SAMENVATTING · OP TENTAMENNIVEAU

Module 3

Hypothesetoetsen

Alles wat je voor de open vragen moet kunnen definiëren, uitleggen, toepassen en beoordelen. Met de modellen als figuur, de vergelijkingen als tabel en de valkuilen gemarkeerd.

WAT ERIN ZIT

01	Hypothesen en het stappenplan	2
02	Beslisregels	2
03	Type I error, Type II error en power	3
04	De vier toetsen	3
05	R-uitvoer interpreteren	4
06	Likert scales en Cronbach's alpha	4
07	Veelgemaakte fouten	5
08	Op het tentamen	5
→	Zo schrijf je het antwoord	5
→	Checklist: kan ik dit uitleggen?	6
→	Overhoor jezelf	6

TENTAMENKANS

p-waarde (overschrijdingskans) ZEKER	97%
t-toets (één steekproef, twee steekproeven, gepaard) ZEKER	95%
Variantieanalyse (ANOVA) en F-toets ZEKER	93%
Chikwadraattoets: aanpassing en onafhankelijkheid ZEKER	92%
Type I- en Type II-fout ZEKER	92%
Likert-item versus Likertschaal (composite score) ZEKER	90%
H0 verwerpen of handhaven (nooit H0 bewezen) ZEKER	90%

Inschatting per onderwerp, op basis van het voorbeeldtentamen en de nadruk in de stof.

KERNBEGRIPPEN

H0 · H1 · alpha · p-value · critical value · rejection region · Type I error · Type II error · power · one-sided · two-sided · t-test · paired t-test · ANOVA

OP HET TENTAMEN

Blok 3

15 punten · 180 min totaal

MC	6 meerkeuzevragen	2.5 punten per vraag
7	Blok 7 (25 punten) bestaat uit drie open vragen over R-output.	200 / 250 / 150 woorden

01 Hypothesen en het stappenplan



FIGUUR 3.1 **Stappenplan hypothesetoets.** Stel H0 en H1 vooraf op, daarna pas de data.

Een **hypothese** is een testbare stelling over een populatie. **H1** (alternatieve hypothese, alternative hypothesis) is de verwachting van de onderzoeker; **H0** (nulhypothese, null hypothesis) is het tegendeel en wordt als startpunt voor waar aangenomen. Je verwerpt H0 alleen bij sterk bewijs; dat beschermt tegen confirmation bias. Formuleer altijd eerst H1 en leid daar H0 uit af. Hypothesen gaan over **populatieparameters** (mu, p), nooit over steekproefgemiddelden.

De richting van H1 bepaalt de toets: **right-tail** (H1: groter dan), **left-tail** (H1: kleiner dan) of **two-sided** (H1: ongelijk, rejection region in beide staarten). Bij een two-sided toets verdeel je alpha over twee staarten (alpha/2 per staart), dus de critical value ligt verder van het midden dan bij een one-sided toets.

Het zevenstappenplan (Neyman-Pearson): (1) H0 en H1 opstellen, (2) de test statistic kiezen, (3) de verdeling onder H0 afleiden, (4) de rejection region intuitief bepalen, (5) alpha kiezen, (6) de critical value(s) vinden, (7) de toets uitvoeren en concluderen in de taal van het probleem (zie dia 11 en 31). Het idee: vergelijk de steekproefuitkomst met de waarde onder H0, gedeeld door de standard error.

02 Beslisregels

- **Critical value**: verwerp H0 als de teststatistiek in de rejection region valt (bij two-sided: |t| groter dan de critical value).
- **p-value**: de kans op een uitkomst zo extreem of extremer dan de waargenomen uitkomst, aangenomen dat H0 waar is. Verwerp H0 als **p <= alpha**. Hoe lager p, hoe sterker het bewijs tegen H0.
- **Confidence interval**: bij een two-sided toets met alpha = 0,05 verwerp je H0 als de waarde onder H0 buiten het 95%-interval ligt.

Conclusiezinnen: "verwerp H0" of "handhaaf H0" (nooit "H0 is bewezen" of "H1 is bewezen"). Een p-value is niet de kans dat H0 waar is, en niet de kans dat je een fout maakt.

03 Type I error, Type II error en power



FIGUUR 3.2 **Type I- en Type II-fouten.** Alpha is de kans op een Type I-fout; power is 1 min de kans op een Type II-fout.

	H0 waar	H0 onwaar
Verwerp H0	Type I error (kans alpha)	correct (power = 1 – beta)
Handhaaf H0	correct	Type II error (kans beta)

Een Type I error is een false positive (een effect zien dat er niet is), een Type II error een false negative (een bestaand effect missen). De **trade-off**: een lagere alpha verkleint Type I maar vergroot Type II. Beide kansen dalen alleen met een **grotere steekproef**. Een gebruikelijke eis voor power is 80%. Meerdere toetsen achter elkaar laten de totale kans op minstens een Type I error oplopen; dat is een reden om ANOVA te gebruiken in plaats van meerdere t-tests.

04 De vier toetsen

t-test: toetst gemiddelden met onbekende populatiestandaardafwijking; de teststatistiek is (gemiddelde min waarde onder H0) gedeeld door de standard error, met t-verdeling en df = n – 1 (one-sample). Varianten:

- one-sample: gemiddelde tegen een vaste waarde;
- two-sample: twee onafhankelijke groepen; **var.equal = TRUE** neemt gelijke varianties aan, anders gebruikt R de **Welch**-toets (gebroken df, bijvoorbeeld 628,47);
- paired: gekoppelde observaties (voor/na, dezelfde dag, dezelfde persoon); je toetst het gemiddelde van de verschillen.

ANOVA: H0 is dat alle groepsgemiddelden gelijk zijn; H1 dat minstens een groep afwijkt. De F-statistiek vergelijkt variatie tussen de groepen met variatie binnen de groepen. Een significante F zegt niet welke groepen verschillen; daarvoor zijn post-

hoc toetsen nodig (zoals Tukey, die per paar een gecorrigeerde p-value geeft). In de uitvoer van **summary(aov(...))** lees je Df, Sum Sq, Mean Sq, F value en **Pr(>F)**.

z-test: bekende populatiestandaardafwijking, of een hypothese over een proportie. De teststatistiek is z = (p-dak – p0) / SE, met SE = wortel(p0 (1 – p0) / n) onder H0.

chi-square test: voor twee categorische variabelen. H0: de variabelen zijn onafhankelijk. De verwachte frequentie (expected frequency) per cel is (rijtotaal x kolomtotaal) / totaal; de statistiek is de som van (waargenomen – verwacht)^2 / verwacht, met df = (rijen – 1) x (kolommen – 1). Ook bruikbaar als goodness-of-fit-toets tegen een verwachte verdeling. Een lage X-squared (en hoge p) betekent: geen bewijs voor samenhang.

Situatie	Toets	R
een gemiddelde tegen een waarde	one-sample t	t.test(x, mu = ...)
twee groepen	two-sample t	t.test(y ~ g)
gekoppelde paren	paired t	t.test(a, b, paired = TRUE)
drie of meer groepen	ANOVA	aov(y ~ g)
twee categorische variabelen	chi-square	chisq.test(tabel)

05 R-uitvoer interpreteren

Zoek in de uitvoer: de teststatistiek (t, F, X-squared), de df, de p-value, het interval en de schatting (sample estimates). Let op de **alternative hypothesis** die R erbij zet ("true mean is not equal to ..."): dat is H1. Controleer eerst of de toets bij de vraag past, dan of p <= alpha, dan de richting van het effect in de gemiddelden. Bij de chi-square test: kijk of het verband praktisch relevant is, niet alleen of het significant is.

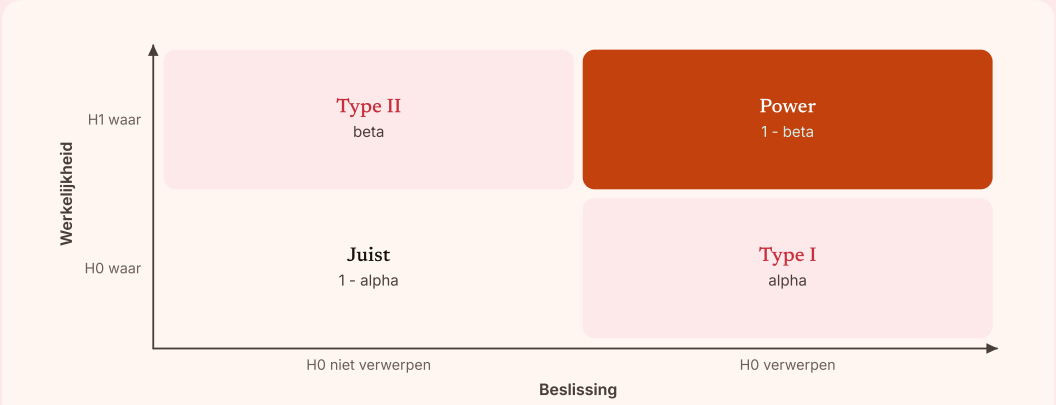
06 Likert scales en Cronbach's alpha

Een **Likert scale** meet een abstract construct met meerdere items, een rating scale (bijvoorbeeld 1 tot 5) en een instructie. Een **reverse coded** item is negatief geformuleerd; je klapt het om (nieuwe score = max + min – score, dus 6 – score bij 1 tot 5) voordat je de schaalscore berekent. Zulke items houden respondenten alert. Andere kwaliteitsmiddelen: **attention checks** en het verwijderen van **flatliners** en te snelle respondenten.

Cronbach's alpha meet interne consistentie (internal consistency), gewoonlijk tussen 0 en 1; 0,7 is de gebruikelijke ondergrens. Alpha stijgt met meer items en met hogere onderlinge correlatie. Een negatieve alpha wijst op een vergeten omkering. Voorbeeld uit de cursusdata: alpha was negatief tot de drie omgekeerde items waren omgeklapt, daarna ongeveer 0,80 voor acht items. De kolom "alpha if item dropped" laat zien welk item de schaal verzwakt.

LET OP

07 Veelgemaakte fouten



FIGUUR 3.3 **Type I- en Type II-fouten.** Alpha is de kans op een Type I-fout; power is 1 min de kans op een Type II-fout.

- ✖ p-value lezen als de kans dat H0 waar is.
- ✖ "Niet significant" lezen als "geen effect"; het kan een gebrek aan power zijn.
- ✖ Een two-sample t-test gebruiken op gekoppelde data, of meerdere t-tests voor drie groepen.
- ✖ Vergeten dat bij een one-sided toets de p-value halveert ten opzichte van two-sided (als het effect in de goede richting ligt).
- ✖ Significantie verwarren met praktisch belang bij grote steekproeven.

08 Op het tentamen

Ongeveer 6 meerkeuzevragen over module 3 (de vragen komen uit de proeftentamens). Het formuleblad geeft de formules voor het confidence interval, de t- en z-statistiek, chi-square en Cronbach's alpha; de F-formule en de p-value-berekening staan er niet op. Je moet een conclusie kunnen trekken uit R-uitvoer van **t.test**, **chisq.test** of **aov**, maar je schrijft geen code.

VOORBEELDANTWOORD

→ Zo schrijf je het antwoord

Vier bouwstenen, in deze volgorde. Elke steen levert punten op; laat er geen weg.

1. Term
Noem het model of begrip met de Engelse vakterm.

2. Definitie
Geef de definitie of de kenmerken, precies.

3. Casusfeit
Koppel aan een concreet feit uit de casus.

4. Conclusie
Sluit af met een oordeel, advies of gevolg.

OPDRACHT 1 · WOORDLIMIET 75 WOORDEN (WOORDLIMIET AANGENOMEN)

Een toets geeft p = 0,03 bij alpha = 0,05. Wat is je conclusie?

TERM De p-value is kleiner dan alpha. DEFINITIE De p-value is de kans op een uitkomst minstens zo extreem als de gevonden uitkomst, als H0 waar is. CASUSFEIT Hier is 0,03 < 0,05, dus de data zijn onwaarschijnlijk onder H0. CONCLUSIE Verwerp H0. Dit is niet de kans dat H0 waar is, en zegt niets over de grootte van het effect.

57 / 75 woorden

KERNBEGRIPPEN

✓ Checklist: kan ik dit uitleggen?

Vink af wat je uit je hoofd in één zin kunt uitleggen, met de Engelse term.

- ☐ H0
- ☐ H1
- ☐ alpha
- ☐ p-value
- ☐ critical value
- ☐ rejection region
- ☐ Type I error
- ☐ Type II error
- ☐ power
- ☐ one-sided
- ☐ two-sided
- ☐ t-test
- ☐ paired t-test
- ☐ ANOVA
- ☐ z-test
- ☐ chi-square test
- ☐ Likert scale
- ☐ reverse coded
- ☐ Cronbach's alpha

FLASHCARDS

? Overhoor jezelf

Dek de rechterkolom af en zeg het antwoord hardop.

Vraag	Antwoord
1 Wat is een hypothese in de context van dit vak?	Een testbare, voorlopige stelling over een verwachte relatie tussen variabelen of over een eigenschap van een populatie, zoals het gemiddelde.
2 Met welke hypothese begin je het best bij het opstellen van een hypothesepaar?	Met H1, de verwachting die je wilt aantonen; H0 volgt daar als het precieze tegendeel uit.
3 Wat is de critical value?	De grens van de rejection region; hij volgt uit alpha en de verdeling van de teststatistiek onder H0.
4 Wat is de p-value?	De kans op een testresultaat zo extreem of extremer als gevonden, onder de aanname dat H0 klopt.

Vraag	Antwoord
5 Wat is een Type I error?	H0 verwerpen terwijl H0 waar is (vals alarm); de kans daarop is alpha.
6 Wat zegt een alpha van 0,05 over de kans op een Type I error?	Op de lange termijn verwerp je H0 in 5% van de gevallen ten onrechte, als H0 waar is.
7 Wat doet het argument var.equal in t.test()?	Het zet de aanname aan of uit dat de variantie in beide groepen gelijk is; dat bepaalt hoe de standard error wordt berekend.
8 Wat toetst een ANOVA?	Of de gemiddelden van drie of meer groepen gelijk zijn; H0 is dat alle groepsgemiddelden gelijk zijn.
9 Wanneer gebruik je een z-test?	Bij een bekende populatiestandaardafwijking of bij een hypothese over een proportie.
10 Wat is H0 bij een chi-square test op onafhankelijkheid?	De twee categorische variabelen zijn onafhankelijk; er is geen samenhang tussen.
11 Wat zijn reverse coded items?	Negatief geformuleerde items die je moet omklappen (bij een 5-puntsschaal 6 min de score) voor je een totaalscore neemt.
12 Wat laat alpha if item dropped zien?	De alpha die je krijgt als je dat ene item weglaat; een hogere waarde dan de totale alpha wijst op een zwak item.