



# Kliniska stabilitetsstudier

En utvärdering av statistiska metoder för longitudinella stabilitetsstudier

*Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet*

Calvin Smith

Jens Ifver

Jonatan Hellgren

Josef Gullholm



# Kliniska stabilitetsstudier

En utvärdering av statistiska metoder för longitudinella stabilitetsstudier

*Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet*

Calvin Smith

Jens Ifver

Jonatan Hellgren

Josef Gullholm

Handledare: Henrik Imberg    Matematiska Vetenskaper  
Per Granath    Mölnlycke Health Care

Institutionen för Matematiska vetenskaper  
CHALMERS TEKNISKA HÖGSKOLA  
GÖTEBORGS UNIVERSITET  
Göteborg, Sverige 2020



## Förord

Den här rapporten är ett resultat av kandidatarbetet *Kliniska stabilitetsstudier*. Arbetet har genomförts under vårterminen 2020 på institutionen för Matematiska vetenskaper vid Göteborgs universitet samt på Mölnlycke Health Care (MHC). Vi som har skrivit arbetet kommer under detta eller kommande år att ta ut en kandidatexamen i Matematisk statistik. Henrik Imberg, doktorand inom tillämpad matematik och matematisk statistik, har varit handledare på universitetet. Kvalitetsingenjör Per Granath har varit vår handledare på MHC.

Vi har under arbetets gång fört en loggbok där vi redovisat vad respektive deltagare arbetat med varje vecka. Denna uppgift har roterats så att var och en har gjort det var fjärde vecka.

Nedan redovisas vilka delar som respektive person har varit ansvarig för, både i utförande samt att skriva i rapporten. För tydlighetens skull vill vi poängtera att det har varit ett mycket gott samarbete, alla har hjälpt varandra med respektive ansvarsområden.

Calvin Smith har varit delansvarig för simuleringsmetoderna och huvudansvarig för sammanfattningen, inledningen samt teoridelen för linjär regression.

Jens Ifver har varit huvudansvarig för härledning och applikation av framtagna metoder och formler för Ppk-beräkning samt teoridelen där betingad och marginell tolkning av mixade modeller behandlas. Jens har varit delansvarig för arbetet om effekten av variation inom batcher på Ppk och p-värde.

Jonatan Hellgren har varit delansvarig för simuleringsmetoderna och huvudansvarig för den populärvetenskapliga rapporten, visualisering av modeller samt teoridelen för mixade modeller.

Josef Gullholm har varit huvudansvarig för kontakt med MHC och arbetet kring effekten av variation inom batcher på Ppk och p-värde samt även även statistisk processkontroll-delen av den teoretiska bakgrunden. Josef har varit delansvarig i att ta fram härledning och applikation av framtagna metoder och formler för Ppk-beräkning.

Vi vill rikta ett stort tack till vår handledare på MHC, Per Granath som visat oss hur MHC arbetar med sina stabilitetsstudier och varit behjälplig när det har behövts.

Vi vill även rikta ett stort tack till vår handledare Henrik Imberg som alltid varit tillgänglig för funderingar samt med en fast hand visat oss i rätt riktning när det behövts.

Institutionen för Matematiska vetenskaper vid Göteborgs universitet, maj 2020.

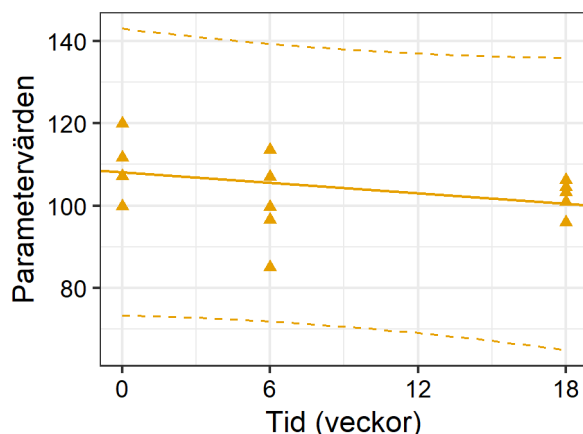
Calvin Smith  
Jens Ifver  
Jonatan Hellgren  
Josef Gullholm

## Populärvetenskaplig presentation

Inom sjukvården kan inga risker tas när det gäller människors liv och hälsa, därför krävs det att materialet som brukas håller en hög standard. För att denna standard ska uppnås krävs det noggranna kontroller som utförs genom stabilitetsstudier. I denna rapport har vi analyserat stabilitetsstudier utförda av Mölnlycke Health Care, ett världsledande företag inom produktion av engångsprodukter för operation och sårbehandling.

Syftet med en stabilitetsstudie är att mäta hållbarheten av en produkt när den åldras, vilket sker genom att observera parametrar från olika tester på ett stickprov av produkterna. Produkterna som undersöks i en stabilitetsstudie kan variera från bilmotorer till läkemedel. I denna rapport har vi analyserat stabilitetsstudier på olika sorters engångsprodukter för operation och sårbehandling, exempelvis kirurgihandskar eller bandagelindor. Parametrar som studeras kan till exempel vara vilket tryck en kirurgihandske klarar av innan den brister eller hur mycket vätska en bandagelinda släpper igenom.

När parametrarna i en stabilitetsstudie är observerade vill vi därefter beräkna dess medelvärde och förväntade variation, för att på så sätt uppskatta produktens hållbarhet. I Mölnlycke Health Cares stabilitetsstudier görs mätningar normalt sett vid upprepade testtillfällen, vilket innebär att vi behöver ta hänsyn till en eventuell tidstrend när vi utvärderar produktens hållbarhet. Detta görs ofta med så kallad regressionsanalys, där vi kan beskriva trenden med hjälp av en linje. Denna linje väljer vi så att den avviker ifrån observationerna så lite som möjligt. Regressionsanalys används ofta inom statistik för att beskriva ett samband mellan en responsvariabel och en eller flera förklarande variabler. I vårt fall utgörs responsvariabeln av de observerade parametrarna samt de förklarande variablerna av tid (i veckor) efter produktion och vilken produktionsbatch produkten tillhör. I Figur 1 illustreras ett exempel på hur en regressionsanalys ser ut.

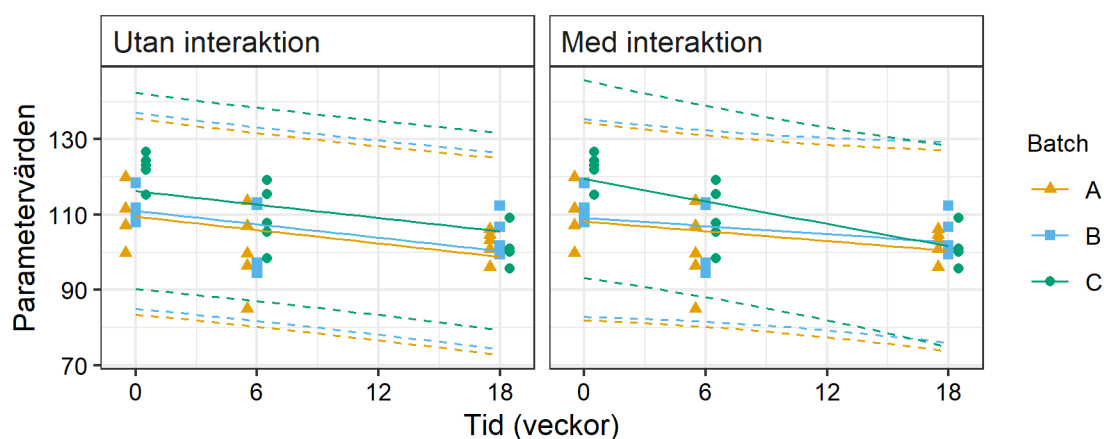


**Figur 1:** I figuren kan vi se fem olika observerade parametervärden per tidstillfälle (0, 6, 18) som illustreras med gula punkter. Trenden som beräknas med regressionsanalys illustreras med en gul linje och som vi kan se, beskriver den väl tidstrenden av observationerna. Dessutom kan vi också se prediktionsintervallet på 99.7% som utgörs av de två streckade linjerna. Specifikationskravet i denna studien är 40, vilket är under prediktionsintervallets undre gräns vid 18 veckor som är cirka 50, och därmed är denna batch godkänd. Parameterobservationerna utgörs i detta fallet, av vilken kraft ett tunt material klarar av innan det brister, vilket kan vara exempelvis en kirurgihandske.

I en stabilitetsstudie vill vi garantera, att en viss andel (ofta 997 av 1000) produkter kommer att uppfylla specifikationskravet under produktens rekommenderade hållbarhetstid. För att göra detta behöver vi beskriva den förväntade fördelningen av framtida observationer. När vi använder oss av regressionsanalys kan vi exempelvis beräkna ett prediktionsintervall, vilket består av en undre

samt övre gräns. Dessa gränser beräknas så att vi kan förvänta oss, att en viss andel av framtida observationer (i vår studie 99.7%) kommer att falla inom dessa gränser. Ett prediktionsintervall kan användas i stabilitetsstudier, genom att en produkt godkänns om specifikationsgränsen är utanför prediktionsintervallets gränser vid tiden för stabilitetsstudiens sista tidstillfälle, vilket innebär att vi förväntar oss att minst 99.7% av produkterna kommer att uppfylla specifikationskravet. I Figur 1 illustreras även ett prediktionsintervall på den tidigare nämnda trenden som två streckade linjer.

För stabilitetsstudier där hållbarheten av en produkt följs upp över tid och där det dessutom finns flera olika källor till variation att ta hänsyn till, är det inte alltid självklart vilken statistisk analys som skall användas. Exempelvis utför ofta Mölnlycke Health Care sina stabilitetsstudier på tre batcher av samma produkt. När vi har flera batcher, så kan vi antingen anta att hållbarheten av de olika batcherna förändras på samma sätt över tid, eller att vissa batcher försämras snabbare eller långsammare än andra. Detta kallas att det finns en interaktion mellan batch och tid och illustreras i Figur 2. När analysen utförs på flera batcher så krävs det att samtliga batcher klarar specifikationskravet för att de ska bli godkända, det vill säga att om specifikationskravet är innanför någon batchs prediktionsintervall så blir ingen batch godkänd.



**Figur 2:** I figuren ser vi hur en regressionsanalys visualiseras när vi har tre batcher där parametervärdena är observerade under tre tillfällen (0, 6, 18). Vi ser regressionsanalysen både utan interaktion samt med interaktion. Vi kan se trendlinjen samt prediktionsintervallet för respektive batch. Dessa blir parallella och förskjutna i y-led när vi använder oss av en analys utan interaktion, men i en analys med interaktion får de även olika lutningar.

Syftet med vår studie var att undersöka hur olika analyser och beräkningar för Mölnlycke Health Cares stabilitetsstudier påverkar resultatet. Detta gjorde vi genom att jämföra analyser utan interaktion, mot analyser med interaktion, men vi har dessutom jämfört enkla mot mer komplexa analyser. Utöver detta undersökte vi hur eventuella mätfel påverkar resultatet. I studien har vi använt oss av olika statistiska metoder och simuleringar för att studera detta. Simuleringar används som ett verktyg för att generera nya likartade utfall av parametervärden.

Resultaten av vår studie visade, att med Mölnlycke Health Cares stabilitetsstudier kan det vara svårt att se en skillnad mellan analyser med och utan interaktion, med avseende på antalet godkända studier. Däremot visade våra simuleringar, att om vi använder analyser utan en interaktion riskerar vi att felberäkna resultatet. Våra resultat visade också att den komplexa analysen är mindre känslig för eventuella mätfel. Sammanfattningsvis kan vi säga, att för den här typen av studiedesign är en mer komplex analys med interaktion att föredra.

## Sammanfattning

Mölnlycke Health Care (MHC) utvecklar och tillverkar produkter inom sårvård och kirurgi. För att säkerställa hållbarheten av sina produkter genomför MHC kliniska stabilitetsstudier. Stabilitetsstudierna utförs med stickprov från ett antal olika produktionsbatcher som sedan testas, med avseende på hållbarhet, över en bestämd tidsperiod. Stabilitetsstudierna analyseras med linjära regressionsmodeller, både med och utan slump effekter. Modellerna innehåller parametrarna tid och produktionsbatch, där produktionsbatch kan betraktas som antingen en slumpmässig parameter eller icke slumpmässig parameter. Från analysen tar man fram ett värde på process-prestationsindex (Ppk), för att avgöra om processen, det vill säga hållbarheten av produkten, är godkänd.

Syftet med arbetet är att undersöka och utveckla statistiska metoder för analys av kliniska stabilitetsstudier med longitudinella data. Vi har särskilt inriktat oss på att utveckla en metod för beräkning av Ppk med longitudinella data samt undersökt effekterna, i form av förändring i Ppk, av att introducera en interaktionsterm mellan tid och batch i MHC:s modeller. Vi har dessutom undersökt hur robusta modellerna är i relation till variationen i datan. Vi har analyserat 13 stabilitetsstudier ifrån MHC. För att vidare analysera effekten av att introducera en interaktionsterm mellan tid och produktionsbatch har vi använt oss av simuleringar.

Resultatet av vårt arbete visar att förändringen i Ppk på genomförda stabilitetsstudier, vid en analys som tar hänsyn till interaktionen mellan tid och batch, jämfört med en enklare analys utan interaktion, är mycket liten. Dessutom visar våra resultat att ökad variation i datan leder till lägre Ppk-skattningar samt högre p-värden för signifikansen med avseende på tid. Vidare är det svårt att hitta en signifikant interaktion mellan tid och batch vid analys av MHC:s stabilitetsstudier. Detta kan bero på att studierna endast innehåller ett fåtal batcher och observationer per batch. För att på ett effektivt sätt kunna identifiera en interaktion så krävs det fler batcher och observationer. Våra simuleringar visar, att sannolikheten att hitta en interaktion mellan tid och batch ökar, ju fler batcher och observationer per batch man har. Simuleringarna visar också, att felaktig inklusion av en interaktionsterm, när en interaktion inte existerar, påverkar Ppk i mindre utsträckning, än om vi felaktigt bortser från en interaktionsterm när det existerar en interaktion.

## Abstract

Mölnlycke Health Care (MHC) is a company that develops and produces products within wound care and surgery. To examine the durability of their products, MHC performs clinical stability studies. The stability studies are performed with samples from a number of different production batches that are then tested, with regards to durability, over a certain period of time. The stability studies are analyzed with linear models, both with and without random effects. The model parameters are time and batch, where batch can be modelled as a fixed effect or a random effect. From the analysis, a value of the Process Performance Index (Ppk) is produced in order to determine if the process, i.e the quality of the product, is approved.

The aim of this study is to examine and develop statistical methods for analysis of clinical stability studies with longitudinal data. We have especially focused on developing a method for computing Ppk with longitudinal data and investigating the effects, with regards to change in Ppk, of introducing an interaction between time and batch in MHC:s models. We have also examined the robustness of the models with respect to the variation in the data. We have analyzed 13 completed stability studies from MHC. To further investigate the effects of an interaction between time and batch, we have used simulations.

The results of our study show, that including an interaction in the analysis of MHC's stability studies, has a very small effect on the Ppk. Our results also show that increased variation in the data leads to lower Ppk estimates and higher p-values for the significance with respect to time. Further, it is difficult to find a significant interaction between time and batch, when analyzing MHC's stability studies. This could be due to the fact, that the studies only contain three batches and a handful of samples per batch. To be effective in detecting a significant interaction, an increased number of batches and samples per batch should be included. Our simulations show that the probability of finding an interaction between time and batch increases when we increase the number of batches and samples per batch. Simulations also show that wrongfully including an interaction term, when there doesn't exist an interaction, doesn't effect Ppk as much as wrongfully excluding an interaction term, when there exists an interaction.



# Innehåll

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Inledning</b>   | <b>1</b>  |
| 1.1      | Bakgrund . . . . .   | 1         |
| 1.2      | Syfte . . . . .  | 1         |
| 1.3      | Delproblem . . . . .   | 1         |
| 1.4      | Metod och avgränsningar . . . . .  | 2         |
| <b>2</b> | <b>Teoretisk bakgrund</b>  | <b>2</b>  |
| 2.1      | Statistisk processkontroll . . . . .   | 2         |
| 2.2      | Linjär regression . . . . .  | 3         |
| 2.3      | Linjära modeller med mixade effekter . . . . .   | 5         |
| 2.4      | Betingad och marginell tolkning av modell med mixade effekter . . . . .  | 6         |
| <b>3</b> | <b>Ppk-beräkning för linjära regressionsmodeller med och utan mixade effekter</b>                              | <b>7</b>  |
| 3.1      | Ppk-beräkning baserat på prediktionsintervall . . . . .  | 7         |
| 3.2      | Standardavvikelse av prediktioner skattade från modeller med mixade effekter . . . . .                         | 8         |
| <b>4</b> | <b>Metod för utvärdering och analys av MHC:s stabilitetsstudier</b>  | <b>9</b>  |
| 4.1      | Data . . . . .   | 9         |
| 4.2      | Modeller . . . . .   | 10        |
| 4.3      | Applikation av metoder för Ppk-beräkning . . . . .   | 10        |
| 4.4      | Beräkningar och simuleringar . . . . .   | 11        |
| 4.4.1    | Effekt av att introducera en interaktion mellan batch och tid i genomförda stabilitetsstudier på MHC . . . . . | 11        |
| 4.4.2    | Effekt av förändrad variation inom batcher på Ppk-beräkningar och signifikanstest . . . . .                    | 11        |
| 4.4.3    | Simulering för att utvärdera möjligheten att identifiera en existerande interaktionseffekt . . . . .           | 12        |
| 4.4.4    | Simulering för att utvärdera effekten av en felaktig modellspecifikation . . . . .                             | 13        |
| <b>5</b> | <b>Resultat</b>  | <b>14</b> |
| 5.1      | Interaktionstermens effekt på Ppk för MHC:s stabilitetsstudier . . . . .                                       | 14        |
| 5.2      | Residualvariationens inverkan på Ppk-beräkningar och signifikanstest . . . . .                                 | 15        |
| 5.3      | Styrkan att identifiera en existerande interaktionseffekt . . . . .  | 17        |
| 5.4      | Konsekvens av felaktig modellspecifikation . . . . .   | 18        |
| <b>6</b> | <b>Diskussion</b>  | <b>18</b> |
| 6.1      | Sammanfattning . . . . .   | 18        |
| 6.2      | Metoddiskussion . . . . .  | 18        |
| 6.3      | Interaktionstermens betydelse . . . . .  | 19        |
| 6.4      | Variation inom batcher . . . . .   | 19        |
| 6.5      | Utvärdering av modeller med och utan slumpeffekter . . . . .   | 20        |
| 6.6      | Konklusion . . . . .   | 20        |
| <b>7</b> | <b>Appendix</b>  | <b>22</b> |

# 1 Inledning

## 1.1 Bakgrund

Mölnlycke Health Care (MHC) utvecklar och tillverkar produkter inom sårvård och kirurgi. För att undersöka kvaliteten av sina produkter utför MHC kliniska stabilitetsstudier. Kliniska stabilitetsstudier utförs med syftet, att undersöka hur hållbarheten av en produkt utvecklas över tid och under olika förutsättningar [1]. Detta görs för att kunna säkerställa hållbarhetstiden hos produkten. En stabilitetsstudies omfattning bestäms utifrån produktens egenskaper och användningsområden, samt vilka förutbestämda specifikationsgränser det finns med avseende på produktens hållbarhet. MHC:s stabilitetsstudier utförs med stickprov från ett antal olika produktionsbatcher som förvaras i en klimatkontrollerad miljö, där olika tester utförs över olika tidsintervall. Syftet med stabilitetsstudierna är att undersöka kvaliteten av produkterna över tid med avseende på ett antal olika parametrar. Efter respektive studie analyseras resultaten för att avgöra om de studerade parametrarna klarar specifikationskraven [2]. Specifikationskraven bestäms utifrån ett värde på process-prestationsindex (Ppk), ett index inom statistisk processkontroll som används för att bedöma kvaliteten av en tillverkningsprocess med avseende på produkternas hållbarhet.

För att analysera stabilitetsstudierna använder sig MHC av linjära regressionsmodeller både med och utan slump effekter. En linjär regressionsmodell med slump effekter tillåter oss att behandla parametrarna i modellen som slumpmässiga variabler. I MHC:s modeller med slump effekter betraktas produktionsbatch som en slumpmässig parameter. Detta innebär att batcher som analyseras antas vara dragna ur en större population av batcher och vi kan då ta hänsyn till variationen av hela populationen. Vi kan då dra generella slutsatser om batchernas effekt på den studerade variabeln. Väljer vi istället att betrakta batch som en icke-slumpmässig parameter antar vi inte längre att batcher är dragna ur en större population. Vi kan då endast ta hänsyn till de specifika batcher som analyseras och deras effekt på den studerade variabeln. I MHC:s modeller, både med och utan slump effekter, antar man att varje batch påverkas av tid på samma sätt. Skulle detta visas grafiskt, där x-axeln är tid och y-axeln är den studerade variabeln så har vi tre olika linjer, en för varje batch, med olika skärningspunkter där samtliga linjer har samma lutning. Introduceras interaktionen mellan tid och batch som en parameter i modellen, så antas till skillnad från tidigare att det finns en batchspecifik tidstrend, alltså att det finns en skillnad i hur varje batch påverkas av tiden. Interaktionen mellan tid och batch kan introduceras både som en slumpmässig och en icke-slumpmässig parameter.

Vid beräkning av Ppk är det vanligtvis ett homogent stickprov av en och samma process. MHC:s stabilitetsstudier består av longitudinell data, det vill säga mätningar på samma objekt över ett tidsintervall. Studierna innehåller dessutom stickprov från ett antal olika produktionsbatcher som antas skilja sig från varandra. Vidare bygger analysen av MHC:s stabilitetsstudier på antagandet att samtliga produktionsbatcher har samma tidstrend. Att introducera en interaktionsterm mellan tid och batch i MHC:s modeller tillåter en mer flexibel modell och innebär att vi kan få ut mer information av analysen. Vi tar då hänsyn till varje batch och dess specifika påverkan av tid, vilket ger möjlighet till djupare analys och förståelse av processens utveckling. Vi kommer i detta arbete att undersöka effekten av att introducera en interaktionsterm mellan tid och batch i MHC:s modeller och hur det påverkar slutsatserna av MHC:s stabilitetsstudier. I kombination med detta kommer vi också att undersöka hur robusta modellerna är i relation till variationen i datan. Vidare kommer vi även att ta fram en metod för beräkning av Ppk anpassat efter longitudinella data.

## 1.2 Syfte

Syftet med arbetet är att utvärdera och utveckla statistiska metoder för analys av kliniska stabilitetsstudier för longitudinella data.

## 1.3 Delproblem

Med anknytning till ovanstående bakgrund och syfte kommer studien att behandla följande delproblem och frågor:

- Presentera en metod för beräkning av Ppk från framtagna modeller och föreslagna analyser anpassade efter MHC:s stabilitetsstudier.
- Vad blir effekten, i form av förändring i Ppk på genomförda stabilitetsstudier på MHC, vid en analys som tar hänsyn till en interaktion mellan tid och batch, jämfört med en enklare analys utan interaktion?
- Hur stor effekt har variationen inom batcher på huruvida parametrarna är signifikant beroende av tid? Hur stor effekt har variationen inom batcherna på Ppk-värdena?
- Vad blir effekten, i form av möjligheten att hitta en interaktion mellan tid och batch givet att det existerar en interaktion, om vi med hjälp av simuleringar ökar antalet batcher och antalet observationer per batch?
- Vad blir effekten på Ppk-beräkning om vi felaktigt inkluderar en interaktionsterm när en interaktion inte existerar?
- Vad blir effekten på Ppk-beräkning om vi felaktigt bortser från en interaktionsterm när det existerar en interaktion?
- Utvärdera fördelar och nackdelar av en linjär modell med och utan slump effekter för denna typ av stabilitetsstudie.
- Utifrån resultaten ge en rekommendation för design och analys av kommande stabilitetsstudier på MHC.

## 1.4 Metod och avgränsningar

För att kunna svara på ovanstående frågeställning och delproblem har vi studerat relevant teori inom statistisk processkontroll [3], linjär regression utan slump effekter [4] samt med slump effekter [5]. Denna teori kommer att presenteras i Kapitel 2. Utifrån den tillgängliga teorin har vi dessutom tagit fram egna beräkningar och härledningar som presenteras i Kapitel 3. Vidare har vi analyserat data från kliniska stabilitetsstudier genomförda på MHC och utifrån detta har en stor del av arbetet ägnats åt simuleringar. Analys och behandling av datan har huvudsakligen skett i programmeringspråket R [6]. Förutom grundläggande funktioner har vi särskilt använt oss av `lme4`-paketet [7] för linjära mixade modeller.

Arbetet behandlar enbart linjära modeller. Andra metoder för att modellera hållbarhet över tid, såsom modeller med icke-linjär tidstrend, samt metoder för att hantera beroenden i longitudinella data, såsom generaliserade minsta-kvadratmetoder och multivariata statistiska metoder, kommer ej att beaktas.

## 2 Teoretisk bakgrund

Vi kommer att beskriva statistisk processkontroll (Avsnitt 2.1) som kommer att användas för att undersöka stabilitetsstudierna. För att kunna analysera stabilitetsstudierna behövs även linjär regression (Avsnitt 2.2) samt teori för linjära modeller med mixade effekter (Avsnitt 2.3 och 2.4). Teorin från Avsnitt 2.1 kommer från [3], Avsnitt 2.2 är hämtad ur [4], teorin från Avsnitt 2.3 och 2.4 är hämtad ur [5].

### 2.1 Statistisk processkontroll

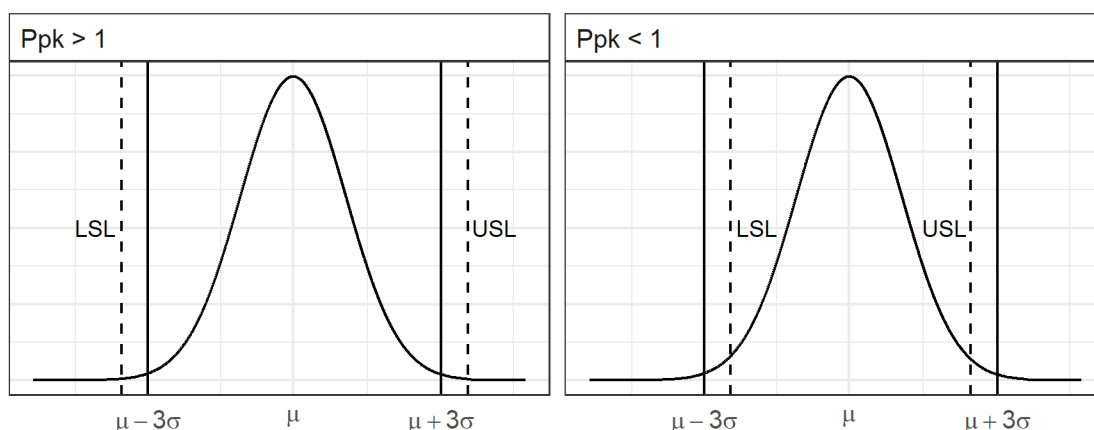
För att kunna tillverka produkter i stor skala, där målet är att alla produkter ska vara identiska, behöver tillverkningsprocesserna som ligger bakom den färdiga produkten kontrolleras genom att utföra olika tester på produkten. För att analysera dessa tester används statistisk processkontroll (SPK) som omfattar en mängd olika statistiska metoder. Dessa metoder används över hela näringslivet för att säkerställa att deras produkter klarar av olika kravställningar, både från kunder

och myndigheter, samt att säkerställa att deras produkter blir maximalt kostnadseffektiva.

Inom SPK används olika kapabilitets- och prestationsindex. Dessa används för att kvalitetssäkra olika produkter genom att undersöka ifall de klarar av leverantörers och/eller lagens gränsvärden för olika testparametrar. Det finns flera olika varianter av index, det vi kommer behandla i denna uppsats är process-prestationsindexet (Process performance index)

$$Ppk = \min\left(\frac{USL - \hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu} - LSL}{3\hat{\sigma}}\right), \quad (1)$$

där  $\hat{\mu}$  är en skattning av stickprovsmedelvärdet,  $\hat{\sigma}$  är den skattade standardavvikelsen av stickprovet, USL betecknar den övre specifikationsgränsen och LSL betecknar den lägre specifikationsgränsen. Specifikationsgränserna kan sättas av till exempel ingenjörerna som designat produkten eller av myndigheter. Uppskattningen av Ppk baseras på att det ska rymmas sex standardavvikelser ( $\sigma$ ) mellan de givna specifikationsgränserna (USL och LSL) om  $Ppk = 1$ , vilket visualiseras i Figur 3. För normalfördelade observerade testvärden ger detta, att specifikationsgränserna ska täcka 99.73% av processens observerade testvärden. Det ger oss att endast 0.27% av värdena kommer vara utanför gränserna, vilket ger oss att maximalt 2,7 av 1000 testvärden inte kommer att uppfylla specifikationskraven.



**Figur 3:** Fördelningskurvan visar hur testvärden från en process fördelar sig (under normalfördelningsantaganden) med  $\mu$  som stickprovsmedelsvärde och  $\sigma$  som standardavvikelse. Om USL och LSL befinner sig utanför de heldragna linjerna betyder det att  $Ppk > 1$  (till vänster). Om USL och LSL befinner sig innanför de heldragna linjerna betyder det att  $Ppk < 1$  (till höger).

## 2.2 Linjär regression

För att kunna studera hur testvärdena i MHC:s stabilitetsstudier utvecklas med avseende på tid och vilken produktionsbatch de tillhör, behövs en metod som på ett enkelt sätt kan förklara förhållandet mellan testvärden och tid, samt förhållandet mellan testvärden och produktionsbatch, men som även kan ta hänsyn till dessa faktorer samtidigt. Vi behöver även en metod som kan ta hänsyn till interaktionen mellan två faktorer, för att avgöra om effekten av en faktor beror på värdet av en annan.

Linjär regression är en metod som beskriver förhållandet mellan en responsvariabel och en eller flera förklarande variabler. Linjär regression är en flexibel metod som kan hantera både kontinuerliga och kategoriska variabler. Givet ett stickprov bestående av  $\{y_i, x_{i1}, \dots, x_{ip-1}\}_{i=1}^n$ , där  $y_i$  är det observerade värdet av responsvariabeln för observation  $i$ ,  $x_{i1}, \dots, x_{ip-1}$  är motsvarande värden för de förklarande variablerna och  $n$  är antalet observationer, kan en linjär regressionsmodell med  $p$  regressionskoefficienter  $\beta_0, \beta_1, \dots, \beta_{p-1}$ , samt  $n$  feltermen  $\varepsilon_i$ , skrivas på formen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{ip-1} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad (2)$$

där  $Y_i$  är responserna för observationen  $i$ ,  $\beta_0$  är interceptkoefficienten som beskriver linjens skärningspunkt med  $y$ -axeln, regressionskoefficienterna  $\beta_1, \dots, \beta_{p-1}$  beskriver relationen mellan de förklarande variablerna  $x_{i1}, \dots, x_{ip-1}$  och responserna  $Y_i$ ,  $\varepsilon_i \sim N(0, \sigma^2)$  är den normalfördelade feltermen för observation  $i$ . Detta ger oss formulerat i matrisform modellen

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

med

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np-1} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (4)$$

där  $\mathbf{Y}$  är en  $n \times 1$ -vektor av responsvärden.  $\mathbf{X}$  är designmatrisen med dimension  $n \times p$ , vars första kolumn består av ettor och resterande kolumner av de oberoende variablerna,  $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$  är en  $n \times 1$  vektor av multivariat-normalfördelade feltermer. Regressionskoefficienterna  $\beta_0, \beta_1, \dots, \beta_{p-1}$  uppskattas med hjälp av minstakvadratmetoden [4] där estimationen av  $\boldsymbol{\beta}$  ges av

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (5)$$

och variansen av  $\hat{\boldsymbol{\beta}}$  ges av

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (6)$$

I detta arbete kommer vi i det enklaste fallet att betrakta en modell på formen

$$Y_{ijk} = \beta_{0i} + \beta_1 t_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2) \quad (7)$$

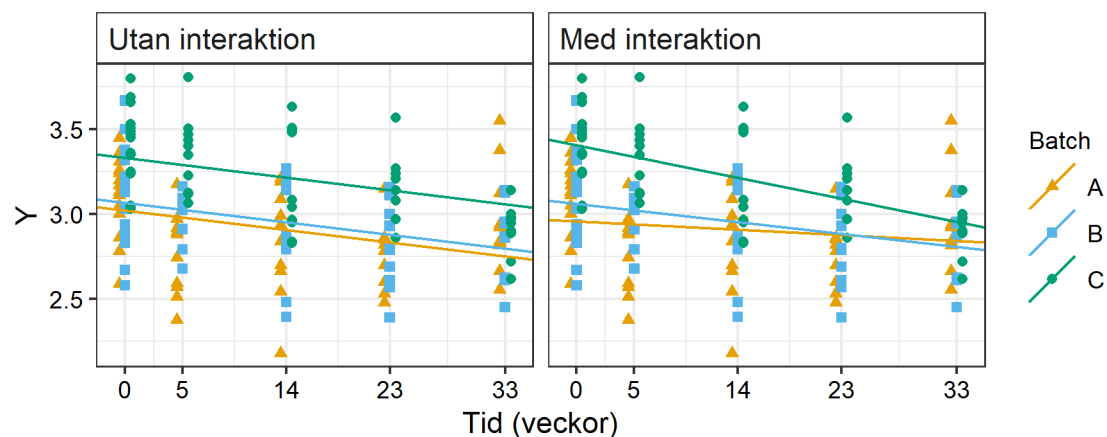
där:

- $Y_{ijk}$  är responserna för batch  $i$  vid provtillfälle  $j$  och observation  $k$ .
- $t_{ij}$  är tidpunkten (i veckor) för batch  $i$  och provtillfälle  $j$ .
- $i \in \{1, \dots, l\}$  där  $l$  är antalet batcher.
- $j \in \{1, \dots, m\}$  där  $m$  är antalet provtillfällen och  $t_j$  är tidpunkten för provtillfället  $j$ .
- $k \in \{1, \dots, q\}$  där  $q$  är antalet observationer för batch  $i$  vid provtillfälle  $j$ .
- $\beta_{0i}$  är interceptkoefficienten för batch  $i$  som beskriver skärningspunkten i  $y$ -led för batch  $i$  och  $\beta_1$  beskriver förändringen i  $y$ -led vid ökning av den förklarande variabeln  $t_{ij}$  för batch  $i$  och provtagningstillfälle  $j$ .

Vi kommer även att introducera en interaktion mellan tid och batch, en sådan modell ser ut på följande vis:

$$Y_{ijk} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2). \quad (8)$$

Till skillnad från (7) så har modellen (8) en batch-specifik tidstrend som betecknas av  $\beta_{1i}$ . I Figur 4 ser vi hur modellen utan interaktion (7) producerar en regressionslinje för varje batch, med olika skärningspunkter, och för modellen med en batch-specifik tidstrend (8) får vi dessutom olika lutningar för varje batch.



**Figur 4:** Figuren visar modellen utan interaktion mellan tid och batch (7) till vänster och modellen med interaktion (8) till höger. I den vänstra figuren har vi tre olika regressionslinjer med samma lutning, en för varje batch. Introducerar vi en interaktion mellan tid och batch får vi dessutom tre olika lutningar, vilket illustreras i den högra figuren.

### 2.3 Linjära modeller med mixade effekter

I klassisk linjär regression beräknad med minstakvadratmetoden (3) betraktas parametrarna  $\beta$  som fixa parametrar, vilket innebär att de sanna parametrarna antas ha ett fixt men för oss okänt värde. Utöver parametrarna inkluderar modellen även en slumpmässig komponent i  $\mathbf{Y}$ , nämligen feltermerna  $\varepsilon$  som inte antar något fixt värde, därför väljer vi att istället modellera dess distribution. På liknande sätt kan vi inkludera en parameter som en slumpmässig effekt i modellen.

I longitudinella studier där det görs upprepade mätningar på ett antal studieobjekt, som är slumpvis dragna ur en större population, är det vanligt att inkludera flera slumpmässiga effekter för att ta hänsyn till eventuella skillnader mellan de olika studieobjekten. Exempelvis är det naturligt att inkludera en term för ett slumpmässigt intercept för respektive studieobjekt. Ett annat alternativ är att inkludera en individuell tidstrend för varje studieobjekt, vilket innebär att vi antar att olika studieobjekt blir påverkade av tiden på olika sätt. Att inkludera en sådan slump effekt hade gett oss en slumpmässig lutning, vilket på så sätt är liknande med att inkludera en interaktionsterm i klassisk regressionsanalys. Inklusion av slump effekter för de olika studieobjekten, ger oss utöver möjligheten att dra slutsatser om de enskilda objekten, dessutom möjligheten att studera variationen både på individnivå och populationsnivå.

En linjär modell med mixade effekter är en modell som inkluderar både fixa och slumpmässiga effekter. Skrivet på matrisnotation blir detta en utökning av (3) med införandet av en ytterligare slump effektsterm  $\mathbf{u}$  och en designmatris  $\mathbf{Z}$ , denna modell ser ut på följande vis:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon, \quad (9)$$

där:

- $\mathbf{Y}$  är en stokastisk responsvektor med väntevärdet  $E(\mathbf{Y}) = \mathbf{X}\beta$ .
- $\beta$  är en okänd vektor med fixa koefficienterna.
- $\mathbf{u}$  är en okänd vektor med slumpmässiga koefficienter, med väntevärde  $E(\mathbf{u}) = \mathbf{0}$  och variansmatris  $\text{Var}(\mathbf{u}) = \mathbf{G}$ .
- $\varepsilon$  är en okänd vektor med de slumpmässiga felen, med väntevärde  $E(\varepsilon) = \mathbf{0}$  och variansmatris  $\text{Var}(\varepsilon) = \mathbf{R} = \sigma\mathbf{I}$ .
- $\mathbf{X}$  är en känd designmatris på liknande sätt som i linjär regression, och skrivs ut explicit i ekvation (4).

- $\mathbf{Z}$  är känd designmatrix liknande  $\mathbf{X}$  fast för de slumpmässiga parametrarna.

Vektorn  $\beta$  samt varians  $\mathbf{G}$  och  $\mathbf{R}$  kan skattas med exempelvis maximum likelihood-metoden.

Om vi nu återgår till exemplet i Avsnitt 2.2 och istället betraktar en mixad modell, där vi väljer att inkludera en slumpeffekt för ett slumpmässigt intercept, då ser vår modell ut på följande vis:

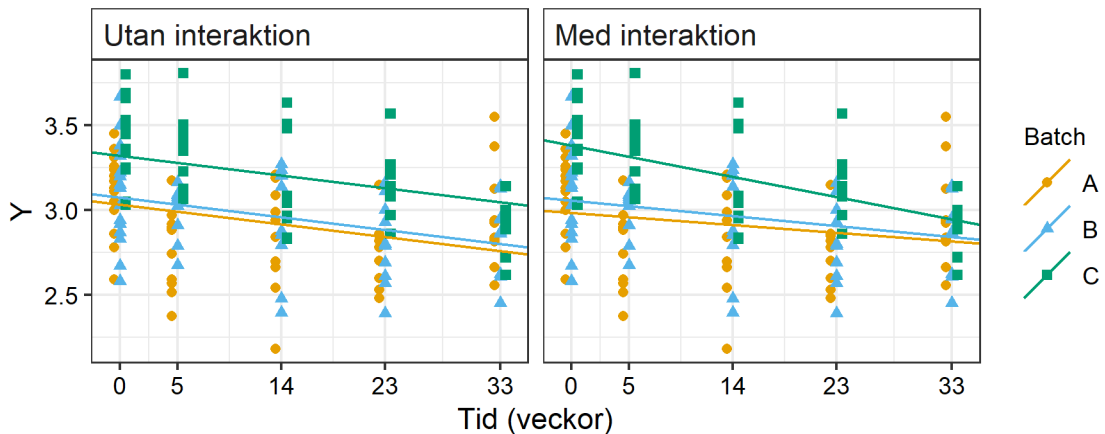
$$Y_{ijk} = \beta_0 + \beta_1 t_{ij} + u_i + \varepsilon_{ijk}. \quad (10)$$

Där  $\beta_0$  är interceptet i y-led för en genomsnittlig batch,  $\beta_1$  är förändringen i y-led per förändring med en tidsenhet och  $u_i$  är den batch-specifika avvikelser från  $\beta_0$ .

Väljer vi dessutom att inkludera en batch-specifik tidstrend för de tre batcherna så kommer vi att inkludera en extra slumpparameter  $u_{1i}$  vilken beskriver den batch-specifika tidstrendens skillnad ifrån den generella tidstrenden  $\beta_1$ . Detta är motsvarande med att inkludera en interaktionsterm i linjär regression (8). Denna modell ser ut på följande vis:

$$Y_{ijk} = \beta_0 + \beta_1 t_{ij} + u_{0i} + u_{1i} t_{ij} + \varepsilon_{ijk}. \quad (11)$$

I Figur 5 visualiseras två modeller för att illustrera skillnaden av modellerna (10) och (11).



**Figur 5:** I denna figur visualeras två olika modeller med mixade effekter skattade på samma data. Den vänstra inkluderar slumpmässigt intercept och den högra inkluderar både slumpmässigt intercept och lutning.

## 2.4 Betingad och marginell tolkning av modell med mixade effekter

Modell (11) innehåller mixade effekter och tillåter en tolkning både på *batchnivå* och på *populationsnivå*. En tolkning på *batchnivå* innebär att vi gör prediktioner och uttalanden om hållbarheten för en specifik batch i det dragna stickprovet och en tolkning på *populationsnivå* innebär att vi gör prediktioner och uttalanden om hållbarheten för den population av batcher som de studerade batcherna är dragna från.

I (9) introduceras två normalfördelade slumpvariabler,  $\mathbf{Y}$  och  $\mathbf{u}$ . Genom att betinga  $\mathbf{Y}$  med  $\mathbf{u}$  kan vi skriva

$$\mathbf{Y}|\mathbf{u} \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \sigma^2\mathbf{I}), \quad (12)$$

med

$$E(\mathbf{Y}|\mathbf{u}) = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \quad (13)$$

$$\text{Var}(\mathbf{Y}|\mathbf{u}) = \sigma^2\mathbf{I},$$

där  $\mathbf{u} \sim N(0, \mathbf{G})$ , vilket givet en prediktion  $\hat{\mathbf{u}}$  av slumpeffekten  $\mathbf{u}$  ger en prediktion av den batch-specifika fördelningen av  $\mathbf{Y}|\hat{\mathbf{u}}$ . Modellen för den betingade fördelningen av  $\mathbf{Y}|\hat{\mathbf{u}}$  kallas för den

*betingade modellen* vilken ger batch-specifika prediktioner  $\mathbf{x}^T\boldsymbol{\beta} + \mathbf{z}^T\mathbf{u}$  av responsen  $y$  givet en prediktion av motsvarande slumpoeffekt  $\mathbf{u}$ . Om vi antar att feltermerna  $\boldsymbol{\varepsilon}$  och slump effekterna  $\mathbf{u}$  är oberoende av varandra kan vi specificera den gemensamma fördelningsfunktionen  $f_{\mathbf{Y},\mathbf{u}}(y, \mathbf{u})$  som produkten mellan den betingade fördelningsfunktionen  $f_{\mathbf{Y}|\mathbf{u}}(y|\mathbf{u})$  och fördelningsfunktionen  $f_{\mathbf{u}}(\mathbf{u})$ ,

$$f_{\mathbf{Y},\mathbf{u}}(y, \mathbf{u}) = f_{\mathbf{Y}|\mathbf{u}}(y|\mathbf{u})f_{\mathbf{u}}(\mathbf{u}),$$

och därefter integrera ut  $\mathbf{u}$  för att erhålla den marginella fördelningsfunktionen  $f_{\mathbf{Y}}(y)$ . Den marginella fördelningsfunktionen av  $\mathbf{Y}$  ger den marginella fördelningen av  $\mathbf{Y}$

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}^T + \sigma^2\mathbf{I}), \quad (14)$$

med den marginella väntevärdesvektorn

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \quad (15)$$

och variansmatrisen

$$\text{Var}(\mathbf{Y}) = \mathbf{Z}\tilde{\mathbf{G}}\mathbf{Z}^T + \sigma^2\mathbf{I} =: \mathbf{V},$$

där  $\tilde{\mathbf{G}}$  är en blockdiagonal matris med  $n$  stycken  $\mathbf{G}$  matriser på diagonalen. Modellen för den marginella fördelningen kallas för den *marginella modellen* och ger populations-specifika prediktioner  $\mathbf{x}^T\boldsymbol{\beta}$  av responsen  $y$  givet en skattning av parametrarna  $\boldsymbol{\beta}$ . När parametrarna i (9) är skattade med maximum likelihood-metoden kan variansen för skattningen av de fixa parametrarna skrivas [8]

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} =: \mathbf{C}. \quad (16)$$

### 3 Ppk-beräkning för linjära regressionsmodeller med och utan mixade effekter

Generellt beräknas Ppk utifrån ett homogent stickprov från en och samma process. MHC:s stabilitets studier innehåller inhomogen longitudinell data och analys av dessa inkluderar tidstrend och batcheffekt, vilket vanligtvis inte tas hänsyn till i Ppk-beräkning. Vi behöver därför ta fram en metod för Ppk-beräkning anpassad efter de modeller som används vid analys av studierna. Vi vill beräkna Ppk för linjära regressionsmodeller både med och utan slump effekter och behöver därför ersätta standardavvikelsen i formel (1), med en standardavvikelse som tar hänsyn till både osäkerhet i de skattade parametrarna och återstående residualvariation.

I detta kapitel presenteras de formler och metoder som vi har tagit fram specifikt för det här projektet. Vi beskriver hur prediktionsintervall kan användas för Ppk-beräkning, samt tar fram formler för skattning av standardavvikelse för en framtida observation från modeller som innehåller både fixa och slumpmässiga effekter.

#### 3.1 Ppk-beräkning baserat på prediktionsintervall

Ett prediktionsintervall skattar möjliga värden för ett framtida utfall baserat på modellen som används och tar hänsyn till både residualvariation och osäkerhet i de skattade parametrarna. Här visar vi med en explicit formel hur prediktionsintervall för linjära regressionsmodeller, både med och utan mixade effekter, kan användas för att beräkna Ppk. Under normalfördelningsantagande har vi att ett  $100(1 - \alpha)\%$ -prediktionsintervall, med signifikansnivå  $\alpha$ , för en ny observation ges på formen

$$I_{\hat{\mu}} = [\hat{\mu} - t_{df, \alpha/2}\hat{\sigma}, \hat{\mu} + t_{df, \alpha/2}\hat{\sigma}]$$

där  $\hat{\mu}$  är det predikterade medelvärdet,  $t_{df}(\alpha/2)$  är t-fördelningens  $\alpha/2$ -kvantil för  $df$  frihetsgrader och  $\hat{\sigma}$  betecknar standardavvikelsen för felet i prediktionen. Antalet frihetsgrader  $df$  beräknas med  $df = n - p$  där  $n$  är antalet observationer i stickprovet och  $p$  är antalet koefficienter i modellen. En tolkning av prediktionsintervallet är, att när de förklarande variabelerna är fixerade, förväntas intervallet innehålla  $100 * (1 - \alpha)\%$  av de förväntade framtida observationerna. Under



ett normalfördelningsantagande motsvarar en konfidensgrad på 99.73% en intervalllängd på  $6\sigma$  och  $t_{df}(0.0027/2) \approx 3$  givet att  $df \geq 30$ . Genom att konstruera ett prediktionsintervall med konfidensgrad 99.73% kan vi använda sambandet

$$I_{\hat{\mu}, \text{övre}} - \hat{\mu} = \hat{\mu} + t_{df, \alpha/2} \hat{\tau} - \hat{\mu} \approx 3\hat{\tau} \quad (17)$$

för att skatta  $\hat{\tau}$  och därmed beräkna Ppk. Genom att kombinera resultatet i (17) med (1) får vi, givet att  $I_{\hat{\mu}, \text{upper}}$  är övre gräns av ett 99.7%-prediktionsintervall, att

$$\begin{aligned} \text{Ppk} &= \min \left( \frac{USL - \hat{\mu}}{I_{\hat{\mu}, \text{övre}} - \hat{\mu}}, \frac{\hat{\mu} - LSL}{I_{\hat{\mu}, \text{övre}} - \hat{\mu}} \right) \\ &\approx \min \left( \frac{USL - \hat{\mu}}{3\hat{\tau}}, \frac{\hat{\mu} - LSL}{3\hat{\tau}} \right). \end{aligned} \quad (18)$$

Med programvara som har stöd för prediktionsintervall kan vi med (18) omvandla ett prediktionsintervall till ett Ppk-värde.

### 3.2 Standardavvikelse av prediktioner skattade från modeller med mixade effekter

I detta avsnitt härleds en metod för att direkt skatta en standardavvikelse  $\tau$  som tar hänsyn till osäkerhet i parametrarna och residualvariation för modellen. Skattningen  $\hat{\tau}$  kan sedan appliceras i beräkning av prediktionsintervall eller direkt på Ppk-beräkning.

Enligt modell (9) får vi för en tänkt framtida observation  $Y_*$  att

$$Y_* = \mathbf{x}_*^T \hat{\boldsymbol{\beta}} + \mathbf{z}_*^T \mathbf{u} + \varepsilon_*, \quad (19)$$

där  $\hat{\boldsymbol{\beta}}$  är en skattning av de fixa effekterna baserat på observerade data,  $\mathbf{u}$  är slump effekterna,  $\mathbf{x}_*^T$  och  $\mathbf{z}_*^T$  är variabelvektorer och  $\varepsilon_*$  är det slumpmässiga felet i den framtida observationen. Variansen av den framtida observationen (19) är

$$\text{Var}(Y_*) = \text{Var}(\mathbf{x}_*^T \hat{\boldsymbol{\beta}} + \mathbf{z}_*^T \mathbf{u} + \varepsilon_*) \quad (20)$$

och om vi antar att termerna  $\hat{\boldsymbol{\beta}}$ ,  $\mathbf{u}$  och  $\varepsilon_*$  är oberoende av varandra, kan (20) utvecklas till

$$\begin{aligned} \text{Var}(\mathbf{x}_*^T \hat{\boldsymbol{\beta}} + \mathbf{z}_*^T \mathbf{u} + \varepsilon_*) &= \text{Var}(\mathbf{x}_*^T \hat{\boldsymbol{\beta}}) + \text{Var}(\mathbf{z}_*^T \mathbf{u}) + \text{Var}(\varepsilon_*) \\ &= \mathbf{x}_*^T \mathbf{C} \mathbf{x}_* + \mathbf{z}_*^T \mathbf{G} \mathbf{z}_* + \sigma^2 = \mathbf{x}_*^T \mathbf{C} \mathbf{x}_* + v_*, \end{aligned} \quad (21)$$

där  $\text{Var}(\mathbf{x}_*^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_*^T \mathbf{C} \mathbf{x}_*$  följer av (16) och  $v_* := \mathbf{z}_*^T \mathbf{G} \mathbf{z}_* + \sigma^2$ . Oberoendeantagandet motiveras med att de fixa parametrarna är skattade av historisk data, medan de slumpmässiga parametrarna samt det slumpmässiga felet, syftar till en ny datapunkt eftersom prediktionsintervallet är tänkt att appliceras på framtida data.

För att applicera resultat (21) på Ppk-beräkning, låter vi en framtida observation vara på formen (19), vilken implicerar det predikterade marginella medelvärdet

$$\hat{Y}_* = \mathbf{x}_*^T \hat{\boldsymbol{\beta}} := \hat{\mu} \quad (22)$$

för en variabelvektor  $\mathbf{x}_*^T$ , där  $\hat{\boldsymbol{\beta}}$  är skattad av historisk data. För att ta hänsyn till osäkerheten i  $\hat{\boldsymbol{\beta}}$  och  $\mathbf{u}$  samt residualvariationen blir skattningen av  $\tau$

$$\sqrt{\text{Var}(Y_*)} = \sqrt{\mathbf{x}_*^T \mathbf{C} \mathbf{x}_* + v_*} := \hat{\tau}. \quad (23)$$

Med (22), (23) och (18) kan Ppk beräknas för modeller med mixade effekter med formel

$$\begin{aligned} \text{Ppk} &= \min \left( \frac{USL - \hat{Y}_*}{3\sqrt{\text{Var}(Y_*)}}, \frac{\hat{Y}_* - LSL}{3\sqrt{\text{Var}(Y_*)}} \right) \\ &= \min \left( \frac{USL - \hat{\mu}}{3\hat{\tau}}, \frac{\hat{\mu} - LSL}{3\hat{\tau}} \right). \end{aligned} \quad (24)$$

Variabelvektorerna  $\mathbf{x}_*^T$  och  $\mathbf{z}_*^T$  innehåller i de stabilitetsstudier som analyserats, information om specifik tidpunkt och batch för respektive prediktion.

I specialfallet för modeller utan slump effekter följer det från (6), för en tänkt framtida observation  $Y_* = \mathbf{x}_*^T \boldsymbol{\beta} + \varepsilon$  att

$$\hat{\tau} := \sqrt{\text{Var}(Y_*)} = \sqrt{\sigma^2(\mathbf{X}^T \mathbf{X})^{-1} + \sigma^2}$$

och

$$\hat{\mu} := \hat{Y}_* = \mathbf{x}_*^T \hat{\boldsymbol{\beta}}.$$

## 4 Metod för utvärdering och analys av MHC:s stabilitetsstudier

I följande kapitel presenteras våra metoder för utvärdering och analys av MHC:s kliniska stabilitetsstudier. Metoderna har implementerats i programmeringsspråket R [6]. För att skatta linjära regressionsmodeller utan slump effekter har vi använt oss av `lm`-funktionen. För linjära regressionsmodeller med slump effekter har vi använt oss av funktionen `lmer`, som är en del av `lme4`-paketet [7]. I första delkapitlet ges en kort beskrivning av en stabilitetsstudie från MHC och dess utformning. Därefter beskriver vi modellerna som kommer att tillämpas i våra metoder. Avslutningsvis går vi igenom metoderna för våra beräkningar och simuleringar.

### 4.1 Data

Vi har använt oss av data från 13 stabilitetsstudier från MHC med totalt 24 testparametrar. Testerna utförs på tre olika batcher (A, B och C) vid tre till sex olika tidpunkter. Varje tidpunkt innehåller minst fem observationer per batch och en stabilitetsstudie kan pågå i upp till 74 veckor. Det som testas är olika typer av mått på hållbarheten av en produkt hos MHC, se Tabell 2. Undersökning av datan visar att observationer i tillgängliga studier är nära normalfördelade, vilket motiverar ett normalfördelningsantagande i de analyser som utförs.

**Tabell 1:** Tabellen visar ett exempel på hur datan från en stabilitetsstudie ser ut. Varje rad avser en observation. Första kolumnen betecknar testtillfälle (veckor), andra kolumnen betecknar vilken batch observationen tillhör och resterande kolumner betecknar vilket test som utförts. Vad olika test-ID avser kan ses i Tabell 2.

| Veckor | Batch | T-229.CD.wet | T-229.CD.dry |
|--------|-------|--------------|--------------|
| 0      | A     | 42.07        | 36.22        |
| 0      | B     | 36.01        | 32.79        |
| 0      | C     | 45.83        | 42.39        |
| 6      | A     | 36.50        | 31.12        |
| 6      | B     | 28.71        | 29.07        |
| 6      | C     | 45.63        | 39.59        |
| 18     | A     | 37.51        | 38.78        |
| 18     | B     | 27.92        | 27.69        |
| 18     | C     | 28.42        | 30.42        |

**Tabell 2:** Tabellen visar olika typer tester av hållbarhet som utförs i MHC:s stabilitetsstudier.

| ID     | Titel   | Syfte  |
|--------|---|--|
| T-229  | Draghållfasthet för ickevävda (nonwoven) material   | Att avgöra hur ickevävda material uppför sig när de utsätts för draghållfasthetstestning   |
| T-238  | Sömstyrka   | Att avgöra sömmars styrka i ett paket  |
| T-1006 | Ludd och andra partiklars generering i torrt tillstånd  | Att mäta hur mycket ett material luddar i torrt tillstånd. Resultatet av testet, dvs luddnings-koefficienten är kumulativt beräknat för partiklar i storleksordningen 3m till 25m då dessa partiklar anses kunna bära mikro-organismer |
| T-1024 | Vattenångas överföringshastighet med Permatran W 101K   | Att få värden av vattenångans överföringshastighet (WVTR) för barriärmaterial  |
| T-1099 | Ledstyrka för flexibla material och produkter vid 254mm/min                                   | Att avgöra styrkan i lederna för laminat, produkter och material   |
| T-1125 | Särbehandling med undertryck (NPWT): Draghållfastheten hos limmade och ickelimmade gränsskikt | Att avgöra maximal styrka av limmade och icke-limmade gränsskikt i NPWT produkter genom draghållfasthetstestning   |
| T-1130 | Särbehandling med undertryck: Delaminering av lim på vätsketransportenhet                     | Att avgöra adesionsegenskaper hos en vätsketransportenhet. Utvärdering med draghållfasthetstestning  |
| T-1179 | Sprängstyrka för tunna material   | Att avgöra maximalt tryck, i rätt vinkel mot ett testobjekt, innan provet sprängs  |
| T-1180 | Särbehandling med undertryck (NPWT): Luftläckage för NPWT-material                            | Att avgöra luftläckage mellan en NPWT komponent och en tillhörande vätsketransportsenhet   |

## 4.2 Modeller

För att åstadkomma en så bred analys som möjligt kommer vi att använda oss av fyra olika modeller i våra beräkningar och simuleringar:

- En linjär modell utan slump effekter (7) med parametrarna *tid* och *batch*.
- En linjär modell utan slump effekter (8) med parametrarna *tid*, *batch* och interaktionen  $tid \times batch$ .
- En linjär modell med slump effekter (10) med *tid* som fix parameter och *batch* som slumpmässig parameter.
- En linjär modell med slump effekter (11) med *tid* som fix parameter och *batch* samt interaktionen  $tid \times batch$  som slumpmässiga parametrar.

Vi kan då jämföra hur en interaktion mellan tid och batch påverkar resultaten av stabilitetsstudierna, både för modeller med och utan slump effekter och även jämföra dessa med varandra. Dessa jämförelser tillåter dessutom en utvärdering av effekten på Ppk, vid beräkning med en felspecifierad modell och möjligheten att identifiera en interaktion.

## 4.3 Applikation av metoder för Ppk-beräkning

Beräkning av Ppk utförs med två olika tillvägagångssätt beroende på om modellen som tillämpas i analysen innehåller slump effekter eller inte. Om analys genomförs med en modell med slump effekter predikteras ett populationsmedelvärde som representerar alla batcher och om analys genomförs med en modell utan slump effekter predikteras ett batch-specifikt medelvärde. I båda fall predikteras medelvärdet för den tidpunkt som de sista mätningarna utfördes för respektive stabilitetsstudie.

Avsnitt 3.2 beskriver metoden som applicerats vid Ppk-beräkning för en modell med slump effekter. Genom en marginell tolkning av modellen predikteras ett populationsmedelvärde  $\hat{\mu}$  med formel (22) för den antagna populationen av batcher. Standardavvikelsen  $\tau$  skattas med formel (23) genom att ta kvadratroten ur summan av variansen för skattningen av de fixa parametrarna, skattningen av variansen för de slumpmässiga parametrarna och den skattade residualvariationen. De två skattningarna  $\hat{\mu}$  och  $\hat{\tau}$  används sedan i formel (24) för att beräkna Ppk för respektive testparameter.

Modeller utan slump effekter ger prediktioner som representerar batch-specifika medelvärden. Stabilitetsstudierna genomförda på MHC innehåller observationer från tre olika batcher vilket betyder att vi måste ta hänsyn till vilken av de tre batcher som enligt analysen presterar sämst. Varje process har bestämda specifikationsgränser som bildar ett intervall för godkända testvärden, se Avsnitt 2.1. Den batch som producerar det batch-specifika medelvärde som ligger närmast en specifikationsgräns för den tidpunkt som respektive stabilitetsstudie avslutats rankas som sämst. För denna prediktion konstrueras ett 99.7%-prediktionsintervall som med formel (17) ger en skattning av standardavvikelsen  $\tau$ . Ppk beräknas sedan med formel (18) där den batch-specifika prediktionen är  $\hat{\mu}$ .

## 4.4 Beräkningar och simuleringar

Vi kommer att anpassa modellerna från Avsnitt 4.2 till datan från MHC:s stabilitetsstudier. Utifrån våra anpassade modeller kommer vi sedan att beräkna Ppk med hjälp av metoderna beskrivna i Kapitel 3. Med hjälp av våra anpassade modeller kommer vi att förändra variationen inom batcher och därefter beräkna Ppk baserat på den modifierade variationen. Dessutom kommer vi att beräkna signifikansen med avseende på tid, när vi har förändrat variationen. För att vidare undersöka hur interaktionen mellan tid och batch påverkar resultaten av stabilitetsstudierna har vi valt att simulera data. Detta eftersom MHC:s stabilitetsstudier är begränsade till tre batcher och fem observationer per batch. Vi kan med simuleringar analysera vad som händer ifall vi ökar antalet batcher och/eller ökar antalet observationer per batch.

Vi vill med hjälp av våra beräkningar och simuleringar svara på följande frågor:

- Vad blir effekten, i form av förändring i Ppk på genomförda stabilitetsstudier på MHC, vid en analys som tar hänsyn till en interaktion mellan tid och batch, jämfört med en enklare analys utan interaktion?
- Hur stor effekt har variationen inom batcherna på Ppk-värdena?
- Hur stor effekt har variationen inom batcher på huruvida testvärdena är signifikant beroende av tid?
- Vad blir effekten, i form av möjligheten att hitta en interaktion givet att det existerar en interaktion, om vi med hjälp av simuleringar ökar antalet batcher och antalet observationer per batch?
- Vad blir effekten av en felaktig modellspecifikation i våra Ppk-skattningar?

### 4.4.1 Effekt av att introducera en interaktion mellan batch och tid i genomförda stabilitetsstudier på MHC

För att undersöka effekten, i form av förändring i Ppk, av att introducera en interaktion mellan tid och batch i genomförda stabilitetsstudier på MHC, kommer vi att beräkna Ppk baserat på modellerna i Avsnitt 4.2. Vi anpassar modellerna till stickproven från stabilitetsstudierna. Sedan använder vi metoden för Ppk-beräkning med prediktionsintervall beskriven i Avsnitt 3.1 och 3.2 för att skatta Ppk för modellerna.

### 4.4.2 Effekt av förändrad variation inom batcher på Ppk-beräkningar och signifikanstest

Vi vill undersöka hur variationen inom batcher påverkar Ppk-beräkningar och signifikanstest. Detta ger oss möjligheten att undersöka hur robusta resultaten av MHC:s stabilitetsstudier är mot störningar i datan.

För att undersöka hur variationen inom batcher påverkar Ppk använder vi modeller utan slump effekt och skalar residualerna. Residualerna är definierade som  $e_i = y_i - \hat{y}_i, i = 1, \dots, n$ , där  $y_i$  är observationen och  $\hat{y}_i$  är modellskattningen vid samma punkt. Den nya datan blir då  $\hat{y}_i = y_i + c * e_i$ ,

där  $c$  är skalningskonstanterna. På så sätt ökar eller minskar vi avståndet mellan observationerna och modellskattningarna och vi erhåller ny data med en förändrad variation. Vi beräknar sedan  $P_{pk}$  för modeller med och utan slump effekter med den nya datan. Vi använder samma data för att se hur  $p$ -värdet, med avseende på tidsparametern  $\beta_1$  förändras, det vill säga om tid är signifikant för hur processens testparameter förändras, med ökad eller minskad variation.

#### 4.4.3 Simulering för att utvärdera möjligheten att identifiera en existerande interaktionseffekt

Givet ett hypotestest

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

med signifikansnivå  $\alpha$  så definieras styrkan av ett hypotestest som

$$P(\text{förförkasta } H_0 | H_1 \text{ sann}) = 1 - P(\text{acceptera } H_0 | H_1 \text{ sann}),$$

det vill säga sannolikheten att vi förförkastar nollhypotesen givet att alternativhypotesen är sann. Detta kan även beskrivas som sannolikheten att vi hittar en effekt givet att det faktiskt finns en underliggande effekt att hitta [4].

Vi kommer att simulera data baserat på MHC:s stabilitetsstudier och att det finns en interaktion mellan *tid* och *batch*. Detta innebär att vi simulerar ny data från en linjär modell, med *tid* som fix parameter samt *batch* och interaktionen  $tid \times batch$  som slumpmässiga parametrar (11), som har anpassats till datan från MHC:s stabilitetsstudier. Genom att vi simulerar data baserat på att det finns en interaktion, så kan vi undersöka hur styrkan, det vill säga hur ofta vi hittar interaktionen, påverkas när vi varierar antalet batcher och observationer per batch. Vidare kan vi även jämföra skillnaden i styrka om vi anpassar den simulerade datan till en modell med eller utan slump effekter.

Styrkan baseras på det tvärsidiga hypotestestet

$$\begin{cases} H_0 : \beta_{batch \times tid} = 0 \\ H_1 : \beta_{batch \times tid} \neq 0 \end{cases}$$

med signifikansnivå  $\alpha = 0.05$ . För att avgöra om interaktionen mellan *tid* och *batch* är signifikant jämför vi två olika modeller: den ena modellen utan interaktion mellan *tid* och *batch* och den andra modellen med interaktion. För att jämföra modellerna använder vi oss av ANOVA [4]. Om  $p$ -värdet är mindre än eller lika med  $\alpha$  så förförkastar vi  $H_0$  och modellen med interaktion anses vara bättre än modellen utan interaktion. Det vill säga, interaktionen mellan *tid* och *batch* är signifikant på signifikansnivå  $\alpha = 0.05$ . Detta gör vi dels för modellerna utan slump effekter (7) och (8), samt modellerna med slump effekter (10) och (11). Vidare har vi för ett hypotestest att styrkan definieras som sannolikheten att vi förförkastar  $H_0$  givet att  $H_1$  är sant. I vårt fall kan detta beskrivas som; givet att det finns en interaktion, vad är sannolikheten att vi hittar den? För att uppskatta styrkan med hjälp av simuleringar använder vi oss av Algoritm 1.

---

**Algoritm 1** *Algoritm för att uppskatta styrka*

---

1. Låt  $i = 1, 2, 3, \dots, N$  där  $N$  är antalet simuleringar som ska köras och låt  $s = 0$ .
  2. Välj antal batcher och observationer per batch som skall simuleras.
  3. Simulera ny data  $d_i$ , baserat på att det finns en interaktion med antal batcher och observationer per batch från (2).
  4. Anpassa en modell med interaktion och en modell utan interaktion till den simulerade datan  $d_i$ .
  5. Gör ett ANOVA test för att jämföra modellerna. Låt  $p_i$  vara motsvarande p-värde.
  6. Om  $p_i \leq \alpha$  förkasta nollhypotesen och  $s = s + 1$ . Annars förkastar vi inte nollhypotesen.
  7. Efter att  $N$  simuleringar har gjorts så uppskattar vi den totala styrkan  $S$  genom  $S = s/N$ .
- 

Har vi exempelvis  $N = 1000$  simulationer och  $S = 80\%$  innebär detta att av 1000 simulationer hittar en signifikant interaktion 800 gånger. Med andra ord så kan vi säga att, givet att det finns en interaktion i datan, så är sannolikheten att vi hittar den 0.8. Vi kan använda algoritmen för att analysera hur styrkan påverkas, om vi varierar antalet batcher och/eller antalet observationer per batch. Vi har använt oss av  $N = 1000$  simuleringar. Vi har valt att variera antalet batcher i den simulerade datan från 3 till 20. Antalet observationer/batch varierar mellan 5, 10, 15 och 20.

#### 4.4.4 Simulering för att utvärdera effekten av en felaktig modellspecifikation

För att svara på hur en felaktig modellspecifikation påverkar Ppk-skattningarna simulerar vi data på två olika sätt: under antagandet att det finns en interaktion och under antagandet att det inte finns en interaktion. En felaktig modellspecifikation innebär att vi inkluderar eller exkluderar en interaktion i analysen på data där en underliggande interaktion inte existerar respektive existerar. Vi beräknar den relativa feltermen  $\Delta$ , genom att jämföra den korrekta modellens Ppk-skattning ( $Ppk_1$ ), med den felaktiga modellens Ppk-skattning ( $Ppk_2$ ). Formeln vi använder för att beräkna den relativa feltermen är:

$$\Delta = \frac{Ppk_2 - Ppk_1}{Ppk_1}.$$

För att uppskatta modellspecifikationernas relativa feltermen använder vi oss av Algoritm 2.

---

**Algoritm 2** *Algoritm för att beräkna effekten av en felaktig modellspecifikation*

---

1. Låt  $i = 1, 2, 3, \dots, N$  där  $N$  är antalet simuleringar som ska köras.
  2. Välj antal batcher och observationer per batch som skall simuleras.
  3. Simulera ny data  $d_i$ , baserat på att det existerar en interaktion (alternativt att det inte existerar en interaktion) med antal batcher och observationer per batch från steg 2.
  4. Anpassa en modell med interaktion och en modell utan interaktion till den simulerade datan  $d_i$ .
  5. Beräkna Ppk utifrån båda modellerna.
  6. Beräkna den relativa feltermen när vi går ifrån en korrekt modell till en felaktig modell.
  7. Efter att  $N$  simuleringar har utförts beräknar vi genomsnittet på feltermen för att få den förväntade feltermen.
-

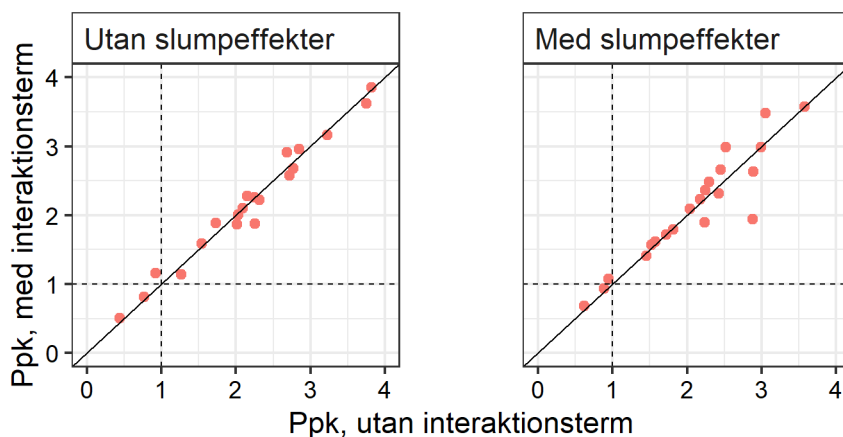
Vi applicerar Algoritm 2 med  $N = 1000$  simuleringar. I steg 3 väljer vi tre batcher och fem observationer per batch för att efterlikna MHC:s stabilitetsstudier.

## 5 Resultat

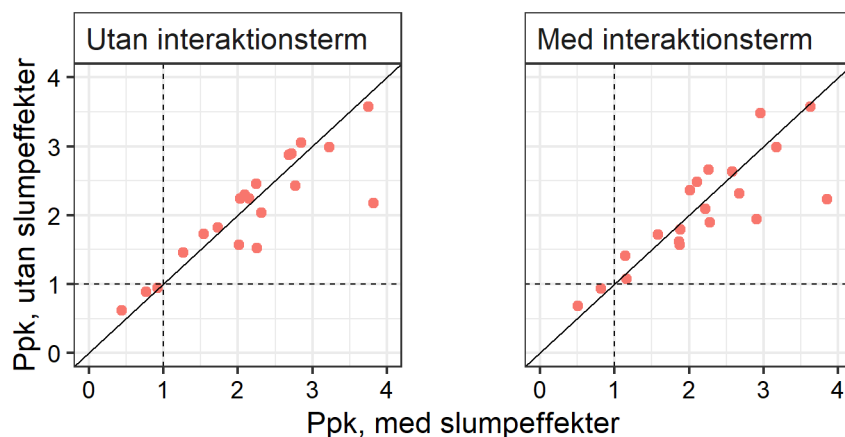
I detta kapitel presenteras resultaten från våra beräkningar och simuleringar. I första avsnittet presenteras resultatet av interaktionens effekt på Ppk för MHC:s stabilitetsstudier, följt av hur variationen inom batcher påverkar Ppk-beräkningar och signifikansen av modellparametrar. Vidare presenterar vi hur möjligheten att identifiera en existerande interaktionseffekt påverkas av att öka antalet batcher och observationer per batch. Till sist presenterar vi hur en felaktig modellspecifikation påverkar Ppk-skattningar. I resultaten som presenteras har vi valt att exkludera höga Ppk-värden, då dessa inte är relevanta för vår frågeställning.

### 5.1 Interaktionstermens effekt på Ppk för MHC:s stabilitetsstudier

I detta avsnitt visas resultatet av Ppk-beräkningar för 13 stabilitetsstudier genomförda på MHC. Det innefattar beräkningarna för 24 testparametrar utförda på modeller utan slump effekter, både med och utan interaktionsterm, samt modeller med slump effekter, både med och utan interaktionsterm. I Figur 6 illustreras skillnaden i Ppk, beräknat på modeller med en interaktionsterm och på modeller utan en interaktionsterm. Pearson-korrelationen mellan Ppk beräknat med en interaktionsterm och Ppk beräknat utan en interaktionsterm, för modeller utan slump effekter är 0.999. Det indikerar en stark positiv korrelation mellan  $x$ - och  $y$ -värden vilket dessutom är synligt i figuren. Motsvarande Pearson-korrelation för modeller med slump effekter är 0.998, vilket också indikerar en stark positiv korrelation. I Figur 7 illustreras skillnaden i Ppk beräknat på modeller med och utan slump effekter. Pearson-korrelationen mellan  $x$ - och  $y$ -värdena är 0.994 samt 0.996 för den vänstra respektive högra grafen, vilket indikerar en stark positiv korrelation.



**Figur 6:** I figuren representerar  $x$ -axlarna Ppk beräknat på modeller utan interaktionsterm och  $y$ -axlarna representerar Ppk beräknat på modeller med interaktionsterm. Den vänstra figuren illustrerar modeller utan slump effekter och den högra figuren illustrerar modeller med slump effekter. Den svarta referenslinjen,  $x = y$ , åskådliggör ingen skillnad mellan  $x$  och  $y$ . De streckade linjerna visar när  $Ppk = 1$ , vilket vanligtvis är det kritiska värdet ( $Ppk > 1$ ). Pearson-korrelationen mellan variablerna i vänstra respektive högra grafen är 0.999 och 0.998.



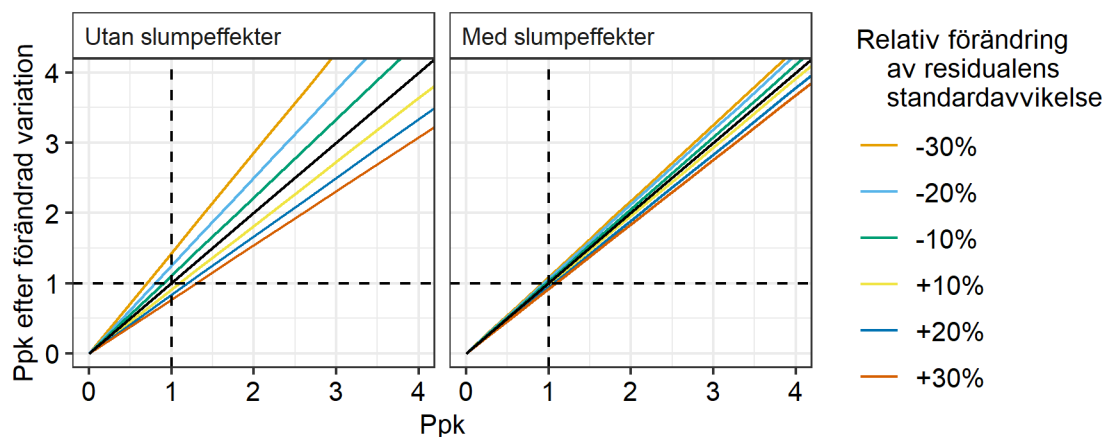
**Figur 7:** I figuren representerar  $x$ -axlarna  $Ppk$  beräknat på modeller utan slump effekter och  $y$ -axlarna representerar  $Ppk$  beräknat på modeller med slumpmässiga effekter. Den vänstra figuren illustrerar modeller utan en interaktion mellan tid och batch, medan den högra figuren illustrerar modeller med interaktion mellan tid och batch. Den svarta referenslinjen,  $x = y$ , åskådliggör ingen skillnad mellan  $x$  och  $y$ . De streckade linjerna visar när  $Ppk = 1$ , vilket vanligtvis är det kritiska värdet ( $Ppk$  godkänt om  $Ppk > 1$ ). Pearson-korrelationen mellan mellan variablerna i vänstra respektive högra grafen är 0.994 och 0.996.

## 5.2 Residualvariationens inverkan på $Ppk$ -beräkningar och signifikans-test

I Figur 8 ser vi hur  $Ppk$  påverkas av en förändrad variation inom batcher för modeller med och utan slump effekter, (11) respektive (8).  $Ppk$ -värdena är beräknade för modeller med interaktion. Motsvarande analys för modeller utan interaktion gav liknande resultat. För grafen över  $Ppk$  beräknat med modeller utan slump effekter, är linjerna dragna mellan  $Ppk$ -värden för respektive förändring i variation. För grafen över  $Ppk$  beräknat med modeller med slump effekter, är istället linjerna en anpassning till  $Ppk$ -värdena, eftersom sambandet mellan residualvariansen och  $Ppk$  inte är lika linjär som för modeller utan slump effekter. I Tabell 3 har vi beräknat hur många testparametrar som har  $Ppk$  under det kritiska värdet, när  $Ppk = 1$  för de båda modellerna. Vi kan se både i Figur 8 och Tabell 3, att för båda modellerna medför ökad variation i datan lägre  $Ppk$ .

I Figur 9 illustreras hur  $p$ -värdet för parametrar i en modell utan slump effekter påverkas av en förändring i variationen inom batcher. Här ser vi att en ökad variation leder till högre  $p$ -värden och en lägre variation leder till lägre  $p$ -värden. I Tabell 4 ser vi hur många av befintliga modeller som får ett  $p$ -värde över det kritiska gränsvärdet 0.05 för tidsparametern vid respektive förändring av residualvariation.

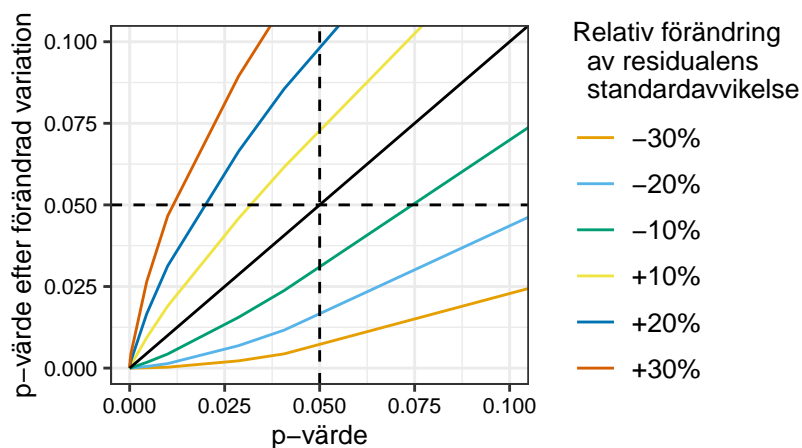




**Figur 8:** Residualvariationens inverkan på Ppk för modeller med slumpeffekter (11) till höger och för modeller utan slumpeffekter till vänster (8), x-axeln visar Ppk beräknat på oförändrad data och y-axeln visar motsvarande Ppk efter att standardavvikelsen i residualtermen skalats ned/upp med 10%, 20% och 30%. Den rätta svarta linjen visar när ingen förändring sker,  $y = x$ . De streckade linjerna visar när  $Ppk = 1$ , vilket vanligtvis är det kritiska värdet ( $Ppk$  godkänt om  $Ppk > 1$ ).

**Tabell 3:** Visar residualvariationens inverkan på Ppk (18) för modeller utan slumpeffekter (8) och modeller med slumpeffekter (11) genom att beräkna hur många testparametrar som fick  $Ppk < 1$ , totalt antal testparametrar är 24.

| Relativ förändring av residualens standardavvikelse      | -30% | -20% | -10% | 0% | +10% | +20% | +30% |
|--|------|------|------|----|------|------|------|
| Antal testparametrar med $Ppk < 1$ (utan slumpeffekter): | 1    | 1    | 1    | 2  | 3    | 3    | 3    |
| Antal testparametrar med $Ppk < 1$ (med slumpeffekter):  | 1    | 1    | 1    | 2  | 3    | 3    | 3    |



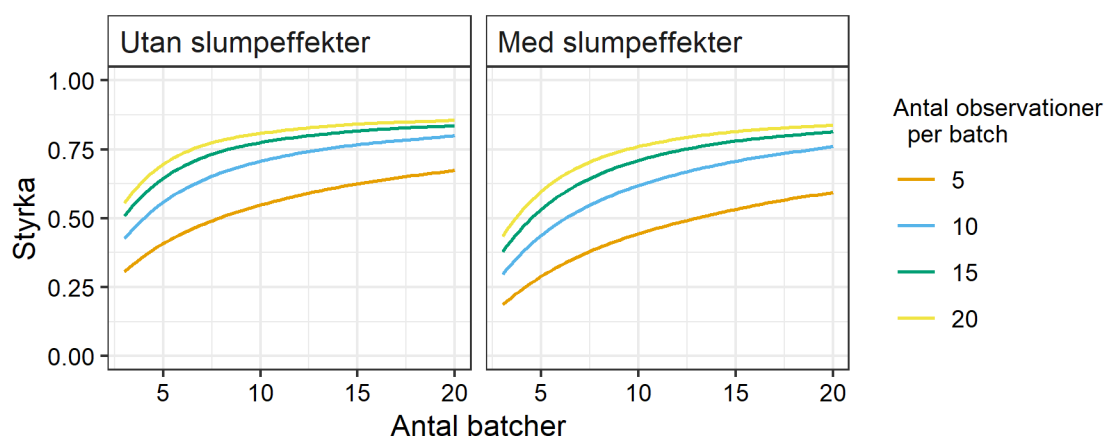
**Figur 9:** Visar residualvariationens inverkan på p-värde för parametrarna i modeller utan slumpeffekter (8) beräknat för befintliga studier från MHC. x-axeln visar p-värde beräknat på oförändrad data, y-axeln visar motsvarande p-värde efter att standardavvikelsen i residualtermen skalats ned/upp med 10%, 20% och 30%. Den rätta linjen avser ingen förändring,  $y = x$ . De streckade linjerna betecknar  $p = 0.05$ , det normala kritiska värdet för p-värde.

**Tabell 4:** Visar residualvariationens inverkan på p-värde för modeller utan slump effekter (8) genom att beräkna hur många tidsparametrar som fick p-värde  $> 0.05$ , totalt antal testparametrar är 24.

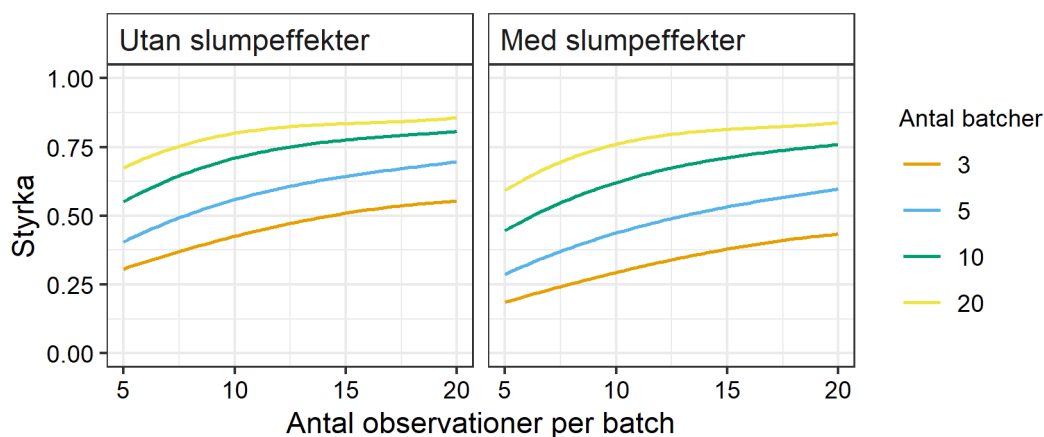
| Relativ förändring av residualens standardavvikelse | -30% | -20% | -10% | 0% | +10% | +20% | +30% |
|---|------|------|------|----|------|------|------|
| Antal testparametrar som har p-värde $> 0.05$ :     | 0    | 0    | 0    | 0  | 2    | 2    | 3    |

### 5.3 Styrkan att identifiera en existerande interaktionseffekt

I Figur 10 ser vi att styrkan blir större i takt med att antalet batcher ökar, dessutom ser vi i Figur 11 en snabbare ökning av styrkan när fler observationer per batch har simulerats.



**Figur 10:** Figuren visar hur styrkan förändras för olika antal batcher och observationer per batch. Antalet batcher visas på x-axeln och antalet observationer per batch illustreras av respektive kurva.



**Figur 11:** Figuren visar hur styrkan förändras för olika antal batcher och observationer per batch. Antalet observationer per batch visas på x-axeln och antal batcher illustreras av respektive kurva.

Styrkan för båda modellerna utvecklas på liknande sätt, med skillnaden att den linjära modellen utan slump effekter (8) har en större initial styrka och ligger sedan konstant högre.

## 5.4 Konsekvens av felaktig modellspecifikation

I Tabell 5 kan vi se de relativa feltermerna beräknade med Algoritm 2 där Ppk-skattningen är baserad på en modell utan slumpeffekter. Samma resultat fast baserat på en modell med slumpeffekter finns i Tabell 6.

**Tabell 5:** I tabellen visas den relativa feltermen för vår Ppk-skattning vid en felaktig modellspecifikation. Ppk-skattningen är baserad på en modell utan slumpeffekter. De tre olika mätvärdena är beräknade på de 24 tillgängliga testparametrarna.

| Simuleringsmetod | Minimum | Maximum | Genomsnitt |
|------------------|---------|---------|------------|
| Med interaktion  | -17.27% | 40.25%  | 4.47%      |
| Utan interaktion | -2.29%  | 9.10%   | -0.59%     |

**Tabell 6:** I tabellen visas den relativa feltermen för vår Ppk-skattning vid en felaktig modellspecifikation. Ppk-skattningen är baserad på en modell med slumpeffekter. De tre olika mätvärdena är beräknade på de 24 tillgängliga testparametrarna.

| Simuleringsmetod | Minimum | Maximum | Genomsnitt |
|------------------|---------|---------|------------|
| Med interaktion  | 0.00%   | 15.65%  | 4.77%      |
| Utan interaktion | -0.90%  | 0.54%   | -0.21%     |

Som vi kan se påverkas modellerna med och utan slumpeffekter på liknande sätt. Vi överskattar Ppk när vi felaktigt exkluderar en interaktion och vi underskattar Ppk när vi felaktigt inkluderar en interaktion. Däremot är överskattningen vid felaktig exkludering av en interaktion mer påtaglig än underskattningen vid en felaktig inklusion. Utöver detta kan vi även se att minimumvärdet och maximumvärdet utgör ett bredare intervall för en modell utan slumpeffekter, jämfört med en modell med slumpeffekter. Tabellerna A1 och A2 visar den relativa feltermen för alla tillgängliga studier och testparametrar.

## 6 Diskussion

### 6.1 Sammanfattning

Utifrån MHC:s stabilitetsstudier har vi tagit fram metoder och formler för beräkning av Ppk. Formlerna (18) och (24) ger estimationer av Ppk genom skattning av prediktionsintervall respektive direkt skattning av standardavvikelsen för en tänkt framtida observation. Resultaten visar att de beräknade Ppk-värdena för MHC:s stabilitetsstudier skiljer sig väldigt lite mellan de olika modellerna. Dessutom kan vi se att en förändring av variationen inom batcher endast påverkar ett fåtal av de befintliga studierna kritiskt och resultaten verkar således robusta. Modeller med slumpeffekter verkar dessutom mer robusta än modeller utan slumpeffekter. Våra simuleringar visar att det är svårt att hitta en signifikant interaktion mellan tid och batch, men att en felaktig inklusion av en interaktionsterm leder till mindre fel i Ppk-skattningen än att felaktigt bortse från interaktionstermen.

### 6.2 Metoddiskussion

Metoden som härleds i Avsnitt 3.1 är framtagen för situationer, där programvaran som används inte erbjuder direkt skattning av standardavvikelsen, som syftar till den framtida observation som tillämpas i Ppk-beräkningen [9]. Istället använder metoden ett prediktionsintervall. Metoden kan dessutom appliceras i situationer, där ett prediktionsintervall är det mest lättillgängliga alternativet. Att förstå all osäkerhet som måste tas hänsyn till i en skattning av  $\tau$ , standardavvikelsen för en tänkt framtida observation, kan vara svårt vilket också motiverar en praktisk användning av metoden. Olika programvaror har dock algoritmer som skiljer sig i skattning av prediktionsintervall och värdet på den Ppk-skattning som erhålls, kommer därför variera beroende av vilken programvara som används. Den vanliga formeln för Ppk bygger på approximationen  $t_{df}(0.0027/2)\hat{\sigma} \approx 3\hat{\sigma}$ ,

där  $\hat{\mu} \pm 3\sigma$  ska innefatta 99.73% av framtida värden. Notera att t-fördelningen används för att ta hänsyn till det faktum, att vi inte känner till den sanna standardavvikelsen  $\sigma$ , utan ersätter denna med en skattning  $\hat{\sigma}$  i Ppk-beräkningen. Vår metod undviker denna approximation genom att använda t-fördelningens kvantil direkt. Detta kan vara fördelaktigt, speciellt för små studier med få frihetsgrader, där approximationen  $t_{df}(0.0027/2) \approx 3$  tenderar att ge stora överskattningar i Ppk.

Metoden som härleds i Avsnitt 3.2 ger en skattning av standardavvikelsen  $\tau$  för en framtida observation, där populationsmedelvärdet  $X\hat{\beta}$  är skattat av observerade data, medan slump effekter  $\mathbf{u}$  och feltermen  $\epsilon$  syftar till en ny datapunkt. I (23) avser  $\hat{\tau}$  alltså en skattning av standardavvikelsen för en framtida observation för en batch som vi ännu inte har observerat. I skattningen av  $\tau$  tar vi hänsyn till osäkerheten av de skattade parametrarna  $\hat{\beta}$ ,  $\mathbf{u}$  och skattningen av residualvariansen  $\sigma^2$  men betraktar  $\sigma^2$  och  $\mathbf{G}$ , kovariansmatrisen för slump effekterna  $\mathbf{u}$ , som kända. I praktiken ersätter vi dessa med motsvarande punktskattningar, men den föreslagna metoden tar inte hänsyn till den extra osäkerhet som detta medför. Resultatet av att skattningen inte tar hänsyn till osäkerheten i varianskomponenterna, kan vi förmoda blir en underskattning av  $\tau$ . Vid jämförelse av skattning gjord med etablerad programvara [9] i kombination med Ppk-formel (18), var skillnaden liten nog att motivera en applikation av metoden i detta projekt. För att erhålla en mer korrekt skattning av  $\tau$ , förslår vi att inkludera upprepad simulering av ny data i metoden. Genom repetitiv skattning av varianskomponenterna hade metoden troligtvis tagit fram en mer korrekt skattning av standardavvikelsen för tänkta framtida observationer. Eftersom vi utgick från den vanliga formeln för Ppk, använder vi i metoden tre standardavvikelser. Ett alternativ hade varit att använda t-fördelningens kvantil istället för approximationen 3. Metoden leder således till mer korrekta skattningar av Ppk.

### 6.3 Interaktionstermens betydelse

Utifrån analyser av stabilitetstudier utförda på MHC kan vi inte se någon trend som antyder att en interaktionsterm har ett särskilt inflytande på skattningen av Ppk, varken för modeller med eller utan slump effekter. En del av förklaringen till detta kan vara att den batch-specifika tidstrenden är svag eller inte existerar. Det vill säga, att stickprov från olika batcher beter sig på liknande sätt över tid. Det kan även bero på att stabilitetsstudierna endast bygger på tre batcher och fem observationer per batch. Urvalet av batcher och observationer är helt enkelt för litet för att man ska kunna hitta en eventuell interaktion mellan tid och batch, även om det faktiskt finns en underliggande interaktion i datan.

Simuleringen av styrka i Avsnitt 5.3 visar att när vi ökar antalet batcher och observationer per batch, så ökar också sannolikheten att vi hittar en interaktion, vilket stödjer argumentet att urvalet i den ursprungliga datan är för litet. Tittar vi på extremfallet, då vi endast har tre batcher och fem observationer per batch, uppnår vi en styrka på mellan 0.24 och 0.26. Det vill säga, trots att vi har simulerat data baserat på att det finns en interaktion, så är sannolikheten att vi faktiskt hittar den ungefär 25%. Konsekvensen av detta blir att man anser att interaktionen inte är signifikant, och i vidare analyser av datan inte tar någon hänsyn till den, trots att den potentiellt kan påverka resultaten.

I relation till detta visar simuleringarna i Avsnitt 5.4 att den relativa feltermen av Ppk-skattningen i genomsnitt är mindre vid en felaktig inkludering av en interaktion än vid ett felaktigt ignorande av en interaktion. Med andra ord är det en större risk att våra resultat påverkas negativt, om vi felaktigt bortser från en interaktionsterm jämfört med om vi felaktigt inkluderar en interaktionsterm i våra analyser. Detta motiverar en tillämpning av modeller med interaktionsterm oavsett om den är signifikant eller inte.

### 6.4 Variation inom batcher

Vi kan se att det är ganska stor skillnad på Ppk beroende på om Ppk skattats med hjälp av modeller med eller utan slump effekter. Modellerna med slump effekter ser ut att vara mer robusta mot en ökad variation inom batcher än modellerna utan slump effekter. Detta kan förklaras av att modeller med slump effekter har fler varianskomponenter än vad modeller utan slump effekter har

och därmed har residualvariationen en mindre påverkan på modeller med slump effekter. Vidare hade det varit intressant att undersöka detta ytterligare för att se om resultatet står sig genom att ändra variationen inom batcher på annat sätt, eller genom att simulera en ökad variation mellan batcher.

## 6.5 Utvärdering av modeller med och utan slump effekter

I våra beräkningar och simuleringar har vi använt oss av modeller både med och utan slump effekter. Finns det några fördelar respektive nackdelar av att analysera stabilitetsstudierna med eller utan slump effekter?

Resultatet visar att för stabilitetsstudierna utförda på MHC är modellvalet av liten betydelse med avseende på beräknade Ppk-värden. Däremot visar våra simuleringar att det relativa felet i Ppk-skattningen när vi simulerar data med interaktion och sedan felaktigt bortser från interaktionen, ligger mellan -17.27% och 40.25% för modellen utan slump effekter. Motsvarande siffror för modellen med slump effekter är 0% och 15.65%. Detta tyder på att en modell utan slump effekter är känsligare för en felaktig modellspecifikation.

När vi simulerade data för att undersöka möjligheten att hitta en interaktion såg vi att den initiala styrkan för en modell utan slump effekter var högre än för modellen med slump effekter. Styrkan låg sedan konstant högre när vi ökade antalet batcher och observationer per batch. Detta leder oss till slutsatsen att en modell utan slump effekter verkar vara bättre på att hitta en signifikant interaktionseffekt när en sådan effekt existerar. En förklaring till detta kan vara att en modell med slump effekter, där man tar hänsyn till variationen i en population av batcher, tenderar att ge prediktioner som är krympta mot det gemensamma populationsmedelvärdet [4]. Detta kan leda till mer konservativa skattningar av interaktionstermens betydelse. Dessutom kan den stora osäkerheten i skattningen av variansparametrarna i en modell med slump effekter leda till att det är svårare att hitta en signifikant interaktion mellan tid och batch.

## 6.6 Konklusion

När stabilitetsstudier innehåller longitudinella data behöver man ta hänsyn till detta vid beräkning av Ppk. Att välja rätt modell för att analysera en stabilitetsstudie kan vara svårt om studien innehåller få batcher och observationer. Simuleringar kan då vara ett kostnadseffektivt sätt att analysera skillnader mellan olika modeller. Huruvida en interaktion mellan tid och batch bör inkluderas i analysen, beror på om den bakomliggande processen motiverar en sådan. Våra resultat visar att det i MHC:s stabilitetsstudier är fördelaktigt att ta hänsyn till en interaktion. Analyser av befintliga studier tyder inte på att det finns någon relevant skillnad mellan modeller med eller utan slump effekter. Att använda sig av en modell med slump effekter tillåter oss däremot att ta hänsyn till variationen av en parameter på populationsnivå, vilket öppnar för mer generella slutsatser. Dessutom visar våra beräkningar och simuleringar, att en modell med slump effekter producerar stabilare resultat och är mer robust för störningar i datan. Utifrån detta förespråkar vi därför en modell med slump effekter och en interaktion mellan tid och batch.

## Referenser

- [1] European Medicines Agency. (2003). *ICH Q1A (R2) Stability testing of new drug substances and drug products*.
- [2] European Medicines Agency. (2003). *ICH Q1E Evaluation of stability data*.
- [3] Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction*. Hoboken, NJ: John Wiley and Sons.
- [4] Rawlings, J.O., Pantula, S.G., & Dickey, D.A. (2001). *Applied Regression Analysis: A Research Tool, 2nd Edition*. New York, NY: Springer-Verlag.
- [5] Galecki, A., & Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. New York: Springer.
- [6] R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- [7] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Fitting Linear Mixed-Effect Models using lme4*. *Journal of Statistical software*, 67(1), 1-48.
- [8] Lynch, B., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Oxford: Sinauer Associates.
- [9] Knowles, J.E., & Frederick, C. (2019). *merTools: Tools for Analyzing Mixed Effect Regression Models*. R package version 0.5.0. URL: <https://CRAN.R-project.org/package=merTools>

## 7 Appendix

**Tabell A1:** Tabellen visar den genomsnittliga relativa feltermen för alla tillgängliga studier och testparametrar vid felaktig modellspecifikation beräknat på 1000 simuleringar och sedan skattad med en modell utan slump effekter. Den vänstra kolumnen är när datan simuleras med en interaktion, samt den högra är när datan är simulerad utan interaktion.

|                          | Med interaktion | Utan interaktion |
|--------------------------|-----------------|------------------|
| Seal.strength,.T-238.(N) | 11.49           | -1.23            |
| Seal.width,.T-238.(mm)   | 4.44            | -2.29            |
| T-229.MD.dry.(N)         | 1.08            | -0.92            |
| T-229.MD.wet.(N)         | 1.95            | -0.87            |
| T-229.CD.dry.(N)         | 2.84            | -0.94            |
| T-229.CD.wet.(N)         | 1.79            | -1.32            |
| T-1099.front.(N)         | 7.02            | -1.92            |
| T-1099.sleeve.(N)        | 2.01            | -1.99            |
| T-1006.PM.Front          | 0.73            | -0.08            |
| T-1006.PM.Reverse        | 0.89            | -0.14            |
| T-1006.Count.Front       | 3.39            | -0.67            |
| T-1006.Count.Reverse     | 11.85           | -0.81            |
| T-1179.Front.Dry         | 3.63            | -0.79            |
| T-1179.Front.Wet         | 3.54            | -1.27            |
| T-1179.Reverse.Dry       | 4.03            | -0.53            |
| T-1179.Reverse.Wet       | 40.25           | -1.64            |
| T-229.MD.wet             | 3.24            | -0.90            |
| T-229.MD.dry             | 2.85            | -1.37            |
| Mean.Load(N)T-238        | 0.71            | -0.53            |
| T-238-Mean.load.[N]      | 1.49            | -0.10            |
| Mean.load.(N)            | 3.30            | -0.46            |
| Mean.Load(N)             | 2.76            | -0.50            |
| T-229.CD.wet             | -17.27          | 9.10             |
| T-229.CD.dry             | 9.18            | -1.94            |
| Minimum                  | -17.27          | -2.29            |
| Maximum                  | 40.25           | 9.10             |
| Genomsnitt               | 4.47            | -0.59            |

**Tabell A2:** Tabellen visar den genomsnittliga relativa feltermen för alla tillgängliga studier och testparametrar vid felaktig modell-specifikation beräknat på 1000 simuleringar och sedan skattad med en modell med slump effekter. Den vänstra kolumnen är när datan simulerats med en interaktion, samt den högra är när datan är simulerad utan interaktion.

|                          | Med interaktion | Utan interaktion |
|--------------------------|-----------------|------------------|
| Seal.strength,.T-238.(N) | 5.26            | 0.13             |
| Seal.width,.T-238.(mm)   | 3.12            | -0.64            |
| T-229.MD.dry.(N)         | 0.75            | -0.49            |
| T-229.MD.wet.(N)         | 3.98            | -0.54            |
| T-229.CD.dry.(N)         | 5.24            | -0.37            |
| T-229.CD.wet.(N)         | 2.14            | -0.90            |
| T-1099.front.(N)         | 13.26           | -0.71            |
| T-1099.sleeve.(N)        | 0.83            | -0.59            |
| T-1006.PM.Front          | -0.00           | 0.00             |
| T-1006.PM.Reverse        | 0.00            | 0.00             |
| T-1006.Count.Front       | 0.00            | 0.00             |
| T-1006.Count.Reverse     | 0.00            | 0.00             |
| T-1179.Front.Dry         | 7.02            | 0.02             |
| T-1179.Front.Wet         | 5.82            | -0.44            |
| T-1179.Reverse.Dry       | 11.60           | -0.04            |
| T-1179.Reverse.Wet       | 15.65           | -0.11            |
| T-229.MD.wet             | 6.21            | -0.18            |
| T-229.MD.dry             | 5.61            | -0.83            |
| Mean.Load(N)T-238        | 0.98            | 0.18             |
| T-238-Mean.load.[N]      | 4.79            | 0.54             |
| Mean.load.(N)            | 8.08            | -0.41            |
| Mean.Load(N)             | 6.13            | -0.04            |
| T-229.CD.wet             | 4.14            | 0.28             |
| T-229.CD.dry             | 3.82            | 0.01             |
| Minimum                  | -0.00           | -0.90            |
| Maximum                  | 15.65           | 0.54             |
| Genomsnitt               | 4.77            | -0.21            |