.050	19.90
	MaxG	0.505	0.399	0.492	1.267	0.208	0.032	31.20
	NPS	0.484	0.161	0.962	3.000	0.003	0.047	21.26
	CPerson ∗ MaxG	−0.045	0.066	−0.248	−0.685	0.495	0.037	27.17
	MaxG ∗ NPS	−0.044	0.046	−0.429	−0.969	0.335	0.025	40.62

TU Delft

Interpretation

- The collinearity statistics for this model output indicate that:
- In our case, all VIF values are ≥ 10; this is a problem
- In our case, all Tolerance values ≤ 0.2; this is a problem
- Hence, structural multicollinearity problems in our case

In exam form

- You could be asked to interpret a multiple regression output as illustrated above:
  - the 3 regression output tables
  - and be asked whether H_0 and H_{alt} must be accepted / rejected based on this
  - the VIF and Tolerance values, and assess whether we have multicollinearity issues (or not)

TU Delft

Good luck!