



Analyzing Data in R



PEOPLE ANALYTICS, 04.10.2020

ALEXANDRA ARNOLD & MANUELA MORF

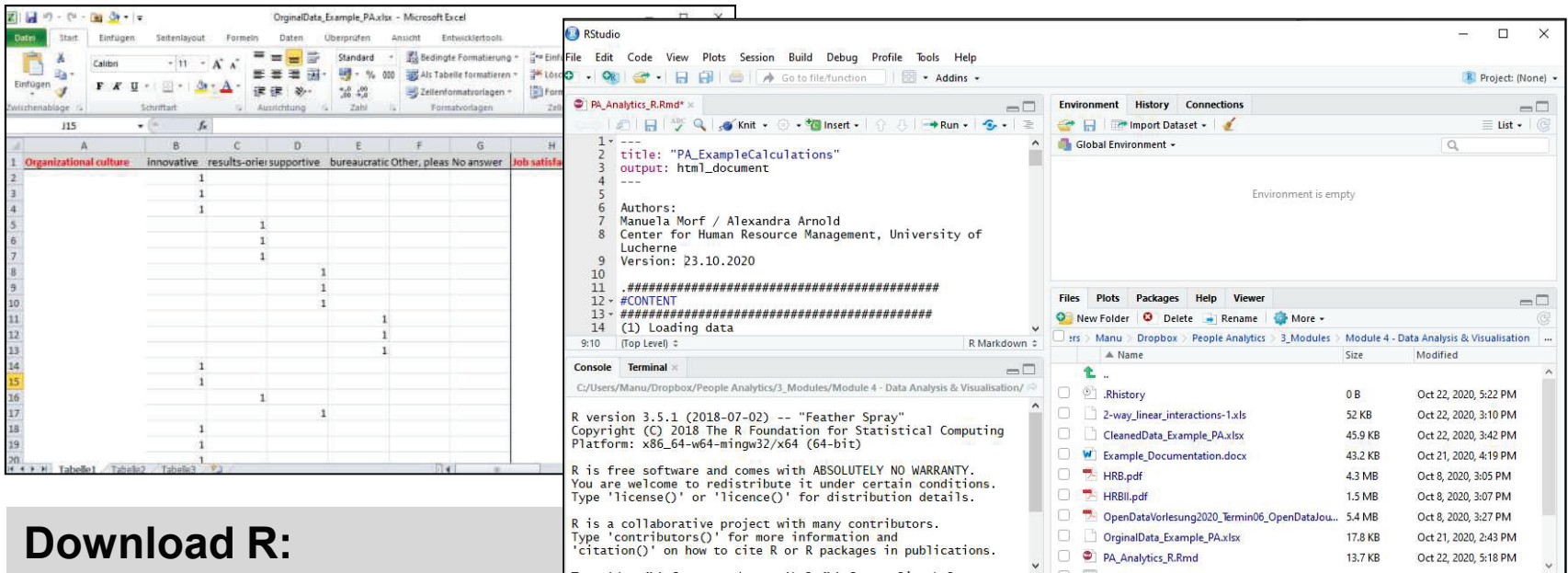
Learning goals

- 1 Knowing how to prepare your data for analysis
- 2 Having your knowledge on basic statistics analyses refreshed

Learning goals

- 1 Knowing how to prepare your data for analysis
- 2 Having your knowledge on basic statistics analyses refreshed

1. We use Excel and R



Download R:

- Software: <https://cran.r-project.org/mirrors.html> (for CH: ETH Zürich)
- User interface (R-Studio): <https://www.rstudio.com/products/rstudio/download/>
- If not sure, choose newest 32-bit version and default settings

Download “Calc” (OpenOffice Excel):

- <https://sourceforge.net/projects/openofficeorg.mirror/>

1. What you finde in the OLAT-Folder

Data and data cleaning:

- Original data file: OriginalData_Example_PA.xlsx
- Cleaned data file (with formula and visualization): CleanedData_Example_PA.xlsx
- Data file ready to upload in R (xlsx and csv): ReadyData_Example_PA.csv
- Codebook and cleaning protocol: Example_Documentation.docx

Codes and Excel graphs:

- R Code file: PA_Analytics_R.Rmd
- Excel graphs: GraphesinExcel.xlsx
- Moderation Template of Dawason: File_jeremydawson.xls (see: <http://www.jeremydawson.co.uk/slopes.htm>)

Templates:

- Correlation table (for home assigment): Template_Correlation Table.docx

1. Our fake data

- **N** = 143 participants
- **Organizational culture**: innovative (**ORC_inno** = 0/1), result-oriented (**ORC_result** = 0/1), supportive (**ORC_sup** = 0/2), bureaucratic (**ORC_bure**=0/1)
- **Job satisfaction** (**JSF**): 1=completely dissatisfied, to 10 = completely satisfied
- **Intention to quit** (**ITQ_scale**): average of items **ITQ_1**, **ITQ_2**, **ITQ_3r** (r stands for reverse coded) with 1= completely disagree to 7 = completely agree
- **Reasons to quit** (multiple answers): (**RTQ_pay** = 1/0), higher position (**RTQ_pos** = 0/1), work-life-balance (**RTQ_wlb** = 1/0), interesting task (**RTQ_intr** =0/1)
- **Employability** (**EMP_scale**): average of items **EMP_1**, **EMP_2**, **EMP_3** with 1= completely disagree to 7 = completely agree
- **Working percentages** (**WPR**): % ranging from 1 to 100
- **Sex**: **MALE** with 1 = male, 0 = female



Garbage in garbage out!

Your survey data needs to be “cleaned” to ensure that it is free of inaccurate or corrupt information.

1. Documentation, documentation, documentation...

Documentation: EMB scale
Janssens, Seb, & Brands, 2003
Mandatory

To what extent do you agree with the following statements?

	1 com- pletely disa- gree	2	3	4	5	6	7 com- plet- ely agre- e	No an- swer ?
EM P_1	☐	☐	☐	☐	☐	☐	☐	☐
EM P_2	☐	☐	☐	☐	☐	☐	☐	☐
EM P_3	☐	☐	☐	☐	☐	☐	☐	☐

Working Percentages WPR
Self-Developed
Non-mandatory

How many working percentages do you have contractually agreed?
--_% contractually agreed working percentages

Sex MALE
Self-Developed
Mandatory

Which sex do you have?
☐ 1 female (=0)
☐ 2 male (=1)
☐ 3 No answer

General Comment GCOM_0

Code book

Data

JSF	ITQ_Scale	ITQ_1	ITQ_2	ITQ_3r	RTQ_pay	RTQ_pos	RTQ
10	1.00	1	1	1	1	0	0
9	1.00	1	1	1	1	0	1
8	1.33	2	1	1	0	0	0
4	4.00	4	4	4	4	1	1
5	5.00	5	5	5	5	1	1
3	4.00	4	4	4	4	1	1
6	2.00	2	2	2	2	0	0
6	2.00	2	2	2	2	1	0
7	3.00	3	3	3	3	0	1
1	6.67	7	7	7	6	0	1
2	7.00	7	7	7	7	0	1
				6	6	1	0
				1	2	1	1
				1	1	0	0

Calculations

```

65
66
67 - #####
68 - # PREDICTING TRANSFORMATIONAL LEADERSHIP
69 - #####
70
71 - ```{r}
72 names(DATA)
73 ```
74
hip, predictor: work engagement
in both models sig

79 Model1.0<-lme(TFL_long~ GC_WEN,random=~|u_id_system,data=DATA,na.action=na.exclude)
80 summary(Model1.0)
81 VarCorr(Model1.0)
82
83 # random slope
84 Model1.3<-lme(TFL_long ~ GC_WEN,random=~GC_WEN|u_id_system,data=DATA,na.action=na.exclude)
85 summary(Model1.3)
86 VarCorr(Model1.3)
87
88 # compare which model is better
89 anova(Model1.0, Model1.3)
90
91 #R square marginal and conditional
92 r.squaredGLMM(Model1.0)

```

(1) Assigning an ID number to each case:

- New variable IDN (if data file contains email addresses, save a separate file with only IDN and email addresses and delete email addresses immediately in OriginalData)

(2) Screening open comments for hints to exclude participants

- Delete answers of ID 108 delete (marked in red) I am currently unemployed, yet answered this survey for my last employment

(3) Doing plausibility checks: Check if everybody indicated to work more than 0%

- Delete answers of ID 28 (marked in red) Indicate

(4) Recoding open comments "other, please specify"

- IDN 43 has mentioned "creative" in ORC → this (tive)
- IDN 109 has mentioned "better boss" in RTQ → this was not recoded, as no other category is applicable

Cleaning protocol

1. The 10 cleaning steps

1	Extract your data file from survey tool
2	Anonymize data and assign ID numbers
3	Assign short names to variables
4	Decide which cases/participants to exclude: Many missing data, hints in open comments, plausibility checks
5	(Re-)code open data
6	Create dummy variables from categorical variables (attention missings!)
7	Recode continuous variables if required (attention missings!)
8	Create scale scores from multiple-item scales (attention missings!)
9	Check (again) if your missings are empty cells
10	Save a .csv file

1. The 10 cleaning steps

- 1 Extract your data file from survey tool
- 2 Anonymize data and assign ID numbers
- 3 Assign short names to variables

The screenshot displays a survey data entry interface. On the left, a table with columns 'IDN' and 'ORC_' is shown. A pink arrow points to the 'IDN' column header, which is highlighted in grey. Below this, a grey box contains the text 'Newly created'. In the center, a pop-up window titled 'Job Satisfaction JSF' is open, showing a list of radio button options for job satisfaction levels from 1 to 11. On the right, a table with columns 'OCR_bure' and 'JSF' is visible. A pink arrow points to the 'JSF' column header, which is highlighted in light blue. The 'JSF' column contains values 10, 9, 8, and 4 for the four rows shown.

IDN	ORC_
	1
	2
	3
	4

Newly created

Job Satisfaction JSF

[Wanous, Reichers, & Hudy, 1997](#)

Mandatory

How satisfied are you overall with your work?

☐ 1 1 completely dissatisfied

☐ 2 2

☐ 3 3

☐ 4 4

☐ 5 5

☐ 6 6

☐ 7 7

☐ 8 8

☐ 9 9

☐ 10 10 completely satisfied

☐ 11 no answer

OCR_bure	JSF
0	10
0	9
0	8
0	4

1. The 10 cleaning steps

- 4 Decide which cases/participants to exclude: Many missing data, hints in open comments, plausibility checks

#IDN = 128; “I am currently unemployed, yet answered this survey for my last employment”



#IDN = 28: WPR = 0
(indicates to not work at all)

1. The 10 cleaning steps

5 (Re-)code open data

Organizational culture	innovativ	results-or	supportiv	bureaucra	Other, ple
	1				creative

“creative” was manually assigned to category “innovative culture”

Statement	Category	competitive	collaborative	...
«we are competitive»	competition	1		
«I like the teamwork here»	collaboration		1	
«We collaborate a lot»	collaboration		1	
...	...			

when coding open questions, you need to create meaningful categories first

1. The 10 cleaning steps

6 Create dummy variables from categorical variables (attention missings!)

Use IF-Function

RTQ_pay	RTQ_pos	RTQ_wlb	RTQ_intw
0	0	1	1
0	1	1	1
0	0	1	1
1	1	0	0
1	1	1	1
1	1	0	0

Other_nls	No answer
0	0
1	1
0	1

No answer =
empty cells

1. The 10 cleaning steps

7 Recode continuous variables if required (attention missings!)

ITQ _1	I will probably look for a new job in the near future.	☐ _1	☐ _2	ITQ_3, high values mean a low intention to quit					
ITQ _2	At the present time, I am actively searching for another job in a different organization.	☐ _1	☐ _2	☐ _3	☐ _4	☐ _5	☐ _6	☐ _7	☐ _8
ITQ _3	I intend to stay with my organization (reverse)	☐ _1	☐ _2	☐ _3	I intend to	ITQ_r			
					7	1			
					7	1			
					7	1			
					4	4			
					3	5			
					4	4			
					6	2			
					5	3			

Re-code: $ITQ_3r = 8 - ITQ_3$

8 is the maximum of the scale (=7) plus 1

Re-code: $ITQ_3r = 8 - ITQ_3$

8 is the maximum of the scale (=7) plus 1

1. The 10 cleaning steps

8 Create scale scores from multiple-item scales (attention missings!)

ITQ_Scale = MEAN(ITQ,1,
ITQ,2,ITQ3)

(in R, you need to use mean incl.
"na.rm=True")

ITQ_Scale ▼	ITQ_1 ▼	ITQ_2 ▼	ITQ_3r ▼
1.00	1	1	1
1.00	1	1	1
1.33	2	1	1
4.00	4	4	4
5.00	5	5	5

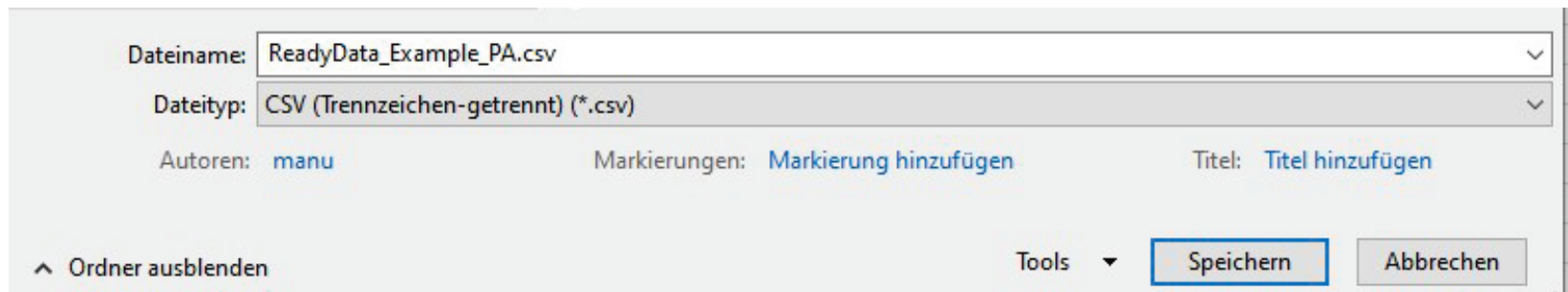
2.00	2	2	
4.00	4	4	4
5.00	5	5	5

1. The 10 cleaning steps

9 Check (again) if your missings are empty cells

10 Save a .csv file

- Ideally you make sure that you have only numeric and categorical data in your file (no long text) before saving a .csv file.



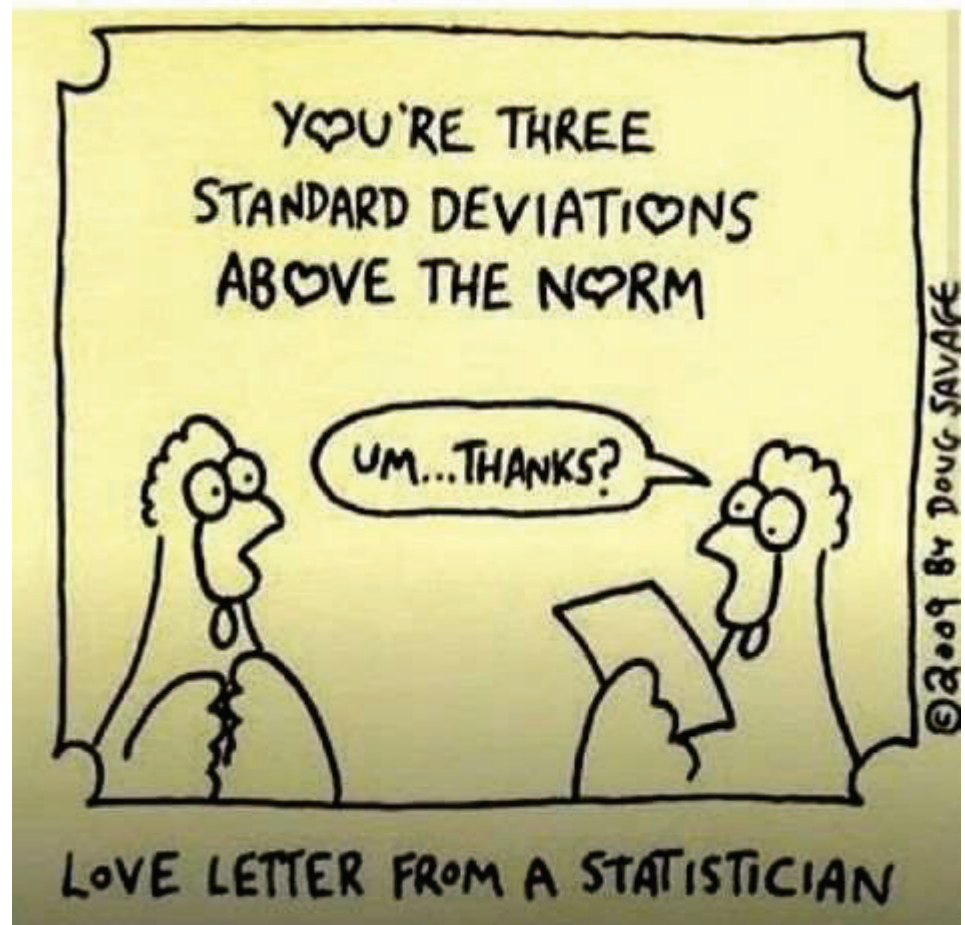
The screenshot shows a file save dialog box with the following elements:

- Dateiname:** ReadyData_Example_PA.csv
- Dateityp:** CSV (Trennzeichen-getrennt) (*.csv)
- Autoren:** manu
- Markierungen:** Markierung hinzufügen
- Titel:** Titel hinzufügen
- Buttons:** Ordner ausblenden, Tools, Speichern (highlighted with a blue border), Abbrechen

Learning goals

- 1 Knowing how to prepare your data for analysis
- 2 Having your knowledge on basic statistics analyses refreshed

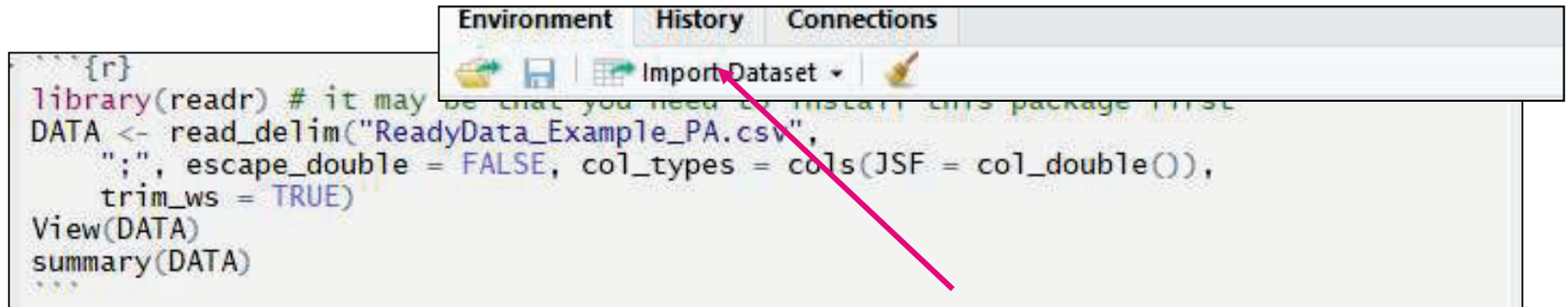
2. How was this again...?



2. Our fake data

- **N** = 143 participants
- **Organizational culture**: innovative (**ORC_inno** = 0/1), result-oriented (**ORC_result** = 0/1), supportive (**ORC_sup** = 0/2), bureaucratic (**ORC_bure**=0/1)
- **Job satisfaction** (**JSF**): 1=completely dissatisfied, to 10 = completely satisfied
- **Intention to quit** (**ITQ_scale**): average of items **ITQ_1**, **ITQ_2**, **ITQ_3r** (r stands for reverse coded) with 1= completely disagree to 7 = completely agree
- **Reasons to quit** (multiple answers): (**RTQ_pay** = 1/0), higher position (**RTQ_pos** = 0/1), work-life-balance (**RTQ_wlb** = 1/0), interesting task (**RTQ_intr** =0/1)
- **Employability** (**EMP_scale**): average of items **EMP_1**, **EMP_2**, **EMP_3** with 1= completely disagree to 7 = completely agree
- **Working percentages** (**WPR**): % ranging from 1 to 100
- **Sex**: **MALE** with 1 = male, 0 = female

2. Reading the data in to R



```
{r}
library(readr) # it may be that you need to install this package first
DATA <- read_delim("ReadyData_Example_PA.csv",
  ";", escape_double = FALSE, col_types = cols(JSF = col_double()),
  trim_ws = TRUE)
View(DATA)
summary(DATA)
```

- You can use the “Import Dataset” function (use .csv file).
- Check with the summary function if the Min, Max, Mean and NA (=Missing) values make sense

2. Deskriptives - the starting point

Variables	1	2	3	4	5	6	7	8
1. Male								
2. Age	.15*							
3. Work percentages	.48***	-.04						
4. Organisational tenure	.15*	.66***	.04					
5. Rank 1	-.45***	-.32***	-.36***	-.27***				
6. Rank 2	.04	-.15*	.01	-.06	-.33***			
7. Rank 3	.26**	.17*	.19*	.10	-.47***	-.31***		
8. Rank 4	.21**	.34***	.21**	.28***	-.34***	-.22**	-.31***	
9. Qualitative overload	.00	-.14*	.06	-.06	.03	.07	-.04	-.06
10. Working hours	.48***	.07	.80***	.09	-.47***	-.03	.20**	.36***

M	0.63	41.49	92.38	14.25	0.34	0.18	0.30	0.18
SD	0.48	10.21	16.27	9.97	0.47	0.38	0.46	0.39

Note. N = 202. Male, Rank 1, Rank 2, Rank 3, and Rank 4 are dummy variables: Their means refer to percentages. Performance-based salary is measured in 1,000 CHF. Mean scale scores were used to calculate correlations.

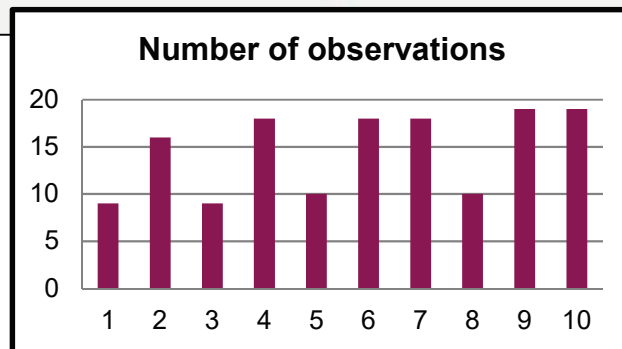
Abbreviation: i-deal, idiosyncratic deal; LMX, leader-member exchange; M, mean; SD, standard deviation.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Source: Morf, M., Bakker, A. B., & Feierabend, A. (2019). Bankers closing idiosyncratic deals: Implications for organisational cynicism. Human Resource Management Journal, 29(4), 585-599.

2. Descriptives - the most important one's

```
{r}  
table(DATA$JSF) # shows you the frequency  
mean(DATA$JSF, na.rm=TRUE) # shows you the mean (na.rm=True, says R to ignore missings  
sd(DATA$JSF, na.rm=TRUE) #shows you the standard deviation  
cor.test(DATA$ITQ_Scale, DATA$JSF, use = "pairwise.complete.obs")# use function tells
```



- **Mean** as
measure of the level:= 5.92
- **Standard deviation** as
measure of variance:= 2.84
- **Correlation** as a measure
of relationship strength: -.092, $p < .05$

```
1 2 3 4 5 6 7 8 9 10  
9 16 9 18 10 18 18 10 19 19  
> mean(DATA$JSF, na.rm=TRUE) # shows you the mean (na.rm=Tr  
ings)  
[1] 5.924658  
> sd(DATA$JSF, na.rm=TRUE) #shows you the standard deviatio  
[1] 2.848074  
> cor.test(DATA$ITQ_Scale, DATA$JSF, use = "pairwise.comple  
  
Pearson's product-moment correlation  
  
data: DATA$ITQ_Scale and DATA$JSF  
t = -28.552, df = 144, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.9430998 -0.8932074  
sample estimates:  
cor  
-0.9218895
```

2. A word about correlations

Correlations (r) can have values between -1 and 1.

- Values between -1 and 0 indicate a negative relationship.
- Values between 0 and +1 indicate a positive relationship.
- As a rule of thumb:
 - Values smaller than 0.20 or greater than -0.20 → weak relationships (no practical relevance)
 - Values between 0.20 and 0.40 or between -0.20 and -0.40 → medium strong relationship (some practical relevance, yet not a priority)
 - Values greater than 0.40 or smaller than -0.40, respectively → strong relationship (high practical relevance)

Bosco, F., Aguinis, H., Singh, K., Field, J., & Pierce, C. (2015).
Correlational Effect Size Benchmarks. *Journal of Applied Psychology*, 100(2), 431-449.

2. What does the data tell us?

1	Frequencies: What is the top reason why employees would quit their current job?
2	Group comparison (two categories): Who is more likely to quit because of a better work-life balance: men or women?
3	Group comparison (more than two categories): In which organizational culture are people the most satisfied?
4	Simple relationships: How is job satisfaction related to intention to quit
5	Moderation (with categorical moderator): Does gender moderate the relationship between job satisfaction and intention to quit?
6	Moderation (with numerical moderator): Does employability moderate the relationship between job satisfaction and intention to quit?
7	Mediation: Does job satisfaction mediate the negative relationship between innovative organizational culture and intention to quit

2. Frequencies

1

Frequencies: What is the top reason why employees would quit their current job?



2. Frequencies

- 1 **Frequencies:** What is the top reason why employees would quit their current job?

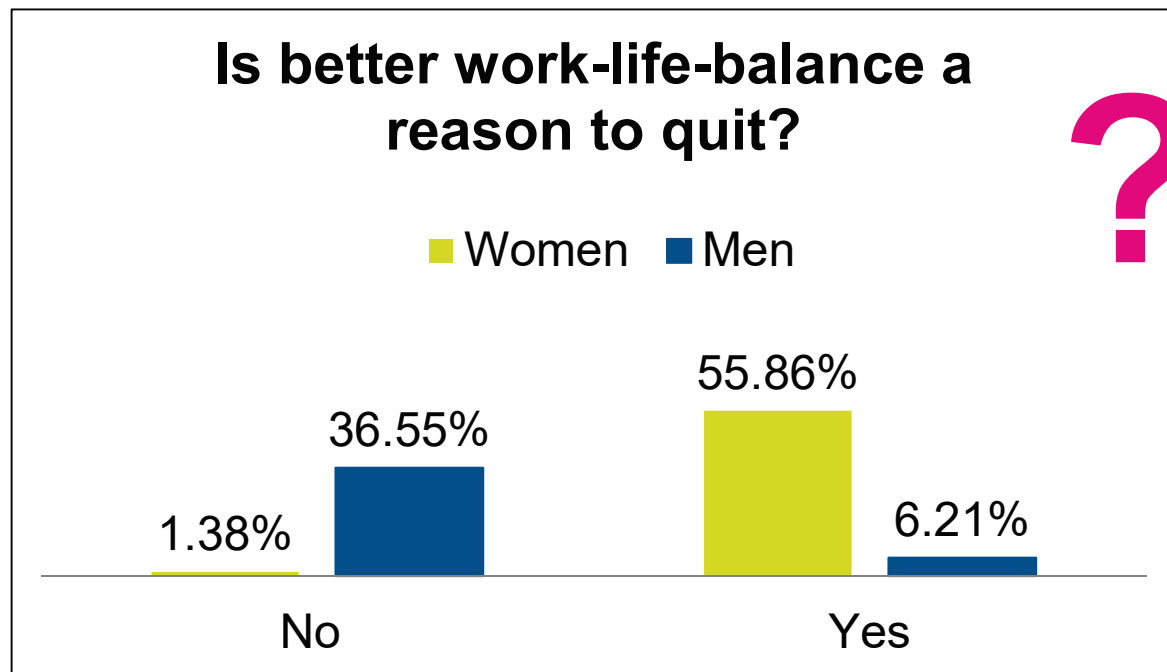
```
## {r cars}
#pay
table(DATA$RTQ_pay) # shows you the absolute numbers
mean(DATA$RTQ_pay, na.rm=TRUE) # shows you the share (%)
```

```
> #pay
> table(DATA$RTQ_pay) # shows you the absolute numbers
 0  1
83 63
> mean(DATA$RTQ_pay, na.rm=TRUE) # shows you the share (%)
[1] 0.4315068
```

- Here you see: 63 participants say that they would leave for higher pay (vs. 83 that say they would not), this equals to 43% of all participants
- If you do this analysis for all categories you find: 63% would quit for better work-life-balance, 58% for more interesting pay, 51% for higher position, and 43% for better pay

2. Group comparison (two categories)

2 Who is more likely to quit because of a better work-life balance: men or women?



2. Group comparison (two categories)

2 Who is more likely to quit because of a better work-life balance: men or women?

```

library(gmodels) # it may be that you need to install this package first
CrossTable(DATA$MALE, DATA$RTQ_wlb) # male will be rows, RTQ_pay will be columen
#To test if differences are statistically significant, use t-tests
t.test(DATA$MALE, DATA$RTQ_wlb)

```

Total Observ

DATA\$MALE	DATA\$RTQ_wlb	0	1	Row Total
0	2	81	83	
	27.610	16.873		
	0.024	0.976	0.572	
	0.036	0.900		
	0.014	0.559		
1	53	9	62	
	36.962	22.588		
	0.855	0.145	0.428	
	0.964	0.100		
	0.366	0.062		
Column Total	55	90	145	
	0.379	0.621		

Welch Two Sample t-test

data: DATA\$MALE and DATA\$RTQ_wlb
 t = -3.4555, df = 289.89, p-value = 0.0006315
 alternative hypothesis: true difference in mea
 95 percent confidence interval:
 -0.31176719 -0.08549308
 sample estimates:
 mean of x mean of y
 0.4246575 0.6232877

- 81 of 83 woman (=97.6%) would like to work from home for a better work-life balance, vs 9 of 62 man (=14.5%)
- T-test suggests that differences are statistically significant

- 81 of 83 woman (=97.6%) would leave for a better work-life balance, vs 9 of 62 man (=14.5%)
- T-test suggests that differences are significant (p < .05)

2. Group comparison (more than two categories)

3 In which organizational culture are people the most satisfied?

Predictor

OCR_inno (1/0)
OCR_result (1/0)
OCR_sup(1/0)
OCR_bure(1/0)

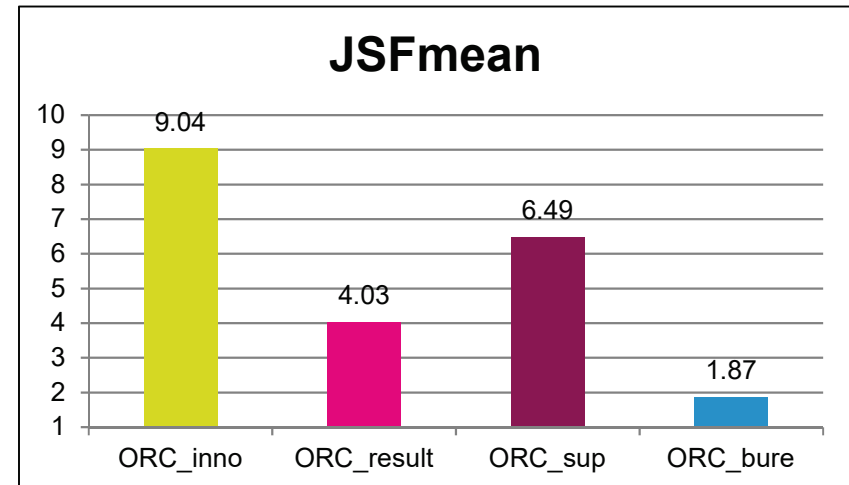
Outcome
JSF (1-10)

2. Group comparison (more than two categories)

3 In which organizational culture are people the most satisfied?

```
library(plyr) # it may be that you need to install this package first
ddply(DATA, ~ ORC_inno, summarise, JSFmean=mean(JSF, na.rm=TRUE), JSFsd=sd(JSF, na.rm=TRUE))
# gives you group mean and standard deviation;
```

	ORC_inno	JSFmean	JSFsd
1	0	4.385417	1.975480
2	1	9.040816	1.274088
3	NA	1.000000	NA



- The average job satisfaction of participants in innovative cultures is 9.04 (SD=1.27) and thus higher when compared to other cultures
- Yet are these values significantly different from each other?

2. Group comparison (more than two categories)

3 In which organizational culture are people the most satisfied?

```
# the baseline OR-201000 = na.omit the model
solution1 <- lm(JSF~ OCR_result + OCR_sup + OCR_bure,DATA, na.action=na.omit)
summary(solution1)
```

	Min	1Q	Median	3Q	Max
	-7.0408	-0.4857	-0.0270	0.5143	5.1250

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.0408	0.1415	63.91	<2e-16 ***
OCR_result	-5.0138	0.2137	-23.25	<2e-16 ***
OCR_sup	-2.5551	0.2192	-11.66	<2e-16 ***
OCR_bure	-7.1658	0.2467	-29.05	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- You can do t-tests for each pair, yet using regressions is faster: do not include your baseline category “innovative culture”:
$$\text{Job Satisfaction} = c + b_1 * \text{result-oriented (vs. innovative)} \\ + b_2 * \text{supportive (vs. innovative)} + b_3 * \text{bureaucratic (vs. innovative)}$$
- The average job satisfaction in result-oriented cultures is -5.01 (=9.04 – 4.03, as in previous slide) lower in innovative cultures, this difference is significant ($p < .05$)

2. Simple relationships

4 How is job satisfaction related to intention to quit

Predictor
JSF (1-10)



Outcome
ITQ_Scale (1-7)

2. Simple relationships

4 How is job satisfaction related to intention to quit

```
# First some descriptives: Look at correlation  
cor.test(DATA$ITQ_Scale, DATA$JSF, use = "pairwise.complete.obs")# use function tells  
how missings should be handled
```

```
Pearson's product-moment correlation  
  
data: DATA$ITQ_Scale and DATA$JSF  
t = -28.552, df = 144, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.9430998 -0.8932074  
sample estimates:  
cor  
-0.9218895
```

- Job satisfaction correlates significantly strongly and negatively with intention to quit (-0.92, $p < .05$, which implies that when job satisfaction increases then intention to quit strongly decreases
- Yet what does “strongly” mean; what if I would like to predict the change?

2. Simple relationships

4 How is job satisfaction related to intention to quit

```
Call:
lm()

residuals:
-1.3
# Estimate a simple regression
solution2 <- lm(ITQ_Scale~JSF,DATA, na.action=na.omit)
summary(solution2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.39425    0.15576   47.47  <2e-16 ***
JSF        -0.67696    0.02371  -28.55  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

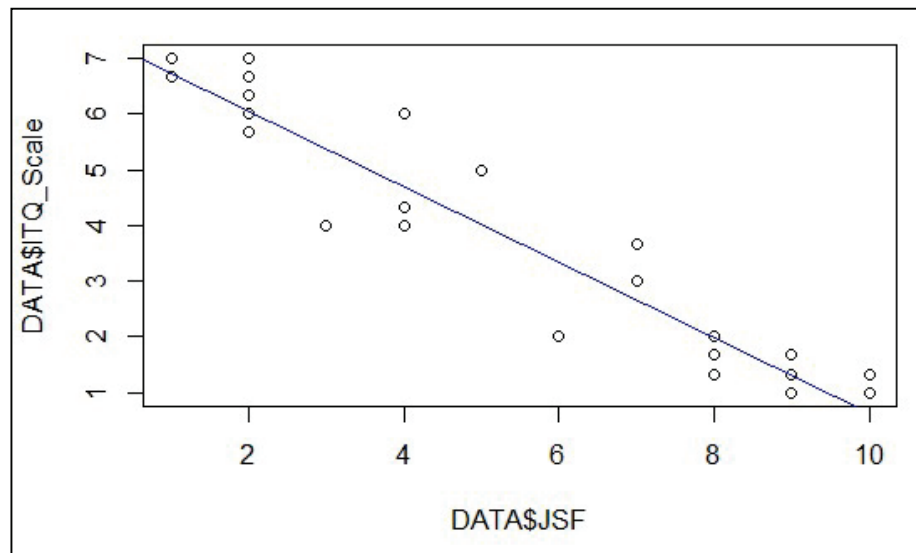
Residual standard error: 0.8131 on 144 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.8499,    Adjusted R-squared:  0.8488
F-statistic: 815.2 on 1 and 144 DF,  p-value: < 2.2e-16
```

- You estimate:
$$\text{Intention to quit} = c + b1 * \text{job satisfaction}$$
- If job satisfaction increases by 1.0 (on a scale from 1-10) then intention to quit decreases by -0.68 (on a scale for 1-7). The relationship is statistically significant ($p < 0.05$). Furthermore, the model predicts 84.99% of the variance in intention to quit.

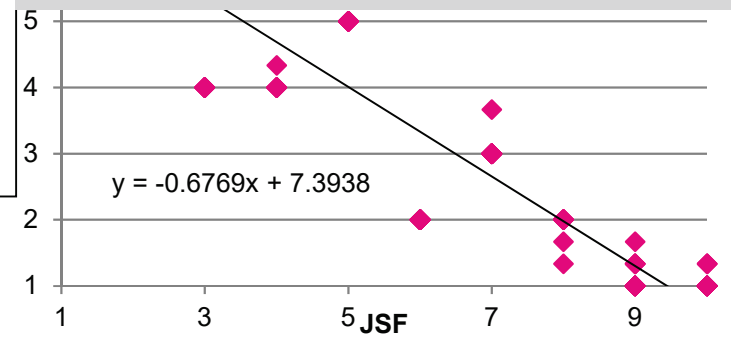
2. Simple relationships

4 How is job satisfaction related to intention to quit

```
# Plot data
plot(DATA$JSF, DATA$ITQ_Scale)
# Add regression line
abline(lm(DATA$ITQ_Scale ~ DATA$JSF, data = mtcars), col = "blue")
```

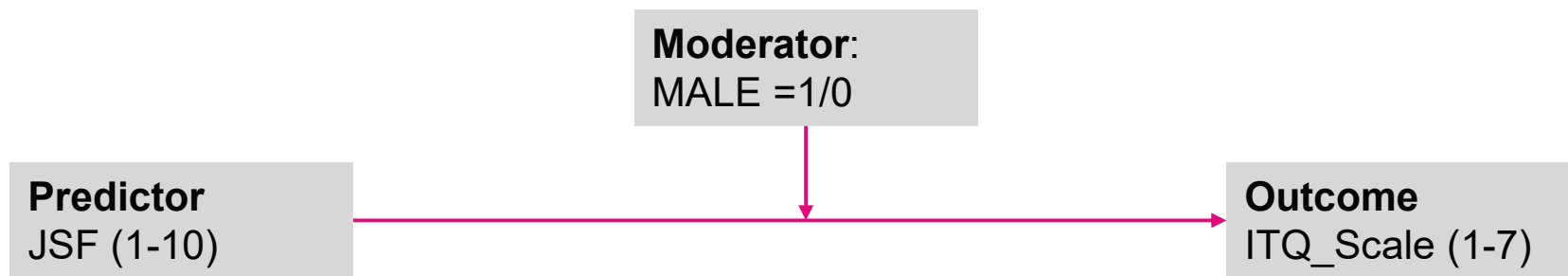


- The intercept reflects the starting point and the b-value the steepness of the estimated line
- We estimated: Intention to quit = $c + b1 * \text{job satisfaction}$
- We obtained: Intention to quit = $7.39 - 0.68 * \text{job satisfaction}$



2. Moderation (with categorical moderator)

- 5 Does gender moderate the relationship between job satisfaction and intention to quit?



- You estimate:
$$ITQ = c + b1 * Predictor + b2 * Moderator + b3 * (Predictor * Moderator)$$

$$\underbrace{b3 * (Predictor * Moderator)}$$

Interaction term

2. Moderation (with categorical moderator)

- 5 Does gender moderate the relationship between job satisfaction and intention to quit?

```
# Estimate a moderation analysis
solution3 <- lm(ITQ_Scale~JSF + MALE + MALE * JSF,DATA, na.action=na.omit)
summary(solution3)
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.45647	0.39974	18.654	< 2e-16	***
JSF	-0.67174	0.04961	-13.541	< 2e-16	***
MALE	0.66019	0.45705	1.444	0.150832	
JSF:MALE	-0.27556	0.07982	-3.452	0.000735	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7634 on 141 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.8693, Adjusted R-squared: 0.8665
F-statistic: 312.5 on 3 and 141 DF, p-value: < 2.2e-16

- Job satisfaction is significantly negatively related to intention to quit ($b = -0.67$, $p < .05$), gender has no significant effect ($b = 0.66$, $p > 0.05$) and the interaction term (JSF:MALE) is significant ($b = -0.27$, $p < .05$). A total of 86.93% of the variance in intention to quit is explained by our model
- Yet, how do we interpret this?

2. Moderation (with categorical moderator)

5 Does gender moderate the relationship between job satisfaction and intention to quit?

- For interpretation, drawing is necessary
- Determine what high and low job satisfaction is in your sample, usually it is mean value plus one standard deviation for high and minus one standard deviation for low
- Estimate intention to quit for 4 data points using the estimated values:
 - Intention to quit = $7.45 - 0.67 * \text{JSF} + 0.00 * \text{MALE} - 0.28 (\text{JSF} * \text{MALE})$

```
JSFmean <- mean(DATA$JSF, na.rm=TRUE)
JSFsd <- sd(DATA$JSF, na.rm=TRUE)
JSFmean - JSFsd # 3.00
JSFmean + JSFsd # 8.76
```

```
JSF + male, high JSF + female)
x <- c(3.00, 8.76, 3.00, 8.76) # JSF values for high and low
m <- c(0, 1, 1, 0) # male vs. female
y <- 7.45647 + -0.67174*x + 0.66019*m + -0.27556 *(x*m) # See results
solution3
(data <- data.frame(x, m = factor(m), y)) %>%
  as.tibble)
```

	Intention to quit	
	Male	Female
high JSF	-0.8815	1.5741
low JSF	4.5335	5.3931

See Excel / R / Or:

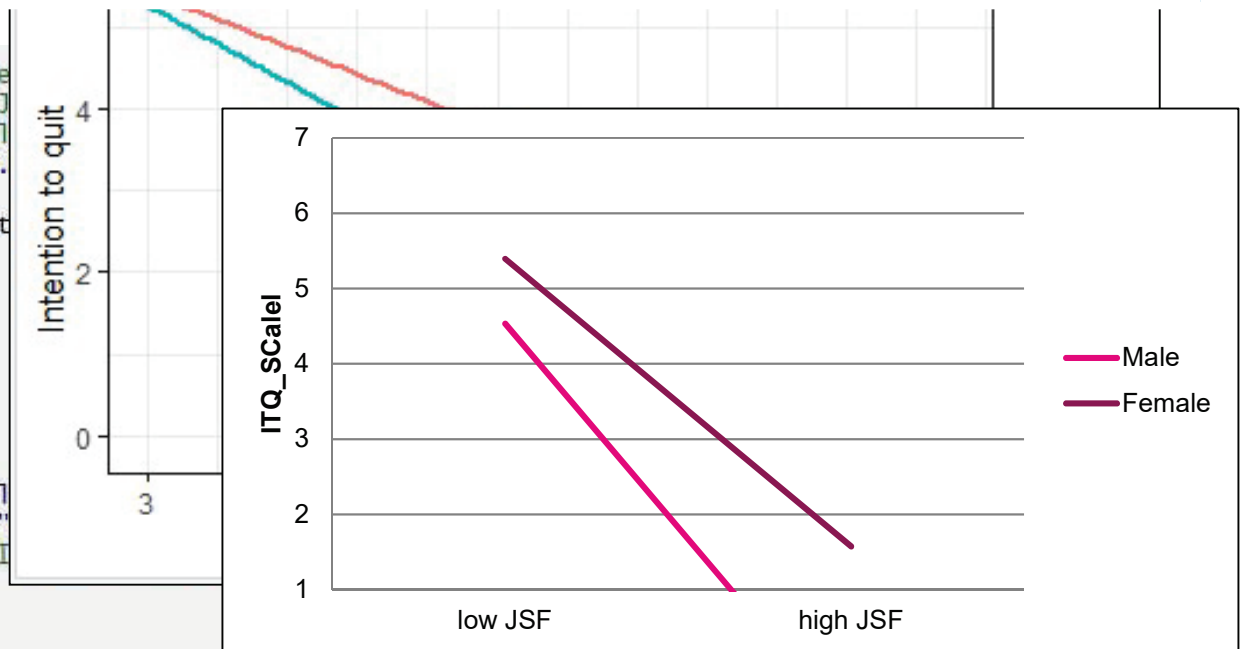
<http://www.jeremydawson.co.uk/slopes.htm>

2. Moderation (with categorical moderator)

5 Does gender moderate the relationship between job satisfaction and intention to quit?

```
library(tidyverse) # you may ne
x <- c(3.00,8.76,3.00,8.76) # J
m<- c(0,1,1,0) # male vs. famal
y <- 7.45647 + -0.67174*x + 0.
results solution3
(data <- data.frame(x, m = fact
as.tibble)

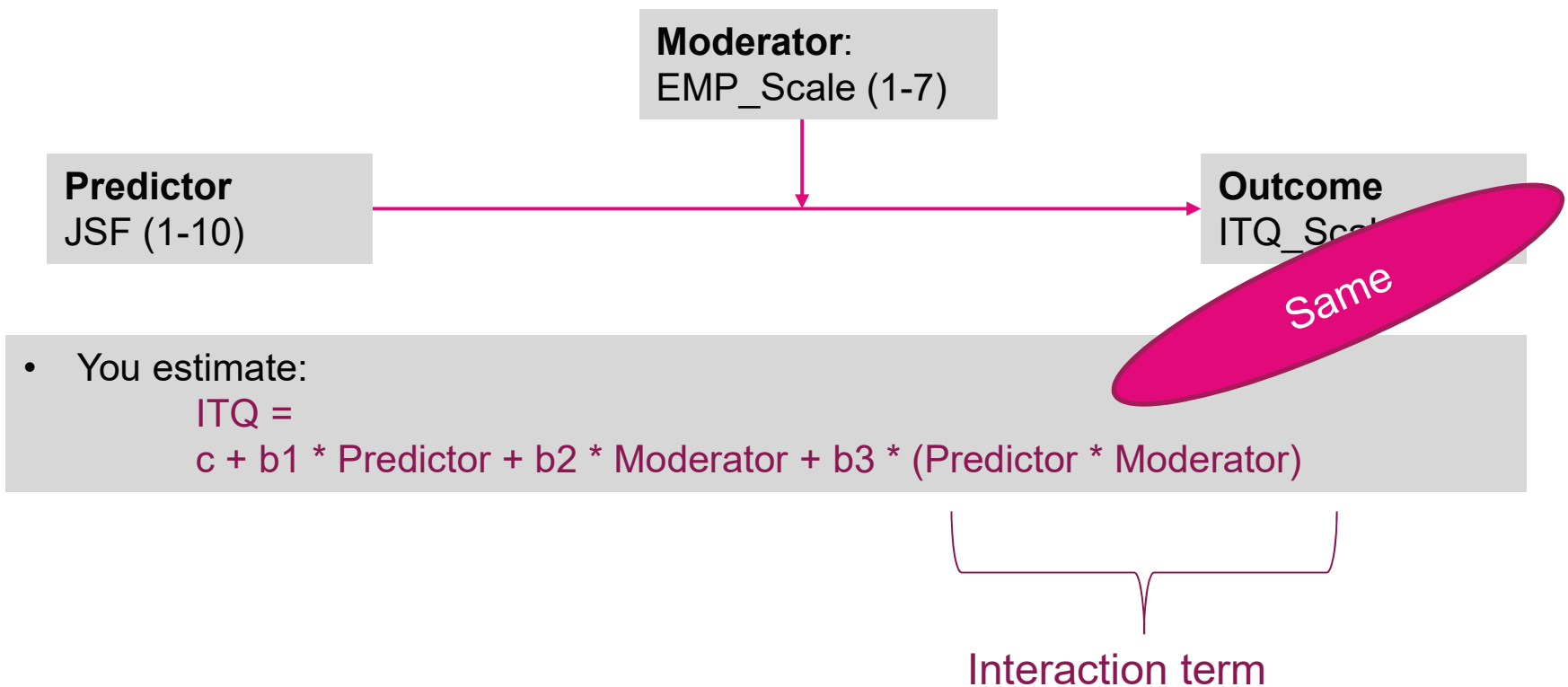
# Plotting the data
basic_plot <- ggplot(data,
  aes(x = x,
      y = y,
      color = m)) +
  theme_bw() +
  labs(x = "Independent variabl
      y = "Dependent variable"
      color = "Moderator") # I
basic_plot +
  geom_smooth(method = "lm")
```



- You can see that the line for men is steeper, indicating that they react with a higher intention to quit when job satisfaction is decreasing

2. Moderation (with numerical moderator)

- 6 Does employability moderate the relationship between job satisfaction and intention to quit?



2. Moderation (with numerical moderator)

- 6 Does employability moderate the relationship between job satisfaction and intention to quit?

```
solution4 <- lm(ITQ_Scale~JSF + EMP_Scale + EMP_Scale * JSF,DATA,  
na.action=na.omit)  
summary(solution4)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.80801	0.48030	22.503	< 2e-16	***
JSF	-1.05720	0.06693	-15.796	< 2e-16	***
EMP_Scale	-0.77467	0.10397	-7.451	8.29e-12	**
JSF:EMP_Scale	0.08685	0.01398	6.213	5.42e-09	***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6812 on 142 degrees of freedom
(1 observation deleted due to missingness)

Multiple R-squared: 0.8961, Adjusted R-squared: 0.8939

F-statistic: 408.2 on 3 and 142 DF, p-value: < 2.2e-16

Same

- Job satisfaction is significantly negatively related to intention to quit ($b = 1.06$, $p < .05$), employability too ($b = -0.77$, $p < .05$) and the interaction term (JSF:EMP_Scale) is significant ($b = 0.09$, $p < .05$). A total of 86.61% of the variance in intention to quit is explained by our model
- Yet, how do we interpret this?

2. Moderation (with numerical moderator)

6 Does employability moderate the relationship between job satisfaction and intention to quit?

- For interpretation, drawing is necessary
- **The only difference:** Determine what high and low job satisfaction is and **also what high and low employability** is, usually it is mean value plus one standard deviation for high and minus one standard deviation for low
- Estimate intention to quit for 4 data points using the estimated values:
 - Intention to quit
$$= 10.81 - 1.06 * \text{JSF} - 0.77 * \text{EMP_Scale} + 0.09 * (\text{JSF} * \text{EMP_Scale})$$

```
JSFmean <- mean(DATA$JSF, na.rm=TRUE) #na.rm tells R that there are missings in the data
JSFsd <- sd(DATA$JSF, na.rm=TRUE)
JSFmean - JSFsd # 3.00
JSFmean + JSFsd # 8.76
```

```
EMP_Scalemean <- mean(DATA$EMP_Scale, na.rm=TRUE) #na.rm tells R that there are missings in the data
EMP_Scalesd <- sd(DATA$EMP_Scale, na.rm=TRUE)
EMP_Scalemean - EMP_Scalesd # 2.98
EMP_Scalemean + EMP_Scalesd # 6.51
```

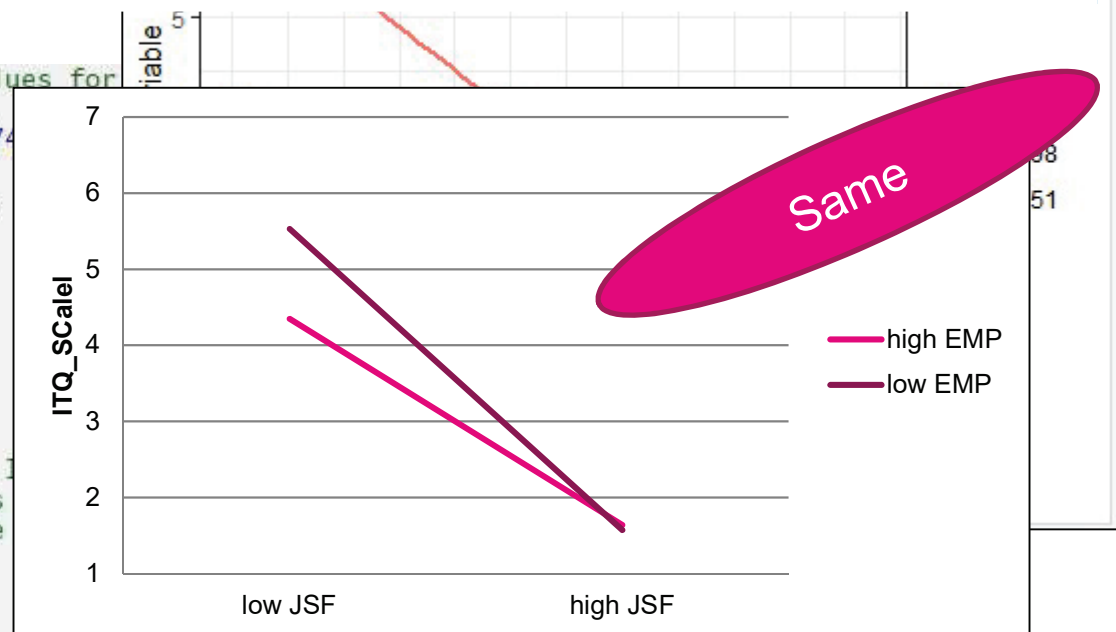
	Intention to quit	
	high EMP	low EMP
low JSF	4.34675	5.5294
high JSF	1.63925	1.571314

2. Moderation (with numerical moderator)

6 Does employability moderate the relationship between job satisfaction and intention to quit?

```
x <- c(3.00,8.76,3.00,8.76) # JSF values for male vs. female
m<- c(2.98,6.51,6.51,2.98) # male vs. female
y <- 10.80801 + -1.05720 *x + -0.774 *m
results solution4
(data <- data.frame(x, m = factor(m),
  as.tibble))

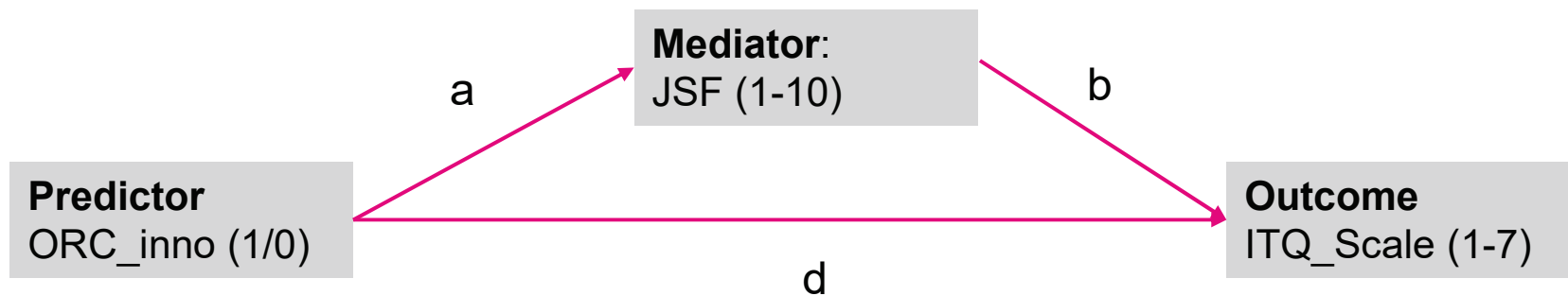
# Plotting the data
basic_plot <- ggplot(data,
  aes(x = x,
    y = y,
    color = m)) +
  theme_bw() +
  labs(x = "Independent variable", # Job Satisfaction
    y = "Dependent variable", # Intention to Quit
    color = "Moderator") # Is here
1 SD, low = mean - 1SD)
basic_plot +
  geom_smooth(method = "lm")
```



- You can see that the line for highly employable individuals is flatter, indicating that they react with lower intention to quit when job satisfaction is decreasing (this is counter-intuitive, however it is still fake data ;-).

2. Mediation

- 7 Does job satisfaction mediate the negative relationship between innovative organizational culture and intention to quit



- Two estimations are of interest:
 $\text{Mediator} = c1 + a * \text{Predictor}$
 $\text{Outcome} = c2 + b * \text{Mediator} + d * \text{Predictor}$
 - Indirect effect of predictor on outcome: $a * b$
 - Total effect of predictor on outcome: $(a*b) + d$
- Estimating two regressions are the key!

2. Mediation

- 7 Does job satisfaction mediate the negative relationship between innovative organizational culture and intention to quit

```
# Test first a model with m = c1 + a*x (m = mediator, x =
predictor)
solution5 <- lm(JSF~ORC_inno,DATA, na.action=na.omit) # here you
test ORC_inno against all other groups
summary(solution5)
```

	Min	1Q	Median	3Q	Max
	-7.0408	-1.0408	-0.0408	0.9592	2.6146

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.3854	0.1808	24.26	<2e-16 ***
ORC_inno	4.6554	0.3110	14.97	<2e-16 ***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.771 on 143 degrees of freedom
(2 observations deleted due to missingness)

- Your a is (=4.66) and statistically significant

```
F-statistic: 224.1 on 1 and 141 degrees of freedom
solution5 <- lm(JSF~ORC_inno,DATA,
test ORC_inno against all other gr
summary(solution5)
```

```
# Test then a model with y = c2 +
outcome, x = predictor)
solution6 <- lm(ITQ_Scale~JSF+ORC_
summary(solution6)
```

2. Mediation

- 7 Does job satisfaction mediate the negative relationship between innovative organizational culture and intention to quit

```
# Test then a model with  $y = c2 + b*m + d*x$  (m = mediator, y = outcome, x = predictor)
solution6 <- lm(ITQ_Scale~JSF+ORC_inno,DATA, na.action=na.omit)
summary(solution6)
```

	Min	1Q	Median	3Q	Max
	-1.3824	-0.4511	0.1684	0.5159	1.3422

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.55614	0.18736	40.329	<2e-16 ***
JSF	-0.72458	0.03832	-18.909	<2e-16 ***
ORC_inno	0.36162	0.22833	1.584	0.115

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Your b is (= - 0.72) and statistically significant
- Your d is (=0.00) as the value is not statistically significant
- Indirect effect: $a * b = -3.37$
- Total effect: $(a*b) + d = -3.37$ (d is here 0)
- Consequently, an innovative culture seems to lower intention to quit through higher levels of job satisfaction

2. A final words about controls

Simple Regression:

- Outcome = $c + b1 * \text{Predictor} + b2 * \text{Control1} + b3 * \text{Control2} \dots$

Moderation:

- Outcome = $c + b1 * \text{Predictor} + b2 * \text{Moderator} + b3 * (\text{Predictor} * \text{Moderator}) + b4 * \text{Control1} + b5 * \text{Control2} \dots$

Mediation:

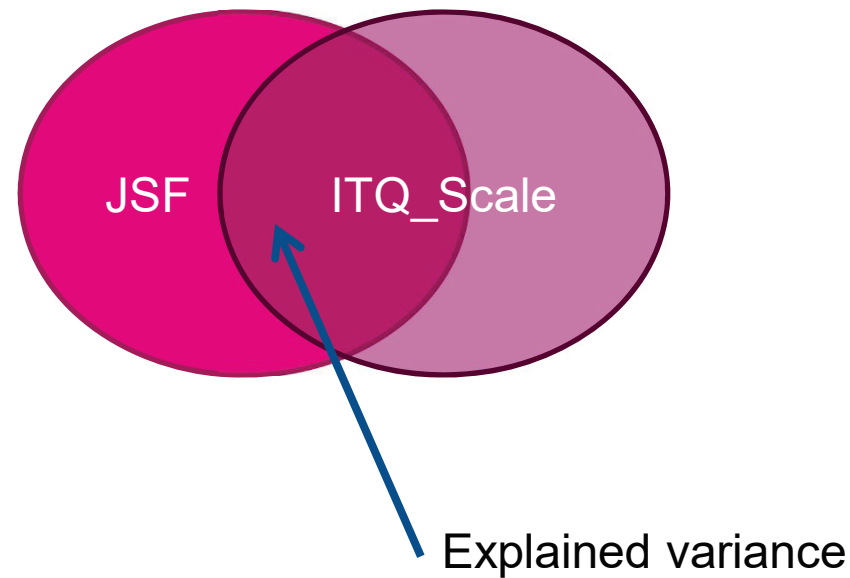
- Mediator = $c1 + a * \text{Predictor} + f1 * \text{Control1} + f2 * \text{Control2} \dots$
- Outcome = $c2 + b * \text{Mediator} + d * \text{Predictor} + g1 * \text{Control1} + g2 * \text{Control2} \dots$

2. A final words about controls

4 How is job satisfaction related to intention to quit

Predictor
JSF (1-10)

Outcome
ITQ_Scale (1-7)



2. A final words about controls

4 How is job satisfaction related to intention to quit

Predictor

JSF (1-10)

Control

WPR (1-100)

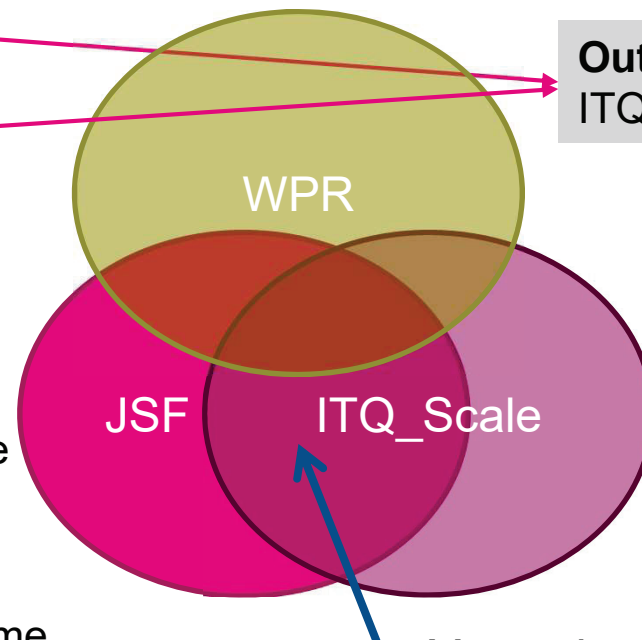
Outcome

ITQ_Scale (1-7)

The problem:

Part-time workers are in general less satisfied, because their jobs are less interesting. Yet they are less likely to have high intentions to quit because finding another part-time job may be harder.

→ Your findings may be biased by part-time work, and therefore you need to control for it



You only are interested in the variance not explained by WPR

2. A final words about controls

4 How is job satisfaction related to intention to quit

our results still hold:

```
solution3 <- lm(ITQ_Scale~JSF + WPR,DATA, na.action=na.omit)
summary(solution3)
```

.action

Residuals:

Min	1Q	Median	3Q	Max
-1.4951	-0.3567	0.2751	0.5160	1.3608

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.756157	0.373671	18.080	<2e-16 ***
JSF	-0.692539	0.024930	-27.780	<2e-16 ***
WPR	0.008165	0.004354	1.875	0.0628 .

Signif. codes:

0 '***' 0.001 '**' 0.01 '.' 0.1 ' ' 1

Residual standard error: 0.8061 on 143 degrees of freedom
(1 observation deleted due to missingness)

- When controlled for working percentages, job satisfaction is still significantly and positively related to intention to quit ($b=-0.69$, $p < .05$). Indeed, working percentages have no significant relationship with intention to quit ($b= 0.01$, $p > .05$) suggesting that you would not even need this control in the model