Oefening II - Correlatie basis

# Basisbegrippen in de Correlation

## 📋 OVERZICHTSTABEL - VRAGEN VOLGENS BLOOM’S TAXONOMIE (25 VRAGEN)

### ONTHOUDEN NIVEAU (5 vragen - 20%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q1 | Wat is correlatie? | Definitie van correlatie herinneren |
| Q2 | Wat is een z-score? | Z-score definitie herinneren |
| Q3 | Belangrijkste maatregelen in verband | Belangrijke correlatiematen herinneren |
| Q4 | Centrum in z-Scores | Centrum van gestandaardiseerde verdeling herinneren |
| Q5 | Eenheden veranderen niet r | Invariantie van correlatie voor eenheden herinneren |

### NIVEAU BEGRIJPEN (13 vragen - 52%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q6 | Variabeletypen voor correlatie | Variabeletypes voor correlatie begrijpen |
| Q7 | Basisprincipes van correlatie versus causaliteit | Verschil correlatie-causatie begrijpen |
| Q8 | Correlatiewaarden interpreteren | Correlatie-interpretatie begrijpen |
| Q9 | Wat correlatie ons vertelt | Richting, sterkte en betekenis van correlatie begrijpen |
| Q10 | Covariantie versus correlatie | Verschil tussen covariantie en correlatie begrijpen |
| Q11 | Visualisatie belang | Belang van visualisatie begrijpen |
| Q12 | Covariantie versus correlatie | Verschil covariantie-correlatie begrijpen |
| Q13 | Wanneer correlatie misleidend kan zijn | Beperkingen van correlatie begrijpen |
| Q17 | Correlatie types | Verschillende correlatie types begrijpen |
| Q18 | Wanneer welke correlatie gebruiken | Correlatie types toepassing begrijpen |
| Q20 | Waarom standardisatie nodig is | Standardisatie noodzaak begrijpen |
| Q21 | Correlatie interpretatie | Correlatie interpretatie begrijpen |
| Q23 | Correlatie veronderstellingen | Correlatie aannames begrijpen |

### TOEPASSEN NIVEAU (3 vragen - 12%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q14 | Pearson vs Spearman voor niet-lineaire relaties | Monotoon versus lineair toepassen |
| Q15 | Richting en kracht | Richting en sterkte bepalen |
| Q16 | Zwak positief | Zwak positieve correlatie herkennen |

### ANALYSEER NIVEAU (1 vraag - 4%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q19 | Impact van uitschieters | Impact van uitbijters analyseren |

### NIVEAU EVALUEREN (1 vraag - 4%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q22 | Statistische declaraties beoordelen | Statistische claims beoordelen |

### CREATIE NIVEAU (2 vragen - 8%)

| Vraagnummer | Vraag Titel | Leerintentie |
| --- | --- | --- |
| Q24 | Ontwerp een correlatiestudie | Correlatiestudie ontwerpen |
| Q25 | Onderzoekshypothese opstellen | Onderzoekshypothese formuleren |

## ONTHOUDEN NIVEAU (5 vragen)

### Vraag Q1 (Onthouden)

**Wat is correlatie?**

Hint: Denk na over hoe twee criminologische variabelen samen bewegen - bijvoorbeeld werkloosheid en misdaadcijfers - niet over oorzaak en gevolg.

1. Een statistische maat voor de sterkte en richting van de relatie tussen twee variabelen “1” = ” ✅ Juist! Correlatie kwantificeert hoe sterk en in welke richting twee kwantitatieve variabelen met elkaar in verband staan. Een positieve correlatie (bijv. r = +0,68 tussen werkloosheid en vermogenscriminaliteit) betekent dat hogere werkloosheid gepaard gaat met hogere criminaliteit, terwijl een negatieve correlatie (bijv. r = –0,55 tussen zichtbaarheid van de politie en wanorde) betekent dat naarmate de ene toeneemt, de andere afneemt. [Lees meer](https://leansixsigmagroep.nl/lean-agile-en-six-sigma/correlatie/?srsltid=AfmBOophFNeGHIK_T7CVXYTkIvaxhEmKCOJDZq3CGaAlWmlS8aPEclbT)”
2. Een maat voor de mate waarin de ene variabele de andere veroorzaakt “2” = ” ❌ Fout. Dit is een veel voorkomende misvatting in onderzoek. Correlatie laat zien dat twee variabelen samen veranderen, maar het bewijst niet dat de ene de andere veroorzaakt. Buurten met een hoge werkloosheid hebben bijvoorbeeld vaak hogere misdaadcijfers, maar dit kan het gevolg zijn van een derde factor, zoals armoede, sociale desorganisatie of zwakke informele controle. [Lees meer](https://leansixsigmagroep.nl/lean-agile-en-six-sigma/correlatie/?srsltid=AfmBOophFNeGHIK_T7CVXYTkIvaxhEmKCOJDZq3CGaAlWmlS8aPEclbT)”
3. Het verschil tussen twee variabelen “3” = ” ❌ Fout. Het nemen van een verschil (Y - X) meet alleen een kloof, niet of de variabelen samen variëren. Het verschil tussen diefstal- en mishandelingspercentages zegt bijvoorbeeld niets over de vraag of deze misdaden samen stijgen of dalen. Correlatie onderzoekt co-variatie - hoeveel beide variabelen afwijken van hun gemiddelden in dezelfde of tegengestelde richting. [Lees meer](https://leansixsigmagroep.nl/lean-agile-en-six-sigma/correlatie/?srsltid=AfmBOophFNeGHIK_T7CVXYTkIvaxhEmKCOJDZq3CGaAlWmlS8aPEclbT)”
4. Een methode om toekomstige waarden te voorspellen “4” = ” ❌ Fout. Dat is regressieanalyse, geen correlatie. Regressie stelt criminologen in staat om bijvoorbeeld het verwachte misdaadcijfer voor een bepaald werkloosheidsniveau te voorspellen. Correlatie vat alleen samen hoe twee variabelen op een bepaald moment in de tijd met elkaar in verband staan - het kan geen voorspellingen doen of oorzakelijk verband bewijzen. [Lees meer](https://leansixsigmagroep.nl/lean-agile-en-six-sigma/correlatie/?srsltid=AfmBOophFNeGHIK_T7CVXYTkIvaxhEmKCOJDZq3CGaAlWmlS8aPEclbT)”

### Vraag Q2 (Onthouden)

**Wat is een z-score?**

**Hint:** Denk na over het vergelijken van misdaadstatistieken in verschillende rechtsgebieden of tijdsperioden - hoe weten we of een misdaadcijfer ongewoon hoog of laag is in vergelijking met het typische patroon?

1. De ruwe score minus het gemiddelde  
   “1” = ” ❌ Fout. Dit geeft u een afwijkingsscore, die aangeeft hoe ver boven of onder het gemiddelde een waarde is, maar geen rekening houdt met de variabiliteit in uw gegevens. Als het gemiddelde misdaadcijfer bijvoorbeeld 50 per 1.000 is en een stad 60 per 1.000 heeft, is de afwijking +10. Maar we kunnen niet beoordelen of dit ‘veel’ is zonder te weten of misdaadcijfers doorgaans variëren met ±2 of ±20 rond het gemiddelde. De z-score standaardiseert dit door te delen door de standaarddeviatie.”
2. Het percentage boven het gemiddelde  
   “2” = ” ❌ Fout. Z-scores worden uitgedrukt in standaarddeviatie-eenheden, niet in percentages. Z-scores kunnen echter worden geconverteerd naar percentielen met behulp van de normale verdeling. z = +1,0 komt bijvoorbeeld overeen met ongeveer het 84e percentiel (wat betekent dat 84% van de waarden onder dit punt valt), maar de z-score zelf is geen percentage. In de criminologie is dit onderscheid van belang bij het interpreteren van risicobeoordelingen of het vergelijken van misdaadstatistieken in verschillende rechtsgebieden.”
3. Een score die aangeeft hoeveel standaarddeviaties een waarde is ten opzichte van het gemiddelde  
   “3” = ” ✅ Juist! Z-scores standaardiseren waarden met behulp van de formule z = (X - μ)/σ, waardoor zinvolle vergelijkingen tussen verschillende variabelen of tijdsperioden mogelijk zijn. Als het geweldsmisdrijfcijfer van een stad bijvoorbeeld z = +2,1 heeft, betekent dit dat het 2,1 standaarddeviaties boven het gemiddelde voor alle steden ligt - een ongewoon hoog percentage. Met Z-scores kunnen we vergelijken ‘hoe ongebruikelijk’ verschillende soorten criminaliteit zijn: een inbraakcijfer met z = +1,5 en een mishandelingspercentage met z = +0,8 duiden beide op bovengemiddelde criminaliteit, maar inbraak is extreem hoog.”
4. Het kwadraat van de originele partituur  
   “4” = ” ❌ Fout. Het kwadrateren van een score zou de schaal drastisch veranderen en informatie verliezen over de vraag of de oorspronkelijke waarde boven of onder het gemiddelde lag. Z-scores behouden de relatieve positie terwijl ze de schaal standaardiseren. Als we de misdaadcijfers in het kwadraat zouden brengen, zou een stad met 100 misdaden per 1.000 10.000 worden - waardoor de betekenis volledig zou veranderen en vergelijkingen onmogelijk zouden worden.”

### Vraag Q3 (Onthouden)

**Een criminoloog heeft een onderzoek uitgevoerd naar verschillende criminaliteitsvariabelen in 45 Belgische gemeenten. Zij heeft gegevens verzameld over politieaanwezigheid, drugsdelicten, geweldsdelicten en vermogenscriminaliteit. Nu moet zij een correlatiematrix rapporteren zijn de belangrijkste maatregelen in correlatieanalyse?**

**Hint:** Bij het bestuderen van de relatie tussen twee misdaadgerelateerde variabelen, welke statistieken heb je dan nodig om de associatie volledig te begrijpen en te rapporteren?

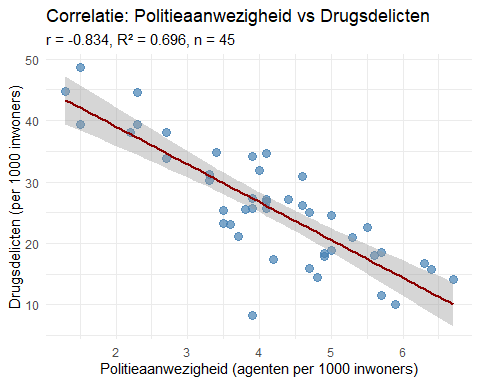
=== DESCRIPTIVE STATISTICS ===

Politieaanwezigheid (agenten per 1000 inwoners): Mean (M): 4.12 Standard Deviation (SD): 1.3 Variance: 1.68

Drugsdelicten (per 1000 inwoners): Mean (M): 25.96 Standard Deviation (SD): 9.58 Variance: 91.86

=== CORRELATION ANALYSIS ===

Correlation coefficient (r): -0.834 Coefficient of determination (R²): 0.696 Explained variance: 69.6 % Sample size (n): 45 p-value: 0 Significance: p < .001 \*\*\*



Saved SVG image: 3.3 Correlatie - Belangrijkste maatregelen/description/media/correlation\_plot\_3.svg with dimensions: 6 x 4

1. De helling en het snijpunt  
   “1” = ” ❌ Fout. Helling en snijpunt zijn regressieparameters, geen correlatiematen. Hoewel regressie en correlatie gerelateerd zijn (helling = r × sy/sx), dienen ze verschillende doelen. Correlatie vat de kracht van associatie samen; Regressie biedt voorspellingsvergelijkingen. In de criminologie zou je correlatie kunnen rapporteren bij het bestuderen of twee soorten criminaliteit de neiging hebben om samen voor te komen, maar regressie bij het voorspellen van toekomstige misdaadcijfers op basis van demografische factoren.”
2. Alleen het gemiddelde en de standaarddeviatie  
   “2” = ” ❌ Fout. Gemiddelden en standaarddeviaties beschrijven individuele variabelen, maar geven niet de relatie tussen twee variabelen weer. Je zou identieke gemiddelden en standaarddeviaties kunnen hebben voor werkloosheids- en misdaadcijfers in twee verschillende onderzoeken, maar totaal verschillende correlaties (r = +0,80 vs. r = -0,20). Correlatieanalyse onderzoekt specifiek hoe variabelen samen variëren, wat andere statistieken vereist dan eenvoudige beschrijvende metingen.”
3. Alleen de correlatiecoëfficiënt (r)  
   “3” = ” ❌ Fout. Hoewel r de primaire maatstaf is, levert het rapporteren van alleen r onvolledige informatie op voor onderzoek. U moet ook weten hoeveel variantie wordt verklaard (R²) en of uw steekproefomvang (n) groot genoeg is om het resultaat te vertrouwen. R = 0,40 tussen aanwezigheid van de politie en misdaadvermindering lijkt bijvoorbeeld zinvol, maar als n = 8 steden, is het resultaat onbetrouwbaar. Professioneel onderzoek vereist het rapporteren van alle drie de statistieken.”
4. Correlatiecoëfficiënt (r), bepalingscoëfficiënt (R²) en steekproefomvang (n)  
   “4” = ” ✅ Juist! Deze drie geven volledige informatie over de relatie. De correlatiecoëfficiënt (r) geeft de sterkte en richting aan (-1 tot +1), R² geeft het verklaarde variantiepercentage aan (r² × 100%) en de steekproefomvang (n) geeft betrouwbaarheid aan. Bijvoorbeeld: ‘De zichtbaarheid en wanorde van de politie vertoonden een matige negatieve correlatie (r = -0,45, R² = 0,20, n = 150 buurten), wat aangeeft dat 20% van de variatie in wanorde kan worden verklaard door de aanwezigheid van de politie.’ Dit geeft lezers alle informatie die nodig is om de bevinding te evalueren.”

### Vraag Q4 (Onthouden)

**Wat betekent een correlatiecoëfficiënt van r = 0?**

**Hint:** Als een criminoloog een correlatie van precies nul vindt tussen twee variabelen, wat zegt dit dan over de relatie tussen deze variabelen?

1. Er is een perfecte positieve relatie “1” = ” ❌ Fout. Een perfecte positieve relatie wordt aangegeven door r = +1,00, niet r = 0. Bij r = +1 betekent dit dat wanneer de ene variabele toeneemt, de andere variabele altijd in dezelfde verhouding toeneemt. Bijvoorbeeld, als er een perfecte positieve correlatie zou zijn tussen het aantal politieagenten en veiligheid (wat in de praktijk nooit voorkomt), zou elke toename van politieaanwezigheid precies evenredig samengaan met verhoogde veiligheid.”
2. Er is een perfecte negatieve relatie  
   “2” = ” ❌ Fout. Een perfecte negatieve relatie wordt aangegeven door r = -1,00, niet r = 0. Bij r = -1 betekent dit dat wanneer de ene variabele toeneemt, de andere variabele altijd in precies dezelfde verhouding afneemt. Bijvoorbeeld, als er een perfecte negatieve correlatie zou zijn tussen politieaanwezigheid en criminaliteit, zou elke toename van politie exact evenredig samengaan met afname van misdaad.”
3. Er is geen lineaire relatie tussen de variabelen “3” = ” ✅ Juist! Een correlatie van r = 0 betekent dat er geen lineaire relatie bestaat tussen de twee variabelen. De variabelen bewegen niet systematisch samen - wanneer de ene variabele toeneemt, is er geen voorspelbaar patroon in de andere variabele. Bijvoorbeeld, er kan een correlatie van r ≈ 0 zijn tussen schoengrootte van politieagenten en hun effectiviteit in misdaadpreventie - deze variabelen zijn onafhankelijk van elkaar. Let op: r = 0 sluit wel andere soorten relaties uit (zoals kromme verbanden).”
4. De data is incorrect gemeten “4” = ” ❌ Fout. Een correlatie van nul duidt niet op meetfouten, maar op het ontbreken van een lineaire relatie. Veel variabelen in de criminologie hebben legitiem geen verband met elkaar. Bijvoorbeeld, de correlatie tussen de kleur van politieauto’s en misdaadpreventie-effectiviteit zou logischerwijs r ≈ 0 zijn - dit zijn gewoon onafhankelijke variabelen. Een correlatie van nul is een zinvolle bevinding die aantoont dat de bestudeerde factoren niet lineair samenhangen.”

### Vraag Q5 (Onthouden)

**Een onderzoeker bestudeert de relatie tussen patrouille-afstanden en politie-responssnelheid in 50 Belgische steden. Haar oorspronkelijke data toont een correlatie van r = -0,65 tussen gemiddelde patrouille-afstand (in meters) en responssnelheid (in minuten). Voor internationale vergelijking met Amerikaanse data (die feet gebruikt, 1 meter = 3,28 feet) moet zij de Belgische afstanden omrekenen van meters naar feet. Wat gebeurt er met de correlatiecoëfficiënt r tussen patrouille-afstand en responssnelheid na deze eenheidsconversie?**

**Hint:** Overweeg hoe eenheidsconversies (meters naar centimeters, of misdaadcijfers per 1.000 versus per 100.000 inwoners) de correlatie tussen variabelen beïnvloeden.

1. r wordt 100× groter  
   “1” = ” ❌ Fout. Eenheden hebben geen invloed op de correlatiemagnitude. Dit is een cruciale eigenschap voor onderzoek waarbij variabelen vaak verschillende schalen gebruiken. De correlatie tussen ‘jaren onderwijs’ en ‘arrestaties per 1.000 inwoners’ zou bijvoorbeeld hetzelfde zijn als de correlatie tussen ‘jaren onderwijs’ en ‘arrestaties per 100.000 inwoners’. De schaalverandering verandert niets aan de onderliggende relatiesterkte. Als veranderende eenheden de correlatie zouden beïnvloeden, zou vergelijkend onderzoek in landen met verschillende meetsystemen onmogelijk zijn.”
2. r wordt 100× kleiner  
   “2” = ” ❌ Fout. Nogmaals, correlatie is schaalinvariant. Of u nu de reactietijd van patrouillewagens meet in seconden, minuten of uren, de correlatie met de opruimingspercentages van misdrijven blijft identiek. Deze eigenschap maakt correlatie bijzonder waardevol voor internationale criminologische vergelijkingen - onderzoekers kunnen gegevens gebruiken die in verschillende eenheden zijn gemeten zonder wiskundige aanpassingen. De relatiesterkte tussen twee variabelen is onafhankelijk van hoe die variabelen worden geschaald.”
3. r blijft precies hetzelfde  
   “3” = ” ✅ Juist! Correlatie is eenheidsvrij en invariant voor lineaire herschaling. Dit komt omdat correlatie beide variabelen tijdens de berekening standaardiseert - het meet hoe variabelen samen bewegen ten opzichte van hun eigen variabiliteit, ongeacht de eenheden. In de criminologie betekent dit dat je correlaties kunt vergelijken tussen onderzoeken die verschillende meetschalen gebruiken. Studies die bijvoorbeeld misdaadcijfers per hoofd van de bevolking versus per vierkante kilometer meten, kunnen nog steeds vergelijkbare correlatiecoëfficiënten hebben met demografische factoren. Deze invariantie-eigenschap maakt correlatie tot een universele maat voor associatie.”
4. r wordt negatief  
   “4” = ” ❌ Fout. Veranderingen in eenheden kunnen de richting van een relatie niet omdraaien. Als een verhoogde aanwezigheid van de politie gepaard gaat met een verminderde criminaliteit (negatieve correlatie), blijft deze relatierichting hetzelfde, of je nu de aanwezigheid van agenten per blok, per vierkante kilometer of per 1.000 inwoners meet. Alleen de schaal verandert, niet het fundamentele patroon. Een sign flip zou duiden op een heel andere substantiële relatie, die eenheidsconversie niet kan veroorzaken.”

## BEGRIJPEN NIVEAU (8 vragen)

### Vraag Q6 (Begrijpen)

**Voor welke soorten variabelen kunt u de Pearson-correlatie berekenen?**

Hint: Overweeg of je de correlatie tussen verschillende soorten variabelen zinvol kunt berekenen - kun je geslacht correleren met misdaadcijfers? Lengte van de gevangenisstraf met kans op recidive?

1. Elk type variabelen  
   “1” = ” ❌ Fout. Er bestaan verschillende correlatiematen voor verschillende variabelecombinaties. Pearson-correlatie is specifiek voor continue variabelen. Andere metingen zijn: Spearman’s rho voor gerangschikte gegevens, phi-coëfficiënt voor twee binaire variabelen, Cramér’s V voor twee categorische variabelen en point-biserial voor één binaire en één continue variabele. Het gebruik van het verkeerde correlatietype kan leiden tot nietszeggende resultaten of ernstige verkeerde interpretaties van relaties in onderzoek.”
2. Een categorische en een continue variabele  
   “2” = ” ❌ Fout. Als u één categorische en één continue variabele hebt, gebruikt u punt-biseriële correlatie (voor binaire categorieën) of eta-correlatie (voor meerdere categorieën). Als je bijvoorbeeld de relatie tussen geslacht (categorisch: man/vrouw) en het aantal arrestaties (continu) bestudeert, zou je punt-biseriële correlatie gebruiken. Als je de relatie tussen het type buurt (categorisch: stedelijk/voorstedelijk/landelijk) en het misdaadcijfer (continu) onderzoekt, zou je ANOVA- of eta-correlatie kunnen gebruiken in plaats van Pearson-correlatie.”
3. Beide variabelen moeten continu zijn (interval- of rationiveau)  
   “3” = ” ✅ Juist! Pearson-correlatie vereist dat beide variabelen worden gemeten op interval- of rationiveau - numerieke variabelen waarbij de afstanden tussen waarden zinvol en consistent zijn. In de criminologie zijn voorbeelden: misdaadcijfers (ratio), strafduur in maanden (ratio), leeftijd bij eerste arrestatie (ratio), aantal eerdere veroordelingen (ratio) en schalen die de houding ten opzichte van de politie meten (interval). Deze variabelen maken zinvolle wiskundige bewerkingen mogelijk die nodig zijn voor het berekenen van gemiddelden, standaarddeviaties en covarianties die de basis vormen van de Pearson-correlatie.”
4. Alleen categorische variabelen  
   “4” = ” ❌ Fout. Pearson-correlatie vereist numerieke variabelen waarbij wiskundige bewerkingen (optellen, aftrekken, middelen) zinvol zijn. Categorische variabelen zoals ‘type misdaad’ (gewelddadig, eigendom, drugs) of ‘gerechtelijke beschikking’ (schuldig, niet schuldig, pleidooi) hebben geen inherente numerieke waarden. Je kunt geen zinvol gemiddelde berekenen van ‘gewelddadige’ en ‘eigendomsdelicten’. Voor categorische variabelen zou je verschillende maten gebruiken, zoals de V- of Chi-kwadraattest van Cramér om associaties te onderzoeken.”

### Vraag Q7 (Begrijpen)

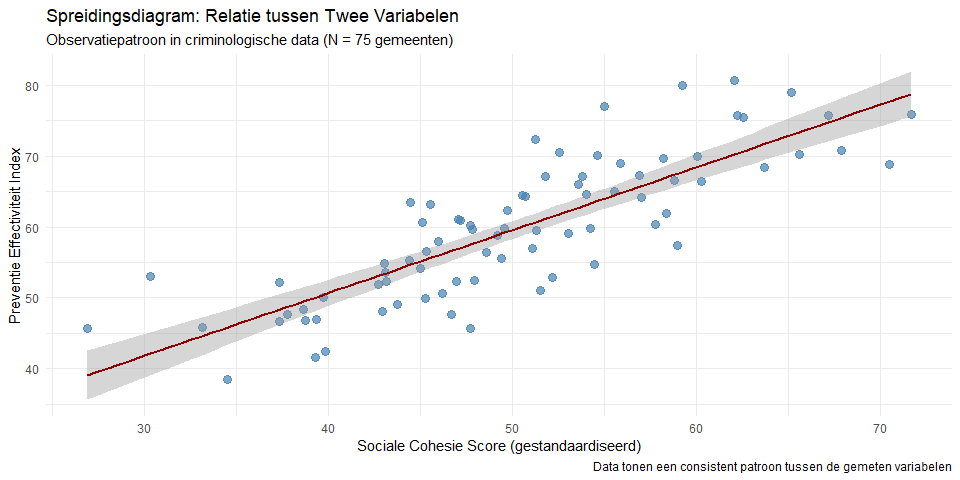
**Waarom bewijst correlatie geen causatie?**

**Hint:** Overweeg waarom een sterke correlatie tussen beveiligingscamera’s en misdaadcijfers, of tussen politieaanwezigheid en criminaliteit, niet noodzakelijk betekent dat het ene het andere veroorzaakt.

1. Derde variabelen, omgekeerde causaliteit of toeval kunnen de relatie verklaren  
   “1” = ” ✅ Juist! Correlatie kan geen causale richting vaststellen of alternatieve verklaringen uitsluiten. Drie hoofdproblemen: (1) Derde variabelen — armoede kan zowel een laag opleidingsniveau als hoge criminaliteit veroorzaken, wat een schijncorrelatie tussen onderwijs en criminaliteit creëert; (2) Omgekeerde causaliteit — verminderen meer politieagenten de criminaliteit, of leidt meer criminaliteit tot het aannemen van meer politie? (3) Toeval — twee variabelen kunnen per toeval correleren, vooral bij kleine steekproeven. Het vaststellen van causaliteit vereist gecontroleerde experimenten, longitudinale ontwerpen, natuurlijke experimenten of geavanceerde statistische technieken die causale effecten kunnen isoleren van deze alternatieve verklaringen.”
2. Correlatie is altijd te zwak om iets te bewijzen  
   “2” = ” ❌ Fout. Zelfs zeer sterke correlaties (r = 0,90 of hoger) bewijzen geen causaal verband. Het probleem is niet de sterkte van de correlatie, maar de logica van causale inferentie. Bijvoorbeeld, u kunt een zeer sterke correlatie (r = 0,85) vinden tussen het aantal beveiligingscamera’s in een buurt en misdaadcijfers, maar dit bewijst niet dat camera’s criminaliteit veroorzaken. De hoge correlatie kan bestaan omdat camera’s worden geïnstalleerd als reactie op hoge criminaliteit, niet omdat camera’s criminaliteit veroorzaken. Correlatiesterkte en causaal bewijs zijn compleet gescheiden kwesties.”
3. Alleen experimenten kunnen relaties aantonen  
   “3” = ” ❌ Fout. Correlatie toont duidelijk aan dat relaties bestaan — het kan alleen niet bepalen of ze causaal zijn. Veel belangrijke relaties kunnen om ethische redenen niet experimenteel worden bestudeerd (je kunt mensen niet willekeurig toewijzen aan criminele carrières). Correlationele studies leveren waardevol bewijs over associaties, risicofactoren en patronen. Hoewel experimenten het sterkste causale bewijs leveren, kunnen observationele studies met technieken zoals instrumentele variabelen, regressiediscontinuïteit of natuurlijke experimenten ook overtuigend causaal bewijs leveren wanneer experimenten niet haalbaar zijn.”
4. Correlatie is hetzelfde als causaliteit  
   “4” = ” ❌ Fout. Dit is precies het misverstand dat onderzoekers moeten vermijden. Media-berichten beweren vaak ‘studies tonen aan dat X Y veroorzaakt’ wanneer het onderzoek alleen correlatie heeft aangetoond. Het vinden dat wijken met meer straatverlichting minder criminaliteit hebben, bewijst bijvoorbeeld niet dat straatverlichting criminaliteit voorkomt — veiligere wijken kunnen simpelweg meer geneigd zijn om straatverlichting te installeren. Het begrijpen van dit onderscheid is cruciaal voor evidence-based strafrechtelijk beleid, aangezien het implementeren van interventies op basis van alleen correlationeel bewijs middelen kan verspillen of zelfs averechts kan werken.”

### Vraag Q8 (Begrijpen)

**Een criminoloog onderzoekt de relatie tussen sociale cohesie in buurten (onderlinge verbondenheid en samenwerking tussen bewoners) en de effectiviteit van misdaadpreventie in 75 Belgische gemeenten. Na het verzamelen van data over gemeenschapsbetrokkenheid en lokale veiligheidsprogramma’s, berekent ze een correlatie van r = 0,75. Hoe moet zij deze bevinding interpreteren voor haar rapport aan het Ministerie van Justitie en Veiligheid?**



**Hint:** Overweeg wat een correlatie van r = 0,75 betekent in termen van relatiesterkte, praktische betekenis voor beleid, en hoe dit zou moeten worden gecommuniceerd aan beleidsmakers.

1. Perfecte negatieve relatie - tegengesteld effect van verwachting  
   “1” = ” ❌ Fout. Deze interpretatie heeft twee grote fouten: (1) r = 0,75 is positief, niet negatief — de waarde ligt boven nul, wat aangeeft dat variabelen in dezelfde richting bewegen; (2) 0,75 is sterk maar niet perfect — een perfecte correlatie zou r = ±1,00 zijn. Een perfecte negatieve correlatie (r = -1,00) betekent dat wanneer sociale cohesie toeneemt, preventie effectiviteit in exact dezelfde verhouding afneemt, zonder uitzonderingen. In de criminologie zijn perfecte correlaties vrijwel onmogelijk vanwege de complexiteit van sociale fenomenen en meetfouten.”
2. Sterke positieve relatie - significante bevinding voor preventiebeleid  
   “2” = ” ✅ Juist! Waarden rond 0,7-0,8 geven sterke positieve associaties aan in sociaalwetenschappelijk onderzoek. Een r = 0,75 betekent dat wanneer sociale cohesie toeneemt, de effectiviteit van preventie programma’s substantieel toeneemt in een voorspelbaar patroon. In de Nederlandse criminologie context kan dit de empirische basis bieden voor investeringen in gemeenschapsopbouw als onderdeel van misdaadpreventie strategieën. De relatie is sterk genoeg om praktisch betekenisvol te zijn voor beleid en voorspelling, waarbij sociale cohesie ongeveer 56% van de variantie in preventie effectiviteit verklaart. Dit suggereert dat gemeenten met sterkere sociale bindingen waarschijnlijk meer succesvol zijn met hun veiligheidsprogramma’s.”
3. Zwakke positieve relatie - minimale beleidsimplicaties  
   “3” = ” ❌ Fout. Een r = 0,75 vertegenwoordigt een sterke relatie, niet zwak. In criminologisch onderzoek worden correlaties rond 0,75 als behoorlijk substantieel beschouwd. Ter vergelijking: de correlatie tussen lengte en gewicht bij volwassenen is typisch rond 0,70-0,80. Als je r = 0,75 vond tussen buurt sociale cohesie en misdaadpreventie effectiviteit, zou dit wijzen op een sterk positief verband — als sociale cohesie toeneemt, neigen preventie programma’s aanzienlijk effectiever te worden. Dit correlatieniveau suggereert dat de variabelen ongeveer 56% van hun variantie delen (0,75² = 0,56), wat aanzienlijk is in sociaalwetenschappelijk onderzoek en belangrijke beleidsimplicaties heeft.”
4. Geen betekenisvolle relatie - resultaten zijn statistisch ruis  
   “4” = ” ❌ Fout. Een r = 0,75 geeft een sterke en betekenisvolle relatie aan, niet de afwezigheid ervan of statistische ruis. ‘Geen relatie’ zou worden aangegeven door r ≈ 0,00. Om r = 0,75 in perspectief te plaatsen: veel belangrijke relaties in de criminologie zijn zwakker dan dit — bijvoorbeeld, de correlatie tussen individuele risicofactoren en daadwerkelijk crimineel gedrag kan r = 0,30-0,50 zijn. Het vinden van r = 0,75 tussen sociale cohesie en preventie effectiviteit zou worden beschouwd als een substantiële en praktisch belangrijke associatie die verdere investering in gemeenschapsgerichte preventie strategieën rechtvaardigt.”

### Vraag Q9 (Begrijpen)

**Wat vertelt een correlatie ons over de relatie tussen twee variabelen?**

**Hint:** Denk aan zowel wat correlatie meet (richting en sterkte) als wat het niet kan vertellen (causaliteit). Wat kan je afleiden uit een correlatiecoëfficiënt?

1. Hoeveel de ene variabele de andere veroorzaakt  
   “1” = ” ❌ Fout. Correlatie meet associatie, niet causaliteit. Dit is een cruciaal onderscheid in onderzoek. Een correlatie tussen eenoudergezinnen en jeugddelinquentie vertelt ons bijvoorbeeld niet dat eenoudergezinnen delinquentie veroorzaken — beide kunnen worden beïnvloed door armoede, buurtfactoren of andere variabelen. Een negatieve correlatie tussen aanwezigheid van de politie en misdaadcijfers kan betekenen: (1) politie vermindert criminaliteit (causaal), (2) gebieden met veel criminaliteit krijgen meer politie (omgekeerde causaliteit), of (3) beide worden beïnvloed door een derde factor zoals rijkdom in de buurt (schijncorrelatie). Verwarring van correlatie met causaliteit leidt tot verkeerd gerichte beleidsmaatregelen en interventies.”
2. Het verschil tussen twee groepsgemiddelden “2” = ” ❌ Fout. Het vergelijken van groepsgemiddelden gebruikt t-tests of ANOVA, niet correlatie. Het vergelijken van gemiddelde recidivecijfers tussen overtreders die een arbeidstraining hebben gekregen versus degenen die dat niet hebben gekregen, zou bijvoorbeeld een t-test gebruiken om het verschil in gemiddelden te onderzoeken. Correlatie onderzoekt hoe twee continue variabelen co-variëren (bijv. hoe individuele recidive-risicoscores gerelateerd zijn aan de daadwerkelijke frequentie van recidive). De statistische benaderingen en interpretaties zijn volledig verschillend — groepsvergelijkingen richten zich op verschillen tussen categorieën, terwijl correlatie zich richt op lineaire relaties tussen continue metingen.”
3. De richting en sterkte van een lineaire relatie tussen twee variabelen  
   “3” = ” ✅ Juist! Correlatie kwantificeert zowel richting als sterkte van hoe twee variabelen samen bewegen in een rechtlijnig patroon. De correlatiecoëfficiënt (r) vangt richting op via het teken (+ of -) en sterkte via de absolute waarde (0 tot 1). Positieve correlaties (+) betekenen dat variabelen in dezelfde richting bewegen: als het inkomen in de buurt toeneemt, nemen vastgoedwaarden toe. Negatieve correlaties (-) betekenen tegengestelde beweging: als de zichtbaarheid van de politie toeneemt, neemt straatcriminaliteit af. De sterkte vertelt hoe consistent deze relatie is: r = -0,80 is een sterkere relatie dan r = -0,30. Het begrijpen van zowel richting als sterkte is cruciaal voor het identificeren van risicofactoren (positieve correlatie met criminaliteit) versus beschermende factoren (negatieve correlatie).”
4. Hoeveel observaties er in de dataset zitten  
   “4” = ” ❌ Fout. De correlatierichting en -sterkte zijn onafhankelijk van de steekproefgrootte (n). Je kunt r = +0,60 hebben met n = 30 steden of n = 3.000 steden — zowel de positieve richting als de matige sterkte blijven hetzelfde. Steekproefgrootte beïnvloedt wel de betrouwbaarheid en statistische significantie van de correlatie, maar niet de richting of sterkte ervan. Grotere steekproeven leveren echter wel stabielere schattingen op van de richting en grootte van de correlatie, daarom proberen criminologische studies vaak veel jurisdicties of tijdsperioden op te nemen wanneer dat mogelijk is.”

### Vraag Q10 (Begrijpen)

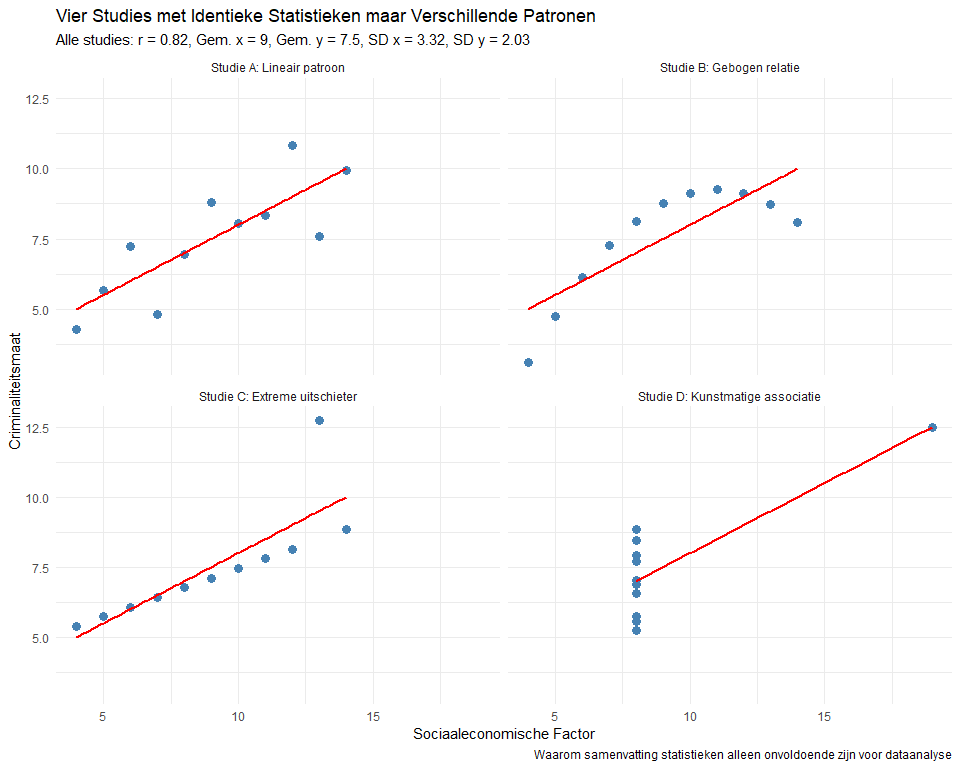
**Wat is het belangrijkste verschil tussen covariantie en correlatie bij het meten van relaties tussen variabelen?**

**Hint:** Denk aan wat er gebeurt wanneer je de meeteenheden van variabelen verandert. Welke maatstaf blijft hetzelfde en welke verandert?

1. Covariantie is altijd positief, correlatie kan negatief zijn  
   “1” = ” ❌ Fout. Beide maatstaven kunnen zowel positief als negatief zijn, afhankelijk van de richting van de relatie tussen variabelen. Covariantie kan variëren van negatief oneindig tot positief oneindig, terwijl correlatie varieert van -1 tot +1. Bijvoorbeeld, de relatie tussen politie-aanwezigheid en criminaliteitscijfers zou zowel een negatieve covariantie als een negatieve correlatie kunnen tonen, wat aangeeft dat meer politie-aanwezigheid geassocieerd is met lagere criminaliteitscijfers. De richting (positief of negatief) hangt af van de onderliggende relatie tussen de variabelen, niet van welke maatstaf je gebruikt.”
2. Covariantie hangt af van meeteenheden, correlatie is eenheidsvrij  
   “2” = ” ✅ Juist! Dit is het fundamentele onderscheid. Covariantie verandert wanneer je de schaal van variabelen wijzigt - als je inkomen meet in euro’s versus duizenden euro’s, krijg je verschillende covariantiewaarden voor dezelfde relatie. Correlatie blijft daarentegen constant ongeacht de meeteenheden omdat het een gestandaardiseerde maatstaf is. Bijvoorbeeld: de correlatie tussen buurtinkomen en criminaliteit blijft r = -0,45 of je inkomen meet in euro’s, duizenden euro’s, of dollars. Deze eenheidsvrije eigenschap maakt correlatie veel praktischer voor vergelijkende analyses en interpretatie, terwijl covariantie afhankelijk is van willekeurige keuzes in meeteenheden.”
3. Correlatie kan alleen waarden tussen 0 en 1 aannemen  
   “3” = ” ❌ Fout. Correlatie kan waarden aannemen tussen -1 en +1, niet alleen tussen 0 en 1. Het bereik van -1 tot +1 omvat zowel negatieve als positieve relaties. Een correlatie van -1 betekent een perfecte negatieve relatie (als de ene variabele toeneemt, neemt de andere perfect voorspelbaar af), 0 betekent geen lineaire relatie, en +1 betekent een perfecte positieve relatie. In criminologisch onderzoek zie je bijvoorbeeld vaak negatieve correlaties tussen sociaaleconomische status en criminaliteitscijfers (r = -0,60), wat aangeeft dat hogere sociaaleconomische status geassocieerd is met lagere criminaliteitscijfers.”
4. Beide maatstaven hebben hetzelfde interpretatiekader  
   “4” = ” ❌ Fout. Hoewel beide de richting van een relatie aangeven, hebben ze zeer verschillende interpretatiekaders. Covariantiewaarden hebben geen gestandaardiseerd bereik en zijn moeilijk te interpreteren zonder context van de oorspronkelijke variabelen - een covariantie van 500 kan sterk of zwak zijn afhankelijk van de schalen van de variabelen. Correlatie heeft daarentegen een vast interpretatiekader: waarden rond ±0,10 zijn zwak, rond ±0,30 zijn matig, rond ±0,50 zijn sterk, en boven ±0,70 zijn zeer sterk. Deze gestandaardiseerde interpretatie maakt correlatie veel praktischer voor onderzoekscommunicatie en vergelijkende analyses.”

### Vraag Q11 (Begrijpen)

**Een onderzoeker presenteert vier verschillende datasets van criminologische studies over de relatie tussen sociaaleconomische factoren en criminaliteit. Alle vier studies rapporteren precies dezelfde correlatiecoëfficiënt (r = 0,82), hetzelfde gemiddelde, en dezelfde standaarddeviatie. De onderzoeker concludeert dat alle studies identieke resultaten tonen. Wat is de belangrijkste methodologische les uit deze situatie?**

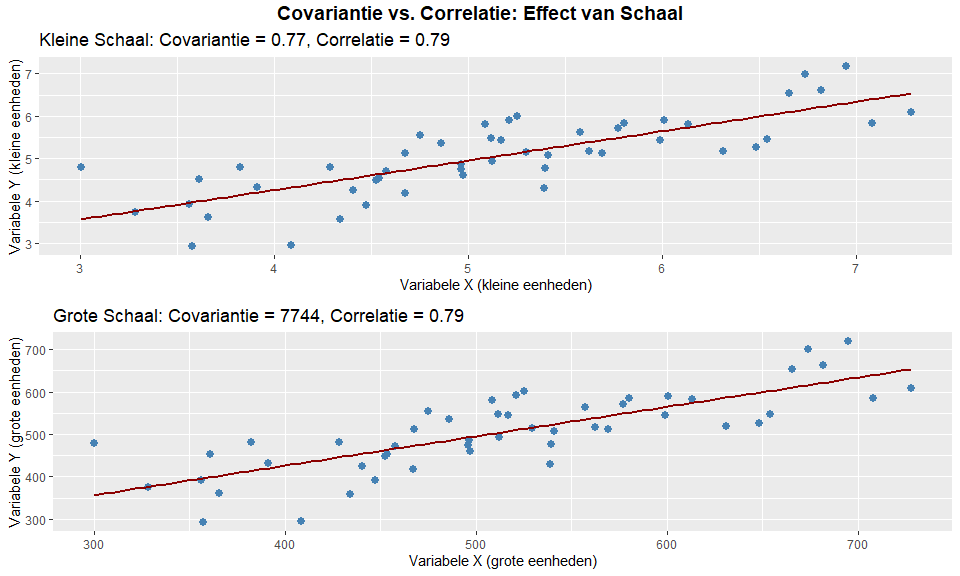


**Hint:** Wanneer statistieken identiek zijn maar patronen verschillen, wat toont dit aan over de beperkingen van samenvatting getallen?

1. Correlatiecoëfficiënten vertellen het complete verhaal  
   “1” = ” ❌ Fout. Deze vier datasets tonen krachtig aan dat alleen correlatiecoëfficiënten diep misleidend kunnen zijn. Alle vier studies delen dezelfde identieke correlatie (r = 0,82), maar ze vertegenwoordigen fundamenteel verschillende relaties: Studie A toont een juiste lineaire relatie, Studie B onthult een gebogen (niet-lineair) patroon, Studie C bevat een invloedrijke uitschieter die de correlatie vertekent, en Studie D toont een kunstmatige associatie door één extreme datapunt. In onderzoek zou het vertrouwen op alleen correlatiecoëfficiënten leiden tot totaal verkeerde interpretaties van de onderliggende relaties.”
2. Visuele dataverkenning is essentieel voor betrouwbare analyse  
   “2” = ” ✅ Juist! Deze demonstratie toont waarom onderzoekers hun gegevens altijd moeten plotten voordat ze conclusies trekken. Alleen door de spreidingsdiagrammen te bekijken kunt u onderscheiden tussen een echte lineaire relatie, niet-lineaire patronen, invloedrijke uitschieters en kunstmatige associaties. In onderzoek kan dit het verschil betekenen tussen het identificeren van een echte trend (zoals de relatie tussen werkloosheid en eigendomscriminaliteit) versus het verkeerd interpreteren van gegevensartefacten. Visuele inspectie helpt onderzoekers de juiste analysemethode te kiezen en potentiële problemen met hun data te identificeren.”
3. Statistische samenvattingen zijn altijd voldoende voor conclusies  
   “3” = ” ❌ Fout. Dit is precies de mythe die deze demonstratie ontkracht. Hoewel correlatiecoëfficiënten en andere samenvattingen waardevolle informatie bieden, kunnen ze identiek zijn voor fundamenteel verschillende datapatronen. Zonder visuele inspectie kunt u een gebogen relatie missen die een ander type analyse vereist, een uitschieter negeren die uw resultaten vertekent, of een hefboompunt over het hoofd zien dat een schijnbare correlatie creëert. In criminologie kunnen dergelijke missers leiden tot verkeerde beleidsaanbevelingen.”
4. Studies met identieke correlaties hebben altijd dezelfde interpretatie  
   “4” = ” ❌ Fout. Dit misverstand ligt aan de basis van waarom deze demonstratie zo’n krachtig onderwijsinstrument is. Dezelfde correlatie kan ontstaan uit volkomen verschillende onderliggende patronen. In onderzoek zou dit betekenen dat verschillende mechanismen (lineaire trends, drempeleffecten, invloedrijke casussen, meetartefacten) allemaal dezelfde samengevatte correlatie kunnen produceren, maar totaal verschillende beleidsimplicaties hebben. Daarom is het cruciaal om altijd de ruwe data te visualiseren voordat interpretaties worden gemaakt.”

### Vraag Q12 (Begrijpen)

**Welke bewering over covariantie en correlatie is waar?**



**Belangrijkste Vergelijking:** - **Covariantie**: Verandert met schaal/eenheden (0,70 vs 7.000), waardoor vergelijkingen moeilijk worden - **Correlatie**: Gestandaardiseerd (-1 tot +1) en eenheidsvrij (0,70 in beide gevallen) - **Interpretatie**: Correlatie maakt zinvolle vergelijking tussen verschillende variabelenschalen mogelijk

**Hint:** Let op eenheden en bereiken.

1. Beide zijn afhankelijk van meetschaal  
   “1” = ” ❌ Fout. Een belangrijk onderscheid tussen deze maatstaven is dat covariantie afhangt van meetschaal terwijl correlatie dat niet doet. Als je eenheden verandert (bijv. van meters naar centimeters), zal de covariantie proportioneel veranderen, maar correlatie blijft onveranderd. Dit is waarom correlatie de voorkeur heeft voor vergelijkende analyse - het biedt een gestandaardiseerde maatstaf die niet wordt beïnvloed door willekeurige keuzes in meeteenheden. Het meten van temperatuur in Celsius versus Fahrenheit zou bijvoorbeeld verschillende covariantiewaarden opleveren, maar identieke correlatiecoëfficiënten bij het onderzoeken van relaties met andere variabelen.”
2. Covariantie is eenheidsvrij en begrensd tussen −1 en +1  
   “2” = ” ❌ Fout. Je hebt covariantie verward met correlatie. Covariantie is NIET eenheidsvrij en is NIET begrensd tussen -1 en +1. Covariantiewaarden hangen af van de meeteenheden (bijv. euro’s, kilogrammen, jaren) en kunnen variëren van negatief oneindig tot positief oneindig. Bijvoorbeeld, de covariantie tussen inkomen (gemeten in duizenden euro’s) en onderwijs (gemeten in jaren) kan 4,5 zijn, maar als inkomen in euro’s werd gemeten, zou de covariantie 4.500 zijn - een heel ander getal dat dezelfde relatiesterkte vertegenwoordigt.”
3. Correlatie is eenheidsvrij en altijd tussen −1 en +1  
   “3” = ” ✅ Juist! Correlatie (r) standaardiseert de relatie door covariantie te delen door het product van standaarddeviaties, waardoor een eenheidsvrije maatstaf ontstaat die altijd tussen -1 en +1 valt. Deze standaardisatie is waarom we correlatiewaarden zinvol kunnen vergelijken tussen verschillende variabelenparen ongeacht hun oorspronkelijke meetschalen. Bijvoorbeeld, r = 0,7 vertegenwoordigt dezelfde relatiesterkte of we nu lengte met gewicht correleren of testscores met studietijd. Deze begrensde, gestandaardiseerde aard maakt correlatie de voorkeursmaat voor het vergelijken van relatiesterkte tussen verschillende variabelenparen.”
4. Beide zijn onbegrensd  
   “4” = ” ❌ Fout. Hoewel covariantie inderdaad onbegrensd is (variërend van negatief oneindig tot positief oneindig), is correlatie specifiek ontworpen om begrensd te zijn tussen -1 en +1. Dit is een fundamenteel verschil tussen deze twee maatstaven. De begrensde aard van correlatie komt voort uit het normalisatieproces, waarbij covariantie wordt gedeeld door het product van standaarddeviaties. Deze wiskundige transformatie zorgt ervoor dat, ongeacht hoe groot de covariantie mogelijk is, de correlatiecoëfficiënt altijd binnen het gestandaardiseerde bereik van -1 tot +1 valt, waardoor het gemakkelijker te interpreteren en vergelijken is tussen verschillende analyses.”

### Vraag Q13 (Begrijpen)

**Wanneer kan een correlatiecoëfficiënt misleidend zijn bij het interpreteren van relaties tussen variabelen?**

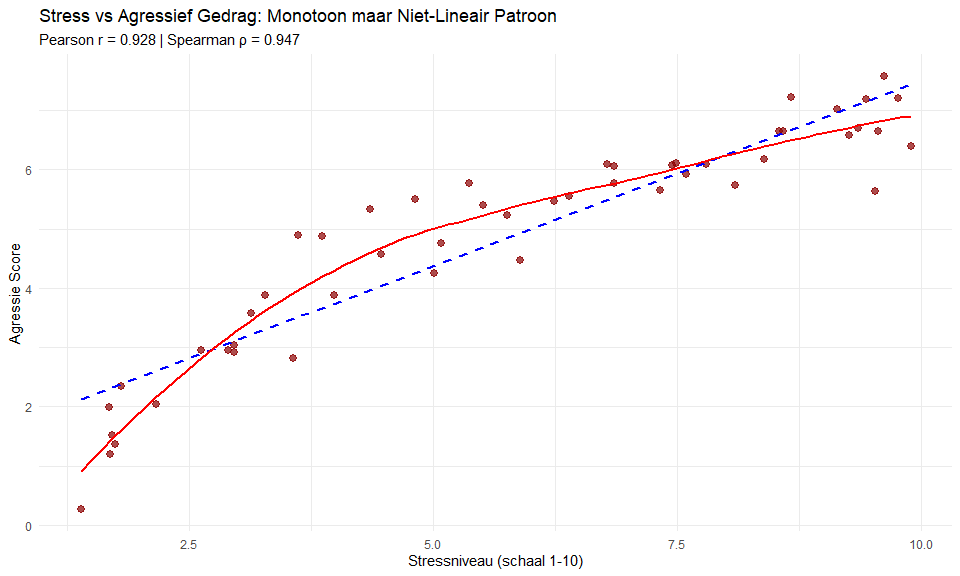
**Hint:** Denk aan situaties waar de numerieke correlatie niet het hele verhaal vertelt over de werkelijke relatie.

1. Wanneer er uitbijters (outliers) in de data zitten  
   “1” = ” ✅ Juist! Uitbijters kunnen correlatiecoëfficiënten drastisch beïnvloeden en misleidende resultaten geven. Een enkele extreme waarde kan een zwakke correlatie sterk doen lijken, of omgekeerd een sterke relatie verzwakken. Bijvoorbeeld, als je de relatie onderzoekt tussen leeftijd en criminaliteit, kan één zeer jonge verdachte met extreem veel arrestaties de hele correlatie vertekenen. Het is daarom essentieel om altijd spreidingsdiagrammen te maken en uitbijters te identificeren voordat je correlaties interpreteert. In criminologisch onderzoek kunnen uitbijters vaak interessante gevallen vertegenwoordigen die aparte aandacht verdienen.”
2. Wanneer de steekproef te klein is  
   “2” = ” ❌ Fout. Hoewel kleine steekproeven minder betrouwbare correlaties opleveren, maken ze de correlatie niet per se misleidend - ze maken het alleen moeilijker om statistisch significante resultaten te krijgen. Een correlatie van r = 0,6 in een kleine steekproef (n=10) kan wel accuraat zijn, maar gewoon niet statistisch significant. Het probleem is gebrek aan power, niet misleiding. Daarentegen kunnen grote steekproeven zeer kleine, praktisch irrelevante correlaties statistisch significant maken, wat ook kan misleiden.”
3. Wanneer beide variabelen normaal verdeeld zijn  
   “3” = ” ❌ Fout. Normale verdelingen maken correlatie juist betrouwbaarder, niet misleidender. Pearson-correlatie werkt het best wanneer beide variabelen normaal verdeeld zijn. Het probleem ontstaat juist wanneer variabelen NIET normaal verdeeld zijn - bijvoorbeeld bij sterk scheve verdelingen of ordinale data. In zulke gevallen kan Pearson-correlatie de werkelijke relatiesterkte onder- of overschatten, en zijn alternatieven zoals Spearman-correlatie geschikter.”
4. Wanneer de correlatie statistisch significant is  
   “4” = ” ❌ Fout. Statistische significantie betekent dat de waargenomen correlatie waarschijnlijk niet door toeval ontstaan is - dit maakt het juist betrouwbaarder, niet misleidender. Echter, statistische significantie zegt niets over praktische significantie. In zeer grote steekproeven kunnen zelfs zeer kleine correlaties (bijv. r = 0,05) statistisch significant zijn maar praktisch betekenisloos. Het is belangrijk om zowel statistische als praktische significantie te overwegen bij interpretatie.”

## TOEPASSEN NIVEAU (3 vragen)

### Vraag Q14 (Toepassen)

**Een onderzoeker onderzoekt de relatie tussen stress en agressief gedrag bij jongeren. De data toont een gebogen patroon: stress verhoogt agressie aanvankelijk snel, maar bij zeer hoge stressniveaus vlakt de toename af (een verzadigingseffect). Het patroon is monotoon toenemend (altijd omhoog gaand, nooit omlaag) maar niet lineair (geen rechte lijn). Welke correlatiecoëfficiënt moet de onderzoeker gebruiken?**



1. Beide correlatiecoëfficiënten zijn even geschikt voor deze situatie  
   “1” = ” ❌ Fout. Pearson en Spearman meten verschillende soorten relaties en geven daarom verschillende resultaten bij niet-lineaire (gebogen) patronen. Pearson correlatie (r = 0,85 in dit voorbeeld) veronderstelt een lineaire (rechte lijn) relatie en onderschat de werkelijke sterkte van de monotone (altijd stijgende) relatie. Spearman correlatie (ρ = 0,96) vangt de consistente rangorde-relatie beter op. In criminologisch onderzoek is dit onderscheid cruciaal - bijvoorbeeld, de relatie tussen leeftijd en bepaalde delicten is vaak gebogen (hoog in adolescentie, afnemend daarna) maar wel monotoon (consistent in één richting) in bepaalde leeftijdsbereiken. Het gebruik van de verkeerde correlatiecoëfficiënt kan leiden tot onderrapportage van belangrijke relaties.”
2. Geen van beide correlatiecoëfficiënten is geschikt voor deze situatie  
   “2” = ” ❌ Fout. Spearman correlatie is juist specifiek ontworpen voor monotone (altijd dezelfde richting) maar niet-lineaire (gebogen) relaties zoals deze. Hoewel Pearson inderdaad minder geschikt is, biedt Spearman een uitstekende oplossing door rangorden te gebruiken in plaats van de werkelijke waarden. Deze aanpak maakt het immuun voor de specifieke vorm van de curve, zolang de relatie maar consistent toenemend of afnemend is. In dit stress-agressie voorbeeld zou Spearman de sterke monotone relatie correct identificeren, terwijl een benadering van ‘geen van beide’ waardevolle informatie over de relatie zou missen.”
3. Pearson correlatie omdat het de standaard correlatiecoëfficiënt is  
   “3” = ” ❌ Fout. Pearson correlatie is inderdaad de meest gebruikte, maar dat maakt het niet automatisch geschikt voor alle situaties. Pearson meet specifiek lineaire (rechte lijn) associaties en zal het gebogen maar monotone (altijd stijgende) patroon onderschatten. In het voorbeeld hierboven geeft Pearson r = 0,85, wat een sterke relatie suggereert, maar mist de werkelijke sterkte van de monotone associatie. Het feit dat stress consistent agressie verhoogt (monotoon patroon) wordt beter vastgelegd door Spearman ρ = 0,96. In criminologisch onderzoek kunnen vele relaties niet-lineair (gebogen) zijn - het automatisch gebruiken van Pearson kan leiden tot het missen van belangrijke patronen in data.”
4. Spearman correlatie omdat het monotone relaties kan meten  
   “4” = ” ✅ Juist! Spearman correlatie is specifiek ontworpen voor monotone (altijd dezelfde richting) relaties, ongeacht of ze lineair (rechte lijn) zijn of niet. Het werkt door variabelen om te zetten naar rangorden en vervolgens de correlatie tussen deze rangorden te berekenen. Hierdoor kan het gebogen patronen vastleggen zolang ze maar consistent in één richting gaan. In dit stress-agressie voorbeeld geeft Spearman ρ = 0,96, wat de sterke monotone relatie correct weergeeft, terwijl Pearson r = 0,85 de relatie onderschat omdat het probeert een rechte lijn door een gebogen patroon te trekken. In criminologische context zijn monotone maar niet-lineaire relaties veel voorkomend - bijvoorbeeld tussen sociale cohesie en criminaliteitscijfers, of tussen leeftijd en bepaalde overtredingspatronen.”

### Vraag Q15 (Toepassen)

**Een criminoloog onderzoekt de relatie tussen sociale binding (attachment) van jongeren met ouders en delinquent gedrag. Ze maakt spreidingsdiagrammen voor verschillende groepen jongeren. In één specifiek diagram ziet ze: een rechte trendlijn die omhoog helt; punten die strak geclusterd zijn rond de lijn met minimale spreiding. Welke beschrijving past het best bij dit patroon?**

**Hint:** Combineer richting (teken) met sterkte (spreiding).



**Hint:** Combineer richting (omhoog = positief) met sterkte (strakke clustering = sterk). Wat vertelt de minimale spreiding rond de lijn?

1. Sterk positief  
   “1” = ” ✅ Juist! Een strakke wolk van punten die dicht geclusterd zijn rond een opwaarts hellende lijn is de visuele kenmerk van een sterke positieve correlatie (typisch r > 0,7). Dit patroon geeft zowel de richting van de relatie aan (positief — als sociale binding toeneemt, neemt delinquent gedrag betrouwbaar toe) als de sterkte ervan (sterk — punten houden zich dicht aan het lineaire patroon met minimale spreiding). De visualisatie toont minimale afwijking van de trendlijn, wat suggereert een zeer voorspelbare relatie waarbij het kennen van de sociale binding-score nauwkeurige voorspelling van delinquent gedrag mogelijk maakt. In dit criminologische voorbeeld zou zo’n sterk patroon kunnen wijzen op een fundamentele relatie waarbij sterke ouder-kind binding paradoxaal samenhangt met meer delinquentie, met minimale invloed van andere variabelen.”
2. Matig positief  
   “2” = ” ❌ Fout. Het beschreven patroon is veel strakker dan ‘matig’ zou suggereren. Bij een matige positieve correlatie (r ≈ 0,3-0,7) zou je meer spreiding rond de trendlijn zien, met punten die verder van de lijn afliggen. De beschrijving van ‘strak geclusterde punten met minimale spreiding’ duidt op een veel sterkere relatie. In criminologisch onderzoek zou een matige relatie betekenen dat sociale binding wel belangrijk is voor delinquentie-niveaus, maar dat er substantiële variatie bestaat die niet door deze ene factor wordt verklaard. Het patroon in de vraag suggereert echter een veel meer deterministische relatie.”
3. Zwak negatief  
   “3” = ” ❌ Fout. Deze optie mist zowel de richting als de sterkte. De richting is duidelijk positief (omhoog hellende lijn), niet negatief. Bovendien beschrijft de vraag een strak patroon, niet een zwak patroon. Een zwakke negatieve correlatie zou een neerwaarts hellende lijn tonen met veel spreiding rond de lijn. In termen van sociale binding en delinquentie zou een negatieve relatie betekenen dat sterkere binding samengaat met minder delinquentie, wat eigenlijk meer theoretisch logisch zou zijn volgens Hirschi’s sociale controle theorie dan de positieve relatie beschreven in dit voorbeeld.”
4. Geen lineair patroon  
   “4” = ” ❌ Fout. De beschrijving wijst juist op een zeer duidelijk lineair patroon. Punten die ‘strak geclusterd zijn rond een rechte trendlijn’ vormen precies de definitie van een sterk lineair patroon. ‘Geen lineair patroon’ zou betekenen dat de punten willekeurig verspreid zijn zonder enige duidelijke trend, of dat ze een gebogen (curvilineair) patroon volgen. In criminologisch onderzoek zou het ontbreken van een lineair patroon suggereren dat er geen directe, voorspelbare relatie bestaat tussen sociale binding en delinquentie, wat in tegenspraak is met de duidelijke trendlijn beschreven in de vraag.”

### Vraag Q16 (Toepassen)

**Een onderzoeker onderzoekt de relatie tussen ouderlijk toezicht en jeugdcriminaliteit in verschillende wijken. Na het berekenen van de correlatiecoëfficiënt, beschrijft ze de relatie als ‘zwak positief’. Welke r-waarde past het best bij deze beschrijving?** > **Hint:** Gebruik algemene interpretatierichtlijnen voor correlatiesterkte. Wat betekent ‘zwak’ in numerieke termen?

1. r = 0,12  
   “1” = ” ✅ Juist! **Interpretatierichtlijnen:** Zwak (|r| = 0,10-0,30), Matig (|r| = 0,30-0,70), Sterk (|r| = 0,70-1,00); Positief (r > 0), Negatief (r < 0). Deze waarde (r = 0,12) valt duidelijk in de categorie ‘zwakke positieve correlatie’ volgens standaard interpretatierichtlijnen. Deze correlatie betekent dat er een kleine maar meetbare relatie bestaat tussen ouderlijk toezicht en jeugdcriminaliteit. De variabelen delen slechts een klein deel van hun variantie (R² = 0,12² = 0,014 = 1,4%), wat betekent dat 98,6% van de variantie in jeugdcriminaliteit niet verklaard wordt door ouderlijk toezicht alleen. In de praktijk zou dit suggereren dat terwijl ouderlijk toezicht wel enige invloed heeft op jeugdcriminaliteit, vele andere factoren (zoals peers, schoolprestaties, buurtfactoren) veel belangrijkere rollen spelen. Bij zwakke correlaties is het essentieel om te overwegen of de relatie statistisch significant en praktisch betekenisvol is.”
2. r = 0,56  
   “2” = ” ❌ Fout. **Interpretatierichtlijnen:** Zwak (|r| = 0,10-0,30), Matig (|r| = 0,30-0,70), Sterk (|r| = 0,70-1,00); Positief (r > 0), Negatief (r < 0). Deze waarde (r = 0,56) valt duidelijk in de categorie ‘matige correlatie’ (0,30-0,70), niet ‘zwak’. Een correlatie van 0,56 zou een duidelijk sterkere relatie aangeven tussen ouderlijk toezicht en jeugdcriminaliteit, waarbij ongeveer 31% van de variantie gedeeld wordt (R² = 0,56² = 0,31). Dit zou suggereren dat ouderlijk toezicht een substantiële, niet slechts zwakke, voorspeller is van jeugdcriminaliteit. In onderzoekstermen zou je deze relatie beschrijven als ‘matig sterk’ in plaats van ‘zwak’, en het zou duiden op een veel meer betekenisvolle praktische relatie dan wat de onderzoeker beschreef.”
3. r = −0,72  
   “3” = ” ❌ Fout. **Interpretatierichtlijnen:** Zwak (|r| = 0,10-0,30), Matig (|r| = 0,30-0,70), Sterk (|r| = 0,70-1,00); Positief (r > 0), Negatief (r < 0). Deze waarde heeft twee fundamentele problemen: (1) De richting is negatief, niet positief - een negatieve correlatie zou betekenen dat meer ouderlijk toezicht samengaat met minder jeugdcriminaliteit, wat theoretisch logischer zou zijn; (2) De absolute sterkte (|-0,72| = 0,72) valt in de categorie ‘sterk’ (>0,70), niet ‘zwak’. Een correlatie van -0,72 zou duiden op een sterke omgekeerde relatie waarbij ouderlijk toezicht een krachtige beschermende factor is tegen jeugdcriminaliteit. Dit zou ongeveer 52% gedeelde variantie betekenen (R² = 0,72²), wat veel sterker is dan de beschreven ‘zwakke’ relatie.”
4. r = 0,93  
   “4” = ” ❌ Fout. **Interpretatierichtlijnen:** Zwak (|r| = 0,10-0,30), Matig (|r| = 0,30-0,70), Sterk (|r| = 0,70-1,00); Positief (r > 0), Negatief (r < 0). Dit is een extreem sterke positieve correlatie (r = 0,93), ver boven de sterke categorie (>0,70), wat het tegenovergestelde is van ‘zwak’. Een correlatie van 0,93 zou betekenen dat ouderlijk toezicht en jeugdcriminaliteit een vrijwel perfecte positieve relatie hebben, waarbij ongeveer 86% van de variantie gedeeld wordt (R² = 0,93² = 0,86). In de praktijk zou dit een bijna deterministische relatie betekenen waarbij ouderlijk toezicht de criminaliteit van jongeren vrijwel perfect voorspelt. Zulke extreme correlaties zijn zeer zeldzaam in sociaal wetenschappelijk onderzoek, en zeker niet wat bedoeld wordt met ‘zwak positief’.”

## ANALYSEREN NIVEAU (4 vragen)

### Vraag Q17 (Begrijpen)

**Welke drie belangrijkste types correlatiecoëfficiënten worden gebruikt in de statistiek en wat zijn hun hoofdkenmerken?**

**Hint:** Denk aan datatypes (interval vs ordinaal) en vorm van de relatie (lineair vs monotoon).

1. Pearson, Spearman, en Kendall - elk geschikt voor verschillende datatypes en relatie-vormen  
   “1” = ” ✅ Juist! Deze drie correlatiecoëfficiënten zijn inderdaad de belangrijkste en elk heeft specifieke toepassingen. Pearson’s r is de meest gebruikte en krachtigste correlatie voor continue data met lineaire relaties - het meet de sterkte van lineaire samenhang tussen variabelen. Spearman’s ρ (rho) gebruikt rangnummers in plaats van werkelijke waarden, waardoor het robuust is voor uitschieters en geschikt voor ordinale data of niet-lineaire maar monotone relaties. Kendall’s τ (tau) is gebaseerd op concordante en discordante paren en werkt bijzonder goed met kleine steekproeven of data met veel gelijke waarden. In onderzoek zou je Pearson gebruiken voor relaties zoals inkomen vs uitgaven, Spearman voor tevredenheidsscores vs gedragsratings, en Kendall voor kleine steekproeven met veel gelijke waarden. Het kiezen van de juiste correlatie hangt af van je datatype en de vorm van de relatie.”
2. Alleen Pearson en Spearman zijn belangrijk; Kendall wordt zelden gebruikt  
   “2” = ” ❌ Fout. Hoewel Pearson en Spearman inderdaad veel gebruikt worden, is Kendall’s τ (tau) een belangrijke derde optie met specifieke voordelen. Kendall’s correlatie is bijzonder waardevol in situaties met kleine steekproeven (n < 30), data met veel gelijke waarden (ties), of wanneer je een interpretatie wilt in termen van waarschijnlijkheid van concordantie. In onderzoek komen deze situaties regelmatig voor - bijvoorbeeld bij het analyseren van gevangenisratings of kleine experimentele groepen. Bovendien heeft Kendall’s τ betere statistische eigenschappen in sommige gevallen en wordt het vaak gebruikt in robuuste statistiek. Het negeren van deze derde optie zou betekenen dat je een waardevolle analytische tool mist die in bepaalde onderzoekssituaties superieur kan zijn aan de andere twee methoden.”
3. Er zijn vele correlatie types, maar de drie hoofdgroepen zijn positief, negatief, en geen correlatie  
   “3” = ” ❌ Fout. Dit antwoord verwarrt correlatie-richtingen (positief, negatief, geen) met correlatie-types (rekenmethoden). Positief, negatief, en geen correlatie beschrijven de richting en sterkte van relaties, niet de statistische methoden die gebruikt worden om ze te berekenen. Dit is een veelvoorkomende conceptuele verwarring bij studenten. De vraag gaat specifiek over de verschillende statistische methoden (Pearson, Spearman, Kendall) die gebruikt worden om correlaties te berekenen, elk met hun eigen wiskundige benadering en geschiktheid voor verschillende datatypes. In onderzoek is het cruciaal om zowel de richting van relaties te begrijpen (positief/negatief) als de juiste statistische methode te kiezen voor je specifieke datatype en onderzoeksvraag.”
4. Correlatie heeft geen verschillende types - alle correlaties worden op dezelfde manier berekend  
   “4” = ” ❌ Fout. Dit misverstand kan leiden tot ernstige analytische fouten in onderzoek. Er zijn inderdaad verschillende types correlatiecoëfficiënten met fundamenteel verschillende berekeningsmethoden en toepassingsgebieden. Zoals getoond in de vergelijkingstabel, berekent Pearson correlatie met werkelijke waarden en veronderstelt lineaire relaties, terwijl Spearman rangnummers gebruikt en alleen monotone relaties vereist, en Kendall werkt met concordante paren. Deze verschillen zijn niet alleen academisch - ze kunnen tot verschillende conclusies leiden met dezelfde data. In onderzoek zou het gebruik van de verkeerde correlatiemethode kunnen leiden tot het missen van belangrijke relaties of het trekken van ongeldige conclusies. Bijvoorbeeld, als je Pearson gebruikt voor sterk scheef verdeelde data terwijl Spearman geschikter zou zijn, kun je zwakkere correlaties vinden dan werkelijk aanwezig zijn.”

### Vraag Q18 (Begrijpen)

**Een criminoloog heeft verschillende datasets en moet de juiste correlatiecoëfficiënt kiezen. Voor welke situatie zou Spearman’s rangcorrelatie (ρ) beter zijn dan Pearson’s correlatie (r)?**



**Hint:** Denk aan wanneer rangnummers beter zijn dan werkelijke waarden.

1. Spearman is altijd beter dan Pearson en zou altijd gebruikt moeten worden  
   “1” = ” ❌ Fout. Dit misverstand zou leiden tot suboptimale analytische keuzes. Pearson correlatie heeft belangrijke voordelen wanneer aan zijn veronderstellingen wordt voldaan: het is krachtiger (heeft meer statistische kracht), geeft interpreteerbare coëfficiënten voor lineaire relaties, en is de basis voor vele andere statistische procedures zoals regressieanalyse. Spearman is specifiek nuttig in bepaalde situaties (niet-lineaire monotone relaties, uitschieters, ordinale data) maar verliest informatie door omzetting naar rangen. In onderzoek zou het blind gebruiken van Spearman voor alle analyses betekenen dat je kracht verliest bij echte lineaire relaties en mogelijk verkeerde interpretaties maakt. De kunst van goede statistiek ligt in het kiezen van de juiste methode voor je specifieke data en onderzoeksvraag, niet in het altijd dezelfde methode gebruiken.”
2. Alleen wanneer je minder dan 10 observaties hebt  
   “2” = ” ❌ Fout. Steekproefgrootte bepaalt niet direct de keuze tussen Pearson en Spearman. Beide correlatiemethoden kunnen gebruikt worden met kleine steekproeven, hoewel hun statistische betrouwbaarheid afneemt. De keuze tussen Pearson en Spearman hangt af van de aard van je data (lineair vs monotoon, aanwezigheid van uitschieters, datatype) niet van de steekproefgrootte. Voor zeer kleine steekproeven zou je Kendall’s τ (tau) kunnen overwegen omdat het betere eigenschappen heeft bij kleine N, maar dit is een andere correlatiemethode. In onderzoek met kleine steekproeven (bijvoorbeeld, interviews met 8 deelnemers) zou je nog steeds kiezen tussen Pearson en Spearman gebaseerd op de vorm van je relatie en datatype, niet alleen op de grootte.”
3. Wanneer er een monotone maar niet-lineaire relatie is, of er extreme uitschieters zijn  
   “3” = ” ✅ Juist! Spearman’s rangcorrelatie is superieur in beide situaties. Bij niet-lineaire maar monotone relaties (panel B) vangt Spearman de sterke monotone trend beter op (ρ = 0,943) dan Pearson (r = 0,878) omdat het rangnummers gebruikt in plaats van werkelijke waarden. Bij extreme uitschieters (panel C) is Spearman veel robuuster (ρ = 0,785) vergeleken met Pearson (r = 0,538) omdat rangnummers de invloed van extreme waarden minimaliseren. In onderzoek kom je regelmatig dergelijke situaties tegen - bijvoorbeeld de relatie tussen sociaaleconomische status en criminaliteit kan niet-lineair zijn (drempeleffecten), en datasets kunnen extreme gevallen bevatten die Pearson correlatie verstoren. Door rangnummers te gebruiken geeft Spearman een betrouwbaardere maat van de onderliggende monotone relatie.”
4. Wanneer de data een perfecte lineaire relatie toont  
   “4” = ” ❌ Fout. Bij perfecte lineaire relaties is Pearson’s r juist superieur aan Spearman’s ρ. Pearson correlatie is specifiek ontworpen voor lineaire relaties en zal in dit geval de maximale kracht hebben en de meest nauwkeurige schatting geven. Zoals getoond in panel A van de visualisatie, zijn beide correlaties vergelijkbaar hoog bij lineaire relaties (Pearson r = 0,961, Spearman ρ = 0,959), maar Pearson heeft de voorkeur omdat het gebruik maakt van de werkelijke waarden in plaats van rangnummers. Spearman wordt juist gebruikt wanneer lineaire veronderstellingen geschonden worden. In onderzoek zou je Pearson kiezen voor relaties zoals ‘uren studie vs. testscores’ waar een rechtlijnige relatie verwacht wordt, en Spearman reserveren voor situaties waar die lineariteit in twijfel wordt getrokken.”

### Vraag Q19 (Analyseren)

**In een positieve trend, welke uitschieter (extreme waarde die afwijkt van het algemene patroon) positie vermindert r het meest?**



**Hint:** Denk aan punten die het lineaire patroon doorbreken.

1. Ver links‑onder (lage X, zeer lage Y)  
   “1” = ” ❌ Fout. Punten links-onder (lage X, lage Y) versterken juist positieve correlaties. **Impact van uitschieters:** Deze positie volgt de algemene trend, waardoor de correlatie sterker wordt. De formule voor Pearson’s r behelst het vermenigvuldigen van afwijkingen van het gemiddelde voor zowel X als Y. Wanneer X onder zijn gemiddelde ligt EN Y onder zijn gemiddelde ligt, is het product positief, wat positief bijdraagt aan de correlatie. De visualisatie toont dit duidelijk — het toevoegen van een laag-laag uitschieter verhoogt r van 0,82 naar 0,94. **Praktische impact:** In criminologisch onderzoek zou dit bijvoorbeeld een wijk met zowel lage sociale cohesie als hoge criminaliteitscijfers vertegenwoordigen - dit versterkt het verband tussen beide variabelen. Deze gevallen bevestigen het verwachte patroon en maken de correlatie sterker, niet zwakker.”
2. Ver rechts‑onder (hoge X, lage Y)  
   “2” = ” ✅ Juist! Een punt in het rechts-onder kwadrant (hoge X, lage Y) vermindert een positieve correlatie maximaal. **Impact van uitschieters:** Deze positie contradiceert de algemene trend en verzwakt de correlatie dramatisch. Deze positie vertegenwoordigt een ernstige schending van het positieve trendpatroon, waarbij X ver boven zijn gemiddelde ligt terwijl Y ver onder zijn gemiddelde ligt. In de correlatieformule creëert dit een grote negatieve productterm die de positieve relatie tegengaat die in de rest van de gegevens wordt gezien. **Kwantitatieve impact:** Zoals getoond in de visualisatie, vermindert deze positie de correlatiecoëfficiënt dramatisch van r = 0,82 naar r = 0,28 - een verlies van ongeveer 65% van de correlatiestrekte! **Praktische betekenis:** In criminologisch onderzoek zou dit bijvoorbeeld een wijk met hoge sociale cohesie maar onverwacht hoge criminaliteit vertegenwoordigen - dit doorbreekt het verwachte verband en suggereert andere verklarende factoren.”
3. Ver rechts‑boven (hoge X, hoge Y)  
   “3” = ” ❌ Fout. Een punt rechts-boven (hoge X, hoge Y) versterkt juist een positieve correlatie in plaats van deze te verminderen. **Impact van uitschieters:** Deze positie versterkt de algemene trend en maakt de correlatie sterker. Wanneer X en Y beide boven hun respectievelijke gemiddelden liggen, vermenigvuldigen hun afwijkingen zich tot een positieve bijdrage aan de correlatiecoëfficiënt. **Kwantitatieve impact:** De visualisatie bevestigt dit — het toevoegen van een hoog-hoog uitschieter verhoogt r van 0,82 naar 0,94, een versterking van ongeveer 15%. **Praktische betekenis:** In criminologisch onderzoek zou dit bijvoorbeeld een wijk met zeer hoge sociale cohesie en zeer lage criminaliteit vertegenwoordigen - dit versterkt het verwachte verband. Hoewel deze punten invloedrijk kunnen zijn en onderzoek verdienen (waarom presteren ze zo goed?), ondersteunen ze het correlatiepatroon in plaats van het te ondermijnen.”
4. Een punt precies op de lijn  
   “4” = ” ❌ Fout. Een punt dat precies op de regressielijn valt heeft bijna geen impact op de correlatiecoëfficiënt. **Impact van uitschieters:** Punten op de lijn hebben minimale tot geen invloed op de correlatiesterkte omdat ze perfect conform zijn aan de lineaire relatie. Deze positie vertegenwoordigt geen uitschieter - het is juist een punt dat het voorspelde patroon perfect volgt. **Kwantitatieve impact:** Dergelijke punten hebben nul of minimale afwijking van het voorspelde model, dus ze veranderen r nauwelijks. **Conceptuele verwarring:** Dit antwoord toont een misverstand over wat constitueert als een ‘uitschieter’. Uitschieters zijn per definitie punten die substantial afwijken van het algemene patroon. **Praktische betekenis:** In onderzoek zijn dit de ‘voorspelbare’ gevallen die exact doen wat het model verwacht - ze bevestigen noch weerleggen de relatie.”

### Vraag Q20 (Begrijpen)

**Waarom gebruikt Pearson’s correlatie gestandaardiseerde waarden (z-scores) in plaats van de oorspronkelijke meetwaarden? Een criminoloog vergelijkt inkomen (in euro’s) met het aantal misdrijven per buurt.**

*Begrip ‘standardisatie’: Standardisatie is het proces waarbij verschillende variabelen worden omgezet naar een gemeenschappelijke schaal. Dit gebeurt door z-scores te berekenen: z = (waarde - gemiddelde) / standaardafwijking. Hierdoor krijgen alle variabelen een gemiddelde van 0 en standaardafwijking van 1, ongeacht hun oorspronkelijke eenheden.* Standardisatie is cruciaal voor het vergelijken van variabelen om betekenisvol te zijn

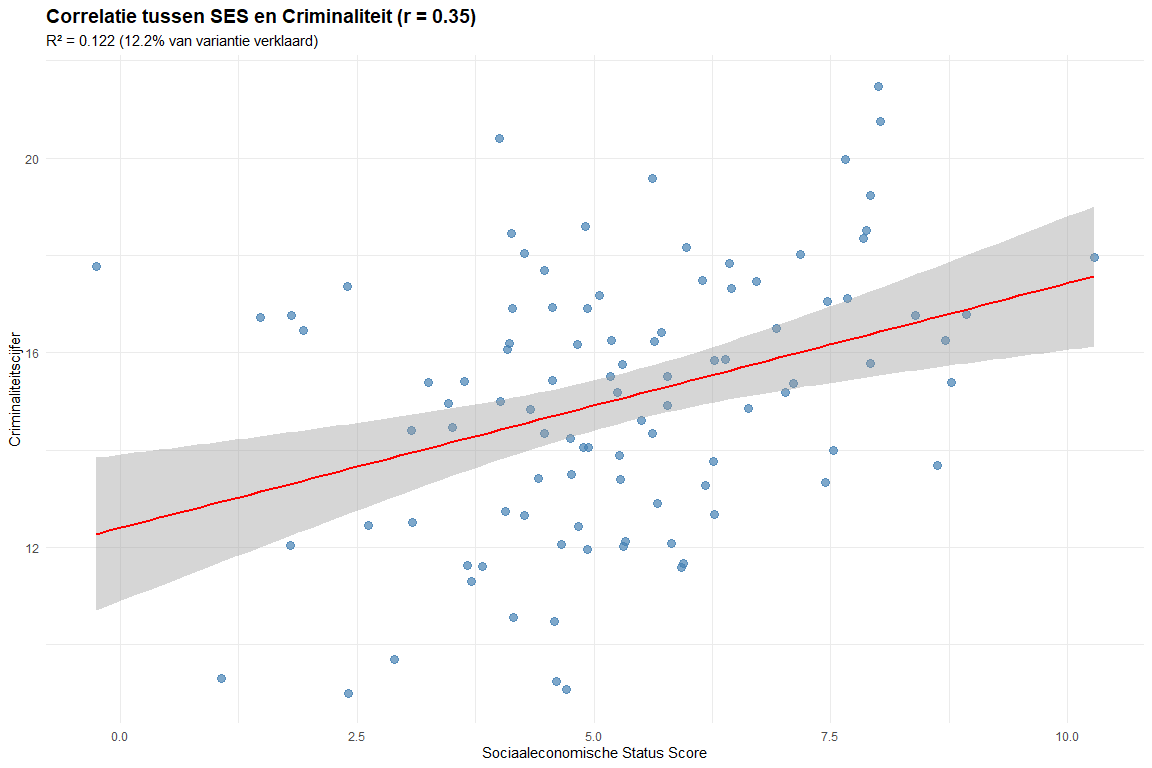
**Hint:** Denk aan wat er gebeurt als één variabele in euro’s en de andere in aantal misdrijven wordt gemeten.

1. Om uitschieters te elimineren en de data normaler te maken  
   “1” = ” ❌ Fout. Standardisatie elimineert geen uitschieters noch maakt het data normaler verdeeld. Z-score transformatie is een lineaire bewerking die de vorm van de verdeling volledig behoudt - als je data scheef verdeeld was of uitschieters had vóór standardisatie, dan blijft dit zo ná standardisatie, alleen op een andere schaal. De visualisatie toont dat alle datapunten dezelfde relatieve posities behouden. Als uitschieters bestonden in de oorspronkelijke data, zullen ze nog steeds uitschieters zijn in de gestandaardiseerde data (alleen uitgedrukt als extreme z-scores). Voor het behandelen van uitschieters zijn andere methoden nodig zoals uitschietertransformaties, robuuste correlaties (Spearman), of uitschietersverwijdering. Standardisatie lost deze problemen niet op - het doel is alleen schaalvergelijkbaarheid.”
2. Om de correlatie te versterken en statistisch significanter te maken  
   “2” = ” ❌ Fout. Standardisatie heeft geen effect op de sterkte of statistische significantie van correlaties. Zoals getoond in de visualisatie hebben zowel de oorspronkelijke data als de gestandaardiseerde data precies dezelfde correlatiecoëfficiënt (r = -0,789). Standardisatie is een lineaire transformatie die alleen de schaal verandert maar niet de onderliggende relaties tussen variabelen. Het doel is niet om correlaties te verbeteren of te versterken, maar om wiskundige berekenbaarheid mogelijk te maken wanneer variabelen verschillende eenheden hebben. In onderzoek zou het verkeerd zijn om standardisatie te gebruiken om correlaties ‘beter’ te laten lijken - het is simpelweg een technische noodzaak voor juiste berekening wanneer verschillende meetschalen worden vergeleken.”
3. Om ervoor te zorgen dat beide variabelen positief zijn voor de berekening  
   “3” = ” ❌ Fout. Correlatie vereist niet dat variabelen positief zijn, en standardisatie maakt variabelen niet automatisch positief. Integendeel, z-scores kunnen zowel positief als negatief zijn, afhankelijk van of de oorspronkelijke waarde boven of onder het gemiddelde ligt. In dit voorbeeld zullen buurten met inkomen onder het gemiddelde negatieve z-scores hebben, terwijl buurten met inkomen boven het gemiddelde positieve z-scores krijgen. Hetzelfde geldt voor misdrijfcijfers. Correlatie kan perfect worden berekend met negatieve waarden - de formule werkt met produkten van afwijkingen, en negatieve waarden dragen gewoon bij aan de berekening. Het feit dat gestandaardiseerde data een gemiddelde van nul hebben (dus zowel positieve als negatieve waarden) is geen probleem maar een kenmerk van de transformatie.”
4. Om variabelen met verschillende eenheden en schalen vergelijkbaar te maken in één correlatiemaat  
   “4” = ” ✅ Juist! Dit is de fundamentele reden voor standardisatie in correlatie. Zonder standardisatie zou het onmogelijk zijn om zinvolle correlaties te berekenen tussen variabelen met verschillende eenheden. In dit voorbeeld heeft inkomen waarden tussen €20.000-€80.000, terwijl misdrijven waarden tussen 10-80 per 1000 inwoners hebben. Deze totaal verschillende schalen maken directe vergelijking onmogelijk. Door beide variabelen om te zetten naar z-scores (met gemiddelde 0 en standaardafwijking 1) kunnen we hun lineaire samenhang meten ongeacht hun oorspronkelijke eenheden. Dit principe stelt criminologen in staat om correlaties te berekenen tussen diverse variabelen zoals inkomen (euro’s), leeftijd (jaren), criminaliteitsscores (punten), woningprijzen (euro’s), etc. Standardisatie zorgt ervoor dat correlatie echt de relatiesterkte meet, niet de willekeurige keuze van meeteenheden.”

## EVALUEREN NIVEAU (3 vragen)

### Vraag Q21 (Begrijpen)

**Een criminoloog rapporteert een correlatie van r = 0,40 tussen sociaaleconomische status en criminaliteitscijfers. Wat is de meest accurate interpretatie van deze correlatie?**



**Hint:** Denk aan richting, sterkte, en hoeveel variantie wordt verklaard.

1. Een sterke positieve relatie waarbij SES 40% van criminaliteit verklaart  
   “1” = ” ❌ Fout. Deze interpretatie maakt twee fouten. Ten eerste is r = 0,40 een matige (niet sterke) correlatie volgens standaard richtlijnen waarbij correlaties van 0,40-0,59 als matig worden beschouwd. Ten tweede verklaart SES niet 40% van criminaliteit - de verklaarde variantie wordt berekend als R² = r², dus R² = 0,40² = 0,16 (16%). In criminologisch onderzoek is het cruciaal om accurate statistische interpretaties te bieden omdat beleidsmakers deze informatie gebruiken voor resource-allocatie. Een r van 0,40 duidt op een nuttige maar beperkte relatie waarbij vele andere factoren (zoals buurtkenmerken, gezinsstructuur, peer invloeden) de overige 84% van de variantie in criminaliteitscijfers verklaren.”
2. Een matige positieve relatie waarbij SES en criminaliteit 16% gedeelde variantie hebben  
   “2” = ” ✅ Juist! Een correlatie van r = 0,40 wordt geclassificeerd als matig sterk en verklaart R² = 0,16 (16%) van de variantie. Dit betekent dat variaties in sociaaleconomische status ongeveer 16% van de waargenomen variatie in criminaliteitscijfers kunnen verklaren, terwijl andere factoren de resterende 84% verklaren. In criminologische termen suggereert dit dat SES een betekenisvolle maar beperkte voorspeller is van criminaliteit. De positieve richting duidt erop dat hogere SES geassocieerd is met hogere criminaliteitscijfers (wat tegen-intuïtief kan lijken maar afhankelijk is van hoe criminaliteit wordt gemeten - bijvoorbeeld, witteboordencriminaliteit vs. straatcriminaliteit). Een matige correlatie biedt nuttige voorspellende waarde voor beleid en interventies maar moet worden aangevuld met andere risicofactoren voor een volledig begrip.”
3. Een zwakke relatie die statistisch niet betekenisvol is  
   “3” = ” ❌ Fout. Een correlatie van r = 0,40 is niet zwak - het valt in de matige categorie volgens Cohen’s richtlijnen waarbij correlaties van 0,20-0,39 als zwak en 0,40-0,59 als matig worden beschouwd. Bovendien impliceert de vraag geen informatie over statistische significantie (p-waarden), dus we kunnen geen uitspraak doen over statistische betekenis. In veel criminologische steekproeven zou een correlatie van 0,40 waarschijnlijk statistisch significant zijn, vooral met redelijke steekproefgroottes. Het is belangrijk om sterkte (effect size) te onderscheiden van statistische significantie - zelfs relatief kleine correlaties kunnen statistisch significant zijn met grote steekproeven, terwijl grote correlaties niet-significant kunnen zijn in zeer kleine steekproeven.”
4. Een perfecte relatie die causaliteit bewijst tussen SES en criminaliteit  
   “4” = ” ❌ Fout. Een correlatie van r = 0,40 is verre van perfect (perfect zou r = 1,00 zijn) en correlatie bewijst nooit causaliteit, ongeacht sterkte. Deze interpretatie maakt fundamentele statistische fouten. Ten eerste toont r = 0,40 een matige relatie met substantiële variabiliteit rond de trendlijn. Ten tweede kan correlatie, zelfs indien perfect, geen causale richting vaststellen - de relatie zou kunnen zijn dat (a) SES criminaliteit beïnvloedt, (b) criminaliteit SES beïnvloedt, of (c) een derde variabele beide beïnvloedt. In criminologisch onderzoek vereist het vaststellen van causaliteit geschikte onderzoeksontwerpen zoals experimentele of quasi-experimentele studies, longitudinale data, of geavanceerde statistische methoden die verstorende variabelen controleren.”

### Vraag Q22 (Evalueren)

**Een nieuwsartikel beweert: “Sterke correlatie (r = 0,65) tussen buurtwaakprogramma’s en verminderde inbraken bewijst dat deze programma’s zeer effectieve misdaadpreventie instrumenten zijn.” Evalueer deze statistische claim.**

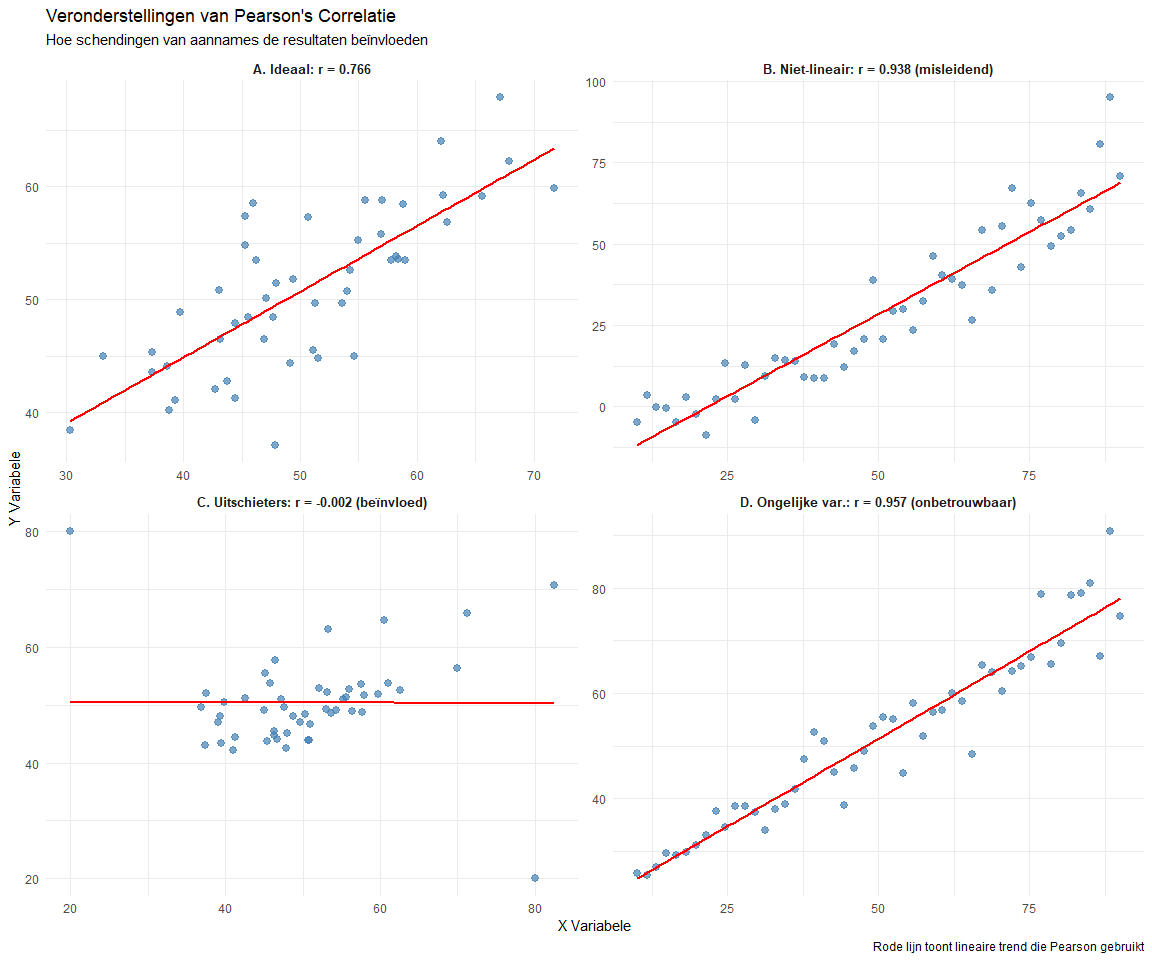
**Hint:** Overweeg correlatie vs. causaliteit, verstorende variabelen en onderzoeksontwerp beperkingen.

1. Ongeldig - correlatie bewijst geen oorzakelijk verband; buurtkenmerken kunnen beide variabelen verklaren  
   “1” = ” ✅ Juist! Verstorende factoren zoals sociaaleconomische status kunnen zowel programmaanwezigheid als misdaadcijfers beïnvloeden. Dit is een klassiek voorbeeld van potentiële verstoring in criminologisch onderzoek. Rijkere buurten hebben vaak zowel meer middelen om waakprogramma’s op te zetten ALS lagere basislijn misdaadcijfers. Andere mogelijke verstorende factoren zijn: bevolkingsdichtheid, residentiële stabiliteit, politieaanwezigheid, of architecturale kenmerken. Zonder controle voor deze variabelen met methoden zoals meervoudige regressie of propensity score matching, kunnen we misdaadvermindering niet toeschrijven aan de waakprogramma’s. Bovendien zou de richting van causaliteit omgekeerd kunnen zijn — buurten met lagere misdaadcijfers hebben misschien meer bewoners bereid om deel te nemen aan waakprogramma’s, in plaats van dat de programma’s de lagere criminaliteit veroorzaken.”
2. Geldige conclusie - sterke correlatie biedt voldoende bewijs van effectiviteit  
   “2” = ” ❌ Fout. Correlatie alleen kan causale effectiviteit niet vaststellen. Zelfs een sterke correlatiecoëfficiënt (r = 0,65) vertelt ons alleen dat de variabelen geassocieerd zijn, niet dat de ene de andere veroorzaakt. Bijvoorbeeld, buurten met sterkere gemeenschapscohesie kunnen onafhankelijk zowel meer buurtwacht participatie ALS lagere misdaadcijfers hebben vanwege informele sociale controle. In de criminologie vereisen beleidsaanbevelingen sterker causaal bewijs dan louter correlatie, typisch uit experimentele of quasi-experimentele ontwerpen (zoals gerandomiseerde gecontroleerde trials of regressie discontinuïteit). Het nieuwsartikel maakt een fundamentele fout in statistische interpretatie die zou kunnen leiden tot verkeerd toegewezen misdaadpreventie middelen.”
3. Ongeldig - r = 0,65 is te zwak om conclusies te ondersteunen  
   “3” = ” ❌ Fout. De correlatiesterkte is niet het hoofdprobleem; causaliteit vs. correlatie wel. In feite zou r = 0,65 over het algemeen als een sterke correlatie worden beschouwd in criminologisch onderzoek. Het fundamentele probleem met de claim van het nieuwsartikel is niet dat de correlatie te zwak is, maar dat correlatie — ongeacht sterkte — geen causaliteit kan vaststellen. Zelfs een perfecte correlatie (r = 1,00) zou niet bewijzen dat buurtwachtprogramma’s misdaadvermindering veroorzaken. Evidence-based criminologie vereist onderscheid maken tussen correlatie en causaliteit door middel van geschikte onderzoeksmethodologie, inclusief gerandomiseerde trials, natuurlijke experimenten, instrumentele variabelen, of statistische controles voor potentiële verstorende factoren.”
4. Geldig als de steekproefgrootte groot genoeg was  
   “4” = ” ❌ Fout. Steekproefgrootte lost het causaliteitsprobleem niet op. Een grotere steekproef zou statistische power verhogen en meer precieze correlatieschattingen leveren, maar het kan een correlationele bevinding niet transformeren in een causale. De logische denkfout in de claim van het nieuwsartikel — dat correlatie causaliteit impliceert — blijft bestaan ongeacht steekproefgrootte. In criminologisch onderzoek is steekproefgrootte belangrijk voor betrouwbaarheid en generaliseerbaarheid, maar het onderzoeksontwerp bepaalt of causale claims gerechtvaardigd zijn. Zonder geschikte ontwerpkenmerken zoals randomisatie, tijd-volgorde vaststelling, of controle voor verstorende variabelen, kunnen zelfs studies met enorme steekproeven geen causale conclusies ondersteunen over programma-effectiviteit.”

### Vraag Q23 (Begrijpen)

**Welke belangrijkste veronderstellingen (aannames) moet je controleren voordat je Pearson’s correlatiecoëfficiënt (r) gebruikt en interpreteert?**

*Begrip ‘veronderstellingen’: Statistische veronderstellingen zijn voorwaarden die vervuld moeten zijn opdat een statistische methode geldig en betrouwbaar is. Als deze aannames geschonden worden, kunnen de resultaten misleidend of incorrect zijn.*



**Kernconcepten:** - **Lineariteit**: Relatie moet rechtlijnig zijn (niet gebogen) - **Continue data**: Interval/ratio meetschaal vereist - **Normaliteit**: Data benadert normale verdeling - **Geen uitschieters**: Extreme waarden kunnen correlatie vertekenen - **Homoscedasticiteit**: Gelijke variantie over alle waarden

**Hint:** Denk aan wat Pearson correlatie meet (lineaire relaties) en welke data daarvoor geschikt is.

1. Alleen lineaire relatie en continue variabelen zijn vereist  
   “1” = ” ❌ Fout. Hoewel lineariteit en continue variabelen inderdaad belangrijke veronderstellingen zijn, zijn er meer aannames die gecontroleerd moeten worden. Normaliteit is belangrijk voor de interpretatie van statistische significantie en betrouwbaarheidsintervallen. Extreme uitschieters kunnen Pearson correlatie drastisch beïnvloeden omdat het gebaseerd is op werkelijke waarden (niet rangnummers zoals Spearman). Homoscedasticiteit (gelijke variantie) beïnvloedt de betrouwbaarheid van correlatieschattingen. In criminologisch onderzoek kan het negeren van deze aannames leiden tot verkeerde conclusies - bijvoorbeeld, uitschieters in geweldsdata kunnen de correlatie tussen sociaaleconomische factoren en criminaliteit vertekenen. Een grondige controle van alle veronderstellingen zorgt voor betrouwbare resultaten.”
2. Alleen dat beide variabelen normaal verdeeld zijn  
   “2” = ” ❌ Fout. Normale verdeling is belangrijk maar niet de enige vereiste. Zelfs als beide variabelen perfect normaal verdeeld zijn, kan Pearson correlatie nog steeds misleidend zijn als andere veronderstellingen geschonden worden. Bijvoorbeeld, twee normaal verdeelde variabelen kunnen een niet-lineaire relatie hebben (zoals een U-vormige curve) die Pearson correlatie zal onderschatten. Of er kunnen extreme uitschieters zijn die, hoewel technisch onderdeel van een normale verdeling, onevenredige invloed hebben op de correlatie. In criminologisch onderzoek zijn variabelen zoals inkomen of criminaliteitscijfers vaak scheef verdeeld, maar zelfs wanneer ze genormaliseerd worden door transformaties, moeten nog steeds andere aannames zoals lineariteit en afwezigheid van invloedrijke punten gecontroleerd worden.”
3. Lineariteit, continue variabelen, normale verdeling, geen extreme uitschieters, en gelijke variantie  
   “3” = ” ✅ Juist! Deze vijf veronderstellingen zijn essentieel voor valide Pearson correlatie interpretatie. Lineariteit is fundamenteel omdat Pearson alleen lineaire relaties meet - gebogen relaties kunnen misleidende correlaties opleveren. Continue variabelen zijn vereist omdat Pearson werkt met werkelijke waarden en afstanden tussen waarden. Normale verdeling is belangrijk voor statistische inferentie (p-waarden, betrouwbaarheidsintervallen). Extreme uitschieters kunnen correlaties dramatisch beïnvloeden omdat Pearson gevoelig is voor extreme waarden. Homoscedasticiteit (gelijke variantie) zorgt voor stabiele correlatieschattingen. Zoals getoond in de visualisatie, leiden schendingen van deze aannames tot misleidende of onbetrouwbare correlaties. In criminologisch onderzoek helpt het controleren van deze veronderstellingen onderzoekers te bepalen of Pearson geschikt is of dat alternatieve methoden (zoals Spearman) beter zijn.”
4. Dat er minstens 30 observaties zijn en alle waarden positief zijn  
   “4” = ” ❌ Fout. Deze zijn geen statistische veronderstellingen van Pearson correlatie. Hoewel grotere steekproeven (N ≥ 30) statistisch wenselijk zijn voor betrouwbaardere schattingen, is dit geen formele veronderstelling van de correlatiemethode zelf. Pearson correlatie kan berekend worden met kleinere steekproeven, hoewel de resultaten minder stabiel zijn. Positieve waarden zijn ook niet vereist - correlatie werkt perfect met negatieve waarden, gemengde positieve/negatieve waarden, of zelfs gecentreerde data (met gemiddelde nul). In criminologisch onderzoek hebben variabelen vaak negatieve waarden (bijvoorbeeld, gecentreerde scores, temperatuurafwijkingen, of veranderingsscores). De werkelijke veronderstellingen gaan over de vorm van relaties en dataverdelingen, niet over arbitraire numerieke eigenschappen zoals positiviteit.”
5. Accepteer r = -0,43 als nauwkeurig omdat het statistisch significant is  
   “1” = ” ❌ Fout. Uitschieters kunnen Pearson correlatie dramatisch beïnvloeden, en statistische significantie betekent niet dat de methode geschikt is. Pearson’s correlatiecoëfficiënt (r) veronderstelt een lineaire relatie en normale verdeling van variabelen, waardoor het zeer gevoelig is voor extreme waarden (niet robuust). In dit geval trekken de drie uitschieters die zichtbaar zijn in het spreidingsdiagram (de oranje driehoeken) waarschijnlijk de correlatiecoëfficiënt naar nul van wat een sterkere negatieve relatie zou zijn. Statistische significantie vertelt ons alleen dat de relatie onwaarschijnlijk te wijten is aan toeval; het valideert niet de keuze van correlatiemethode. In criminologisch onderzoek kunnen uitschieters belangrijke speciale gevallen vertegenwoordigen, maar hun onevenredige wiskundige impact op Pearson’s r kan het algemene patroon verdoezelen.”
6. Bevraagteken de robuustheid van de correlatie; adviseer rapportage van zowel Pearson als Spearman correlaties  
   “2” = ” ✅ Juist! Uitschieters kunnen Pearson r vervormen; Spearman correlatie zou robuuster zijn voor extreme waarden. Door data om te zetten naar rangen vermindert Spearman’s correlatie (ρ of rho) de invloed van uitschieters en biedt een nauwkeurigere maat van de monotone relatie tussen onderwijs en arrestatiecijfers. In dit geval onthult Spearman’s correlatie (-0,67) een sterkere negatieve relatie dan Pearson’s (-0,43), wat suggereert dat onderwijs’ beschermende effect tegen arrestaties wordt onderschat door de niet-robuuste Pearson coëfficiënt. Het rapporteren van beide metrieken wordt beschouwd als beste praktijk in criminologisch onderzoek wanneer data uitschieters bevatten. Deze robuuste benadering biedt transparantie en stelt lezers in staat te begrijpen hoe gevoelig de bevindingen zijn voor analytische keuzes en extreme waarden.”
7. Verwerp de gehele analyse omdat uitschieters alle correlationele studies ongeldig maken  
   “3” = ” ❌ Fout. Uitschieters maken analyse niet ongeldig maar vereisen robuuste behandeling. Hoewel uitschieters niet-robuuste correlatiecoëfficiënten beïnvloeden, maken ze de gehele analyse niet waardeloos. In plaats daarvan signaleren ze de noodzaak voor robuuste statistische benaderingen. In de criminologie vertegenwoordigen uitschieters vaak betekenisvolle gevallen (bijv. individuen met unieke kenmerken) in plaats van meetfouten. Juiste robuuste benaderingen zijn: (1) robuuste correlatiemethoden gebruiken zoals Spearman’s of Kendall’s tau; (2) resultaten rapporteren zowel met als zonder uitschieters; (3) variabelen transformeren; of (4) robuuste regressietechnieken gebruiken. Volledige verwerping zou waardevolle data verspillen. Criminologen moeten uitschieters herkennen en geschikte robuuste analytische strategieën selecteren.”
8. De correlatie is zeker te zwak om betekenisvol te zijn ongeacht uitschieters  
   “4” = ” ❌ Fout. r = -0,43 vertegenwoordigt een matige relatie die praktisch belangrijk zou kunnen zijn, en robuuste analyse toont een nog sterkere relatie. In criminologisch onderzoek kunnen zelfs matige correlaties substantiële implicaties hebben voor beleid. Een correlatie van -0,43 duidt erop dat ongeveer 18% van de variantie in arrestatiecijfers wordt verklaard door opleidingsniveau — een opmerkelijke bevinding. In de criminologie worden correlaties boven 0,3 vaak als betekenisvol beschouwd. Wanneer geanalyseerd met robuuste methoden die minder gevoelig zijn voor uitschieters (zoals de Spearman correlatie van -0,67), lijkt de relatie zelfs sterker. Het wegwuiven van matige correlaties als ‘te zwak’ riskeert het over het hoofd zien van belangrijke criminologische relaties die beleid kunnen informeren.”

## CREËREN NIVEAU (2 vragen)

### Vraag Q24 (Creëren)

**Ontwerp een correlationele studie om de relatie tussen social media gebruik en antisociaal gedrag bij adolescenten te onderzoeken. Welke belangrijke ontwerpelementen zou je opnemen?**

**Hint:** Overweeg variabelen, meting, steekproeftrekking, controles en ethische overwegingen.

1. Enquête bij 100 tieners over uren social media gebruik en zelf-gerapporteerde antisociale daden  
   “1” = ” ❌ Onvolledig. Dit ontwerp heeft verschillende methodologische zwakheden: (1) De steekproefgrootte (N=100) is relatief klein voor het detecteren van potentieel subtiele effecten in gedragsonderzoek; (2) Simpele zelf-gerapporteerde uren faalt om de complexe, multidimensionale aard van social media gebruik vast te leggen (verschillende platforms, actief vs. passief gebruik, inhoudstypen, etc.); (3) Geen controlevariabelen betekent dat elke waargenomen correlatie te wijten kan zijn aan talloze verstorende factoren (sociaaleconomische status, ouderlijk toezicht, persoonlijkheidskenmerken, etc.); (4) Cross-sectioneel ontwerp voorkomt het vaststellen van temporele voorrang, wat noodzakelijk is voor causale inferentie; (5) Geen validatiematen voor zelf-gerapporteerd antisociaal gedrag verhoogt meetfouten. Een robuuste correlationele studie vereist het aanpakken van deze methodologische uitdagingen.”
2. Longitudinaal ontwerp dat social media gebruik meet (meerdere platforms, tijd), antisociaal gedrag (meerdere maten), controleert voor leeftijd, SES en familiefactoren, met juist ethisch toezicht  
   “2” = ” ✅ Juist! Dit uitgebreide ontwerp incorporeert belangrijke methodologische sterktes: (1) Longitudinale meting staat toe veranderingen over tijd te volgen en temporele voorrang vast te stellen (ging social media gebruik vooraf aan gedragsveranderingen?); (2) Het meten van meerdere aspecten van social media gebruik (tijd besteed, platformtypen, inhoudsbetrokkenheid) vangt de complexiteit van digitaal gedrag; (3) Het gebruiken van meerdere gedragsmaten versterkt constructvaliditeit en vermindert meetfouten; (4) Controleren voor belangrijke covariaten zoals leeftijd, sociaaleconomische status en familiefactoren helpt alternatieve verklaringen uit te sluiten; (5) Ethisch toezicht zorgt voor deelnemersbescherming, geïnformeerde toestemming, data privacy en gepaste risicomanagement. Dit ontwerp maximaliseert de validiteit van potentiële correlationele bevindingen terwijl het methodologische zwakheden minimaliseert.”
3. Experimentele manipulatie waarbij sommige tieners worden toegewezen aan hoog social media gebruik  
   “3” = ” ❌ Ongepast. Hoewel experimentele ontwerpen causaliteit definitiever kunnen vaststellen dan correlationele ontwerpen, heeft deze benadering ernstige gebreken: (1) Ethische zorgen met potentieel schadelijke manipulatie — onderzoeksetische principes verbieden het blootstellen van deelnemers, vooral minderjarigen, aan interventies die risico kunnen verhogen voor negatieve uitkomsten; (2) Externe validiteitsproblemen, omdat kunstmatig toegewezen social media gebruik verschilt van natuurlijke, zelf-geselecteerde gebruikspatronen; (3) Ecologische validiteitsproblemen, omdat gecontroleerde experimentele condities de echte wereld social media betrokkenheid slecht vertegenwoordigen; (4) Retentie-uitdagingen bij het handhaven van compliantie met toegewezen condities over betekenisvolle tijdsperioden; (5) De vraag vroeg specifiek om een correlationeel ontwerp, niet een experimenteel. Experimentele ontwerpen hebben hun plaats, maar moeten ethisch en gepast worden geïmplementeerd.”
4. Eenvoudige correlatie tussen totale schermtijd en politiecontacten  
   “4” = ” ❌ Te simplistisch. Deze benadering heeft meerdere conceptuele en methodologische zwakheden: (1) Totale schermtijd is een te algemene maat, die alles omvat van huiswerk tot videogames, niet specifiek social media gebruik; (2) Politiecontacten vertegenwoordigen alleen de meest extreme en gedetecteerde antisociale gedragingen, waardoor de overgrote meerderheid van minor en ongedetecteerde gedragingen wordt gemist; (3) Beide maten zijn zeer vatbaar voor selectiebias, meetfouten en derde-variabele problemen; (4) Geen overweging van belangrijke modererende factoren (leeftijd, geslacht, sociaaleconomische context) die beide variabelen beïnvloeden; (5) Mist meetvaliditeit voor zowel de onafhankelijke als afhankelijke variabelen. De resulterende correlatie zou moeilijk betekenisvol te interpreteren zijn vanwege deze fundamentele beperkingen.”

### Vraag Q25 (Creëren)

**Welke van de volgende hypotheses over de relatie tussen werkloosheid en diefstalcijfers op wijkniveau is het best geformuleerd voor correlatie-onderzoek?**

**Hint:** Een goede hypothese specificeert: variabelen, richting, sterkte, en analyse-eenheid.

1. “Werkloosheid heeft invloed op criminaliteit”  
   “1” = ” ❌ Fout. **Te vaag en ontoetsbaar.** Deze hypothese mist: (1) Specificatie van variabelen - hoe meet je werkloosheid en criminaliteit? (2) Richting - positief of negatief verband? (3) Analyse-eenheid - individueel, wijk, of stadniveau? (4) Verwachte sterkte. Zonder deze details kan de hypothese niet betekenisvol getest worden.”
2. “Wijken met een hoger werkloosheidspercentage zullen hogere diefstalcijfers per 1000 inwoners hebben”  
   “2” = ” ✅ Juist! **Goed geformuleerde hypothese.** Bevat de essentiële elementen: (1) **Variabelen:** Duidelijk gedefinieerd (werkloosheidspercentage, diefstalcijfers per 1000 inwoners) (2) **Richting:** Positieve relatie voorspeld (hoger werkloosheid → meer diefstal) (3) **Analyse-eenheid:** Wijkniveau (4) **Testbaarheid:** Kan getest worden met correlatie-analyse. Dit is hoe echte onderzoekshypotheses worden geschreven - ze voorspellen relaties tussen variabelen, niet specifieke correlatiecoëfficiënten.”
3. “Werkloze mensen stelen meer”  
   “3” = ” ❌ Fout. **Verkeerde analyse-eenheid en te simplistisch.** Deze stelling gaat over individueel gedrag (werkloze mensen), terwijl de vraag wijk-niveau analyse vereist. Ook suggereert het causaliteit in plaats van correlatie, en mist het specificatie van hoe variabelen gemeten worden.”
4. “Werkloosheid veroorzaakt alle diefstal”  
   “4” = ” ❌ Fout. **Te extreem en wetenschappelijk onhoudbaar.** Deze absolute claim (alle diefstal) is: (1) Onrealistisch - diefstal heeft meerdere oorzaken (2) Ontoetsbaar - zou worden weerlegd door één tegenvoorbeeld (3) Causaal in plaats van correlationeel (4) Mist operationalisatie van variabelen.”