



Broodje Statistiek

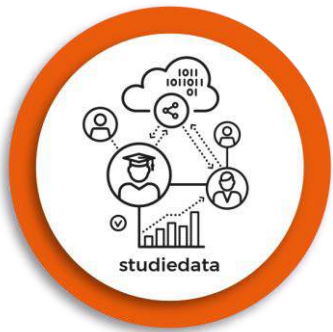
Webinar Statistisch Handboek Studiedata



Versnellingsplan
Onderwijsinnovatie
met ICT



16-12-2020



Zone Studiedata – projecten en producten

Het Statistisch Handboek Studiedata is een initiatief van de zone Studiedata, onderdeel van het Versnellingsplan ICT in het onderwijs.

- **Website**
- Routekaarten best practices
- **Quickscan studiedata**

Communicatie
en awareness

Scholing

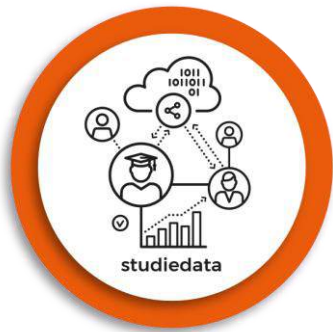
- **Statistisch Handboek Studiedata**
- Leergang data science
- Licentie DataCamp

- Functieprofielen
- **Referentiekader privacy & ethiek**
- Whitepaper Regie op studiedata

Governance en
bestuurlijk

IT en organisatie

- Nationaal Cohortonderzoek Onderwijs hoger onderwijs
- Whitepaper real-time studiedata
- **Simulatiedataset**
- Handleiding onderzoekers

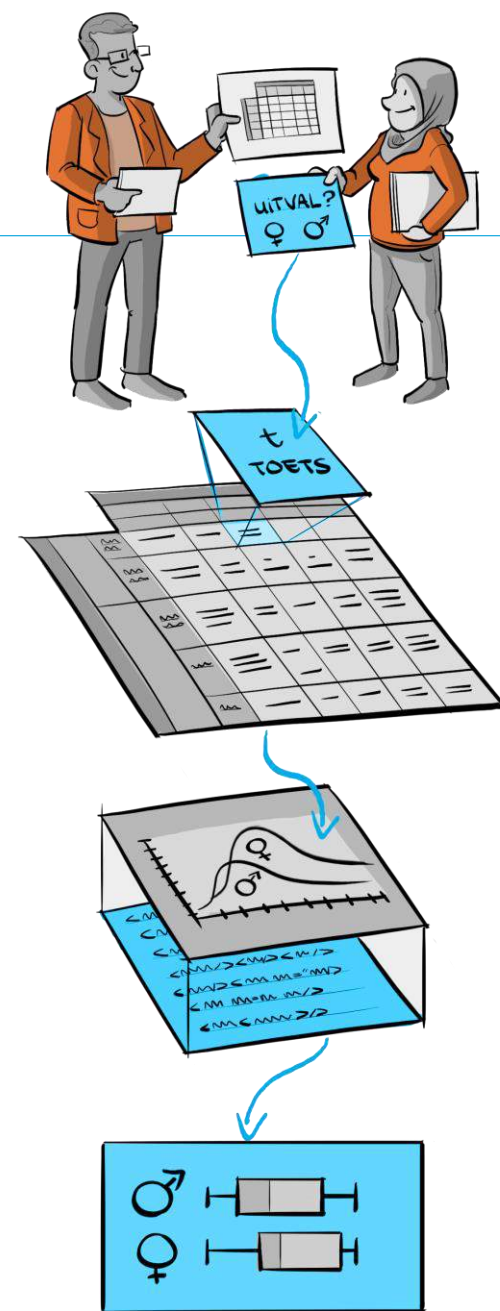


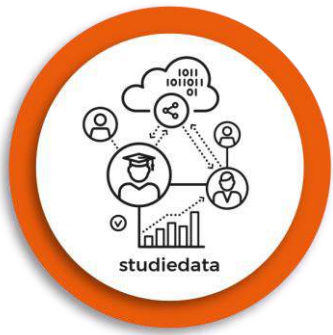
Mentimeter

Voor het volgende onderdeel maken we gebruik van Mentimeter

Ga naar www.menti.com

Vul in **93 01 67 9**

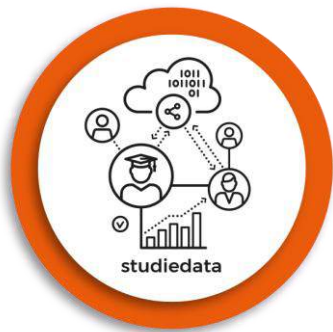




Opzet webinar

- 1 Regressiemodellen
- 2 Casus uit de praktijk
- 3 Het Statistisch Handboek in de praktijk gebruiken





Update Statistisch Handboek Studiedata

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|--|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoets I | Wilcoxon signed rank toets I / Tekentoets II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhappkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |

| | | Onafhankelijke variabelen | | | | |
|------------------------|------------------------------|---|---------------------------------------|--|-------------------------------------|-------------------------------------|
| | | cross-sectioneel | | longitudinaal | | hiërarchisch |
| | | categorische variabelen | continue en categorische variabelen | categorische variabelen | continue en categorische variabelen | continue en categorische variabelen |
| Afhankelijke variabele | continu (interval en ratio) | Multipale lineaire regressie / Factoriele ANOVA | Multipale lineaire regressie / ANCOVA | Factoriele repeated measures ANOVA / Mixed ANOVA | Multilevel lineaire regressie | Multilevel lineaire regressie |
| | gecensureerd continu | Survival analyse | Survival analyse | | | |
| | binair | Logistische regressie / Loglineaire analyse | Logistische regressie | Multilevel logistische regressie | Multilevel logistische regressie | Multilevel logistische regressie |
| | nominaal | Multinomiale regressie / Loglineaire analyse | Multinomiale regressie | Multilevel multinomiale regressie | Multilevel multinomiale regressie | Multilevel multinomiale regressie |
| | ordinaal | Ordinale regressie | Ordinale regressie | Multilevel ordinale regressie | Multilevel ordinale regressie | Multilevel ordinale regressie |
| | Meerdere continue variabelen | MANOVA | MANCOVA | | | |



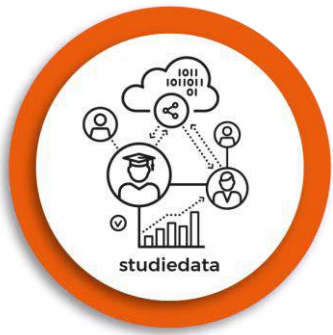
Opzet webinar

1

1 **Regressiemodellen**

2 Casus uit de praktijk

3 Het Statistisch Handboek in de praktijk gebruiken

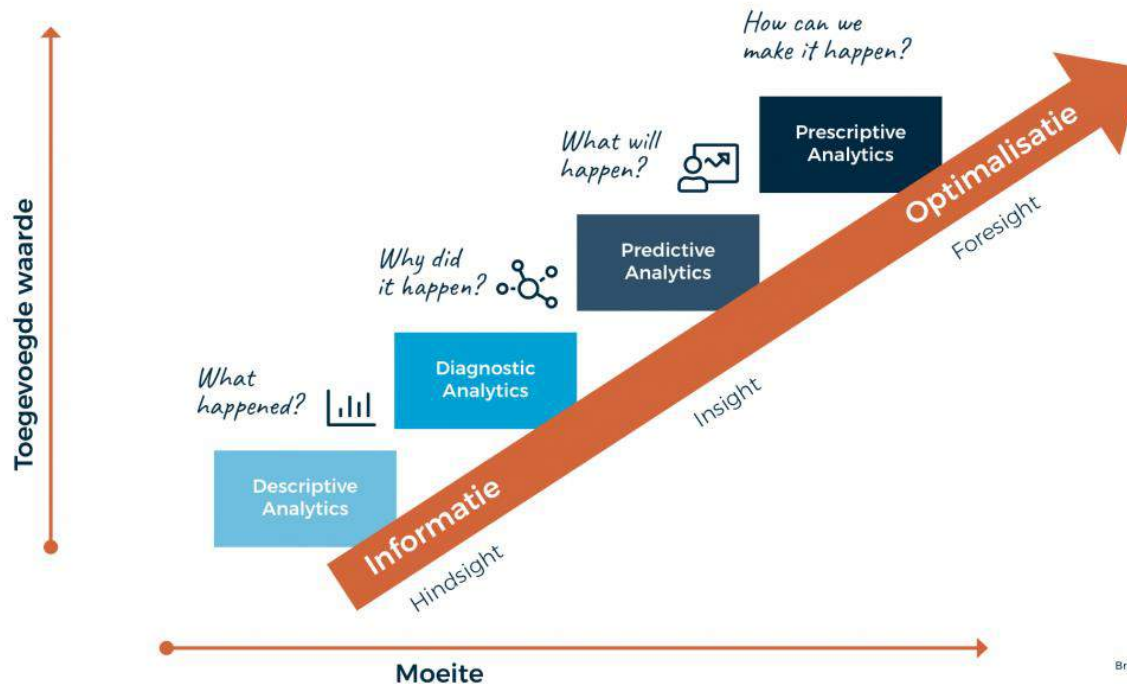


Regressiemodellen - Introductie

1

Positionering Statistisch Handboek binnen Analytics

- De eerste en tweede toetsmatrix zijn te plaatsen onder Diagnostic Analytics
- Regressiemodellen kunnen ook geplaatst worden onder Predictive Analytics



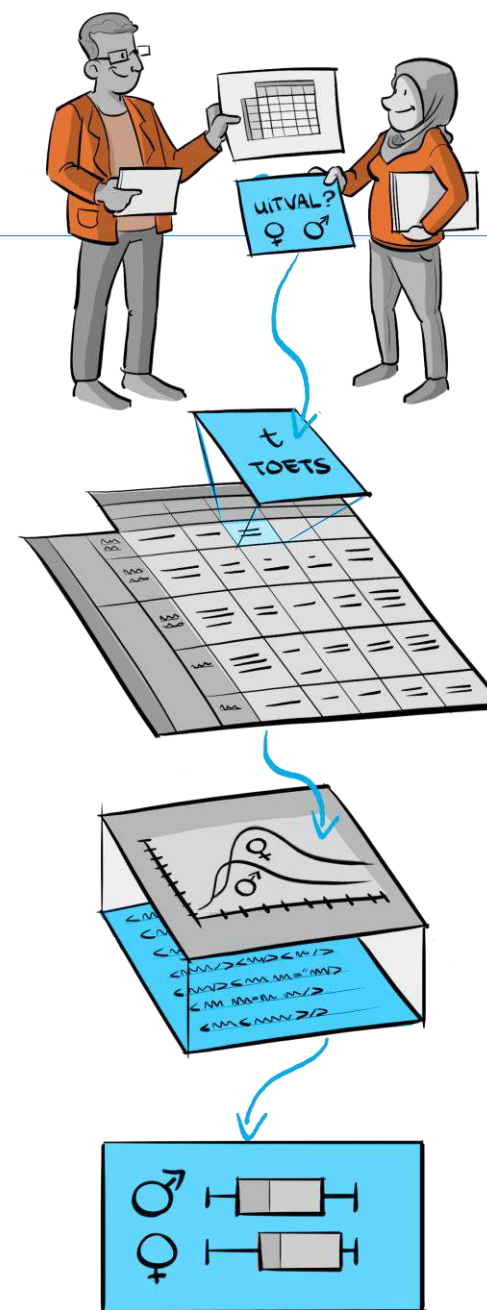


Mentimeter

Voor het volgende onderdeel maken we gebruik van Mentimeter

Ga naar www.menti.com

Vul in **93 01 67 9**





Regressiemodellen - Definitie

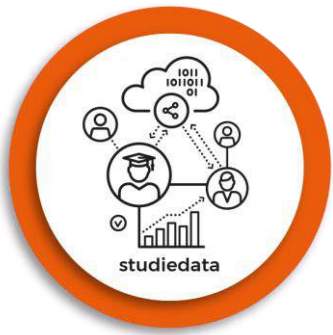
1

In een regressiemodel wordt een afhankelijke variabele y voorspeld met een of meerdere predictors x

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

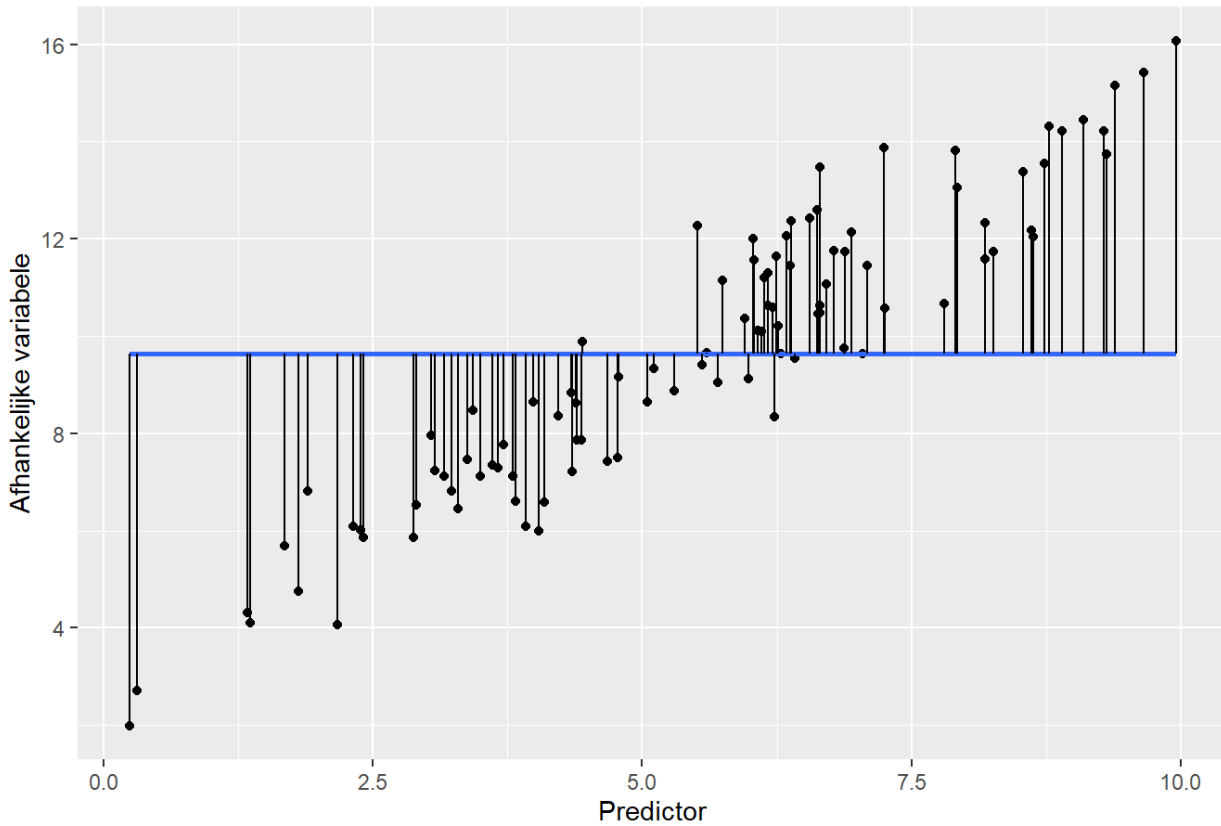
Een praktisch voorbeeld is een regressiemodel waarbij het gemiddelde cijfer in jaar 1 van een opleiding voorspeld wordt op basis van het geslacht en het gemiddelde eindexamencijfer op de middelbare school

$$\text{Gemiddeld Cijfer}_i = \beta_0 + \beta_1 \text{Geslacht}_i + \beta_2 \text{Eindexamencijfer}_i + \varepsilon_i$$

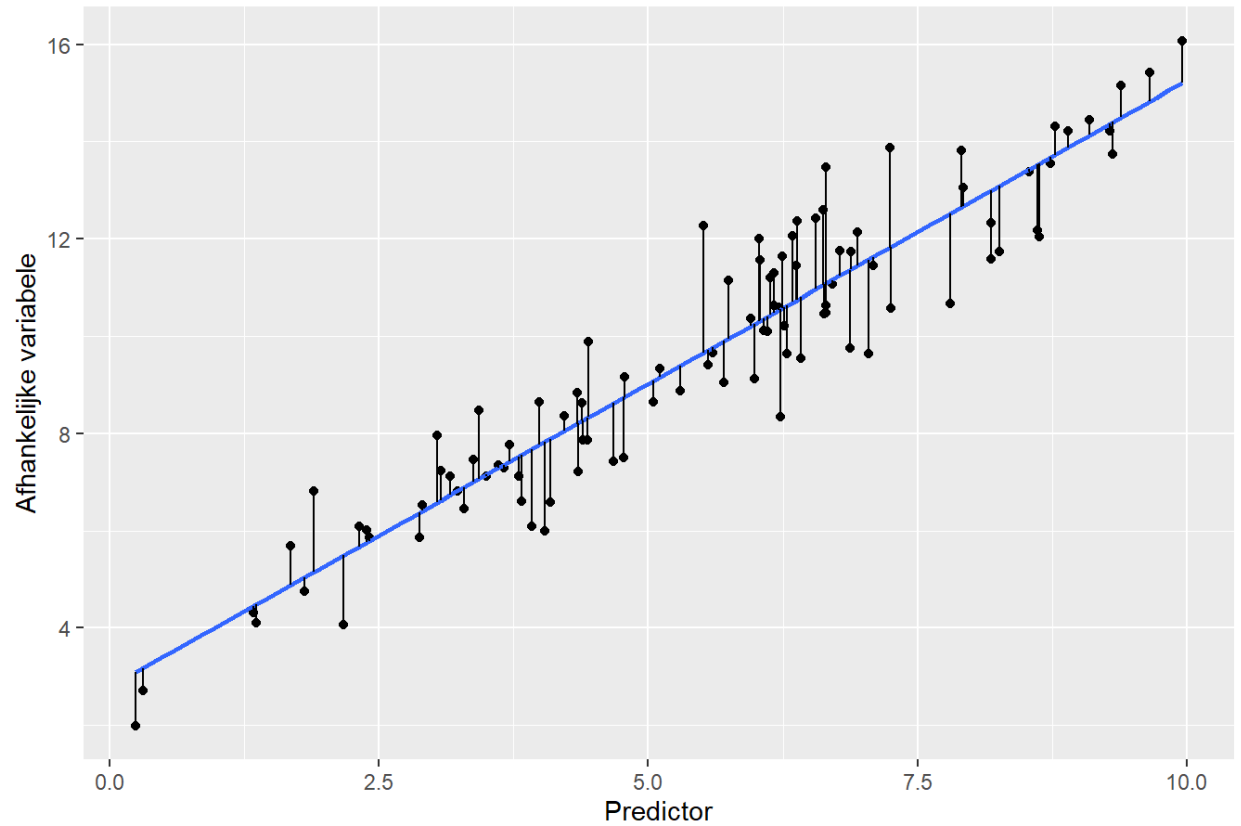


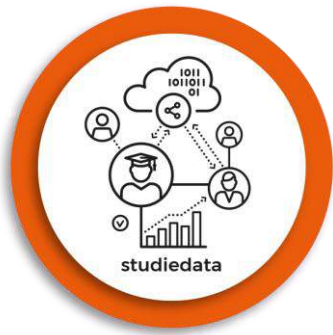
Regressiemodellen – Visueel

Interceptmodel



Regressiemodel





Regressiemodellen - Definitie

1

In een regressiemodel wordt een afhankelijke variabele Gemiddeld Cijfer voorspeld met de predictors Geslacht en Eindexamencijfer

$$\text{Gemiddeld Cijfer}_i = \beta_0 + \beta_1 \text{Geslacht}_i + \beta_2 \text{Eindexamencijfer}_i + \varepsilon_i$$

- $\text{Gemiddeld Cijfer}_i$: de afhankelijke variabele Gemiddeld Cijfer voor deelnemer i
- β_0 : de intercept
- β_1 : de regressiecoëfficiënt van predictor Geslacht _{i}
- Geslacht_i : de predictor Geslacht voor deelnemer i , man of vrouw
- ε_i : de residu van deelnemer i , verschil tussen de voorspelling van het regressiemodel en de observatie



Regressiemodellen - Soorten

Het type regressie-analyse dat gebruikt wordt hangt af van het meetniveau van de afhankelijke variabele:

- Continu: multipele lineaire regressie
- Binair: logistische regressie
- Nominaal: multinomiale logistische regressie
- Ordinaal: ordinale logistische regressie

| | | Onafhankelijke variabelen | | | | |
|------------------------|------------------------------|---|---------------------------------------|--|-------------------------------------|-------------------------------------|
| | | cross-sectioneel | | longitudinaal | | hiërarchisch |
| | | categorische variabelen | continue en categorische variabelen | categorische variabelen | continue en categorische variabelen | continue en categorische variabelen |
| Afhankelijke variabele | continu (interval en ratio) | Multipele lineaire regressie / Factoriele ANOVA | Multipele lineaire regressie / ANCOVA | Factoriele repeated measures ANOVA / Mixed ANOVA | Multilevel lineaire regressie | Multilevel lineaire regressie |
| | gecensureerd continu | Survival analyse | Survival analyse | | | |
| | binair | Logistische regressie / Loglineaire analyse | Logistische regressie | Multilevel logistische regressie | Multilevel logistische regressie | Multilevel logistische regressie |
| | nominaal | Multinomiale regressie / Loglineaire analyse | Multinomiale regressie | Multilevel multinomiale regressie | Multilevel multinomiale regressie | Multilevel multinomiale regressie |
| | ordinaal | Ordinale regressie | Ordinale regressie | Multilevel ordinale regressie | Multilevel ordinale regressie | Multilevel ordinale regressie |
| | Meerdere continue variabelen | MANOVA | MANCOVA | | | |

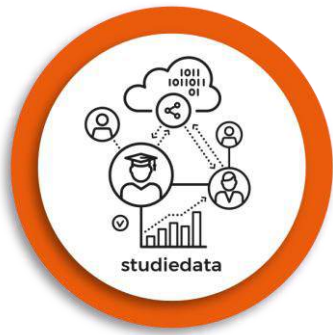


Regressiemodellen - Nut

1

Regressiemodellen zijn te gebruiken voor meerdere toepassingen

- Het toetsen van relaties tussen variabelen waarbij gecontroleerd wordt voor andere variabelen
- Het kwantificeren van het effect van de ene variabele op de andere variabele
- Het zo goed mogelijk voorspellen van een afhankelijke variabele



Regressiemodellen - Nut

1

Toetsen van relaties tussen variabelen waarbij gecontroleerd wordt voor andere variabelen

Stel dat je wilt onderzoeken of er verschillen zijn tussen mannen en vrouwen wat betreft de cijfers van een bepaalde opleiding

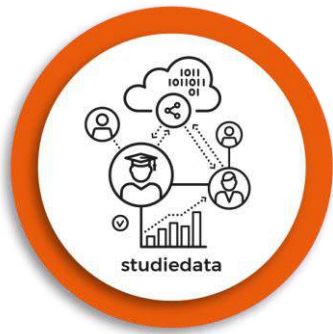
$$\text{Gemiddeld Cijfer}_i = \beta_0 + \beta_1 \text{Geslacht}_i + \varepsilon_i$$

Deze opleiding wordt gekozen door vrouwen met een relatief hoog gemiddeld eindexamencijfer en mannen met een relatief laag gemiddeld eindexamencijfer

In dat geval zijn zowel de predictor Geslacht als de predictor Gemiddeld Eindexamencijfer relevante variabelen

$$\text{Gemiddeld Cijfer}_i = \beta_0 + \beta_1 \text{Geslacht}_i + \beta_2 \text{Eindexamencijfer}_i + \varepsilon_i$$

Met dit model kan er voor beide predictors getoetst worden of ze een significante voorspeller zijn van het gemiddelde cijfer



Regressiemodellen - Nut

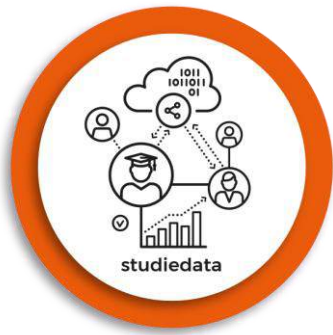
1

Het kwantificeren van het effect van de ene variabele op de andere variabele

De coëfficiënten van de regressievergelijking geven de sterkte van de relatie tussen de predictor en afhankelijke variabele weer

$$\textit{Gemiddeld Cijfer}_i = 2,14 + 0,20 \textit{ Geslacht}_i + 0,72 \textit{ Eindexamencijfer}_i + \varepsilon_i$$

- Het verschil tussen mannen en vrouwen qua gemiddeld cijfer is 0,20 punten
- Bij een toename van het gemiddelde eindexamencijfer met één punt, neemt het gemiddelde cijfer toe met 0,72 punten



Regressiemodellen - Nut

1

Het zo goed mogelijk voorspellen van een afhankelijke variabele

Met behulp van de regressievergelijking kan de afhankelijke variabele voorspeld worden met behulp van de predictors

$$\widehat{\text{Gemiddeld Cijfer}}_i = 2,14 + 0,20 \text{ Geslacht}_i + 0,72 \text{ Eindexamencijfer}_i$$

Wat is het voorspelde gemiddelde cijfer in studiejaar 1 voor een man met een gemiddeld eindexamencijfer van 7,8?

$$\widehat{\text{Gemiddeld Cijfer}}_i = 2,14 + 0,20 * 1 + 0,72 * 7,8 = 7,956$$

Het voorspellen van variabelen kan nuttig zijn, bijvoorbeeld om prognoses te maken voor beleidsdoeleinden. De Vrije Universiteit heeft in het verleden een prognosemodel gemaakt om uitval van studenten te voorspellen



Vragen



Opzet webinar

2

1 Regressiemodellen

2 **Casus uit de praktijk**

3 Het Statistisch Handboek in de praktijk gebruiken



Casus uit de praktijk – Casus

2

“Stel dat het bij een bacheloropleiding mogelijk is om in te stromen met een vwo diploma en met een propedeuse in het hbo, de zogenaamde hbo-p instroom. Dan zou de opleiding zich kunnen afvragen of er een verschil is tussen het studiesucces van beide groepen. Naar aanleiding van deze vergelijking zouden bijvoorbeeld aanpassingen gedaan kunnen worden in de begeleiding van studenten als dat nodig is. Hoe zou je dit kunnen aanpakken met een statistische analyse met behulp van het Statistisch Handboek Studiedata?”



Casus uit de praktijk - Nut

2

Hoe is studiesucces te definiëren? Idee:

- Gemiddeld cijfer
- Aantal studiepunten
- Uitval (wel of geen uitval)

Voorbeeld: data van eerste studiejaar

Wat zijn de juiste statistische toetsen voor deze drie uitkomstvariabelen?

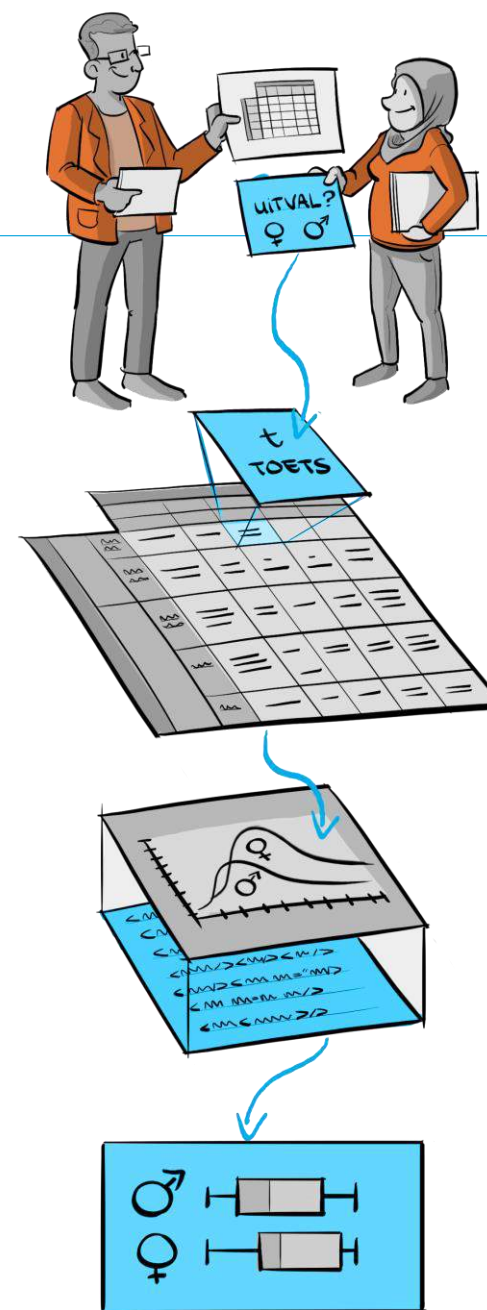


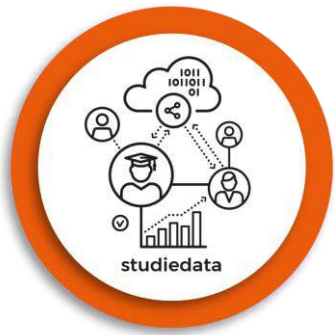
Mentimeter

Voor het volgende onderdeel maken we gebruik van Mentimeter

Ga naar www.menti.com

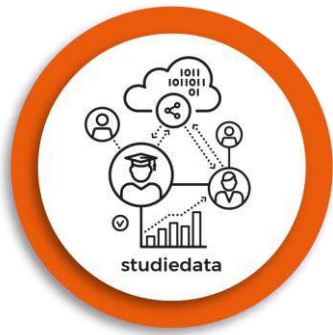
Vul in **93 01 67 9**





Casus uit de praktijk – Gemiddeld cijfer

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|--|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoets I | Wilcoxon signed rank toets I / Tekentoets II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |



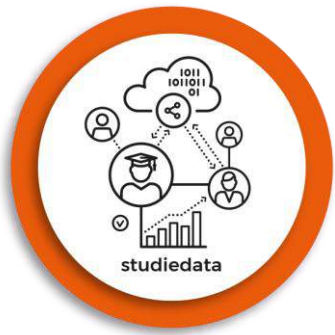
Casus uit de praktijk – Gemiddeld cijfer

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|---|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoes I | Wilcoxon signed rank toets I / Tekentoes II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |



Casus uit de praktijk – Aantal studiepunten

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|--|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoets I | Wilcoxon signed rank toets I / Tekentoets II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |



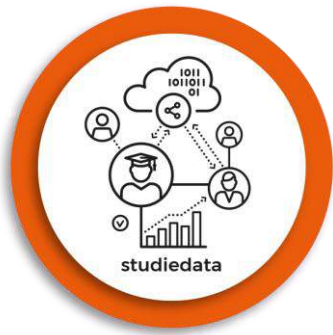
Casus uit de praktijk – Aantal studiepunten

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|---|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoes I | Wilcoxon signed rank toets I / Tekentoes II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |



Casus uit de praktijk – Uitval

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|--|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoets I | Wilcoxon signed rank toets I / Tekentoets II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |
| | | | | | | | |



Casus uit de praktijk – Uitval

| | | | Onafhankelijke variabele | | | | |
|------------------------|-----------------------------|-----------------------|--|---|--|---|--|
| | | | 1 groep | 2 groepen | | >2 groepen | |
| | | | met referentie | gepaard | ongepaard | gepaard | ongepaard |
| Afhankelijke variabele | continu (interval en ratio) | normaal verdeeld | One sample t-toets | Gepaarde t-toets | Ongepaarde t-toets | Repeated measures ANOVA | One-way ANOVA |
| | | niet normaal verdeeld | Tekentoeets I | Wilcoxon signed rank toets I / Tekentoeets II | Mann-Whitney U toets I / Mood's mediaan toets | Friedman's ANOVA I | Kruskal Wallis toets I |
| | categorisch | binair (2 waarden) | Chi-kwadraat toets voor goodness of fit en binomiaaltoets | McNemar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher's exact toets | Cochran's Q toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | nominaal (>2 waarden) | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Bhapkar toets | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I | Multilevel multinomiale logistische regressie | Chi-kwadraat toets voor onafhankelijkheid en Fisher-Freeman-Halton exact toets I |
| | | ordinaal | Chi-kwadraat toets voor goodness of fit en multinomiaaltoets | Wilcoxon signed rank toets II | Mann-Whitney U toets II | Friedman's ANOVA II | Kruskal Wallis toets II |



Vragen



Opzet webinar

3

Regressiemodellen

Casus uit de praktijk

Het Statistisch Handboek in de praktijk gebruiken

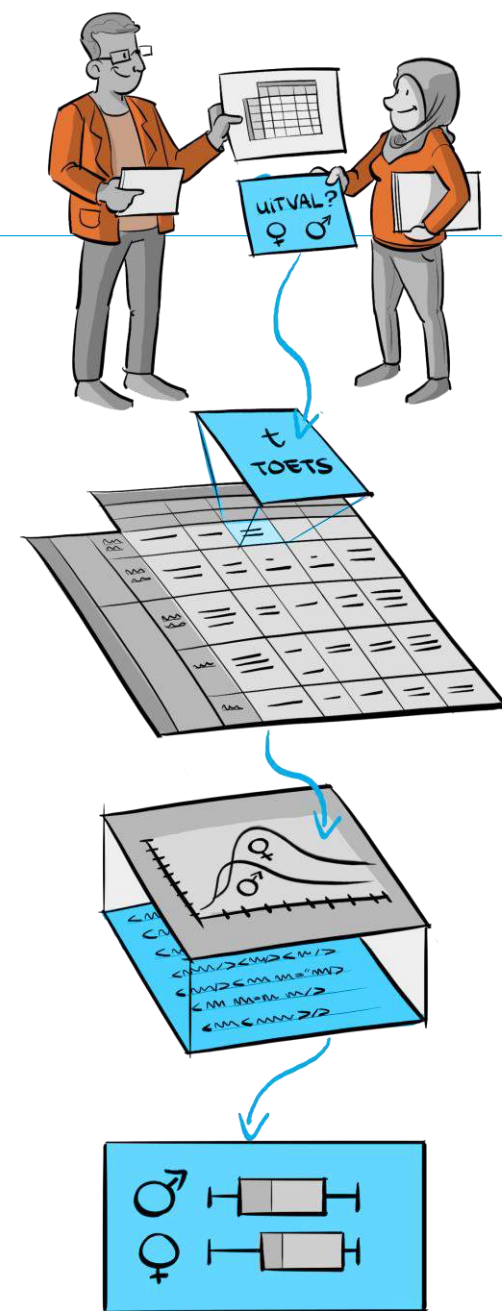


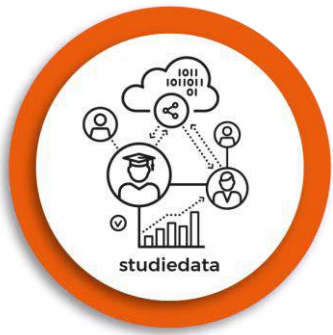
Mentimeter

Voor het volgende onderdeel maken we gebruik van Mentimeter

Ga naar www.menti.com

Vul in **93 01 67 9**





Dank voor jullie aanwezigheid!

Voor verdere vragen en/of opmerkingen omtrent het Statistisch Handboek Studiedata, mail gerust naar e2.griffioen@vu.nl

Het Statistisch Handboek Studiedata is te vinden op www.sh-studiedata.nl

Meer informatie over de zone Studiedata is te vinden op <https://versnellingsplan.nl/zones/studiedata/>



Bijlage 1. Versiegeschiedenis

V0.1 – Eerste versie

V0.3 – Versie na feedback TB en toevoeging slides Iris