

Analyse af gang

Bachelorprojekt i Matematik

Frederik Leander Schultz - BHN280



UNIVERSITY OF
COPENHAGEN

Københavns universitet - Science

11. Januar 2019

Abstract

This paper describes the angle between the upper leg and lower leg when people walk. Furthermore it investigates if there is a significant difference between this angle in the two legs. It also describes the principle of Limits of Agreement and calculates this for the pointwise observations. In addition, this paper describes and calculates a confidence band for both pointwise and joint observations

Indhold

1	Indledning	2
2	Data	2
3	Idégenerering	3
4	Manipulation af data	4
4.1	Udpluk af data	4
4.2	Vinkeludregninger	4
4.3	Gennemsnitsskridt	5
4.4	Estimat af vinkler	6
5	Analyse	10
5.1	Idéer	10
5.2	Permutationstest	11
5.2.1	Eksempel	12
5.2.2	Permutationstest på forskel mellem ben	13
5.3	Prædiktions- og Konfidensintervaller	14
5.3.1	Limits of Agreement (LoA)	14
5.4	95 % konfidensbånd for gennemsnittet af forskellen	18
6	Fremtidige analyser	21
7	Konklusion	22

1 Indledning

I dette projekt analyseres den vinkel der kommer imellem overbenet og underbenet når man går. Først vil der blive gjort tanker om hvordan disse udregninger bliver udført, og hvilke dele af data der skal arbejdes med. En udplukning af data der skal benyttes til at udregne denne vinkel vil blive udført, og der forklares hvilke metoder der er brugt til dette. En manipulation vil nu hjælpe med at producere et estimat af hvert individs knævinkel. En permutationstest blive beskrevet og benyttet til at undersøge hvorvidt der er signifikant forskel på hvert individs dominerende og ikke-dominerende ben. Et Limits of Agreement bånd vil blive beskrevet og beregnet og til sidst benyttes bootstrapproceduren til at laves konfindensinterval.

2 Data

Data til dette projekt er stillet til rådighed af Luis Bueno Ernesto Salazar fra Sao Carlos, Brazil.

16 personer er hver blevet udstyret med såkaldte "reflective markers" på 9 udvalgte steder på ben og fod (9 på hvert ben), hvorefter de er blevet instrueret i at gå på et løbebånd, som kørte med konstant hastighed. Vha. lys og kameraer (OPTOTRAK-systemet) måles positionen af de 9 reflective markers med en frekvens på 100 Hz (10 millisekunder) over en periode på ca. et minut. Hver person har fået foretaget målinger på begge ben.

Derved kigges der på 32 forskellige datasæt, hvor 16 af dem er for hver persons højre ben og de resterende 16 er deres venstre ben.

I hvert af disse datasæt er der 28 søjler, hvor den første søjle er Frame (dette vil blive erstattet med tid i projektet), der holder styr på tidspunktet for målingen. De resterende søjler indeholder observationer af benet set fra 2 forskellige retninger, nemlig vertikal (lodret / vertical), transversal (fra side-til-side / medio-lateral) og longitudinal (frem-tilbage / antero-postero). F.eks angiver søjle 2-4 den 3-dimensionale position i rummet svarende til den reflective marker som er placeret ved den anatomiske position Metatarsal Head, sidder på foden.

Herved er der 32 datasæt, hver med 27 positioner med 7000 observationer i hver søjle. Alle forsøgspersoner er raske mænd mellem 17 og 46 år der er mellem 1.64-1.86 m. høj, og har en vægt mellem 64 og 100 kg.

3 Idégenerering

Da dette datasæt blev stillet til rådighed og der skulle overvejes hvordan data skulle arbejdes med, blev det ret hurtigt besluttet at kigge på den vinkel, der er imellem underbenet og overbenet i forhold til tiden imens man går. Derfor skulle der først overvejes, hvordan man fik data til at være på en måde hvorpå man kan arbejde med dette. Først og fremmest skal der tages et valg om hvordan man vil udregne denne vinkel, og to måder kom i tankerne da dette blev overvejet.

1. Man kan udregne vinklen ved at lave to vektorer, en vektor der følger underbenet og en vektor der følger overbenet, ud fra de reflective markers der er placeret på de respektive steder, og benytte vektorregning til at udregne den ønskede vinkel.
2. Man kan udregne vinklen ved at udtrække tre punkter (et ved anklen, et ved knæet og et ved hoften), og derved lave en trekant imellem disse punkter og herefter benytte trekantsberegninger til at udregne den ønskede vinkel.

Der er fordele og ulemper ved begge metoder. Metode 1 har den fordel at der ikke skal så mange udregninger skal udregne denne vinkel, men for at man kan bruge det skal vektorene først defineres. Metode 2 har den fordel at det er nemt at finde den trekant der skal bruges ud fra data, dog skal der laves en del mere arbejde for at få udregnet vinklen. Metode 2 benyttes i dette projekt, men man kunne ligeså godt have benyttet metode 1.

4 Manipulation af data

4.1 Udpluk af data

Da hvert datasæt er bygget op sådan at der er 9 reflective markers på hvert ben, er det første der skal gøres at udtrække de punkter der skal benyttes. Dette gøres ved at lave 16 nye datasæt der kun indeholder det data der skal bruges. Søjle 1, 5, 7, 8, 10, 14 og 16 trækkes ud for hvert datasæt, hvor søjle 1 beskriver tiden i frames og de resterende beskriver positionerne for anklen, knæet og hoften set fra siden. For at kunne arbejde med dem, ses disse søjler som punkter i et koordinatsystem og hvert punkt kan beskrives som $p_i = (x, y), i \in \{ankel, knæ, hofte\}$, hvor x er Antero-posterior positionen, og y er den vertikale position. Med disse punkter dannes en trekant med punkterne $p_{ankel}, p_{knæ}, p_{hofte}$. Herefter benyttes pythagoras formel til at udregne længderne mellem punkterne til hver tid.

$$\begin{aligned} |p_{ankel} - p_{knæ}| &= \sqrt{(x_{ankel} - x_{knæ})^2 + (y_{ankel} - y_{knæ})^2} \\ |p_{knæ} - p_{hofte}| &= \sqrt{(x_{knæ} - x_{hofte})^2 + (y_{knæ} - y_{hofte})^2} \\ |p_{ankel} - p_{hofte}| &= \sqrt{(x_{ankel} - x_{hofte})^2 + (y_{ankel} - y_{hofte})^2} \end{aligned}$$

Nu hvor længderne i den førnævnte trekant er beregnet, kan man udregne vinklen

4.2 Vinkeludregninger

For at udregne vinklen mellem over- og underben bruges formelen:

$$V(t) = \left(\pi - \arccos \left(\frac{a(t)^2 - b(t)^2 - c(t)^2}{2 \cdot b(t) \cdot c(t)} \right) \right) \frac{180}{\pi} \quad (1)$$

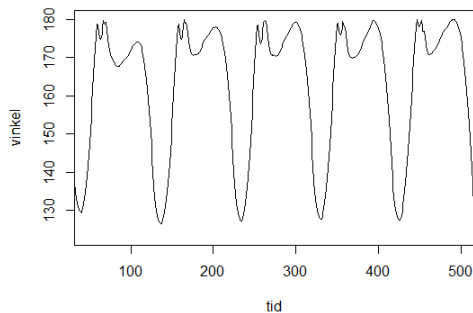
Hvor

- $a(t)$ = Længden mellem anklen og hoften
- $b(t)$ = Længden mellem knæet og hoften
- $c(t)$ = Længden mellem anklen og knæet
- t = tid

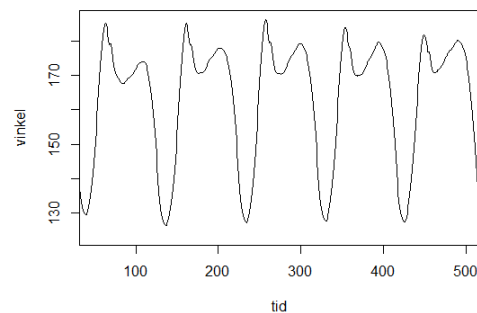
Her var der dog problemer, da det ikke var alle værdier der gav mening, og vi fandt ud af at dette skyldtes overekstendering, hvilket vil sige at knæ-vinklen er større end 180 grader (se figur (a) nedenfor). Måden vinklen udregnes på fungerer altså ikke, hvis vinklen er over 180 grader, da den side i trekanten der bliver dannet af anklen og hoften kommer til at ligge på den forkerte side af knæet og derved give en forkert vinkel. For at få dette fikset, blev der lavet en funktion der tjekkede om personen overekstenderede. Dette blev gjort ved at tjekke om hældningen for linjen der følger overbenet er større end hældningen for linjen der følger underbenet og dette udregnes således: $\frac{y_1 - y_2}{x_1 - x_2}$.

Hvis enten under- eller overbenet er fuldstændig lodret vil denne hældning være uendelig og der laves en funktion der sætter hældningen til at være 0 hvis hældningen er ∞ .

Hvis hældningen af linjen der følger overbenet er større end hældningen af linjen der følger underbenet til tid t , sættes udregnes vinklen som i formel (1), og hvis det modsatte er sandt kan formel et trukket fra 180 benyttes. Når vinklen udregnes på denne måde fås de rigtige værdier (Se figur (b) nedenfor).



(a) Her ses et plot hvor der ikke er taget højde for overekstendering



(b) Her ses et plot, hvor der er taget højde for overekstenderingen

4.3 Gennemsnitsskridt

Herefter ville det være optimalt hvis data kunne reduceres, da dette ville gøre det nemmere at arbejde med. Først laves en funktion der finder alle de lokale minima hvor vinklen er mindre end 140 grader, og derved kan disse punkter bruges som

”starten” af skridtet.

Derfor vælges de 10 skridt i midten af datasættet, da vi regner med at testpersonerne her, har vænnet sig til forsøget og derved vil der være mindst mulig støj.

4.4 Estimat af vinkler

Nu er data kommet på formen (t_j, y_j) , hvor t er tiden, y er vinklen og j er frame, og der ses på $\mathbb{E}y_j = f(t_j)$. Da det ligner, ud fra plots som figur (b), at vinklen i knæet er periodisk, vil signalet kunne estimeres ved brug af Fourier-rækker:

Antag at vinklerne kommer fra en lineær normal model på formen:

$$y_j = f(t) + \epsilon_j,$$

hvor $\epsilon \sim \mathcal{N}(0, \sigma^2)$ iid og

$$f(t) = \sum_{k=1}^K \beta_k \phi_k(t),$$

er en funktion der estimerer data, hvor β_k er en basis og $\phi_k = \sin(\frac{2\pi}{p}kt_j) + \cos(\frac{2\pi}{p}kt_j)$

.

Nu vil middelværdierne for den lineære normale model være:

$$\mathbb{E}(y) = \begin{pmatrix} \mathbb{E}y_1 \\ \vdots \\ \mathbb{E}y_n \end{pmatrix} = \begin{pmatrix} f(t_1) \\ \vdots \\ f(t_n) \end{pmatrix} = \{A\beta \mid \beta \in \mathbb{R}^{k+1}\} = \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_k(t_1) & 1 \\ \vdots & & \vdots & \vdots \\ \phi_1(t_n) & \cdots & \phi_k(t_n) & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \\ \text{intercept} \end{pmatrix}$$

$$\text{Hvor } A = \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_k(t_1) & 1 \\ \vdots & & \vdots & \vdots \\ \phi_1(t_n) & \cdots & \phi_k(t_n) & 1 \end{pmatrix} \text{ er designmatricen.}$$

Fourier-rækker er på formen

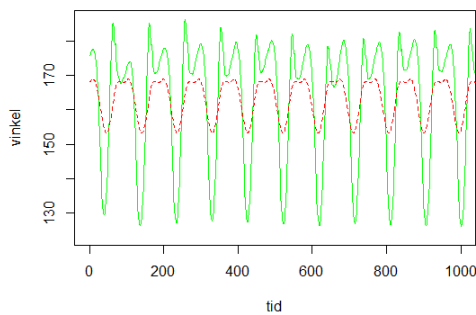
$$f(t) = \sum_{k=1}^K \left(a_k \cos\left(\frac{2\pi}{p}kt\right) + b_k \sin\left(\frac{2\pi}{p}kt\right) \right)$$

Hvor t er tiden, p er perioden og K er antal sinus- og cosinusfunktioner. K kan varieres for at få fourierrækken til at passe bedre til værdierne.

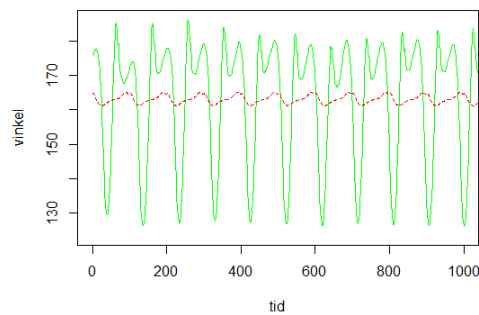
Nu skal der overvejes to ting:

- Hvad skal perioden være?
- Hvor mange led skal der være i fourierrækken?

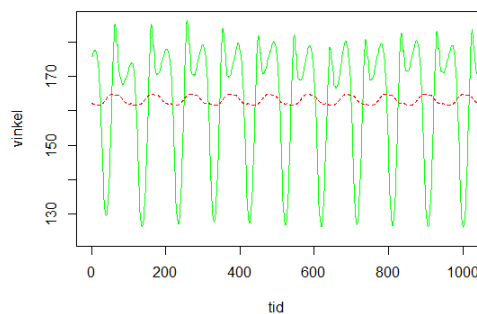
I forhold til den første overvejelse, ses der ud fra plots, at hvert skridt har en periode på omkring 100 frames og derfor prøves nogle værdier heromkring, nemlig $p = 95, p = 100, p = 105$ og K sættes til at være 10, for at få et overblik over hvor godt disse approximere data:



(a) Her ses en fourierrække med periode 95, sat ovenpå de udregnede vinkler



(b) Her ses en fourierrække med periode 100, sat ovenpå de udregnede vinkler



(c) Her ses en fourierrække med periode 105, sat ovenpå de udregnede vinkler

Her ses det tydeligt at disse approksimationer ikke er særlig gode, dog er en periode på 95 bedst. For at få disse til at passe bedre, vil vi nu forsøge at få hvert skridt til

at være lige langt i forhold til tiden, for hvis dette er tilfældet ved vi at perioden er lig 1.

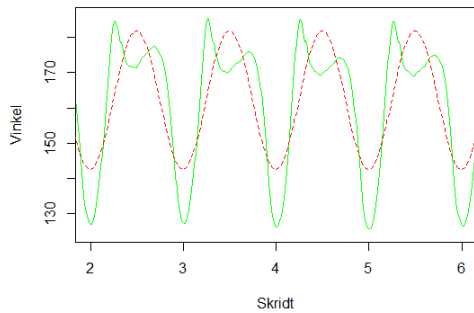
Dette gøres ved at hive i tidsaksen, sådan at et skridt varer én tidsenhed. Hvis det n' 'te skridt varer m frames, ændres hver tidsenhed til at være $n + \frac{i}{m}$, hvor i er $1, 2, 3, \dots, m$.

F.eks hvis skridt nummer 2 varer 135 frames vil tidsaksen mellem 2 og 3 ændres til at være

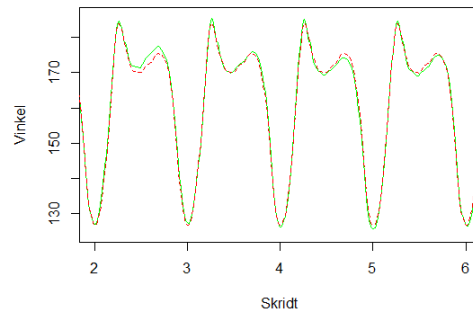
$$\left[2 + \frac{1}{135}, 2 + \frac{2}{135} \dots 2 + \frac{134}{135}, 3 \right]$$

Nu hvor hvert skridt varer en tidsenhed, sættes det nu igen ind i en fourierrække, hvor p sættes lig 1, og det overvejes hvor mange led der skal være i fourierrækken.

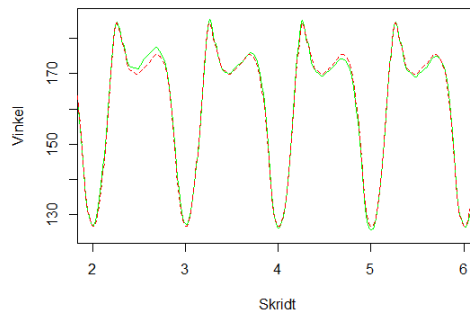
Igen prøves nogle forskellige:



(a) Her ses en fourierrække med ét led sammenlignet med de vinkler den skal approksimere



(b) Her ses en fourierrække med 10 led sammenlignet med de vinkler den skal approksimere



(c) Her ses en fourierrække med 15 led sammenlignet med de vinkler den skal approksimere

I de overstående plots, er den grønne kurve de udregnede vinkler, og den røde stiplede kurve er fourierapproximationen.

Det ses at fourierrækken med et enkelt led ikke er godt nok til at estimere vinklerne, da det slet ikke passer med kurven for vinklerne.

Fourierrækken med 10 led passer godt i forhold til kurven for vinklerne.

Fourierrækken med 15 led passer også godt, men da den ikke passe bedre en den med 10 led benyttes den fourierrækken med 10 led for ikke at overestimere data, og en periode på 1 tidsenhed benyttes da det ses at det passer godt til kurverne for

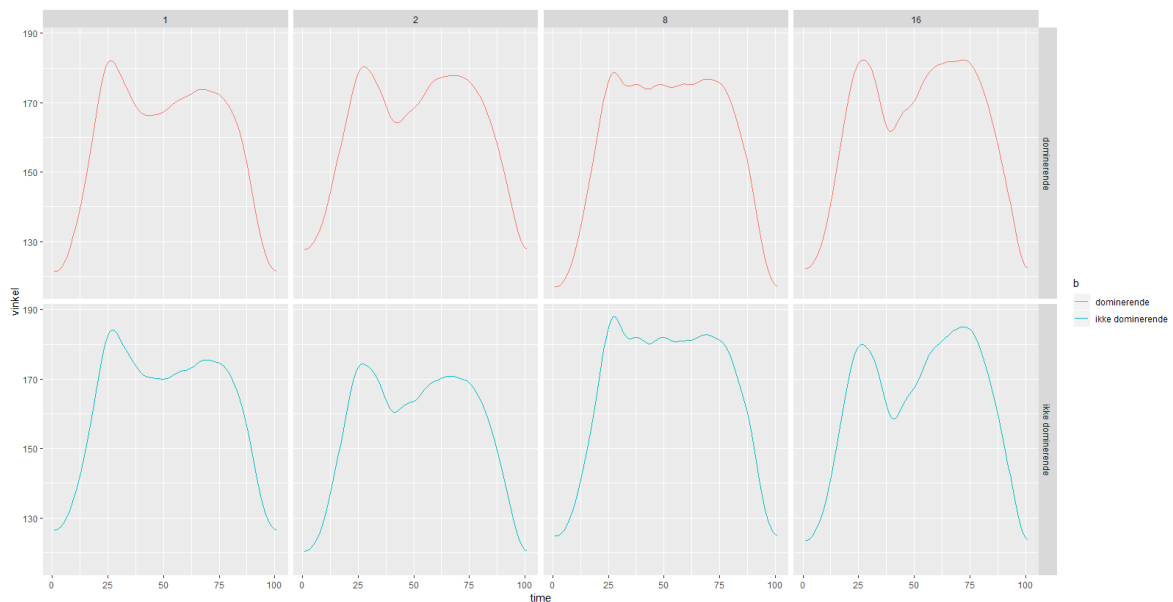
vinklerne. Denne approksimation benyttes fremover:

$$f(t) = \sum_{k=1}^{10} (a_k \cos(2\pi kt) + b_k \sin(2\pi kt))$$

5 Analyse

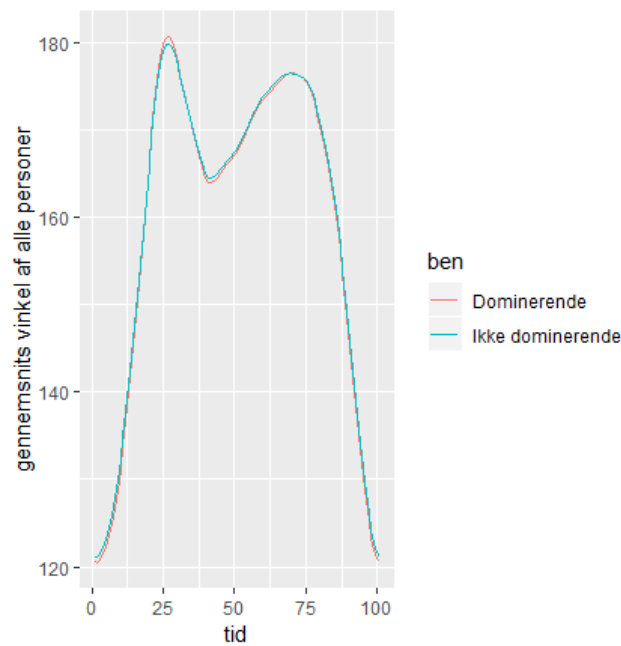
5.1 Idéer

Nu sættes alle datasæt sammen og der ses på et plot af nogle af personers dominerende ben og ikke dominerende ben:



Figur 4: Her ses vinklerne for 4 af personers dominante og ikke dominante ben.

Ud fra dette plot tyder det på at der ikke er særlig stor forskel på det dominerende og ikke-dominerende ben fra person til person og det ser også sådan ud for de øvrige personer. Herved kigges der på gennemsnittet af vinklen af hvert ben for alle personerne og derved fås dette plot:



Figur 5: Her ses en sammenligning af gennemsnittet af vinklen alle personernes dominante og ikke dominante ben

Det ses tydeligt ud fra dette plot at der ikke er særlig stor forskel på gennemsnitsvinklen i forhold til tiden. Altså kan man opstille en nul-hypotese som lyder på at der ikke er nogen signifikant forskel på gennemsnitsvinklen i de to ben.

5.2 Permutationstest

Permutationstest er en metode hvorpå man kan teste en nul-hypotese, hvor fordelingen af teststørrelsen bestemmes ved at beregne teststørrelsen igen og igen for forskellige permutationer af observationerne.

Denne metode kan fx. bruges til at undersøge om der er signifikant forskel på to grupper. Det gøres ved at finde en passende teststørrelse, og derefter resample observationerne hvor man har byttet rundt på labels indenfor de to grupper, og benytter de nye værdier af teststørrelsen til at finde en kritisk værdi for testet.

Lad $X = (X_{i=1,\dots,n})$ komme fra en fordeling og opstil en nul-hypotese:

$$H_0 : \mu_1 = \mu_2 \text{ imod } \mu_1 \neq \mu_2$$

Hvor μ_j angiver middelværdien for den j 'te gruppe.

I dette tilfælde vil det være oplagt at bruge forskellen på middelværdien af de to grupper som teststørrelse:

$$T(x) = \|\hat{\mu}_1(x) - \hat{\mu}_2(x)\|_2$$

Nu kunne det overvejes hvad fordelingen af teststørrelsen $T(x)$ er under H_0 , men dette er ikke nødvendigt i en permutationstest, da alle teststørrelserne vil komme fra samme fordeling. Hvis man under H_0 antager at X_i 'erne følger samme fordeling i de to grupper, kan man betragte en mængde G af permutationer, som er bijektive afbildninger: $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$, hvorom der gælder at $X_\sigma = (X_{\sigma(i)})$ vil have samme fordeling som X for alle $\sigma \in G$. Derved bør det gælde at den observerede teststørrelse $T(x)$ ikke er ekstrem i forhold til $T(x_{\sigma_1}), \dots, T(x_{\sigma_n})$. Dette betyder at teststørrelsen vil være en del af acceptmængden. Det kan dog ske at teststørrelsen vil være i den kritiske mængde og derved må man forkaste hypotesen.

5.2.1 Eksempel

Antag at man har X_1, \dots, X_n uafhængige identisk fordelte stokastiske variable, en permutation $\sigma \in G$ og en teststørrelse $T(x)$.

Nu opstilles en nulhypotese, f.eks kunne det tænkes at

$$\begin{aligned} \mathbb{E}X_1 &= \dots = \mathbb{E}X_n = \mu_1 \\ \mathbb{E}X_{n+1} &= \dots = \mathbb{E}X_m = \mu_2 \end{aligned}$$

og derved kan nul-hypotesen opstilles ved $H_0 : \mu_1 = \mu_2$ mod $\mu_1 \neq \mu_2$.

Under denne hypotese vil (X_1, \dots, X_n) have samme fordeling som $(X_{\sigma_1}, \dots, X_{\sigma_n})$ og (X_{n+1}, \dots, X_m) have samme fordeling som $(X_{\sigma_{n+1}}, \dots, X_{\sigma_m})$, for et godt valg af G . Altså skal man overveje hvad G skal være. Det kunne fx. være at X_1, \dots, X_n var observationer af kvinder og $X_{n+1} \dots X_m$ var observationer af mænd, og derved kunne det tænkes at $\sigma \in G$ skulle permutere kvinder og mænd indbyrdes. Herved kan man finde den observerede teststørrelse $T(x)$ og sammenligne den med $T(x_{\sigma_1}), \dots, T(x_{\sigma_m})$ og se om den observerede testværdi er ekstrem.

5.2.2 Permutationstest på forskel mellem ben

Da vi ikke ved meget om observationerne, ville denne metode være god at bruge, da vi kan antage at alle de approksimerede funktioner kommer fra den samme fordeling indenfor de to ben.

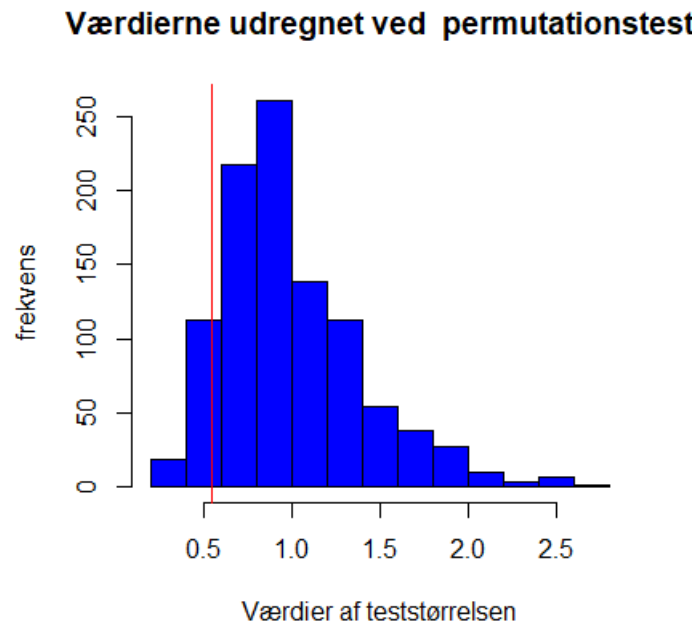
Denne metode benyttes til at undersøge hvorvidt der er signifikant forskel på vinklen af testpersonernes dominante og ikke-dominante ben. Dermed ses observationerne som $X = (X_1, \dots, X_{16})$ og $Y = (Y_1, \dots, Y_{16})$ stokastiske variable, hvor X er vinkelobservationerne for de 16 personers dominante ben og Y for deres ikke-dominante ben. Dermed vil stikprøvefunktionen være $\|\hat{\mu}_X - \hat{\mu}_Y\|_2$. Da det er forskellen på vinklerne mellem benene vi gerne vil kigge på, vælges $T(x)$ til at være L^2 -størrelsen af gennemsnittet af vinklerne for hvert ben, for hver person:

$$T(x) = \sqrt{\frac{\sum_{i=1}^{101} (X(t_i) - Y(t_i))^2}{101}} = 0.5481253$$

Nu opstilles nulhypotesen ved $H_0 : \mu_X = \mu_Y$ imod $H_1 : \mu_X \neq \mu_Y$ og derved findes en mængde G af permutationer σ sådan at $X_\sigma = (X_{\sigma(i)})$ har samme fordeling som $X = (X_i)$ og $Y_\sigma = (Y_{\sigma(i)})$ har samme fordeling som $Y = (Y_i)$, for alle $\sigma \in G$.

Derved laves en mængde G af permutationer der sørger for at personernes ben ikke byttes rundt med andre personer. Dette løses ved at lave to "puljer": k_1 og k_2 . Herved kan der laves en funktion, der sørger for, at hvis en persons dominante ben fx. kommer i k_1 , så skal personens andet ben ind i k_2 , og derved undgå problemet hvor en persons ben begge kommer i samme "pulje".

Efter dette er gjort, permuteres k_1 og k_2 indbyrdes 5000 gange, og hver gang udregnes teststørrelsen $T(x_{\sigma_1}), \dots, T(x_{\sigma_{5000}})$.



Figur 6: Her ses fordelingen af teststørrelsen udregnet ved permutationstest, hvor den røde streg er den observerede teststørrelse

P -værdien udregnes nu ved at tage antallet af teststørrelser der er større end den observerede teststørrelse delt med 5000, da dette er antallet af teststørrelse der er beregnet og vi får en P -værdi på $1 - \frac{486}{5000} = 0.9082$. Da P -værdien skal være over 0,05 for at ligge i acceptmængden, kan nul-hypotesen ikke forkastes.

5.3 Prædiktions- og Konfidensintervaller

Nu hvor det er bestemt at der ikke er nogen signifikant forskel på de to ben, kunne det være interessant at prædiktere et interval der siger noget om hvor stor forskel der bør være på en ny testperson, og derved også sige noget om hvor stor forskel der må være på vinklerne, før der er noget galt. Derudover vil vi kigge på et konfidensinterval for gennemsnittet af forskellen mellem vinklerne.

5.3.1 Limits of Agreement (LoA)

Limits of Agreement er en metode man typisk bruger til sammenligne to forskellige metoder til at måle specifikke ting. Man bruger LoA til at bestemme hvor stor

forskel der er på de to målemetoder, og hvor godt man kan sammenligne dem. Mere specifikt giver denne metode et estimat af et interval der angiver hvor stor forskel der bør være på de to målemetoder, og hvor ”enige” de er. Begrebet er i første omgang udviklet i univariate numeriske målinger: Lad $X_{i,j}$ være stokastiske variable hvor i er metoden og j er objektet man vil måle, og kig på $Y_j = X_{1,j} - X_{2,j}$. Hvis nu man regner med at Y_j er uafhængige, identisk fordelte stokastiske variable, og antager at Y_j er normalfordelt kan man lave et 95 % LoA således:

$$\mathbb{E}Y_j \pm 1.96 \cdot \text{SD}(Y_j)$$

I praksis erstattes middelværdi og standardafvigelse med deres empiriske værdier som er beregnet ud fra de observerede data. Metoden er motiveret af en antagelse om, at Y_j 'erne er normalfordelte, således at intervallet vil indeholde ca. 95 % af observationerne.

Dog er der forskel på om man vil kigge LoA punktvis, eller joint. Dette kommer an på hvilken hypotese man har lyst til at teste. Hvis man vil kigge på LoA for enkelte målinger kan man benytte det punktwise interval, men hvis man vil kigge på et interval for hele funktioner, så kigger man på et joint bånd.

Hvis man fx. vil kigge på LoA båndet for den maksimale vinkel for hvert ben, ville det være bedst at kigge på det punktwise bånd, da det er enkelte punkter der kigges på, men hvis man vil opstille en hypotese om at de to målinger er meget enige for hele funktioner, bliver man nødt til at kigge på det joint interval.

Når man har med funktionelle observationer at gøre, gøres dette derfor på en lidt anden måde. Da vi har med funktionelle observationer at gøre ønskes det derfor at lave et LoA til hvert tidspunkt. Dette kan gøres ved

$$\mathbb{E}Y_j(t) \pm 1.96\text{SD}(Y_j(t))$$

hvor t er tiden.

Herved kan man få et bånd om ens funktionelle observationer og derved bestemme hvor stor enighed der er imellem de to observationer for funktionelle data. Hvis dette LoA interval er lille betyder det at de to metoder er meget enige og omvendt hvis intervallet er stort betyder det at de ikke er enige.

Hvis man nu har fundet en ny og billigere metode som måler en specifik ting, kan man benytte LoA til at se hvor ”enig” den nye metode er med den gamle, og hvis de er ”enige” nok (intervallet er lille nok), kan man benytte den nye metode. Hvornår intervallet er lille nok, variere i forhold til hvad man kigger på.

Da man, pga. tidligere test, kan se det dominerende og ikke-dominerende ben som to målinger af samme ben kan man bruge LoA til at give et bud på hvor meget en måling på en ny person bør afvige fra de målinger der allerede er lavet, før der er noget galt f.eks at der er en skade på det ene ben. Forskellen mellem de to målinger på samme individ kan benyttes til at estimere variansen på forskellen mellem to målinger på en given person.

I dette projekt benyttes LoA altså til at lave et ”usikkerhedsinterval” for forskellen mellem et par af målinger på en ny person.

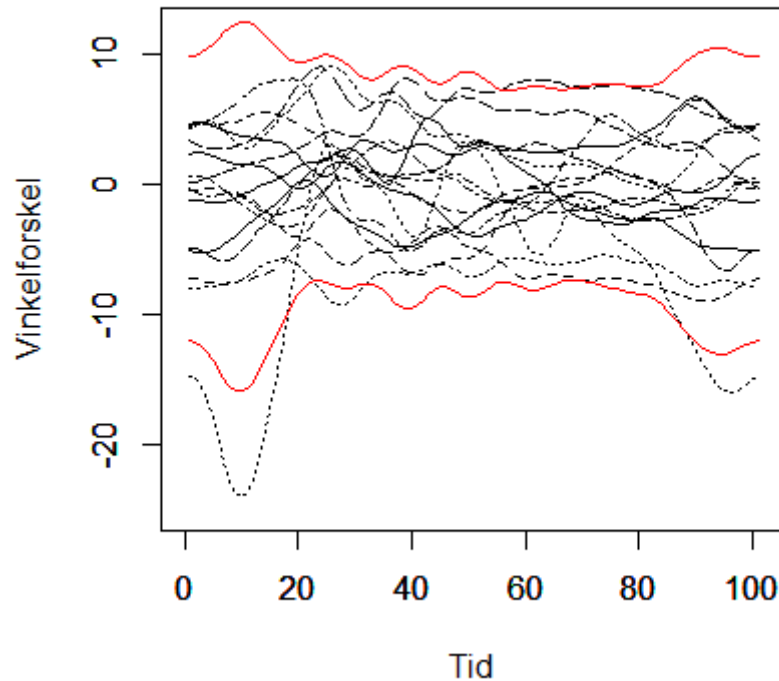
For at finde dette LoA bånd udregnes forskellen mellem vinklen på de to målinger for hver person samt middelværdien og standardafvigelsen til hver tid for disse. Derved kan punktvis LoA bånd fremstilles. Det vil sige at hvis man kigger på hvert tidspunkt, hvor meget må målingen på den nye person så afvige fra de observerede data. Lad $Z = X_{1,j} - Y_{1,j}$, hvor X er det ene ben og Y det andet. Så kan det punktwise LoA bånd udregnes således:

$$\mathbb{E}Z_j(t) \pm 1.96 \cdot \text{SD}(Z_j(t))$$

Da datasættet kun indeholder 16 observationer vil det være svært at vurdere et par ting:

1. Om intervallerne / LoA til hvert tidspunkt indeholder 95 % af observationerne
2. Om det simultane område indeholder 95 % eller evt. færre af det samlede antal kurver.

Dog kan man ud fra plottet nedenfor se på hvor stor en del af observationerne der ligger helt indenfor intervallet.



Figur 7: Alle forskelle med joint LoA bånd

Udfra ovenstående plot ses det tydeligt at over 95 % af disse værdier ligger punktvis indenfor det røde interval og det er udregnet at resten ligger indenfor intervallet. Dog er der forskel på om man interesserer sig for et punktvis eller joint LoA bånd. Det ses at der er 2 af kurverne der ikke ligger helt inden for intervallet (en under og en over), hvilket vil sige at der kun er $100 - \frac{2}{16} \cdot 100 = 87.5\%$ af disse observationer der ligger inden for dette bånd. Hvis man vil have et 95 % joint LoA bånd skal man altså overveje hvordan man skal lave dette.

Da der her ikke kigges på særlig mange observationer er dette selvfølgelig ikke nok til at sige noget særlig præcist om dette interval, men hvis man nu havde 200 observationer i stedet for 16, kunne man overveje hvor stor en procentdel der er inden for dette interval. Typisk vil det gælde at procendelen af kurver der ligger helt inden for intervallet vil være en del under 95 % hvis man kigger på det punktwise LoA bånd.

5.4 95 % konfidensbånd for gennemsnittet af forskellen

Nu hvor vi har dannet os et indtryk af en prædiktion af hvor meget forskellen mellem vinklen bør afvige på en ny testperson, kunne det være interessant at kigge på et konfidensinterval for gennemsnittet af forskellen af vores observerede data. Dette ligner meget LoA som tidligere er beskrevet, dog er det muligt at sige noget om dette konfidensinterval med større præcision.

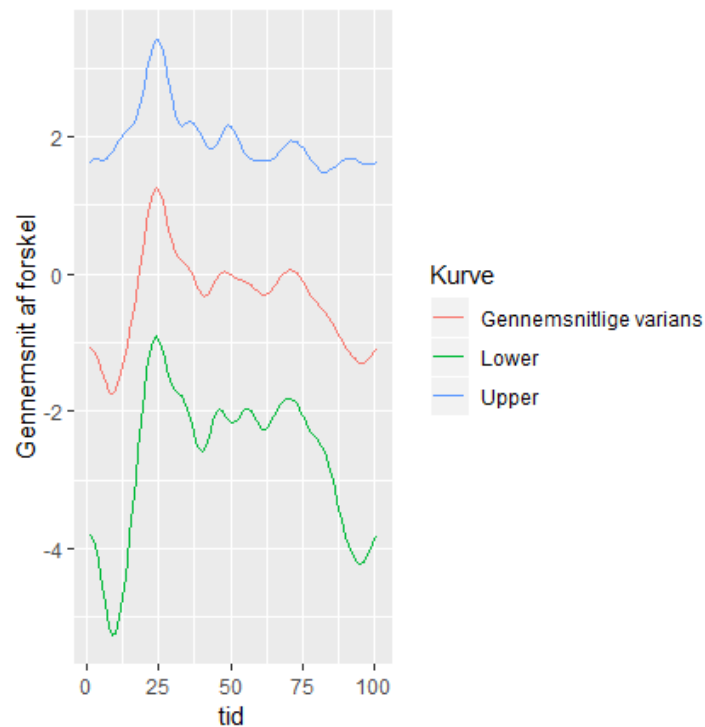
Lad $Y_j = \mathbb{E}(X_{1,j} - X_{2,j})$. Herved kan et 95 % konfidensinterval for gennemsnittet mellem de to forskelle udregnes således:

$$Y_j \pm 1.96 \cdot \text{SD}(Y_j) \cdot \frac{1}{\sqrt{n}},$$

hvor n er antal observationer.

Dette gøres ved at benytte bootstrap proceduren til at få produceret nye datasæt. Bootstrap proceduren er en resampling metode, hvor man samler ens datasæt. Dette kan både gøres med og uden tilbagelægning, men i dette projekt benyttes bootstrap med tilbagelægning. Altså trækker man observationer fra et datasæt en af gangen og herefter lægger man dem ind i datasættet igen. Det vil sige at man kan få den samme observation ud flere end en gang i det resamplede datasæt. Derved vil det resamplede datasæt højst sandsynligt ikke være det samme som det man startede med, og derved kan man sige noget om variationen i et gennemsnit beregnet på baggrund af datasæt bestående af nogle funktioner.

Denne procedure benyttes altså på datasæt hvor man har udregnet gennemsnittet af forskellen mellem de to målinger for hver person, og der bruges tilbagelægning, så gennemsnittet af forskellen for en person godt kan fremkomme flere gange i det samme datasæt. Når dette er gjort udregnes hvor stor en procentdel der ligger i det 95 % konfidensbånd der blev beskrevet før, og det udregnes at omkring 95 % af de simulerede værdier ligger inden for dette interval punktvis.



Figur 8: Her ses det punktvis 95 % konfidensbånd

Igen er der dog forskel på om man interressere sig for et punktvis eller et joint konfidensbånd.

Da dette også vil være interessant at kigge på ses der nu på hvor mange af funktionerne der ligger helt indenfor båndet. Dette gøres ved at tjekke om observationerne til hver tid ligger helt inde i intervallet, og det udregnes at kun omkring 80 % af kurverne ligger helt inden for dette interval tidligere blev udregnet.

Da man gerne vil have at der er 95 % sandsynlighed for at kurverne ligger helt inden for det interval, prøves at lave nogle nye intervaller for at se hvor stort dette skal være.

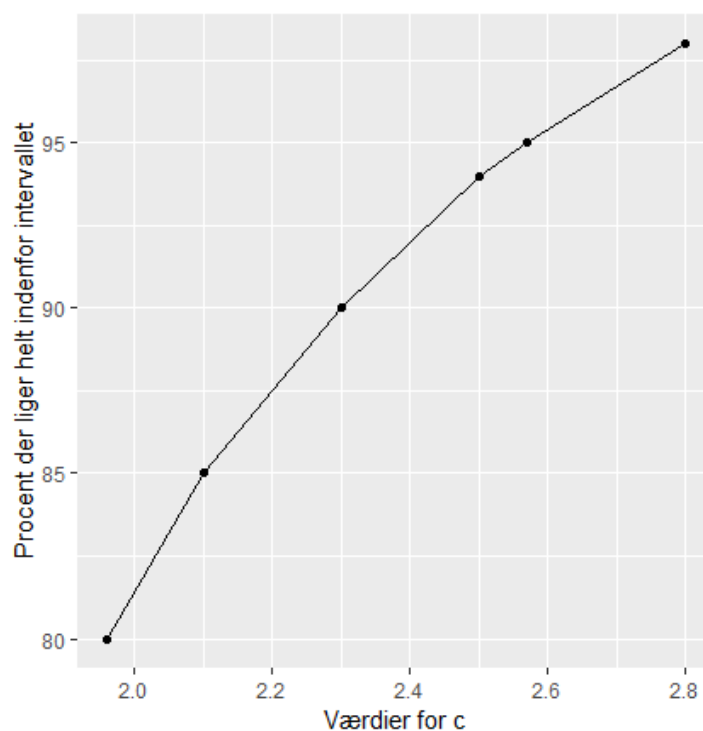
Hvis nu at $Y_j(t) = \mathbb{E}(X_{1,j}(t) - X_{2,j}(t))$, kan dette interval udregnes ved:

$$Y_j(t) \pm c \cdot \text{SD}(Y_j(t)) \cdot \frac{1}{\sqrt{n}}$$

hvor n er antal observationer. og vi kan kigge på hvad c skal være.

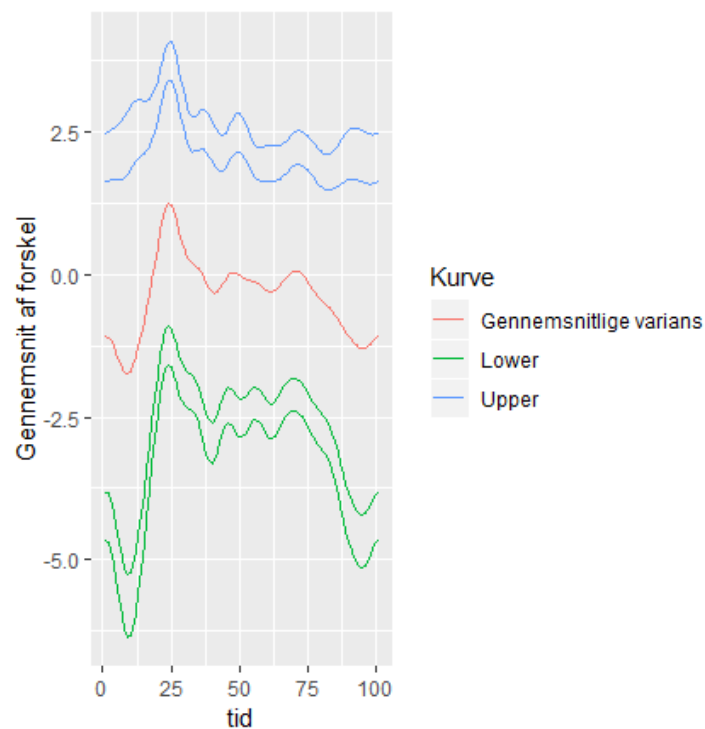
Hvis man sætter $c = 1.96$ er det som nævnt før kun omkring 80% af kurverne der ligger indenfor intervallet, og der prøves nogle større værdier for c . Hvis man sætter $c = 2.3$ fås det at det er omkring 90% af dem der ligger inden for intervallet, og det

ses til sidst at 95% af kurverne vil ligge inden for intervallet hvis man sætter $c = 2.57$



Figur 9: Her ses hvor stor en procentdel af kurverne der ligger helt indenfor intervallet i forhold til hvad c er

Nedenfor ses et plot af gennemsnittet af forskellen mellem vinklerne på de 16 personers ben, hvor begge konfidensbånd er indsat.



Figur 10: Her ses både det punktvisse og det joint konfidensinterval

I figuren ovenfor ses de to intervaller. Det indre interval (altså den nederste blå- og den øverste grønne kurve) er det punktvisse konfidensbånd, mens det ydre interval er det joint konfidensbånd.

6 Fremtidige analyser

Hvis man har disse resultater og gerne ville gå videre med dem, hvad kunne man så kigge på?

Man kunne se om der var en signifikant forskel imellem vinklerne imellem de forskellige personer, hvilket også kunne gøres ved permutationstest. Derudover kunne man se på om ting som vægt, højde og BMI havde en indflydelse på hvordan vinklen ser ud.

Hvis man nu lavede et nyt forsøg, hvor man vidste at personerne havde haft gamle skader i benet, så kunne man se på hvilken effekt dette ville have på forskellen af vinklerne imellem benene. Det kunne også være interessant at kigge på vinklen, hvis

testpersonerne løb på løbebåndet i stedet for at gå.

7 Konklusion

Ved at benytte permutationstest kan det konkluderes at der ikke er signifikant forskel på vinklen imellem over- og under-ben på det dominerende og det ikke-dominerende ben.

Det kan også konkluderes, ud fra det udregnede Limits of Agreement bånd, der er udregnet, at hvis en ny test-person skulle lave det samme forsøg, så vil forskellen på vinklen i de to ben med 95 % sandsynlighed ligge i dette interval, hvis der ikke er andre effekter der spiller ind, som en skade på et ben. Dog kunne denne metode ikke sige noget særlig præcist om de observerede data, da der ikke var nok observationer. Efter man har udregnet de to konfidensbånd er pointen nu, at hvis man fx. vil teste hypotesen om, at forskellen i vinkler mellem de to ben er $= 0$ til alle tidspunkter, så vil man kunne bruge et test baseret på, om nul-funktionen er helt indeholdt i intervallet. Med denne fremgangsmåde er det vigtigt, at man har styr på niveauet af testet forstået således, at man ved at der er 95 % sandsynlighed for, at hele kurven ligger inden for intervallet. Dette kræver at man benytter et interval med en korrekt simultan dækningsgrad og ikke blot det punktvis bånd med for lav dækningsgrad. Ved at betragte figur 10 ses det, at man godkender hypotesen om, at middelforskellen mellem vinklerne kunne være $= 0$ til alle tidspunkter.

Litteratur

- [1] J. Røislien , L. Rennie , I. Skaaret.

Functional limits of agreement: A method for assessing agreement between measurements of gait curves.

2012

- [2] Ernst Hansen.

Introduktion til Matematisk Statistik Bind 2.

3. udgave, 2012