**Toetsende Statistiek**Docent: Manfred te Grotenhuis | Vakcode: SOW-MTB2009

***Deze samenvatting is afkomstig uit het collegejaar 2017-2018. Het kan zijn dat sommige onderdelen iets verschillen van de huidige tentamenstof. Let hier op! Dit document bevat een samenvatting van de hoorcolleges.***

**Cursusomschrijving:**

* Toets op gemiddelde en proportie;
* De samenhang tussen kwalitatieve variabelen: tabelanalyse;
* Toetsen voor vergelijking van twee onafhankelijke groepen;
* Toetsen voor vergelijking van twee individueel gekoppelde groepen;
* Vergelijkingstoetsen voor meer dan twee onafhankelijke groepen;
* Samenhang tussen kwantitatieve variabelen
* Inleiding regressie-analyse.

**Doelstellingen:**

De student:

* is in staat om gegevens uit grootschalig veldonderzoek voor te bereiden op statistische analyses met behulp van SPSS;
* is in staat statistische toetsen toe te passen die gangbaar zijn in sociaalwetenschappelijk onderzoek;
* kan de genoemde statistische procedures toepassen met behulp van SPSS, waarbij de nadruk ligt op de SPSS-commandotaal;
* is in staat de uitkomsten van SPSS te interpreteren en te rapporteren;
* kan systematisch en analytisch sociaalwetenschappelijk onderzoek te beoordelen.

# 

# **College 1**

## Definities

Statistisch toetsen: verdeling in de steekproef lijkt op die in de populatie. De verdeling van leeftijd is ongeveer een driehoek, met veel jongeren en weinig ouderen in de steekproef. De steekproevenverdeling ziet er uit als een normaalverdeling.

*μ is gemiddelde in de populatie, σ(x) = standaardafwijking in de populatie*

*is gemiddelde in de steekproef, s = standaardafwijking in de steekproef*

*SE = standaardfout, hoeveel een steekproef afligt van de gemiddelde steekproef*

BEREKENING STANDAARDFOUT IN STEEKPROEF = σ / √n

n = omvang steekproef

### Wanneer de populatie onbekend is

Gebruik dan **en s! SE=s/ √n

## Enkelzijdige hypothesetoetsing

Onderzoekshypothese 🡪 verandering

Nulhypothese 🡪 gelijk gebleven

Verwerpingsgebied 🡪 alfa 🡪 0,05

overschrijdingskans 🡪 p. als p-waarde kleiner dan a, dan nulhypothese verwerpen.

## Betrouwbaarheidsintervallen

90% betrouwbaarheidsinterval is 1,65 standaardfout van gemiddelde af. 95% betrouwbaarheidsinterval is 2 (1,96) standaardfout van gemiddelde af. Hoe groter de steekproef, hoe smaller de bi wordt.

BEREKENING BI

1. Stel de breedte vast (90/95)
2. Bereken 1,65 of 1,96 \* standaardfout
3. Dus **+/- (1,65/1,96 \* standaardfout)

## Standaardaanpak

1. Bereken iets in de steekproef waarmee je een hypothese kunt toetsen
2. Bereken de bijbehorende standaardfout
3. Bereken de p-waarde
4. Vergelijk de p-waarde met a 🡪 SPSS houdt geen rekening met enkelzijdige toetsen
5. Bereken het betrouwbaarheidsinterval

# **College 2**

Meetniveau’s

* Nominaal
  + Volgorde maakt niet uit
* Ordinaal
  + Volgorde maakt uit 🡪 rangorde
  + Je mag de mediaan en modus uitrekenen
* Interval
  + Rangorde met zelfde afstand
* Ratio
  + Rangorde met zelfde afstand en 0-punt
* Dichotoom
  + 2 categorieën

t-waarde = hoeveel SE ligt het gevonden gemiddelde af van H0-hypothese. Vanaf t=3 is het significant

BEREKENEN T-WAARDE

Absolute afstand/SE = t-waarde

**/SE = t

## Standaardaanpak toets op gemiddelde

1. Bereken een gemiddelde
2. Bereken de bijbehorende standaardfout
3. Bereken de p-waarde
4. Vergelijk de p-waarde met alfa
5. Bereken het betrouwbaarheidsinterval

Toets in SPSS met T-TEST

## Standaardaanpak toets op proportie

1. Bereken de proportie
2. Bereken de p-waarde
3. Vergelijk p-waarde met alfa
4. Bereken een bepaalde betrouwbaarheidsinterval
   1. DIT KAN ALLEEN ALS p1 +/- 3 SE TUSSEN 0 EN 1 LIGT

Toets in SPSS met NPAR TESTS.

# **College 3**

## Onafhankelijke groepen

* 2 aselecte steekproeven

Aanpak gemiddelde:

1. Maak een tabel met gemiddelden per groep (((means y(afhankelijk) by x(onafhankelijk))))
2. Kijk naar de standaardafwijkingen. Verschillend? Dan kan je niet pooled (gemiddelde) standaardafwijking nemen
3. Gebruik een T-TEST als n=30 of meer per groep
4. Gebruik de rij equal variances assumed als de Sig. groter of gelijk is aan 0,05

Aanpak proportie:

1. Maak een tabel met proporties en standaardafwijking
2. Gebruik een T-TEST
3. Standaardafwijkingen spelen geen rol, dus let alleen op de bovenste rij

## Afhankelijke groepen

* SE is kleiner, want minder toeval in de data
* Relationele groepen, een groep op twee tijdstippen

Aanpak gemiddelde

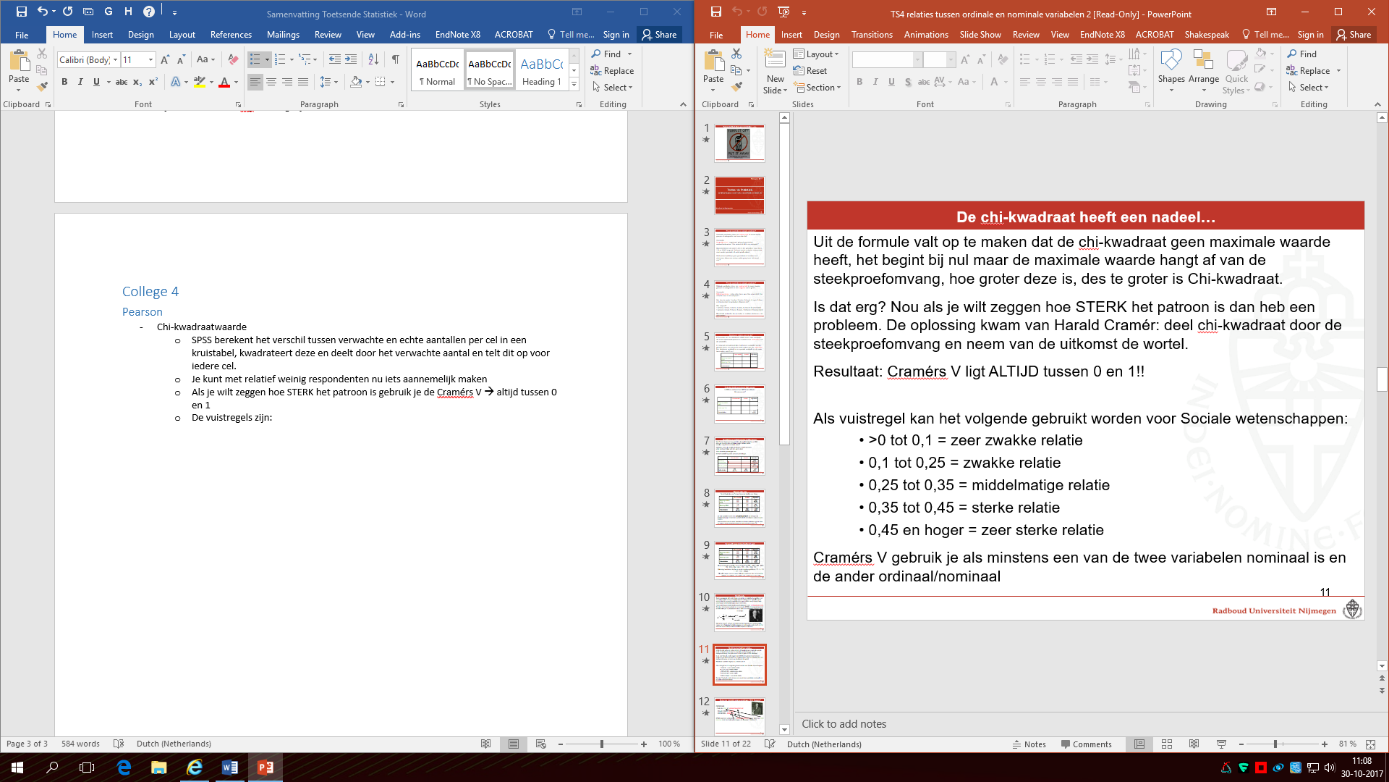
1. Bereken het verschil tussen de groepen
2. Maak een tabel met gemiddelde en standaardafwijking (tabel over het verschil)
3. Toets toepassen als: n=30 of meer en er sprake is van afhankelijkheid
4. Gebruik een T-TEST

Aanpak proporties

1. Gebruik een kruistabel met marginale resultaten
2. Gebruik de toets van MCNEMAR om significantie te berekenen
3. Kijk dan naar Exact Sig. en vergelijk dit met alfa

# **College 4**

## Pearson

* Chi-kwadraatwaarde
  + SPSS berekent het verschil tussen verwachte en echte aantallen per cel in een kruistabel, kwadrateert deze en deelt door het verwachte aantal en telt dit op voor iedere cel.
  + Je kunt met relatief weinig respondenten nu iets aannemelijk maken
  + Als je wilt zeggen hoe STERK het patroon is gebruik je de Cramérs V 🡪 altijd tussen 0 en 1
  + De vuistregels zijn:
  + Wordt gebruikt als één van de variabelen nominaal is en de andere ordinaal

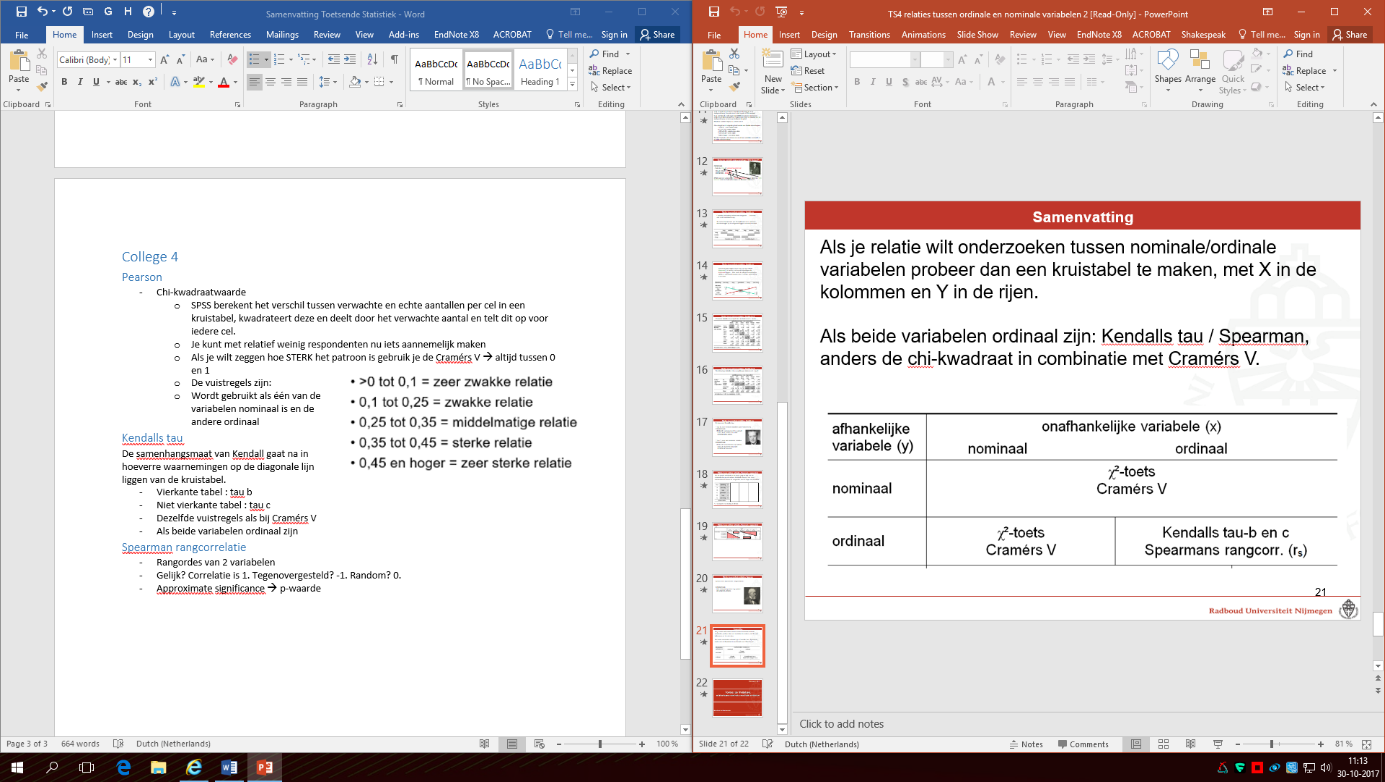
## Kendalls tau

De samenhangsmaat van Kendall gaat na in hoeverre waarnemingen op de diagonale lijn liggen van de kruistabel.

* Vierkante tabel : tau b
* Niet vierkante tabel : tau c
* Dezelfde vuistregels als bij Cramérs V
* Als beide variabelen ordinaal zijn

## Spearman rangcorrelatie

* Rangordes van 2 variabelen
* Gelijk? Correlatie is 1. Tegenovergesteld? -1. Random? 0.
* Approximate significance 🡪 p-waarde

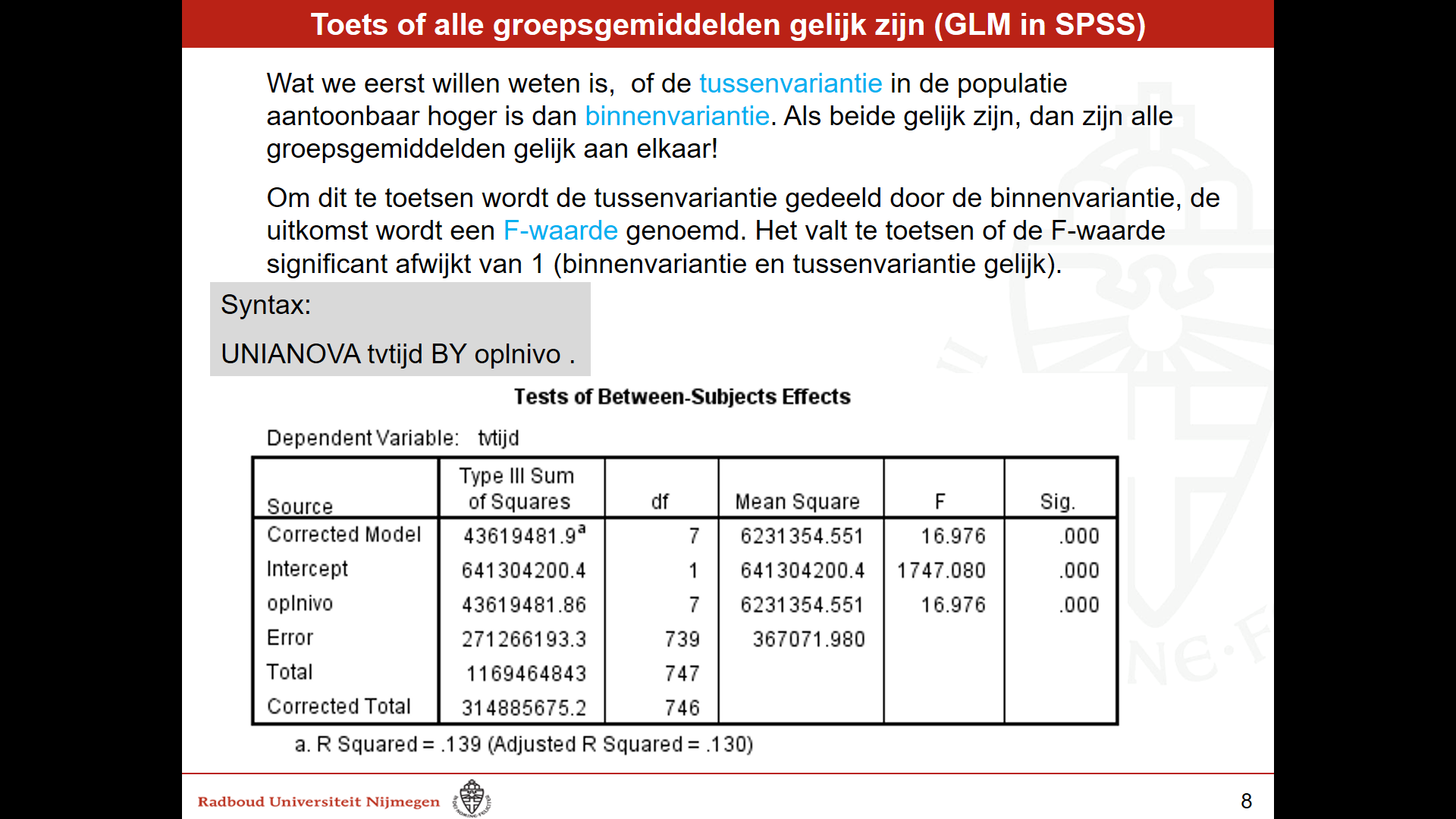


Je mag altijd een meetniveau lager meten, maar nooit hoger. Gebruik Pearson alleen als beide variabelen ordinaal zijn.

# **College 5**

## Vergelijken van 2 of meer gemiddelden.

ANOVA (analysis of variances)

* Kijken naar de variantie **tussen**en **binnen** groepen.
* De verhouding tussen beide varianties is van belang bij de toetsing van de nulhypothese
* Nulhypothese is dat alle groepsgemiddelden gelijk zijn aan elkaar
* Binnenvariantie laag en tussenvariantie hoog 🡪 individuen liggen dicht bij groepsgemiddelde en groepsgemiddelde liggen ver uit elkaar
* Binnenvariantie hoog en tussenvariantie laag 🡪 individuen liggen ver van hun groepsgemiddelde en groepsgemiddelden liggen dicht bij elkaar
* Om tusssen en binnenvariantie uit te rekenen gebruik je de f-waarde. Er is geen bovenwaarde en hij komt amper onder de 1 uit.
* F-waarde is het getal achter oplnivo in de F kolom.
* Binnenvariantie is mean square achter oplnivo
* Tussenvariantie is mean square achter error
* Als de sig. achter de f-waarde kleiner of gelijk is aan 0,05, dan wordt de H0-hypothese verworpen

## Hoeveel wijken de groepsgemiddelden dan van elkaar af?

* Mogelijkheid 1
  + Rekening houden met *kanskapitalisatie*
  + Dit is de toets van Bonferroni
  + Gebruik je als je nog geen hypothesen hebt
* Mogelijkheid 2
  + Simple
  + Dit is toetsen ten opzichte van een referentie categorie
  + Dus level 1 vs. Level 3
  + Level 2 vd. Level 3 enz.
* Mogelijkheid 3
  + Repeated
  + Alleen voor ordinale variabelen
  + Dit is toetsen ten opzichte van de volgende categorie
  + Dus level 1 vs. Level 2
  + Level 2 vs. level 3 enz
* Design
  + Groepen van het experiment

## Meerdere variabelen

Je kunt controlevariabelen toevoegen. UNIANOVA y BY x z WITH z. de eerste z is van nominaal of ordinaal niveau, de tweede z van interval of ratio niveau.

Als je groep\*voormeting aanmaakt 🡪 voormeting verschilt per groep

## Volgorde van toetsen

1. Toets zonder controlevariabelen
2. Dan verstorende variabelen invullen
3. Toets variabelen hetzelfde in alle groepen via design
4. Niet significant? Maak uiteindelijke toets

# **College 6**

3 voorwaarden voor causale interpretatie

1. Er moet samenhang tussen x en y zijn
2. Tijdsvolgorde: x komt voor y
3. De samenhang is niet het resultaat van een andere variabele

Interpretatiemodel

* X 🡪 z 🡪 y

Schijnrelatiemodel

* Z verklaart de relatie tussen x en y, die dus eigenlijk niet bestaat

Hybridemodel

* Geen zuivere interpretatie, dus er blijft nog een kleine directe relatie over

Suppressiemodel

* Z onderdrukt de relatie tussen x en y

Interactiemodel

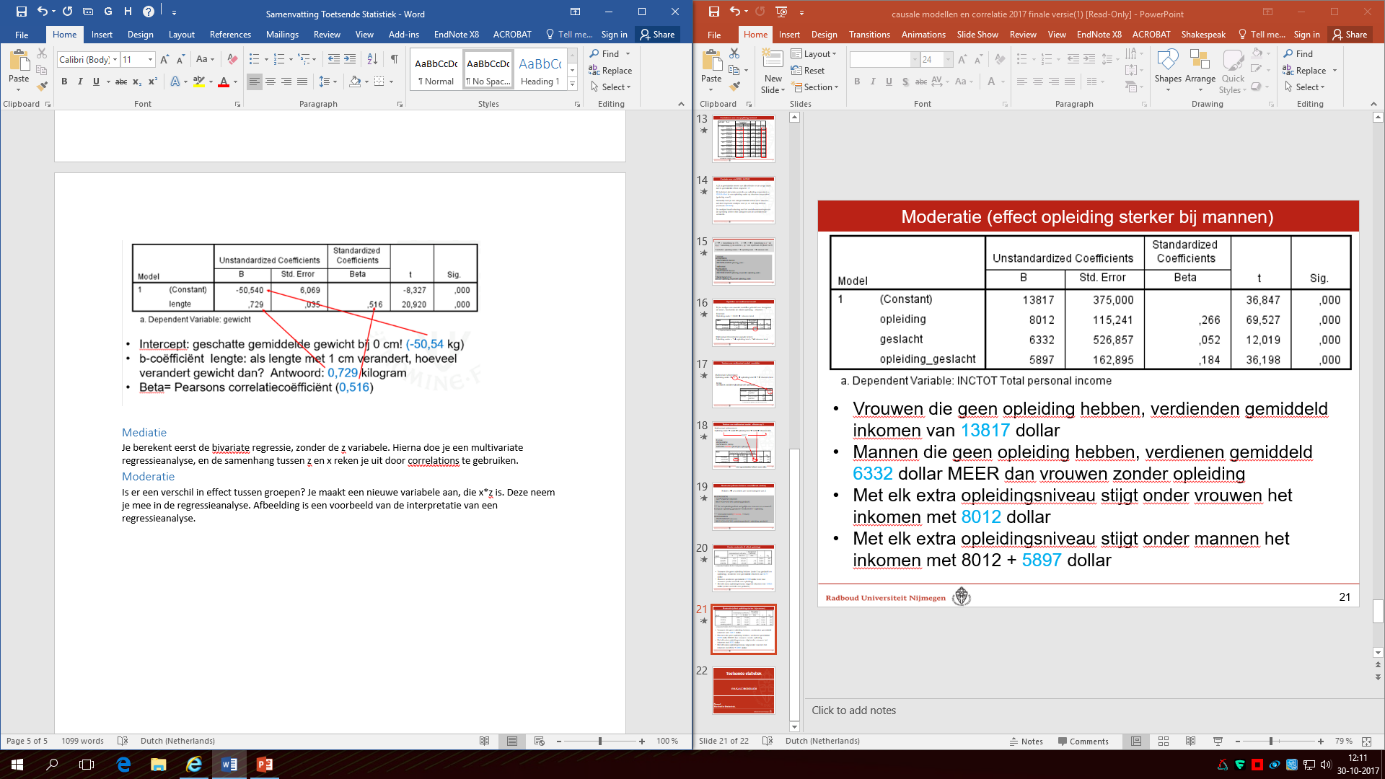
* Z is anders voor verschillende groepen

## 

## Mediatie

Je berekent eerst de bivariate regressie, zonder de z variabele. Hierna doe je een multivariate regressieanalyse, en de samenhang tussen z en x reken je uit door correlations te gebruiken.

## Moderatie

Is er een verschil in effect tussen groepen? Je maakt een nieuwe variabele aan, die x\*z is. Deze neem je mee in de regressieanalyse. Afbeelding is een voorbeeld van de interpretatie van een regressieanalyse.