

# Methoden in de psychologie: 2020-2021

## PC-sessie 2: Lineaire regressie - interacties

### Opgave

---

In deze oefenles maken we gebruik van de statistische software R via het programma RStudio.

De commando's die je kunt gebruiken, zijn terug te vinden in het document met voorkennis (**voorkennis\_R**, zie Ufora), de nota's van **PC sessie 1** en de cursusnota's. Zorg dat je deze zaken goed doornomen hebt!

Om je op weg te helpen voor deze specifieke oefenles, hebben we het script **Opgave\_script\_PC2.R** op Ufora ter beschikking gesteld.

### Oefening 1: Predictoren voor stress

In deze oefening maken we opnieuw gebruik van de data uit een studie die de invloed van een aantal variabelen op stress nagaat ( $n = 377$ ). Zie **PC sessie 1** voor de uitleg over de data en de verschillende variabelen.

```
# path waar je databestand staat correct specificeren  
# of working directory in R wijzigen  
# datastress<-read.csv("DataStress.csv", sep=";", dec=".")  
  
# Data kunnen ook als volgt ingelezen worden:  
datastress<-read.csv("http://www.da.ugent.be/datasets/DataStress.csv",  
                    sep=";", dec=".")  
class(datastress)
```

```
[1] "data.frame"
```

```
dim(datastress)
```

```
[1] 377  11
```

```
head(datastress)
```

	SEX	EDUCLEV	CHI	AGE	DECAUT	JOBINS	PSYDEM	TSOSUP	DEPRES	SKILLUT
1	man	laag	2	59.45893	40	3.0	34	30	20	0
2	man	hoog	6	46.40298	48	3.0	34	28	19	0
3	man	hoog	3	38.74179	40	6.0	32	24	32	-2
4	man	middel	0	52.78679	40	4.5	25	14	18	-2
5	vrouw	middel	7	39.89203	36	4.5	24	15	25	0
6	vrouw	middel	2	36.72416	44	6.0	26	30	26	0

  

	STRESS
1	41.43302
2	99.78582
3	43.82572
4	16.74125
5	147.29815
6	80.11587

We starten met het lineair regressiemodel met `stress` als uitkomstvariabele en vijf predictoren, namelijk `chi`, `age`, `depres`, `decaut` en `tsosup`.

1. Ga na of een model met een interactie tussen `age` en `decaut` de variantie in de uitkomstvariabele `stress` beter kan verklaren dan een model zonder deze interactie.
  - (a) Welk lineair model schatten we om na te gaan of de interactie bijdraagt tot het model? Schrijf dit neer en interpreteer de parameters van het model.
  - (b) Wat is de nulhypothese die hoort bij de toets die we gaan uitvoeren?
  - (c) Voer de toets uit in R. Rapporteer de geobserveerde toetsingsgrootte en formuleer een conclusie.
  
2. In een volgende stap staan we stil bij mogelijke problemen met multicollineariteit in het bovenstaande model.
  - (a) Bekijk de correlaties tussen de verschillende predictoren (zonder de interactie). Wat merk je?
  - (b) Maak zelf een variabele aan die in bovenstaand model codeert voor de interactie tussen `age` en `decaut`. Bereken de correlatie tussen deze variabele en respectievelijk de variabelen `age` en `decaut`. Wat stel je vast?
  - (c) Ga na of er potentiële problemen zijn met multicollineariteit in het bovenstaande model? Wat zijn je bevindingen?
  - (d) Maak 2 nieuwe variabelen aan: `age_c` en `decaut_c` die de gecentreerde scores van respectievelijk `age` en `decaut` bevatten. Maak ook een variabele aan die codeert voor de interactie tussen `age_c` en `decaut_c`.  
Als je de correlatie tussen beide gecentreerde variabelen en hun interactieterm beschouwt, wat stel je vast?
  
3. Vervang in het oorspronkelijk model met de interactie tussen `age` en `decaut` de variabelen `age` en `decaut` door de respectievelijke gecentreerde variabelen.

- (a) Welk lineair model schatten we om na te gaan of de gecentreerde interactie bijdraagt tot het model? Schrijf dit neer en interpreteer de parameters van het model.
- (b) Voer de toets uit in R en formuleer je conclusie.
- (c) Ga opnieuw na of er mogelijks problemen zijn met multicollineariteit.

## Oefening 2: pijneducatie

In deze oefening maken we opnieuw gebruik van de data uit de studie rond pijneducatie  $n = 121$ . Zie de cursunota's en **PC sessie 1** voor de uitleg over de data en de verschillende variabelen.

```
# path waar je databestand staat correct specificeren
# of working directory in R wijzigen
# pijneducatie<-read.csv("pijneducatie.csv", sep=";", dec=".")

# Data kunnen ook als volgt ingelezen worden:
pijneducatie<- read.csv("http://www.da.ugent.be/datasets/pijneducatie.csv",
                        sep=";", dec=".")
class(pijneducatie)
```

```
[1] "data.frame"
```

```
dim(pijneducatie)
```

```
[1] 121  5
```

```
head(pijneducatie)
```

	Buig	Gender	Leeft	Conditie	Dep
1	-10.863835	Man	39.51317	Rugpijneducatie	33.214520
2	1.699991	Vrouw	40.25680	Rugpijneducatie	23.221516
3	5.249518	Man	35.34418	Rugpijneducatie	35.309551
4	9.263604	Vrouw	33.27902	Rugpijneducatie	15.787603
5	9.523696	Man	39.63532	Rugpijneducatie	8.411790
6	31.783503	Vrouw	25.09226	Rugpijneducatie	1.546317

Fit een model voor de uitkomst Buig (i.e. verschilscore in voorover buigen) met een interactie tussen de nominale predictoren Conditie en Gender, waar ook gecontroleerd wordt voor Dep en Leeft. We gaan na of er een effect is van Conditie op Buig. Aangezien Conditie zowel als hoofdeffect als in een interactie met Gender in het model opgenomen is, beschouwen we toetsen voor de volgende effecten in het model: hoofdeffect van Conditie en interactie-effect tussen Conditie en Gender is.

In bordsessie 2 komen we ook op deze oefening terug.

1. Schrijf het bovenstaande lineair model neer.
2. Geef de nul- en alternatieve hypothesen die horen bij de toets voor het hoofdeffect en de toets voor het interactie-effect.
3. Voer de toetsen uit in R en formuleer een conclusie.
4. Maak een effectenplot om het interactie-effect tussen **Conditie** en **Gender** te interpreteren.