Moderation analysis presentation

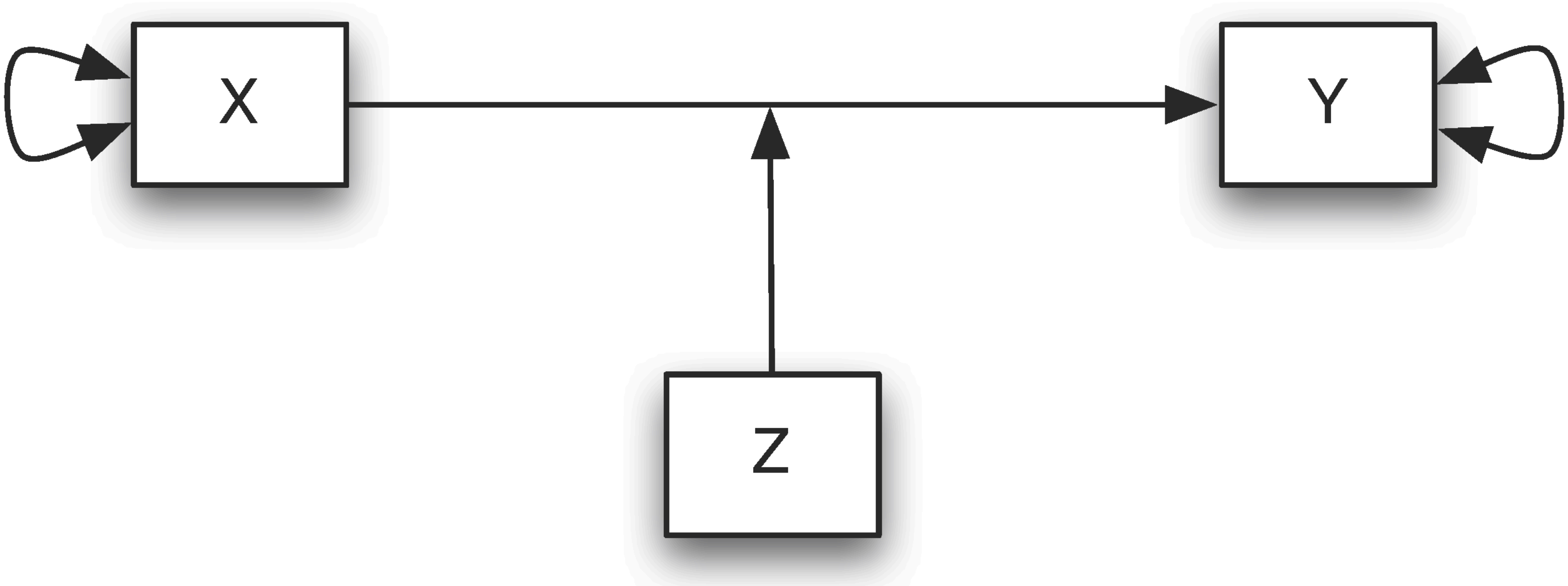
Learn what moderation analysis is and how to run the analysis in R

UZH
IKMZ

1 Moderation analsyis

1.1 Definition

Moderation analysis allows us to test for the influence of a third variable, Z, on the relationship between variables X and Y. Rather than testing a causal link between these other variables, moderation tests for when or under what conditions an effect occurs.



UZH
IKMZ

1.2 Interaction

Moderators can strengthen, weaken, or reverse the nature of a relationship. Moderation can be tested by looking for significant interactions between the moderating variable (Z) and the independent variable (X).

Nota bene: Moderators (when) are conceptually different from mediators (how/why) but some variables may be a moderator or a mediator depending on your question.

UZH
IKMZ

1.3 Similarity with ANOVA

The moderation is an interaction and therefore comparable to the interaction in ANOVA. If X and moderator are dichotomous, the moderation corresponds to a 2x2 ANOVA.

However, a moderator moderates the causal relationship from X to Y. The scale level can be dichotomous, categorical or metric. Furthermore, a moderator must be causally independent of X.

1.4 Illustration

See Igartua and Hayes (2021) for a complete discussion.

Outcome: Feeling toward Immigrants

Mean Feeling toward Immigrants (Higher = More Positive) as a Function of Similarity and Narrative Voice Conditions

	Similarity (X)		
Narrative Voice (Z)	**Dissimilar**	**Similar**	**Unweighted Mean**
Third-Person	Y1 = 51.545	Y2 = 47.098	Y12 = (Y1+Y2)/2 = 49.321
	n1 = 112	n2 = 112	
First-Person	Y3 = 47.009	Y4 = 55.117	Y34 = (Y3+Y4)/2 = 51.063
	n3 = 108	n4 = 111	
Unweighted Mean	Y13 = (Y1+Y3)/2 = 49.277	Y24 = (Y2+Y4)/2 = 51.108	

n: 443

df: 439

Main effect of Similarity (X): Y24 - Y13 = 51.108 – 49.277 = 1.831, F(1, 439) = 0.558, p = .455.

Main effect of Voice (Z): Y12 - Y34 = 51.063 – 49.321 = 1.742, F(1, 439) = 0.505, p = .478.

Interaction (XZ): (Y4 - Y3) - (Y2 - Y1) = (55.117 – 47.009) – (47.098 – 51.545) = 12.555, F(1, 439) = 6.563, p = .011.

Moderation Analysis of a 2X2 Design Using a Main Effects (–0.5/0.5 Coding) and a Simple Effects (0/1 coding) Parameterization.

	Main effect			Simple effect		
	Coef.	**S.E.**	**p**	**Coef.**	**S.E.**	**p**
Constant (a)	50.192	1.225	< 0.01	51.545	2.436	< 0.01
Similarity (X) -> b1	1.831 = Y24 - Y13	2.450	.455	–4.446 (when Z=0, Y2-Y1)	3.445	.198
Narrative Voice (Z) -> b2	1.742 = Y34 - Y12	2.450	.478	–4.535 (when X=0, Y3-Y1)	3.447	.193
XZ -> b3	12.555 = (Y4 - Y3) - (Y2 - Y1)	4.900	.011	12.555	4.900	.011
Multiple R		.131			.131	

Conditional (simple) effects of Similarity

	Effect	**SE**	**p**
Third-Person Voice	–4.446 = Y2 - Y1	3.445	.198
First-Person Voice	8.108 = Y4 - Y3	3.485	.020

Nota bene: simple effect parametrization implies that b1 and b2 represetn two of the four simple effects in the 2X2 ANOVA rather than the main effect of X and Z

1.5 How-To

Technically, moderations (interactions) are linked multiplicatively in the regression analysis: A x B.

Statistically, the moderator and X must always be considered “in isolation” (not just as moderation or interaction).

Moderation analysis can be conducted by adding one (or multiple) interaction terms in a regression analysis. For example, if Z is a moderator for the relation between X and Y, we can fit a regression model:

$$Y = \beta_0 + \beta_1 * X + \beta_2 * Z + \beta_3 * XZ + \epsilon$$

$$Y = \beta_0 + \beta_2 * Z + (\beta_1 + \beta_3 * Z) * X + \epsilon$$

Thus, if β_3 is not equal to 0, the relationship between X and Y depends on the value of Z, which indicates a moderation effect.

1.6 Binary moderator

$$Y = \beta_0 + \beta_2 * Z + (\beta_1 + \beta_3 * Z) * X$$

If Z is a dichotomous/binary variable (e.g. gender) the above equation can be written as:

$$\beta_0 + \beta_1 * X + \epsilon$$

for male (Z=0)

$$\beta_0 + \beta_2 + (\beta_1 + \beta_3) * X + \epsilon$$

for female (Z=1)

UZH
IKMZ

2 Interpretation and centering

UZH
IKMZ

2.1 Interpretation

If X and/or moderator become significant, main effects are present. If the moderation term becomes significant, there is a moderation effect. The (possibly significant) influences of X and the moderator are then so-called “conditional” effects.

The value 0 usually has no meaningful meaning (e.g. in rating scales 1 to 5 there is no zero at all). Therefore, it is a good practice to centering means subtracting the overall mean from each value.

UZH
IKMZ

2.2 Effect of centering on the interpretation

The change in meaning must be taken into account in the interpretation:

- influence of the predictor on Y with an average expression of the moderator
- influence of the moderator on Y with an average expression of the predictor

UZH
IKMZ

2.3 Variable's effect as a function of a moderator

Let's assume that X and Z are either dichotomous or continuous and the outcome variable Y is a continuous dimension suitable for analysis with linear regression, we have the following equations:

$$\hat{Y} = a + b_1 * X + b_2 * Z + b_3 * XZ = a + (b_1 + b_3 * Z) * X + b_2 * Z$$

In this representation, the weight for X is not a single number but, rather, a function of Z: $b_1 + b_3 * Z$.

The output is sometimes called the simple slope of X or the conditional effect of X.

The coefficients b_1 and b_2 may or may not have a substantive interpretation, depending on how X and Z are coded or, in the case of dichotomous X and Z, what two numbers are used to represent the groups in the data.

2.4 Correct interpretations

- b_1 is the conditional effect of X on Y when Z = 0.
- b_1 is the estimated difference in Y between two cases in the data that differ by one unit in X but have a value of 0 for Z.
- b_2 is the conditional effect of Z on Y when X = 0.
- When X = 0, the conditional effect of Z reduces to b_2.
- When Z = 0, X's effect equals b_1; when Z = 1, X's effect equals $b_1 + b_3$; when Z = 2, X's effect is $b_1 + 2b_3$, and so forth.

2.5 Multicollinearity

The moderation term is formed with moderator and X. However, the moderator and X are also contained individually in the regression equation. This often leads to multicollinearity (i.e. low tolerances or high VIF values).

If moderator and X are centered, the symptoms of multicollinearity are superficially defused. However, the multicollinearity itself remains.

Multicollinearity means that predictors are (too strongly) related to each other. Since the moderation term also consists of A (or B), it can easily correlate with A (or B).

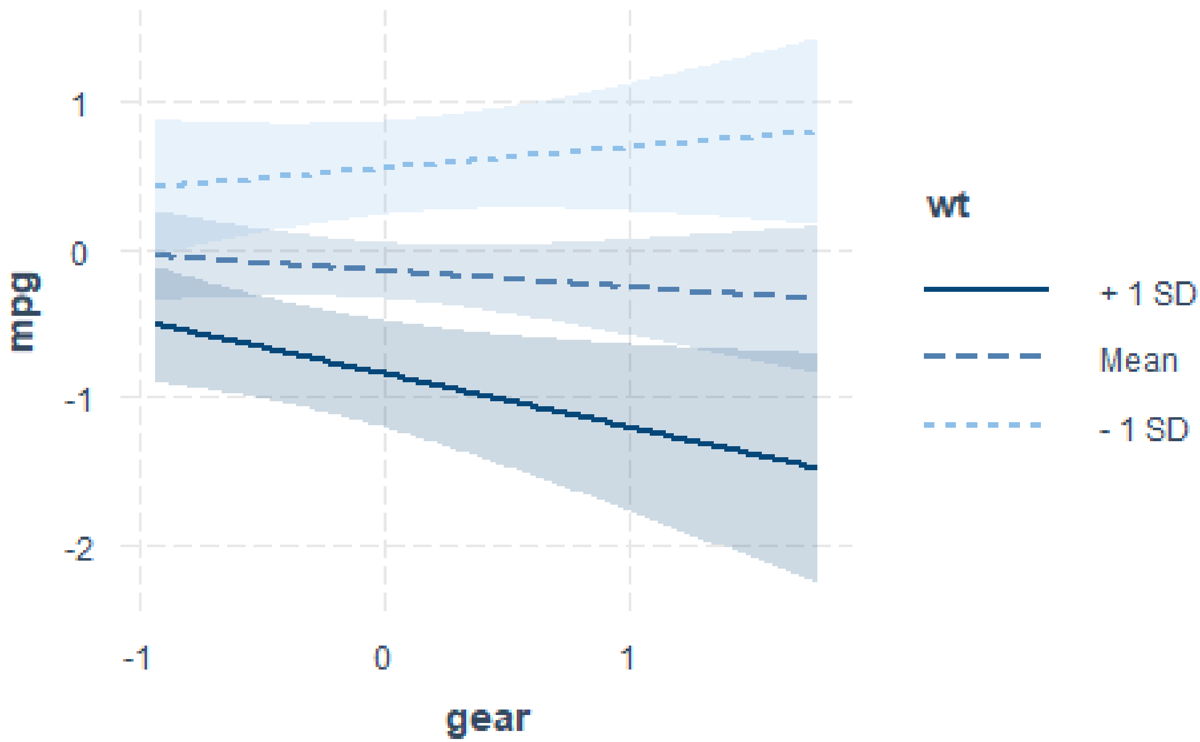
UZH
IKMZ

3 Visualization of the moderation effect

3.1 Simple slopes

When the moderation effect becomes significant, it needs to be "illustrated" in order to make it interpretable. To do so, we can rely on simple slope analysis: comparison of the regression lines for low, medium and high levels of the moderator.

Typically, we use the mean value of the moderator, as well as the values + and - 1 SD are used, but theoretically any values can be considered.

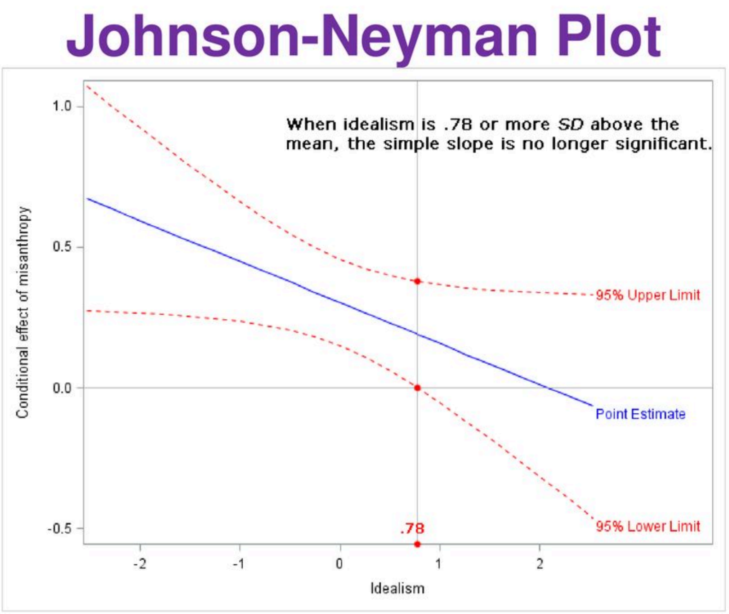


UZH
IKMZ

3.2 Significance range: Johnson & Neyman

This method suggests to conduct comparison of the regression equation for many characteristics of the moderator to identify areas of significance.

It is more useful than simple slopes for metric moderators, since illustrations result in less loss of information in the moderator's levels. The effect (b) of the X on Y is now illustrated in the diagram as a function of the moderator (not to be confused with a regression line!), as well as the confidence interval for this effect:



UZH IKMZ

4 Steps in the analysis

4.1 Five main steps

A moderation analysis typically consists of the following steps:

- compute the interaction term X*Z
- fit a multiple regression model with X, Z, and XZ as predictors
- test whether the regression coefficient for XZ is significant or not
- interpret the moderation effect
- display the moderation effect graphically.

UZH
IKMZ

5 Example in R

5.1 Example: news attention as moderator

For example, trust in politicians (e.g., trust in national parties) could moderate the effect of citizens' media diet (e.g., attention to political news) on political participation in election.

We will be using the Selects Swiss Panel Election Study 2019 and the following variables:

- participation (in W3): W3_f11100
- trust in politicians (in W3, assumed similar in W2): W3_f12800c
- media diet (in W2): W2_f13400a-f

5.2 Prepare the data

Let's load the data and select the relevant variables:

```
library(foreign)
db <- read.spss(file=paste0(getwd(),
                "/data/1184_Selects2019_Panel_Data_v4.0.sav"),
                use.value.labels = T,
                to.data.frame = T)
sel <- db |>
  dplyr::select(W3_f11100,W3_f12800c,
                W2_f13400a,W2_f13400b,W2_f13400c,W2_f13400d,
                W2_f13400e,W2_f13400f) |>
  stats::na.omit() |>
  dplyr::rename("particip"="W3_f11100",
                "party_trust"="W3_f12800c",
                "tv"="W2_f13400a",
                "newsp"="W2_f13400b",
                "freen"="W2_f13400c",
                "socmed"="W2_f13400d",
                "online"="W2_f13400e",
                "radio"="W2_f13400f")
```

5.3 Recoding and creation of media score

```
# recode participation
sel$particip <- ifelse(sel$particip=="Voted",1,0)
sel$particip <- as.factor(sel$particip)
# trust as numeric
sel$party_trust <- as.character(sel$party_trust)
sel$party_trust[sel$party_trust=="Full trust"] <- "10"
sel$party_trust[sel$party_trust=="No trust"] <- "0"
sel$party_trust <- as.numeric(sel$party_trust)
# trust as binary
sel$party_trust_b <- ifelse(sel$party_trust>=6, "yes","no")
# reverse scale for media attention and build score
sel$tv <- (as.numeric(sel$tv)-4)*(-1)
sel$newsp <- (as.numeric(sel$newsp)-4)*(-1)
sel$freen <- (as.numeric(sel$freen)-4)*(-1)
sel$socmed <- (as.numeric(sel$socmed)-4)*(-1)
sel$online <- (as.numeric(sel$online)-4)*(-1)
sel$radio <- (as.numeric(sel$radio)-4)*(-1)
sel$medatt <- (sel$tv+sel$newsp+sel$freen+sel$socmed+sel$online+sel
```

5.4 Variable centering

When Z and X are numeric, it is important to mean center both your moderator and your independent variable to reduce multicolinearity and to make interpretation easier. Centering can be done using the scale() function, which subtracts the mean of a variable from each value in that variable.

```
# scale trust and medatt
sel$party_trust <- as.numeric(scale(sel$party_trust, center=T, scal
sel$medatt <- as.numeric(scale(sel$medatt, center=T, scale=F))
```


UZH
IKMZ

5.5 Moderation model

We can compute moderation analysis directly with the glm() function but we can also rely on the PROCESS macro developed by Hayes. You can download PROCESS for R here.

```
mod1 <- glm(particip ~
              party_trust + medatt +
              party_trust*medatt,
            data=sel,
            family = "binomial")
stargazer::stargazer(mod1, type="text", si
```

```
=============================================
                     Dependent variable:
                 ----------------------------
                          particip
---------------------------------------------
party_trust            0.095*** (0.022)
medatt                 0.895*** (0.084)
party_trust:medatt     -0.071* (0.041)
Constant               1.450*** (0.042)
---------------------------------------------
Observations                3,940
Log Likelihood            -1,902.390
Akaike Inf. Crit.          3,812.780
=============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
# source("./process/PROCESS v4.3 for R/prc
# sel$particip <- as.numeric(sel$particip)
# process(data = sel,
#         y = "particip",
#         x = "party_trust",
#         w ="medatt",
#         model = 1) # 1: for simple moder
```

See output on next slide.

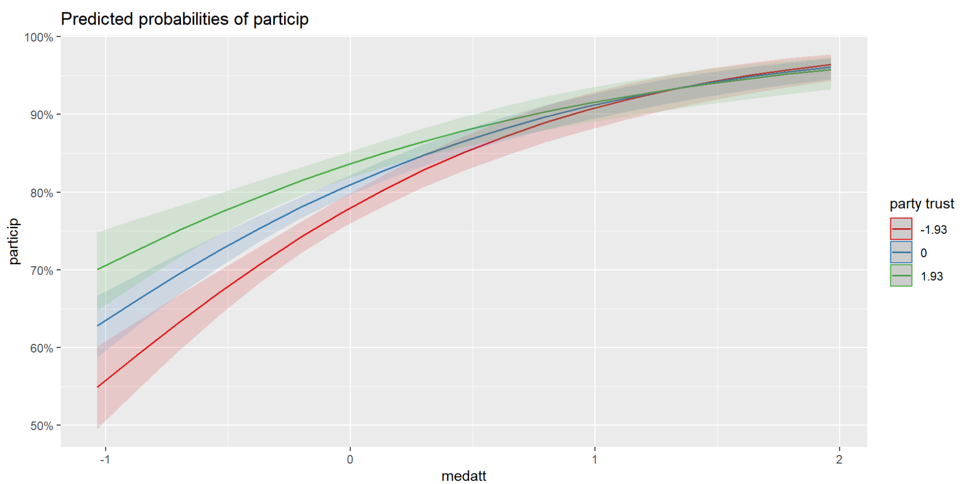
5.6 PROCESS output

```
******************** PROCESS for R Version 4.3.1 ********************

           Written by Andrew F. Hayes, Ph.D.  www.afhayes.com
   Documentation available in Hayes (2022). www.guilford.com/p/hayes3

*********************************************************************

PROCESS is now ready for use.
Copyright 2020-2023 by Andrew F. Hayes ALL RIGHTS RESERVED
Workshop schedule at http://haskayne.ucalgary.ca/CCRAM


******************** PROCESS for R Version 4.3.1 ********************

           Written by Andrew F. Hayes, Ph.D.  www.afhayes.com
   Documentation available in Hayes (2022). www.guilford.com/p/hayes3

*********************************************************************

Model : 1
    Y : particip
    X : medatt
    W : party_trust

Sample size: 3940


*********************************************************************
Outcome Variable: particip

Coding of binary Y for logistic regression analysis:
   particip  Analysis
     0.0000    0.0000
```

5.7 Visualizations

The plot_model() function will automatically plot the simple slopes (1 SD above and 1 SD below the mean) of the moderating effect:

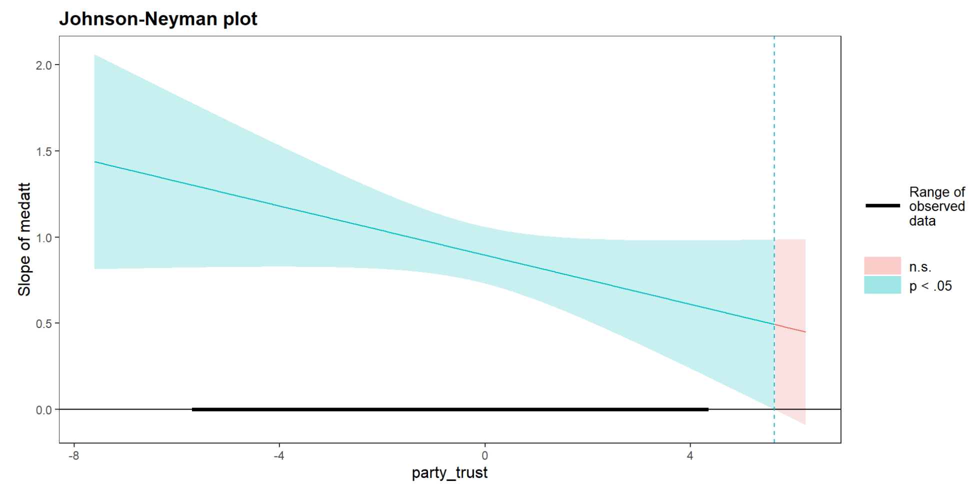
```
sjPlot::plot_model(mod1, type = "pred",
                   terms = c("medatt", "pa
```




```
# library(interactions)
# interact_plot(mod1, pred = "medatt",
# modx = "party_trust", interval = TRUE)
```

```
library(interactions)
johnson_neyman(mod1, pred="medatt", modx="
```

```
JOHNSON-NEYMAN INTERVAL

When party_trust is INSIDE the interval [-97.90,
5.64], the slope of medatt
is p < .05.

Note: The range of observed values of party_trust is
[-5.68, 4.32]
```

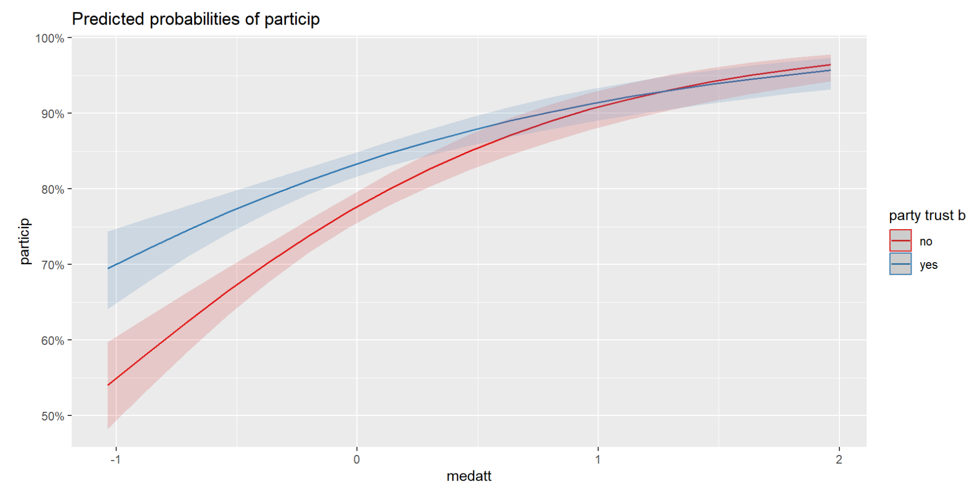


UZH IKMZ

5.8 Moderator as categorical variable

We use a dichotomous moderator and check whether we obtain similar results:

	Dependent variable:
	particip
party_trust_byes	0.364*** (0.084)
medatt	1.048*** (0.119)
party_trust_byes:medatt	-0.287* (0.167)
Constant	1.247*** (0.061)
Observations	3,940
Log Likelihood	-1,904.162
Akaike Inf. Crit.	3,816.324
Note:	*p<0.1; **p<0.05; ***p<0.01



UZH
IKMZ