

Tentamen Onderzoeksvaardigheden 1

Onderdeel: Statistiek 1

25-1-2007, 9:00 - 12:00

NW & I

Docent: A.R.T. Donders

Het onderdeel Statistiek 1 heeft een openboek tentamen. Het tentamen bestaat uit 19 meerkeuzevragen en een case met 6 open vragen. Voor alle vragen kun je (maximaal) 4 punten halen. In totaal kun je 100 punten krijgen. Het aantal punten dat je verdient hebt wordt door 10 gedeeld om je eindcijfer te geven.

Schrijf op ieder vel dat je inlevert je naam en je studentnummer.

Lees de vraag zorgvuldig door!

Laat bij de open vragen goed zien hoe je aan je antwoorden komt en geef waar gevraagd een korte maar duidelijke argumentatie.

Bij de vragen staan regelmatig allerlei berekende gegevens. Daarbij is de notatie van het boek aangehouden. Maak zoveel mogelijk gebruik van deze gegevens, maar denk eraan dat er soms ook gegevens bij staan die je niet kunt gebruiken voor de beantwoording van de vraag!

Sommige vragen zijn makkelijker/moeilijker dan andere opgaven. Raak niet teveel tijd kwijt aan een moeilijke opgave! Mocht je niet zo snel uit een opgave komen, ga dan door met de volgende opgave en ga wanneer je alle opgaven gemaakt hebt, weer verder met deze moeilijke opgave.

Banken doen hun best om met allerlei nieuwe betalingsmethoden het leven voor de consument aangenamer te maken, terwijl ze zelf meer winst hopen te maken. Deze ontwikkeling zou je laten kunnen beginnen met de introductie van het pinnen. In dit tentamen draait het om de nieuwe (en oude) betalingsmethoden en wat consumenten daar van vinden.

1. Hieronder staan de bedragen (afgerond op hele euro's) van 6 willekeurig gekozen pintransacties. Wat is de mediaan van deze observaties?

7, 15, 22, 24, 27, 1705

- a. 23
- b. 300
- c. van 22 tot 24
- d. de mediaan is in dit geval niet gedefinieerd

2. Welke van onderstaande uitspraken is juist?

- a. Het gemiddelde kan alleen gebruikt worden indien een variabele bij benadering normaal verdeeld is.
- b. De mediaan geeft een beter beeld van de locatie van een variabele wanneer er sprake is van uitschieters.
- c. Een variabele kan nooit meer dan één modus hebben.
- d. Het gemiddelde kan alleen bij continue variabelen berekend worden.

3. Stel dat bekend is dat de kans dat een persoon pint indien hij of zij een uitgave boven of gelijk aan 20 euro doet gelijk is aan 0.5 en dat de kans dat een persoon pint als hij of zij een uitgave onder de 20 euro doet gelijk is aan 0.3. Verder is bekend dat 60% van alle transacties gaat om bedragen onder 20 euro. Wanneer je een persoon bij kassa ziet pinnen, hoe groot (ongeveer) is dan de kans dat het gaat om een uitgave boven of gelijk aan 20 euro?

- a. 0.40
- b. 0.20
- c. 0.47
- d. 0.53

4. Stel dat de uitgaven die met behulp van pin betaald worden goed beschreven kan worden met een normale verdeling met als gemiddelde (μ) 70 euro en als

standaarddeviatie (σ) 20 euro (NB dit is niet echt realistisch; waarschijnlijk zal deze variabele scheefverdeeld zijn met een grote groep lage bedragen (de dagelijkse boodschappen) en een relatief dikke staart (de dure aanschaffen zoals kleding en TV's etc); het is waarschijnlijker dat het logaritme van de uitgaven normaal verdeeld is, maar dat maakt zowel de omschrijving als de vragen die gesteld kunnen worden alleen maar lastiger; vandaar deze aanname). Wat is de kans dat een willekeurig gekozen pinbetaling groter is dan 80 euro?

- a. 0.500
- b. 0.000
- c. 0.309
- d. 0.050

5. Uit alle pintransacties (zie vraag 4: $\mu = 70$ euro en $\sigma = 20$ euro) wordt een steekproef (willekeurig en onafhankelijk) van 16 transacties. Wat is de kans dat het gemiddelde bedrag van deze 16 transacties kleiner is dan 66 euro?

- a. 0.421
- b. 0.800
- c. 0.788
- d. 0.212

6. Bij vragen 4 en 5 werd het populatiegemiddelde van de uitgaven aangeven met μ . Wat is de kansverdeling van μ ?

- a. de normale verdeling
- b. de (Student) t-verdeling
- c. de binomiale verdeling
- d. μ heeft geen verdeling

7. Het ontwikkelen van nieuwe betalingsmethodes kost geld en aangezien banken geen filantropische instellingen zijn, zouden ze dat geld graag terugverdienen. Eén nieuwe ontwikkeling is betalen met je mobiele telefoon. Er zijn al wel beperkte mogelijkheden voor betalen met je mobiele telefoon (bv voor parkeergelden en denk ook aan belspelletjes en SMS-diensten), maar het gaat hierbij dan om een brede toepassing waarbij je bijvoorbeeld in een restaurant of bij een bioscoop of in een winkel je rekening kan voldoen met je mobieltje. Er zijn al een aantal proefprojecten geweest en volgens de producenten van de technologie (o.a. VISA, Philips en Sony) waren de consumenten laaiend enthousiast. Maar zou men ook bereid zijn om te betalen voor deze dienstverlening? Om dat te onderzoeken legde een onderzoekster deze vraag ("Hoeveel

ben je bereid om per maand te betalen voor de mogelijkheid om te kunnen betalen met je mobieltje?") voor aan een willekeurig gekozen steekproef. Daarnaast vroeg ze hoeveel de deelnemers kwijt waren aan hun huidige betalingswijzen (bv. de kosten voor creditcards, pin- en/of chippassen). Het verschil tussen deze twee kosten noemde ze de meerprijs die men wilde betalen voor het mobiele betalingsgemak. Vervolgens onderzocht ze of er een verband was tussen de gemeten meerprijs en het inkomen van de persoon. Wat is de meest geëigende toets in deze situatie?

- a. de gepaarde t-toets
- b. een toets op de correlatiecoëfficiënt
- c. een t-toets voor twee onafhankelijke steekproeven
- d. een (non-parametrische) Wilcoxon toets

8. Om te onderzoeken of er een verschil zit in geslacht vroeg ze aan 25 getrouwde (hetero)stellen welke meerprijs ze bereid waren te betalen. De partners moesten onafhankelijk van elkaar antwoorden en waren dus niet op de hoogte van het antwoord van de ander. Er was een grote spreiding in gezinsinkomen en betalingsgedrag tussen de stellen. De gemiddelde meerprijs die de mannen bereid waren om te betalen bedroeg 4 euro met een variantie van 2 en voor de vrouwen bedroeg het gemiddelde 3.20 euro met een variantie van 2. De gepoolde variantie was dus 2. Ze berekende ook nog de verschillscores (man – vrouw). Het gemiddelde hiervan bedroeg 0.80 euro met een variantie van 2.56. Wat is de uitkomst van een toets om te kijken of er verschil in meerprijs tussen mannen en vrouwen zit?

- a. De toetsingsgrootte t is gelijk aan 2 met 48 vrijheidsgraden. Er is geen significant resultaat en we hebben dan ook geen verschil kunnen aantonen.
- b. De toetsingsgrootte t is gelijk aan 2.5 met 48 vrijheidsgraden. Er is een significant resultaat en we hebben dan ook verschil kunnen aantonen.
- c. De toetsingsgrootte t is gelijk aan 2 met 24 vrijheidsgraden. Er is geen significant resultaat en we hebben dan ook geen verschil kunnen aantonen.
- d. De toetsingsgrootte t is gelijk aan 2.5 met 25 vrijheidsgraden. Er is een significant resultaat en we hebben dan ook verschil kunnen aantonen.

9. Wat is een p-waarde (overschrijdingskans)?

- a. De kans dat de nulhypothese juist is.
- b. De kans dat de alternatieve hypothese juist is.
- c. De kans op het gevonden resultaat of een nog extremer resultaat gegeven dat de nulhypothese juist is.
- d. De kans op het gevonden resultaat of een nog extremer resultaat gegeven dat de nulhypothese juist is.

10. Tevens is gevraagd waar men de grootste behoefte had aan betalen met de mobiele telefoon. De antwoordmogelijkheden waren: “tijdens het uitgaan (bioscoop, café, restaurant)”, “in de supermarkt”, “onderweg (in het Openbaar Vervoer, bij tankstations, op parkeerplaatsen)”. Om te onderzoeken of er een samenhang is tussen het feit dat iemand een relatie heeft en de locatie waar de grootste behoefte bestaat tot het betalen met de mobiele telefoon, moet men gebruik maken van:

- a. de t-toets voor gepaarde waarnemingen
- b. de chi-kwadraat (χ^2) toets voor kruistabellen (onafhankelijkheid).
- c. de toets voor twee proporties.
- d. de t-toets voor onafhankelijke steekproeven.

11. Bij een kleine pilot kregen 13 personen een palmtop mee. Telkens wanneer ze ergens zouden willen betalen met hun mobiele telefoon voerden ze het bedrag in. Zo kon na een maand nagegaan worden hoeveel ieder persoon betaald zou hebben indien overal de mogelijkheid tot betalen met de mobiele telefoon aanwezig zou zijn. Er deden 8 personen uit Utrecht en 5 personen uit Nijmegen aan de pilot mee. De uitgaven (afgerond op hele euro's) waren:

Voor de personen uit Utrecht: 1739, 181, 181, 175, 80, 64, 52, 48

Voor de personen uit Nijmegen: 455, 330, 320, 290, 180

Men wil nagaan in welke stad mensen meer zullen betalen met hun mobiele telefoon om zo een zo gunstig mogelijke stad voor een echt onderzoek te vinden. Welke van volgende uitspraken is juist?

- a. Aangezien de data niet normaal zijn verdeeld, wordt gebruik gemaakt van een Wilcoxon toets. De toetsingsgrootte W is gelijk aan 22. De p -waarde is 0.066. Er is geen verschil aangetoond.
- b. Aangezien de data niet normaal zijn verdeeld, wordt gebruik gemaakt van een Wilcoxon toets. De toetsingsgrootte W is gelijk aan 22. De p -waarde is 0.033. Er is wel een verschil aangetoond; de mediane besteding in Nijmegen is hoger.
- c. De toetsingsgrootte z is 0. De p -waarde is 1. Er is geen verschil aangetoond.
- d. Aangezien de toetsingsgrootte z gelijk is aan 0 wordt er gebruik gemaakt van een Wilcoxon toets. De toetsingsgrootte W is gelijk aan 22. De p -waarde is 0.155. Er is geen verschil aangetoond.

Dit soort moderne betalingsmethoden hebben nog een ‘bijwerking’: het wordt steeds beter mogelijk om na te gaan waar een individu zijn/haar geld aan uitgeeft. Indien het echt mogelijk is om het bestedingsgedrag van individuen in kaart te brengen, ontstaan er voor de overheid talloze nieuwe beleidsinstrumenten. Zo zou men de accijnzen die iemand betaald hiervan kunnen laten afhangen. Iemand de eerste fles wijn iemand in een week koopt, zou een lagere accijns kunnen hebben dan de tweede, derde etc. Dit omdat een matige consumptie van wijn volgens sommige onderzoekers goed voor de gezondheid zou zijn (NB dit is nog geen uitgemaakte zaak!), terwijl een overmatige consumptie zeker slecht is. Maar je zou ook kunnen denken aan de mogelijkheden om dit soort informatie te gebruiken voor terrorismebestrijding. Tot voor kort waren dit soort opties niet alleen technisch onmogelijk maar ook een sociaal taboe, gezien het belang dat toen gehecht werd aan de eigen privacy. Een onderzoeker is echter van mening dat de cultuur omgeslagen is. Hij vermoedt dat er tegenwoordig een grote mate van “naar buiten treden” bestaat. Alhoewel dat niet samenvalt met de geijkte betekenis die binnen de psychologie aan dit begrip gegeven wordt, noemt hij deze trek: extravertie. Mensen zijn volgens hem tegenwoordig dus extravert. Hij baseert zich op het succes van sites als “www.youtube.com”, programma’s als Idols en The X-factor en de hedendaagse hang naar bling-bling, opvallen en jezelf uiten. Hij nam enquêtes af bij 100 personen waarin hij

de extraversie bepaalde (continue schaal) als mede de angst voor terrorisme ('angst' ook een continue schaal) en de bereidwilligheid om de privacy op te geven ('privacy' ook een continue schaal, waarbij geldt: hoe hoger de score hoe minder men bereid is om de privacy op te geven). Daarnaast werden de gebruikelijke standaardgegevens gevraagd

12. Eerst onderzocht hij op er een verschil zit tussen mannen en vrouwen voor wat betreft hun extraversie. De variabele "vrouw" is 0 ("nee") voor mannen en 1 ("ja") voor vrouwen. De uitkomst van de analyse staat hieronder vermeldt:

Group Statistics

	vrouw	N	Mean	Std. Deviation	Std. Error Mean
extraversie	nee	52	43.1086	9.62728	1.33506
	ja	48	39.1664	9.05262	1.30663

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
extraversie	Equal variances assumed	.167	.684	2.105	98	.038	3.94213	1.87272	.22579	7.65847
	Equal variances not assumed			2.110	97.963	.037	3.94213	1.86807	.23499	7.64927

Welke van onderstaande uitspraken is juist?

- Er is geen significant verschil tussen mannen en vrouwen aangezien de p-waarde van 0.684 groter is dan de α van 0.05.
- Er is geen significant verschil tussen mannen en vrouwen aangezien het gemiddelde verschil van 3.94 in het betrouwbaarheidsinterval ligt.
- Er is geen significant verschil tussen mannen en vrouwen aangezien de p-waarde van 0.38 groter is dan de α van 0.05.
- Er is wel een significant verschil tussen mannen en vrouwen aangezien de p-waarde van 0.038 kleiner is dan de α van 0.05.

13. Het 95% betrouwbaarheidsinterval rond de gemiddelde extraversiescore van de vrouwen loopt ongeveer van 36.5 tot 41.8. Welke van onderstaande uitspraken is juist?

- De kans dat het populatiegemiddelde van de vrouwen in dit interval ligt is gelijk aan 95%.
- Indien er heel veel steekproeven met een omvang van 48 uit de populatie vrouwen getrokken wordt dan zullen 95% van de steekproefgemiddeldes tussen deze twee grenzen liggen.
- Geen van beide bovenstaande uitspraken is correct.
- De uitspraken bij **a.** en **b.** zijn beide correct.

14. Om te onderzoeken hoe de relatie tussen de extraversie van een persoon en de hoeveelheid angst voor terrorisme die iemand ervaart, er uitziet, voert men een regressieanalyse uit. Delen van de output staan hieronder:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	51.224	3.522		14.542	.000
	angst	-.258	.088	-.285	-2.943	.004

a. Dependent Variable: extraversie

Welke van onderstaande uitspraken is niet correct?

- De (Pearson productmoment) correlatiecoëfficiënt r tussen angst en extraversie is significant en negatief.
- De multiple correlatiecoëfficiënt R tussen angst en extraversie is significant en positief.
- Als de variabele angst met 1 toeneemt, zal de variabele extraversie 0.258 lager worden.
- Als de variabele extraversie met 1 toeneemt, zal de variabele angst 0.258 lager worden.

15. De formule voor het 95% betrouwbaarheidsinterval voor β (de populatieregressiecoëfficiënt) is gelijk aan $b \pm t_{0,025} \times SE(b)$. Welke van onderstaande uitspraken is niet correct.

- De 'b' in bovenstaande formule is -0.258.
- De ' $t_{0,025}$ ' in bovenstaande formule is -2.943.
- De 'SE(b)' in bovenstaande formule is 0.088
- Het 95% betrouwbaarheidsinterval dat met bovenstaande formule bepaald wordt, zal de waarde '0' niet bevatten.

16. Wat is de interpretatie van de waarde die in de rij '(Constant)' en de kolom 'B' staat?

- Dit is de voorspelde waarde voor extraversie wanneer de variabele angst gelijk is aan het steekproefgemiddelde.
- Dit is de voorspelde waarde voor extraversie wanneer de variabele angst gelijk is aan 0.
- Dit is het gemiddelde van de variabele extraversie.
- Deze waarde is niet te interpreteren.

17. Om te onderzoeken of het verband tussen angst en extraversie afhangt van het geslacht voert men de volgende analyse uit. Eerst wordt een variabele aangemaakt die gelijk is aan het product van de variabele angst en vrouw; deze variabele heet 'vrouw_x_angst'. Dan worden er drie regressiemodellen gefit: het eerste model bevat angst als onafhankelijke variabele, de tweede bevat angst en vrouw en de derde angst, vrouw en vrouw_x_angst. Delen van de output staan hieronder vermeldt:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	51.224	3.522		14.542	.000
	angst	-.258	.088	-.285	-2.943	.004
2	(Constant)	52.455	4.757		11.026	.000
	angst	-.305	.150	-.337	-2.039	.044
	vrouw	1.210	3.127	.064	.387	.700
3	(Constant)	53.431	5.945		8.987	.000
	angst	-.337	.189	-.372	-1.778	.079
	vrouw	-2.334	13.219	-.123	-.177	.860
	vrouw_x_angst	.086	.311	.218	.276	.783

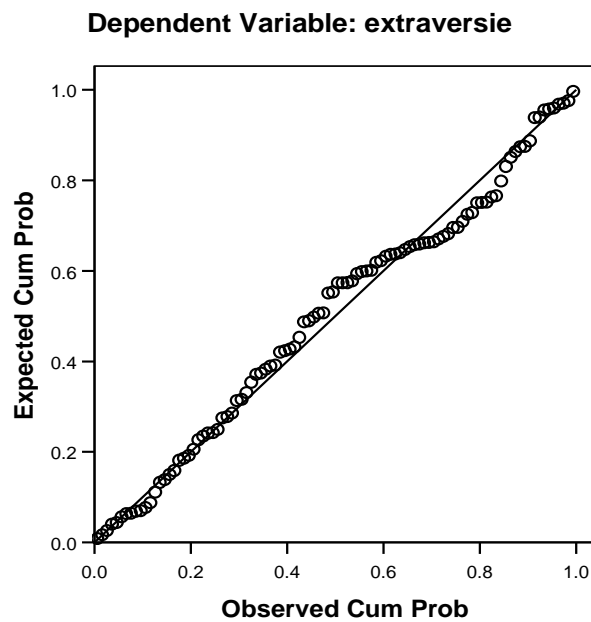
a. Dependent Variable: extraversie

Welke van onderstaande uitspraken is juist?

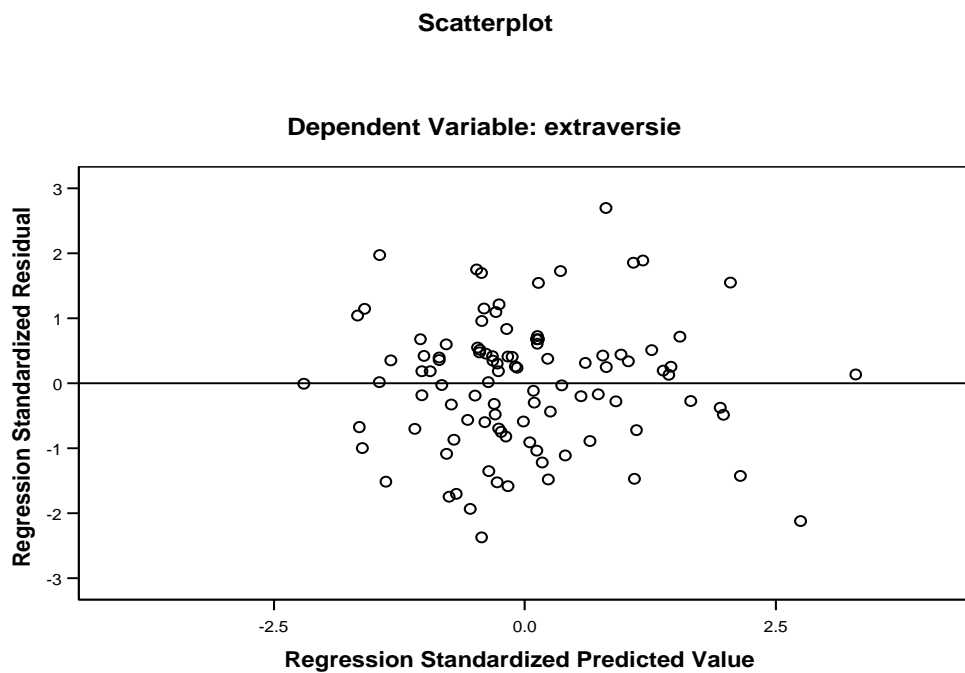
- a. Er zijn geen verschillen in extraversie tussen mannen en vrouwen indien gecorrigeerd wordt voor angst.
- b. Er zijn verschillen in extraversie tussen mannen en vrouwen indien gecorrigeerd wordt voor angst
- c. Op basis van deze analyse is het duidelijk dat er geen uitspraak gedaan kan worden over eventuele verschillen in extraversie tussen mannen en vrouwen.
- d. Geen van bovenstaande alternatieven is juist.

Er zijn ook nog twee plaatjes gemaakt:

Normal P-P Plot of Regression Standardized Residual



Figuur 1: Uitkomst van de analyse.



Figuur 2: Nog een uitkomst van de analyse.

18. Op basis van welke van deze figuren kan een uitspraak gedaan worden over het feit of er sprake is van homoscedasticiteit?

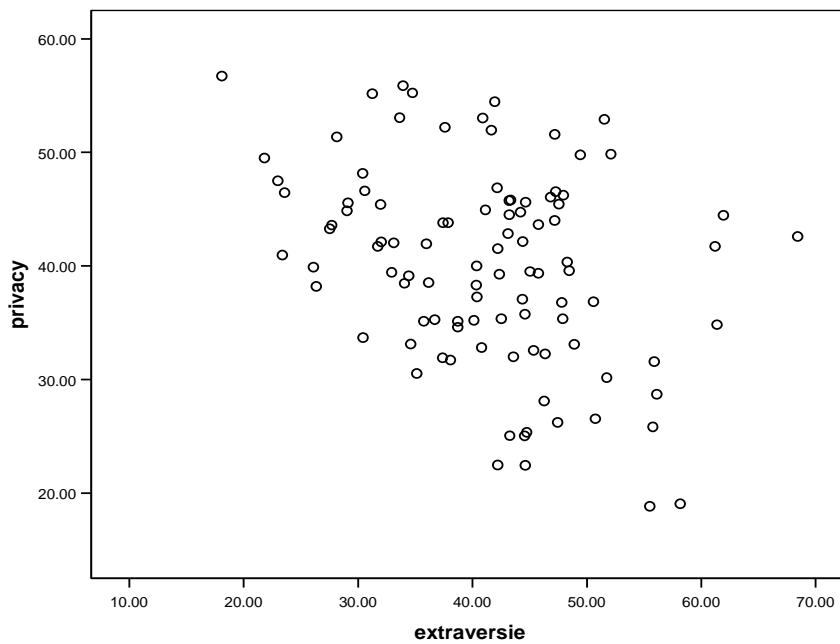
- a. Figuur 1
- b. Figuur 2
- c. Beide figuren.
- d. Geen van beide figuren

19. Op basis van welke van deze figuren kan een uitspraak gedaan worden over het feit of er sprake is van een lineaire relatie?

- a. Figuur 1
- b. Figuur 2
- c. Beide figuren.
- d. Geen van beide figuren

Case

Het ging de onderzoeker natuurlijk om de relatie tussen extraversiteit en de bereidheid om privacy op te geven. Als eerste maakt de onderzoeker een scatterplot (strooidiagram) en voert hij een eerste regressieanalyse uit. Stukken uit de output staan hieronder vermeldt; privacy staat hierbij voor het belang dat men hecht aan privacy; hoe hoger deze score, hoe minder men geneigd is om de privacy op te geven.



Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.354 ^a	.125	.116	8.11050

a. Predictors: (Constant), extraversie

b. Dependent Variable: privacy

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53.213	3.622		14.691	.000
	extraversie	-.321	.086	-.354	-3.743	.000

a. Dependent Variable: privacy

- a. (4 punten) Druk in woorden uit hoe de relatie tussen privacy en extraversie eruit ziet (dus iets als: “als variabele XXX met 1 toeneemt dan wordt variabele YYYY zus en zoveel groter/kleiner”).
- b. (4 punten) Wat is de waarde van de Pearson productmoment correlatiecoëfficiënt?

Vervolgens neemt de onderzoeker ook de variabele angst (angst voor terrorisme) in het model op. Wederom volgen stukken van de output.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.707 ^a	.500	.490	6.15994

a. Predictors: (Constant), angst, extraversie

b. Dependent Variable: privacy

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	80.368	4.205		19.111	.000
	extraversie	-.486	.068	-.536	-7.157	.000
	angst	-.525	.062	-.639	-8.538	.000

a. Dependent Variable: privacy

- c. (4 punten) De waarde van “Std. Error of the Estimate” (dit is de residuele standaarddeviatie) is in deze analyse flink wat kleiner dan in de vorige analyse. Kun je hier een verklaring voor geven?

De onderzoeker neemt ook nog een interactievariabele op. Deze variabele (genaamd “interactie”) is gelijk aan het product van extraversie en angst. Wederom volgt output:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	51.359	10.353		4.961	.000
	extraversie	.200	.235	.220	.852	.396
	angst	.226	.254	.275	.891	.375
	interactie	-.018	.006	-1.029	-3.043	.003

a. Dependent Variable: privacy

d. (4 punten) Geef de formule voor de regressievergelijking; dat is dus iets als 'variabele x' = a + b₁ × 'variabele y' + b₂ × 'variabele z' etc., maar dan ingevuld (wat is variabele x, wat is a etc.).

e. (4 punten) Wat is de voorspelde waarde voor privacy voor de onderstaande 4 personen:

persoon	angstscore	extraversiescore
A	20	20
B	20	60
C	60	20
D	60	60

f. (4 punten) Beschrijf in woorden wat de uitkomst van deze laatste analyse is. Hoe zit de relatie tussen extraversie en privacy eruit?

Sleutel

1 a

2 b

3 d

4 c

5 d

6 d

7 b

8 d

9 c

10 b

11 a

12 d

13 c

14 d

15 b

16 b

17 a

18 b

19 b