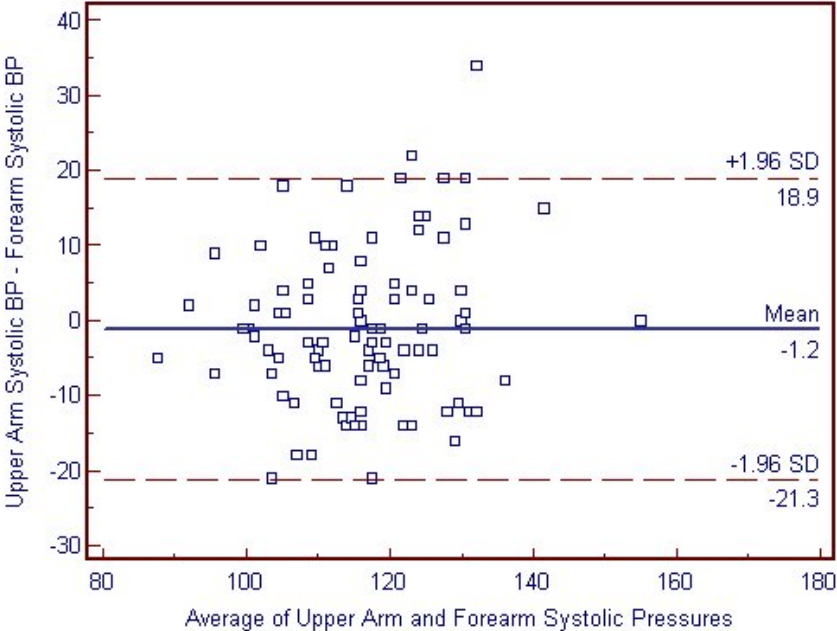


Inhoud	
Types design	
Epidemiologie	
Beschrijvende statistiek	
Gemiddelden vergelijken	
Correlatie / regressie	
Regressie validatie	
Powercalculatie	
Types design	
Prospectief	
Case-control	
Cross-sectioneel	Definitie steekproef op 1 bepaald punt in de tijd
Cross-validatie	Definitie 2 groepen: 1 controle + 1 interventie → na enige tijd controle en interventie situatie omwisselen
Epidemiologie	
epidemiologie	Definitie Leer van het voorkomen van ziektes
incidentie	Definitie (nieuwe gevallen / groep) per tijdseenheid
prevalentie	Definitie (aantal gevallen / groep) op een bepaald tijdstip ≈ incidentie x ziekte duur
Causaliteit	Evidence based Bast praktisch Logisch/plausibel
Evidence category	Indeling Ia = meta-analyse en gerandomiseerde gecontroleerde trials Ib = minimaal 1 gerandomiseerde gecontroleerde trial IIa = minimaal 1 gecontroleerde studie (zonder randomisatie) IIb = minimaal 1 quasi-experimentele studie III = observationele / beschrijvende studies IV = mening van experts / werkgroepen
mortaliteit	Definitie incidentie van overlijdens
prognose	Definitie ziektebeloop zonder effectieve therapie
Placebo effect	Waarde bestaat bijna niet, enkel bij pijn een beetje
Onderzoeksfasen	Definitie alleen bij medicijnen Fase I kleine groep mensen testen op nevenwerkingen Fase II Is het medicijn ook daarwerkelijk werkzaam in mensen Fase III vergelijking met het standaard medicijn Fase IV onderzoek nadat het medicijn is geregistreerd
lethaliteit	Definitie mortaliteit / incidentie
confounder	Definitie factor die ook afhankelijk is van de oorzakelijke factor maar geen invloed heeft op de bestudeerde variabele bv. Vuuraansteker komt vaker voor bij longkanker, vuuraansteker is de confounder, roken is de oorzakelijke factor
Neg predictieve waarde	Definitie = test echt neg / totaal neg testen = $TN / (TN + FN)$
Pos predictieve waarde	Definitie = test echt pos / tot pos testen = $TP / (TP + FP)$
Se (sensitiviteit)	Definitie = test pos / tot echt pos = $TP / (TP + FN)$
Sp (specificiteit)	Definitie = test neg / tot echt neg = $TN / (TN + FP)$ Afhankelijk van indien prevalentie erg laag is moet de specificiteit erg hoog zijn anders heeft de test geen waarde

Bivariaat	<p>Definitie 2 variabelen die elkaar beïnvloeden</p>
Beschrijvende statistiek	
Gemiddelde = μ (mu) of \bar{x}	<p>Definitie μ = populatiegemiddelde \bar{x} = steekproef gemiddeld</p> <p>Verdeling enkel zinvol bij een normale verdeling, anders is een mediaan beter</p>
percentiel	<p>Definitie hoeveel procent van de bevolking onder deze waarde komt P98 = 98% van de bevolking komt onder de waarde, 2% komt er boven P50 = 50% van de bevolking komt onder de waarde, 50% komt er boven P2 = 2% van de bevolking komt onder de waarde, 98% komt er boven</p> <p>Verdeling zinvol bij normale en niet-normale verdeling</p>
Studentverdeling = t-verdeling	<p>Definitie = normaal verdeling waarvan de SD onbekend is de sample SD wordt gebruikt</p>
e = grondgetal natuurlijk log	<p>Definitie = limiet van $n \rightarrow \infty (1 + 1/n)^n = 1+1+1/2+1/6+1/24...$</p>
Standard error of the mean = SE(M)	<p>Definitie = SD van een steekproef = SD steekproef / wortel aantal metingen = SD gebaseerd op een schatting</p> <p>Nut wordt kleiner naarmate de steekproef toeneemt (SD niet)</p>
Variantie = σ^2 (sigma)	<p>Definitie uitgedrukt in kwadraat dus moeilijk te vergelijken met het gemiddelde, gebruik hiervoor de SD</p> <p>Voor steekproef $\text{variantie}^2 = 1/(n-1)$ som van (alle steekproef uitkomsten - \bar{x})²</p>
Standaarddeviatie = σ (sigma) = SD	<p>Definitie = wortel van de variantie = SD van een populatie (=onafh van het aantal metingen) deze weet je eigenlijk nooit, dan moet iedereen in de populatie gemeten worden</p> <p>Normale verdeling 1 SD = 68,3% van de populatie bevindt zich tussen de -1 SD en +1SD 2 SD = 95,4% van de populatie bevindt zich tussen de -2 SD en +2SD</p> <p>Niet-normale verdeling</p>
Verdeling	<p>Parametrische verdeling volgens een patroon, bv normale verdeling, uniform, skewed etc.</p> <p>Nonparameterische verdeling geen patroon</p> <p>Skewed normale verdeling met een asymmetrische piek</p> <p>Normale = Gauss verdeling natuurlijk logaritmische functie, met het gemiddelde en de SD als variabele parameters als SD \approx gemiddelde dan is er eigenlijk nooit sprake van een normale verdeling</p> <p>Uniform alle mogelijkheden zijn gelijk verdeeld</p>
Nominale schaal	<p>Definitie = tekst schaal bv provincie indeling, nationaliteit,</p>
Betrouwbaarheidsinterval	<p>Definitie = \bar{X} (steekproef gemiddelde) \pm SEM * 1,96 buiten de 95% betrouwbaarheidsinterval = significant = $p < 0,05$ bij gepaarde t-test dan is het significant als het hele interval boven of onder 0 ligt</p>
Variabele soort	<p>Soorten scale = continue schaal ordinal = vaste waarden, bv 1,2,3 nominaal = categorische, bv geslacht</p>
(LR) Likelyhood ratio	<p>Definitie LR+ = Se / (1-Sp) LR- = (1-Se) / Sp</p> <p>Interpretatie 1 = test heeft geen waarde</p>

	<p>>1 = hoe hoger hoe beter de test <1 = hoe lager hoe beter de test voor de omgekeerde relatie</p>
Gemiddelden vergelijken	
X ² (Chi-kwadraat)	<p>Nut bepalen of 2 populaties met nominale waarden verschillen</p> <p>Vrijheidsgraden (df) aantal onderzochte groepen -1</p>
ANOVA one-way	<p>Definitie =ANalysis Of VAriance</p> <p>Nut Meer dan 2 gemiddelden vergelijken bv lengte in 3 sporten ivm 2 sporten metingen moeten onafhankelijk zijn</p>
ANOVA repeated measurements	<p>Definitie =ANalysis Of VAriance</p> <p>Nut Meer dan 2 middelen vergelijken die afhankelijk van elkaar zijn</p> <p>SPSS General linear model → repeated measures → factor bv tijd of load, levels is aantal metingen, measure is parameters benoemen → define → de variables overschrijven met de gemeten waarden → options: compare means met Bonferroni correctie</p>
ANOVA Bonferroni correctie	<p>Definitie probleem: als je bv 20 hypothesen gaat testen is de kans dat 1 significant (p=0,05) is al behoorlijk groot door toeval. oplossing: significantie niveau/20 als significantie niveau gebruiken = 0,05/20 = 0,0025</p>
t-test gepaard	<p>Definitie zie t-test</p> <p>Uitkomst</p> <ul style="list-style-type: none"> • gemiddeld verschil (+ SD) met steekproef 2 • p-waarde = sig = kans dat het verschil buiten de normale verdeling van de 1^e steekproef ligt (>0,05 = sig verschil) • betrouwbaarheidsinterval moet boven OF onder de nul liggen, maar niet over de nul
t-test ongepaard	<p>Definitie zie t-test</p> <p>SPSS kolom maken met "code": 1 = populatie 1, 2 = populatie 2 uitslagen van populatie 1 EN 2 in dezelfde kolom onder elkaar zetten met de juiste code ervoor compare means → independent t-test → grouping variabele is "code" → define groups: 1 en 2 invullen → dependent list van alle te vergelijken variabelen</p>
p-waarde	<p>Definitie = kans dat de waarde van een steekproef buiten het als normaal gedefinieerde gebied van de populatie valt (links, rechts of tweezijdig) een kleine p-waarde = als niet normaal beschouwen = significant verschillend (is >2 SD erbuiten)</p> <p>Een vs. Tweezijdig 1p = éézijdig 2p = tweezijdig</p> <p>Normale p-waarden kleiner dan 0,01-0,05 = 1-5% = significant</p> <p>Normale verdeling p-waarde = (gemiddelde – gemeten waarde) / SD</p>
Significantie	<p>Formule +/- significant verschil grens = A * (wortel ((σ¹²/n1) + (σ²²/n2)))</p> <p>Als μ₁-μ₂ buiten de significant verschil grenzen ligt (of groter of kleiner) = significant A = significantie uit tabel (5%=1,96 , 1%=2,58)</p>
t-test = t-toets	<p>Nut 1 aantonen of het steekproef gemiddelde verschilt van de populatie de populatie moet normaal verdeelt zijn en het gemiddelde en de SD moet bekend zijn van de populatie</p> <p>Nut 2 aantonen of 2 steekproef gemiddelden (uit de zelfde populatie) verschillen Het gemiddelde en de SD van de populatie hoeven dan niet bekend te zijn A) onafhankelijke/ongepaarde steekproeven van dezelfde populatie → beide steekproeven dezelfde SD B) gepaarde steekproeven → de 2^e steekproef is het verschil (bv gewicht voor en gewicht na)</p>
Correlatie / regressie	
CI	<p>Definitie Confidence interval, maat voor de variabiliteit van een voorspeller</p>

<p>ICC</p>	<p>Definitie = interclass correlation coefficient ≈ Pearson's r, maar gemiddeld en SD wordt voor de groepen samengenomen</p> <p>Nut gebruikt voor test-hertest betrouwbaarheid</p> <p>LET OP ! correlatie van belang voor groepsniveau, als het om de interpretatie per individu gaat is de SEE ook erg belangrijk</p>
<p>95%CI</p>	<p>Definitie 95% Confidence Interval = maat voor spreiding van de voorspellingen t.o.v. geobserveerd kan per variabele in de $ax+b$ vergelijking of voor de r als totaal</p> <p>SPSS SPSS berekent de 95% voor de a en de b in $ax+b$, maar niet voor de r als totaal (kan met http://faculty.vassar.edu/lowry/rho.html)</p>
<p>SEE</p>	<p>Definitie Standard Error of Estimate = Standard Deviation of the Residuals = Residual Standard Deviation</p> <p>Nut maat voor spreiding van de voorspellingen t.o.v. geobserveerd</p>
<p>Bland-Altman plot</p>	<p>Doel Te gebruiken om een meetmethode te valideren met een andere meetmethode, waarbij beide methoden niet direct te kalibreren te zijn. Dus er kan spreiding zitten in beide meetmethoden. Nu ga je er vanuit dat het gemiddelde van de beide lijsten met waarde het juiste resultaat is.</p> <p>Uitvoering gemiddelde van totaal van beide lijsten met waarden SD uitrekenen van de het totaal van beide lijsten met waarden</p> <p>Limits of Agreement 1,96 SD boven en onder het gemiddelde = 95% = limits of agreement = binnen dat interval liggen 95% van alle verschillen van beide methoden per methode moet de clinicus beoordelen of de limits of agreement klinisch significant zijn (want de correlatie is sowieso al significant)</p>  <p>The figure is a Bland-Altman plot. The y-axis is labeled 'Upper Arm Systolic BP - Forearm Systolic BP' and ranges from -30 to 40. The x-axis is labeled 'Average of Upper Arm and Forearm Systolic Pressures' and ranges from 80 to 180. A solid horizontal line represents the mean difference, labeled 'Mean' with a value of -1.2. Two dashed horizontal lines represent the 95% limits of agreement, labeled '+1.96 SD' with a value of 18.9 and '-1.96 SD' with a value of -21.3. The plot contains numerous data points represented by small squares, mostly clustered between 100 and 140 on the x-axis and -10 and 20 on the y-axis.</p>
<p>Multi-collineair</p>	<p>Definitie als 2 variabelen in een multi-regressie analyse ook onderling goed correleren</p>
<p>Residuals</p>	<p>Definitie verschillen tussen geobserveerde en voorspelde waarden</p>
<p>r = Pearson's coefficient</p>	<p>Doel bestaat er een samenhang tussen 2 lijsten met waarden ?, geen regressie formule, is hetzelfde als enkelvoudige lineaire regressie</p> <p>Notatie kleine letter r geen R</p> <p>SPSS correlate → bivariate → Pearson's coefficient</p>
<p>Regressie meervoudig lineair (multi-variate)</p>	<p>Nut meerdere onafhankelijke variabelen</p> <p>SPSS Analyze → regression → linear → meerdere onafhankelijke variabelen invullen</p>

	<p>Method: Stepwise, zet alle mogelijk variabelen in 1 block selection variable: bv geslacht 1=man → selecteert alleen de mannen</p> <p>LET OP: wanneer er een variable met een missing value wordt bijgeplaatst dan wordt de regressie op minder value's gedaan en verandert de regressie</p> <p>Significantie zie enkelvoudige lineaire regressie per onafhankelijke variabele staat de significantie genoemd, moet kleiner zijn dan 0.05 bij meer variabelen moeten de R² stijgen en de Standard Error of Estimate kleiner worden de significantie moet <0.05 zijn anders verandert de regressie niet significant door de parameter</p> <p>Stepwise (SPSS: method) de volgorde van parameters is van belang van de significantie, de R2 en de SEE veranderen niet, wel de significantie, met stepwise controleert SPSS dit automatisch</p>
<p>Partial Least Squares (PLS)</p>	<p>Nut helpt om te kiezen uit een grote groep variabelen (die ook onderling afhankelijk zijn) om te bepalen waarmee je de regressie gaat uitvoeren</p> <p>SPSS analyse → regression → partial least regression</p> <p>Variance Importance in the Projection (VIP) >0,95 = goede variabele <0,95 = variabele heeft weinig effect</p> <p>Latent factors aantal factoren die niet gemeten zijn, maar wel een invloed hebben op de meting</p>
<p>Constante</p>	<p>CAVE Zonder constante is de R en R2 vals heel erg hoog deze waarden zijn niet te vergelijken met constante.</p>
<p>Regressie enkelvoudig lineair (uni-variate)</p>	<p>Nut samenhang tussen waarden in een regressie-formule vatten</p> <p>Lineair $Y = A + Bx + U$ Y = afhankelijke variabele x = onafhankelijke variabele U = storingsterm A = constante B = (ongestandaardiseerde) coëfficiënt (in SPSS staan A en B beide onder B) = regressie coëfficiënt (gestandaardiseerde coefficient = β = omgerekend als gesteld is dat de variatie 1 is) Significant: ANOVA om te bekijken of het een significant deel van de variatie beschrijft of niet</p> <p>SPSS Analyze → regression → linear</p> <p>Uitkomst tussen -1 en +1 -1 = volledig omgekeerd evenredig 0 = geen samenhang 1 = volledig evenredig</p> <p>Significantie afh van het aantal paren bij een groot aantal paren, kan ook een lage r-waarde wel significant zijn</p> <p>Significantie R^2 = % van verklaarde variantie (moet dus zo hoog mogelijk zijn) Standard Error of Estimatie = maat voor variantie, moet zo laag mogelijk zijn (wordt ook lager bij meer personen) significantie moet kleiner zijn dan 0.05</p>
<p>Regressie validatie</p>	
<p>Cross-validatie</p>	<p>Definitie met andere data een regressie formule valideren interne validatie: de regressie toepassen op subsets van originele data externe validatie: de regressie formule toepassen op nieuwe data</p>
<p>Bootstrapping</p>	<p>Definitie vorm van cross validatie regressie wordt gevalideerd door verschillende data samples te nemen en te vergelijkingen met de originele data</p> <p>Aantal samples hoe hoger, hoe beter, 200 is meestal voldoende</p> <p>Resultaat zijn de parameters significant of niet de constante is meestal niet significant maar zonder constant is de R en R2 vals heel hoog</p> <p>SPSS Analyze → regression → linear → bootstrap</p>

Jack-kniving	
Powercalculatie	
Sample grootte bij vergelijking van 2 kwantitatieve proporties	<p>Definitie formule om 2 "ja/nee" groepen te vergelijken</p> <p>Formule $n = (A+B)^2 \times \frac{(p1(1-p1)) + (p2(1-p2))}{(p1-p2)^2}$ </p> <p>n = sample grootte (van groep 1 OF 2, totaal te onderzoeken is dus 2n) A = significantie uit tabel (5%=1,96 , 1%=2,58) B = power uit tabel (80%=0,84 , 90%=1,28 , 95%=1,64) p1 = kans op "ja" in groep 1 p2 = kans op "ja" in groep 2</p>
Sample grootte bij vergelijking van 2 gemiddelden (Beck Scale)	<p>Definitie formule om 2 (continue) gemiddelden te vergelijken</p> <p>Formule $n = (A+B)^2 \times 2 \times (\sigma^2 / KV^2)$ </p> <p>n = sample grootte (van groep 1 OF 2, totaal te onderzoeken is dus 2n) A = significantie uit tabel (5%=1,96 , 1%=2,58) B = power uit tabel (80%=0,84 , 90%=1,28 , 95%=1,64) σ = verwachte standaarddeviatie KV = minimaal klinische relevant verschil</p>
Kwalitatief onderzoek	
Doel	<p>Doelen</p> <ul style="list-style-type: none"> • gedachten/problemen van patiënten boven water halen die niet in de standaard vragenlijsten staan • bestaande vragenlijsten (kwantitatief) onderzoek bevestigen/bijstellen