

Deel 1

Psychologie | Rijksuniversiteit Groningen

2026 - 2027

RGPS-1106-003 | €16,95

Tentamengericht | Overzichtelijke structuur

Sinds 1994 | Beoordeeld met een 8,2



| Slim
Academy



Content editors gezocht!

Zoek jij een flexibele bijbaan die perfect aansluit op je studie én die je vanuit huis kunt doen?



Bij Slim Academy verdien je geld met iets wat je toch al doet: leren. Door het samenvatten en nakijken van je eigen lesstof herhaal je alles wat jij voor je tentamen moet weten terwijl je werkt!

Wat bieden wij:

- Een mooie vergoeding
- Een **flexibele bijbaan** waarbij je werkt wanneer het jou uitkomt
- Je werkt **vanuit huis**: Geen reistijd, gewoon lekker in je eigen omgeving
- Een mooie aanvulling op je CV
- **Gratis toegang** tot al ons studiemateriaal
- Een baan die perfect aansluit bij je studie



Als **Content Editor** draag je bij aan de kwaliteit van onze samenvattingen door ze te controleren op inhoud en ze up-to-date te maken.



Slim
Academy

Geïnteresseerd? Stuur je cv en motivatie naar klantenservice@slimacademy.nl

Welkom bij Slim Academy!

Voor je ligt de **literatuur en hoorcollege samenvatting** van het vak **Statistiek 2**, dat je volgt in blok 1 van het tweede jaar van de bachelor Psychologie aan de Rijksuniversiteit Groningen. In dit vak leer je over regressie, correlaties en relaties in de statistiek.

Door de literatuur en hoorcolleges te bundelen, kunnen wij ervoor zorgen dat je het materiaal voor vakken die **twee periodes duren** op tijd thuis ontvangt. In Deel 1 wordt alle stof uit blok 1 behandeld, in Deel 2 vind je de stof uit blok 2 terug. Zo blijf je altijd goed voorbereid! Voor een optimale voorbereiding op je tentamen raden we je aan gebruik te maken van ons volledige aanbod. Voor dit vak zijn de onderstaande producten beschikbaar.

Product	Inhoud	Content	Versies
Samenvatting Deel 1	Bevat zowel de literatuur als hoorcolleges. De literatuur bestaat uit alle opgegeven tentamenstof. De hoorcolleges zijn gebaseerd op de belangrijkste punten die de professor heeft behandeld.	Week 1 t/m 6	✓ App ✓ Papier
Samenvatting Deel 2	Bevat zowel de literatuur als hoorcolleges. De literatuur bestaat uit alle opgegeven tentamenstof. De hoorcolleges zijn gebaseerd op de belangrijkste punten die de professor heeft behandeld.	Week 7 t/m 12	✓ App ✓ Papier
Oefenvragen	Interactief oefenen in de leeromgeving. Zo kun je nagaan of je de stof volledig beheerst en weet je beter wat je kunt verwachten op het tentamen.	• 12 sets; • 250+ vragen	✓ App
Flashcards	Kaarten die je helpen om definities, schema's en kernbegrippen snel te herhalen, zodat je op een actieve manier leert en de stof sneller en beter onthoudt.	• 12 sets; • 150+ kaarten	✓ App

Contact

Heb je vragen over ons aanbod, de levering van je samenvattingen of iets anders? Neem gerust contact met ons op via **06 81 60 19 60** of stuur een e-mail naar **klantenservice@slimacademy.nl**.

Feedback geven

Kan je ons vertellen wat er mis is met de geselecteerde tekst?

Verstuur feedback

Fout gevonden?

Geef het door en wij zorgen dat de samenvatting snel wordt bijgewerkt!

Hoe werkt het?

Selecteer de tekst

Klik op het vlaggetje

Verstuur je feedback

Deze samenvatting is alleen voor persoonlijk gebruik en mag niet worden gekopieerd of verspreid. De tekst is opgesteld naar eigen inzicht van de auteur en bedoeld als ondersteuning bij het studeren. Raadpleeg altijd de originele bronnen.

Digitale en papieren updates

Bij Slim Academy ontvang je **iedere week een digitale update van de samenvatting**. Aan het begin van het blok zijn nog niet alle lessen geweest, daarom vullen wij de samenvatting wekelijks aan zodra onze auteurs de nieuwste lesstof hebben gevolgd. Zo blijft de inhoud **actueel, betrouwbaar en van hoge kwaliteit**. Je hoeft dus niet te wachten tot het einde van het blok: elke week komt er een nieuw deel beschikbaar. In het onderstaande overzicht zie je **wanneer welke stof digitaal beschikbaar is**.

Later in het blok bundelen we alle wekelijkse updates tot één overzichtelijke, **papieren samenvatting**. Deze wordt op **6 oktober 2026** verstuurd en bevat **week 1 t/m 4**. Omdat **week 5 en 6** pas later in de lessen worden behandeld, is dit onderdeel **uitsluitend digitaal in de app** beschikbaar. Komt er later in het blok nog nieuwe stof bij? Geen zorgen! Zodra die verschijnt, vatten we alles snel samen en sturen we een update in de app.

De verplichte tentamenstof komt uit het boek van Moore, McCabe & Craig (2021). Introduction to the Practice of Statistics. 10th edition. ISBN 978-1-319-38366-4. en Agresti: Agresti (2018). Statistical Methods for the Social Sciences. 5th Global Edition. ISBN 978-1-292-22031-4.

Stof		Datum
Week 1	Week 1: Simpele lineaire regressie 1	14-09-2026
Week 2	Week 2: Simpele lineaire regressie 2	21-09-2026
Week 3	Week 3: Inferentie voor regressie en correlatie	28-09-2026
Week 4	Week 4: Multivariate relaties	05-10-2026
Week 5	Week 5: Meervoudige lineaire regressie 1	12-10-2026
Week 6	T.B.A	23-10-2026

JOIN DE WHATSAPP GROEP

- ✓ Chat met jouw mede-studenten
- ✓ Stel al jouw (studie)vragen aan onze studie-experts
- ✓ Krijg extra oefenvragen om jouw kennis te testen
- ✓ Krijg gratis voorbeeldsamenvattingen





We wensen je veel succes met studeren en het behalen van je tentamens!

Team Slim Academy

Inhoudsopgave

Welkom bij Slim Academy!	1
Digitale en papieren updates	2
Inhoudsopgave	3
Week 1: Simpele lineaire regressie	4
College week 1	4
Literatuur week 1	7
Week 2: Simpele lineaire regressie 2	17
College week 2	17
Literatuur week 2	21
Week 3: Inferentie voor regressie en correlatie	25
College week 3	25
Literatuur week 3	29
Week 4: Multivariate relaties	31
College week 4	31
Literatuur week 4	34
Week 5: Meervoudige lineaire regressie 1	37
College week 5	37
Literatuur week 5	40

Week 1: Simpele lineaire regressie

College week 1

Introductie

In dit hoofdstuk staat college 1 van statistiek 2 centraal, waarin het regressiemodel behandeld wordt en de schatters ervan. Hoofdstuk 2 en 10.1 uit het M&M boek worden behandeld in dit college.

Het regressiemodel

Bij **simpele lineaire regressie** wordt gekeken naar hoe variabele x invloed heeft op variabele y . Als x verandert, verandert y ook. Bij simpele lineaire regressie zijn er twee belangrijke variabelen: de *response* en de *explanatory variable*.

- De **response variable** is ook wel de **afhankelijke variabele** (y);
- De **explanatory variable** is ook wel de **predictor** en **onafhankelijke variabele** (x).

Met de *predictor* (x) probeer je de afhankelijke variabele (y) te voorspellen. Beide variabelen zijn continue variabelen, dus de variabelen kunnen allerlei waarden aannemen in een bepaald interval. Belangrijke doelen van regressie zijn, naast kijken of er een lineaire relatie bestaat tussen x en y , het voorspellen van y met x en het bestuderen van de relatie tussen y en x .

Het statistische model voor lineaire regressie gaat ervan uit dat voor elke waarde van x , y normaal verdeeld is. Het gemiddelde van y wordt dan zo genoteerd: μ_y , afhankelijk van x . Dit betekent dat verschillende waarden van x zorgen voor verschillende gemiddelde waarden van y . De gemiddelde waarde van y , wordt ook wel de *mean response* genoemd. Als er een lineaire regressie uitgevoerd wordt, wordt ervan uitgegaan dat deze gemiddelde μ_y op een rechte lijn liggen. Daar hoort de volgende formule bij:

$$\mu_y = \beta_0 + \beta_1 x$$

Waarbij de variatie in y , gemeten met de standaarddeviatie σ , gelijk is voor alle waarden van x .

Dit idee kan geschreven worden als: **DATA = FIT + RESIDUAL**:

- **FIT** = de gemiddelden in elke subpopulatie;
 - $\beta_0 + \beta_1 x_i$;
- **RESIDUAL** = de afwijking van de data tot de lijn met populatiegemiddelden;
 - Ook wel afgekort als: ϵ ;
 - Een assumptie die bij *residuals* hoort, is dat de *residuals* normaal verdeeld zijn, met het gemiddelde van 0 en de standaarddeviatie σ ;
 - Variatie in y door andere bronnen dan x zijn de *unexplained source of variance*, ook wel de *error*, *residual* en *noise* genoemd.

Als de $\text{DATA} = \text{FIT} + \text{RESIDUAL}$ formule wordt geschreven in statistische termen ziet het er dan zo uit: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

Waarbij:

- $\beta_0 + \beta_1 x_i$ de gemiddelde reactie is wanneer $x = x_i$;
- ϵ_i zijn de *residuals* die verwacht worden bij een normale verdeling met het gemiddelde van 0 en de standaardafwijking σ ;
- De parameters bij dit model zijn:
 - Hoe hoog de lijn ligt: *intercept* β_0 ;
 - Hoe steil de helling is: *slope* β_1 ;
 - Hoe erg de spreiding rondom de regressielijn is: regressie standaard deviatie σ .

Een belangrijke vraag over data is de vraag: hoe goed de data bij het model past. Het stuk van de data dat past, wordt de fit genoemd en het stuk wat niet past, wordt de *residual* genoemd. Vandaar het volgende: $\text{DATA} = \text{FIT} + \text{RESIDUAL}$.

De *residual* geeft aan hoever het punt van de regressielijn aflight. Hoe dichter het punt bij de regressielijn ligt, hoe beter het model bij de data past. Andersom geldt hetzelfde, hoe verder het punt van de regressielijn ligt, hoe slechter het model bij de data past.

ϵ_i zijn de residuen waarbij aangenomen wordt dat ze normaal verdeeld zijn, met een gemiddelde van 0 en een standaarddeviatie σ .

Schatten van de regressieparameters

Bij statistiek 1a is geleerd dat *ordinary least squares (OLS) estimation*, de lijn schat om de associatie tussen y en x samen te vatten of te beschrijven. Dus om de regressielijn te schatten is de volgende formule te gebruiken:

$$\hat{y} = b_0 + b_1x$$

Waarbij:

- b_0 = schatting van de intercept parameter β_0 ;
- b_1 = schatting van de helling parameter of *slope* parameter β_1 ;
- s = schatting van de regressie standaarddeviatie σ .

Bij dit vak wordt de OLS-regressie lijn gebruikt als basis voor inferentie over een populatie van waaruit een (random) steekproef getrokken is, maar hier gaat het in het volgende hoofdstuk verder over.

De OLS-methode geeft de 'beste' lijn in termen van de kleinste fout in voorspelling van de kleinste som van de gekwadrateerde residuen. Dus het doel van OLS is: $\min \sum_i e_i^2$

Waarbij de *residuals* op de volgende manier berekend worden: $e_i = y_i - \hat{y}_i$

Om de geschatte regressielijn nog wat verder te verklaren:

Geschatte OLS-regressielijn: $\hat{y} = b_0 + b_1x$

$$b_1 = r \times \frac{SD(y)}{SD(x)}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Waarbij:

- r de correlatie tussen x en y in de steekproef is;
- $SD(x)$ en $SD(y)$ de standaarddeviatie van x en y in de steekproef zijn;
- $\bar{y}$ en $\bar{x}$ de steekproefgemiddeldes van x en y zijn.

Interpretatie van regressiecoëfficiënten

Bij de formule van de geschatte (OLS) regressielijn, $\hat{y} = b_0 + b_1x$, vertelt de b_1 wat over de richting van de regressielijn:

- $b_1 > 0 \rightarrow$ stijgende lijn, ook wel een positieve relatie tussen x en y ;
- $b_1 = 0 \rightarrow$ horizontale lijn, ook wel geen relatie tussen x en y ;
- $b_1 < 0 \rightarrow$ dalende lijn, ook wel een negatieve relatie tussen x en y .

σ schatten

σ^2 is de variantie van de residuen. s is de schatting van de regressie standaarddeviatie σ . s is ook de standaarddeviatie van de residuen in de steekproef, ook wel de regressie standaardfout. s^2 is de variantie van residuen. Deze variantie kun je berekenen door eerst het residu te berekenen voor elk individu en daarna de variantie te berekenen. Dit doe je met de volgende formule:

$$s^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

Gebruik van software

In deze cursus gaat er gebruik gemaakt worden van de SPSS en JASP software output. Het is daarbij belangrijk dat je in je achterhoofd houdt dat in de output alleen de eerste drie cijfers achter de komma gegeven worden. Je maakt dus gebruik van afgeronde getallen. Dit kan de uitkomst beïnvloeden.

Verkenning en beschrijving van de data

De data moet altijd eerst beschreven en verkend worden voordat er een analyse op gedaan wordt. Dus moeten er eerst descriptieve resultaten worden gegeven per variabelen. Dit zijn de gemiddeldes, standaarddeviaties en steekproefgrootte. Dit kan gedaan worden door grafische middelen zoals boxplots en histogrammen. Daarna moeten er descriptieve resultaten gegeven worden over de samenhang tussen de twee variabelen, zoals de correlatie. Dit kan gedaan worden aan de hand van een *scatterplot*. Dit is belangrijk om te checken, omdat het geen zin heeft om een lineair model te fitten, als het verband niet ongeveer lineair is. Een lineair model veronderstelt namelijk een rechte lijn relatie, en als die er niet is, zal het model misleidende of onjuiste conclusies opleveren. Hoe dit verder en beter uitgezocht kan worden, wordt in latere hoofdstukken uitgelegd.

Slim Samengevat!

- Bij **simpele lineaire regressie** zijn er twee variabelen belangrijk: de *response* (**afhankelijke variabele**, y) en de *explanatory* (**onafhankelijke variabele**, x).
- Het model wordt beschreven als **DATA = FIT + RESIDUAL**, waarbij FIT de gemiddelden in subpopulaties zijn en RESIDUAL de afwijkingen van de data tot de lijn met populatiegemiddelden.
- De **OLS-methode** (*ordinary least squares*) wordt gebruikt om de beste lijn te schatten door de kleinste som van de gekwadrateerde residuen te bereiken.
- De geschatte regressielijn is $y = b_0 + b_1x$, waarbij b_0 de **intercept**, b_1 de **helling** en s de schatting van de regressiestandaarddeviatie zijn.
- Regressiecoëfficiënten, zoals b_1 , geven de richting en sterkte van de relatie tussen y en x aan (positief, negatief of geen relatie).