



**LUNDS**  
UNIVERSITET

# **Kvantitativ modellering av celltillväxt vid exponering av giftiga substanser – en analys av holografiska experiment**

My Nguyen

Viktoria Antunovic

Kandidatuppsats, 15 hp

Statistiska institutionen, Lunds universitet

Handledare: Peter Gustafsson

VT 2024

# Sammanfattning

Denna uppsats är skriven på uppdrag av ett företag som har utvecklat ett mikroskop för att stödja forskning av celler. Mikroskopet är det första som kan ta detaljerade bilder av celler vid olika utvecklingsfaser och är avsedd att hjälpa forskare att förstå cellers beteende, utveckling och tillväxt under cellens livstid.

De experiment och den förberedande statistiska behandlingen av mätningarna som vår analys grundar sig på har gjorts av företaget.

Huvudsyftet med uppsatsen är att undersöka och belysa de tillhandahållna cellexperimentens statistiska egenskaper samt även för att avgöra om det finns någon statistiskt säkerställd behandlingseffekt på cellernas tillväxt/tillväxthastighet när de odlas i ett medium med tillsatta giftiga substanser i olika doser. Det är också av intresse att analysera och modellera samt kvantifiera cellernas tillväxthastighet (deras celldelningsintensitet) under en längre tidsperiod (48 timmar) och hur den påverkas av aktuell dos. Studien, analysen och modelleringen görs med hjälp av enkel och multipel regressionsanalys.

Resultaten i uppsatsen har bland annat betonat vikten av kontrollgruppen i varje experiment. Jämförelser mellan kontrollgrupperna avslöjar även experimenteffekter samt korrelationen mellan initialt cellantal och tillväxthastighet, där färre initialceller leder till högre tillväxt. Residualanalyser har påvisat att modellen uppfyller antaganden för kontrollgruppen och lägre doser, medan högre doser uppvisar negativa tillväxtkoefficienter. En möjlig förklaring är att giftpåverkan hämmar celltillväxten för högre doser och att celldelningsprocessen fungerar annorlunda vilket inte ger tillförlitlig mäteffekt. Statistiken och modelleringen i experimentet bör därför baseras på celldelningsintensiteten, särskilt under logfasen av celldelningscykeln. En logaritmisk transformation av cellantalet,  $\log(Y)$  är särskilt lämplig för att fånga den exponentiella tillväxtfasen (log-fasen) hos en cellkultur. För låga doser av giftet/preparatet kan effekterna på celldelningsintensiteten kvantifieras enkelt med hjälp av linjär regressionsmodell. Den gemensamma regressionsmodellen för dos och kontroll indikerade en korrelation mellan position och dos, vilket ledde till att variabeln position

uteslöts och endast tid och dos användes som förklarande variabler. Resultatet blev en mindre tillfredsställande skattning av tillväxthastigheterna. En möjlig förklaring till detta kan vara den varierande påverkan av giftmängden vilket komplicerade kvantifieringen av effekten.

Det är värt att notera att för samtliga experiment är den skattade tillväxthastigheten avsevärt lägre vid de höga doserna, dos tre och fyra, än för kontrollgrupperna samt experimenten vid dos ett och två.

*Nyckelord: Celltillväxt, Cellantal, Kontrollgrupp, Dos, Position, Brunn, Regressionsanalys*

## Abstract

This essay has been written at the request of a company that has developed a microscope to support cell research. The microscope is the first to capture detailed images of cells at different stages of development and is intended to help biologists understanding cell behavior, development and growth of cells throughout their lifetime.

The experiments and the preparatory statistical treatment of the measurements on which our analysis is based have been done by the company.

The main purpose of the essay is to investigate and highlight the statistical properties of the provided cell experiments and to determine if there is any statistically significant treatment effect on cell growth/growth rate when applying toxic substances in different doses. It is also of interest to analyze model and quantify the growth rate of the cells (their cell division intensity) over a longer period of time (48 hours) and how it is affected by the current dose. The study, analysis and modeling are done using simple and multiple regression analysis.

The results of the thesis emphasize, among other things, the importance of the control group in each experiment. Comparisons between the control groups also reveal experimental effects as well as the correlation between initial cell count and growth rate, where fewer initial cells lead to higher growth rate. Residual analyses have shown that the model meets assumptions for the control group and lower doses, while higher doses exhibit negative growth coefficients. A possible explanation is that the toxic effect inhibits cell growth at higher doses and that the cell division process functions differently, leading to unreliable measurement effects. Therefore, the statistics and modeling in the experiment should be based on cell division intensity, especially during the log phase of the cell division cycle. A logarithmic transformation of the cell count,  $\log(Y)$ , is particularly suitable for capturing the exponential growth phase (log phase) of a cell culture. For low level of the toxin/preparation, the effects on cell division intensity can be quantified using a linear regression model. The regression model for dose and control indicated a correlation between position and dose, leading to the exclusion of the position variable and using only time and dose as explanatory variables. This resulted in a less



satisfactory estimation of growth rates. A possible explanation for this could be the varying impact of the toxin amount, which complicates the quantification of its effect.

It is noteworthy that, in all experiments with higher doses, dose three and four, the estimated growth rate is significantly lower than in the control groups and the experiments at dose one and two.

*Keywords: Cell growth, Cell count, Control group, Dose, Position, Well, Regression analysis*

# Innehållsförteckning

<b>1 Inledning</b> .....	<b>7</b>
<b>2 Teori</b> .....	<b>10</b>
2.1 Celltillväxt ur ett biologiskt perspektiv.....	10
2.1.1 Cellcykeln .....	10
2.1.2 Tillväxtfaser .....	13
2.2 Regressionsanalys.....	14
2.2.1 Enkel linjär regression .....	14
2.2.2 Multipel Regressionsmodell.....	15
2.3 Minsta kvadratmetoden .....	16
2.4 Analys av modellerna.....	17
2.4.1 Multikolinjäritet .....	17
2.4.2 Justerad förklaringsgrad, $R_{adj}^2$ .....	17
2.5 Residualanalys .....	18
2.5.1 Leverage och Outliers .....	18
2.5.2 Konstant varians .....	19
2.5.3 Normalfördelning.....	19
2.6 Chi-två test.....	19
2.7 Konfidensintervall för skattad parameter.....	20
<b>3 Experimentell uppställning</b> .....	<b>21</b>
3.1 Experiment 1 .....	22
3.2 Experiment 2 .....	22
3.3 Experiment 3 .....	23
3.4 Positionernas betydelse .....	23
3.4 Komplexitet i experimenten .....	23
<b>4 Statistisk modellering av celltillväxten</b> .....	<b>25</b>
4.1 Kontrollgrupp .....	25
4.1.1 Enkel linjär regression .....	27
4.1.2 Multipel linjär regression .....	31
4.2 Analys av doser .....	33
4.2.1 Separata regressionsmodeller för var sin dos.....	34
4.2.2 Gemensam regressionsmodell för alla doser.....	36
4.2.3 Konfidensintervall för tillväxthastigheten - doseffekt .....	38
<b>5 Diskussion</b> .....	<b>41</b>
Möjligheter och vidare forskning.....	42
<b>7 Bilagor</b> .....	<b>46</b>

# 1 Inledning

Inom biologi är forskning om celler ett brett och viktigt ämnesområde. Vare sig det gäller människo- eller djurceller är det lika viktigt att forskningen går framåt för att vi efterhand bättre ska kunna förstå cellers beteende, tillväxt och utveckling.

Mikroskopets historia sträcker sig ända bak till 1400-talet men det var först under det senare 1600-talet som mikrobiologins pionjärer Robert Hooke och Anthony von Leeuwenhoek utvecklade mikroskop som kunde användas för mikroorganismer. Hook var först med att upptäcka cellen och utgav boken "Micrographia". Ett par år senare observerade Leeuwenhoek strukturen hos blodceller och var också den första med att upptäcka bakterien. Sedan dess har många forskare utvecklat och förbättrat mikroskopet och idag har vi även mikroskop som inte är mycket större än en tjock penna (Fields, 2019); (Microscope, u.å).

Företaget har utvecklat ett mikroskop som förenklar den dagliga forskningen om celler och erbjuder mer tydliga och exakta mätningar. Företagets VD och grundare beskriver förenklat mikroskopets funktion genom att den skickar in ljusvågor från två olika håll genom cellerna, det bildas då mönster när ljusvågorna möter cellen och skapar en bild av cellen. Cellerna förblir opåverkade av ljuset då ingen energi absorberas (Källa 1) i cellen. I en intervju med grundaren förklarar han att det inte är ett s.k. mikroskop utan en cytometer. Enligt grundaren är den avgörande skillnaden att "Mikroskop avbildar celler. Cytometrar mäter celler.", (Källa 2). Därför kommer vi härnäst referera till mikroskopet som en cytometer.

Fördelarna med cytometern är att det genererar visuella element i form av videor och sekvenser av bilder. På så sätt går det att visuellt följa hur cellerna uppför sig vid de olika faserna under deras livscykel. Cytometern genererar även större mängder av kvantitativa data om hur cellodlingar förändras med en hög tidsupplösning över lång tid, både för individuella celler och hela populationen. Detta ger tillgång till fler mätpunkter och öppnar för en mer omfattande och djuplodande forskning. Med hjälp av digital holografi kan cellodlingar alltså analyseras utan att något förstörande preparat eller strålning behöver användas (Källa 1).

## Bild på maskin

*Figur 1: Bild över maskin*

### *Cell Proliferation*

Cell Proliferation, celltillväxt (cellförökning) på svenska, ligger till grund för att kunna förstå hur bakterier, virus och sjukdomar angriper och utvecklas. (Wexler, 2014) förklarar att celltillväxten är en väldigt komplex process, vilket kräver en djup förståelse av de bakomliggande mekanismerna. Det är därför viktigt att kunna förstå och fastställa hur cellcykeln påverkas av olika tillväxtfaktorer. Wexler påpekar vidare att särskild uppmärksamhet riktas mot att förstå processen när celltillväxten "skadas" av giftiga substanser eller medel.

Enligt ursprunget initiativtagarna till denna uppsats, har man vanligtvis inom forskningen inte tillgång till så många mätpunkter per tidsenhet vid studier av celltillväxt. Analyser utformas vanligtvis på så sätt att endast ett fåtal punkter väljs ut och används för att beräkna normalvärde och standardavvikelser, därefter beräknas skillnaden i cellantalet för de valda tidpunkterna.

Med hjälp av den välutvecklade tekniken hos cytometern ges tillgång till större datamängder vilket väcker intresset för hur celltillväxten skulle kunna mätas på ett mer utförligt/tillfredsställande sätt. För att kunna besvara denna fråga är det viktigt att undersöka om metoden för datainsamling uppfyller nödvändiga statistiska egenskaper. Därför är vårt fokus i uppsatsen att analysera och validera uppställningen av de tre experiment som har utförts av företaget.

Varje experiment innehåller en kontrollgrupp som är en renodling av celler samt fyra cellkulturer med olika doser av det giftiga ämnet tillsatt. Med utgångspunkt från uppläggen och resultaten av de tre experimenten ska vi i uppsatsen försöka bringa klarhet kring tillvägagångssättet för att bland annat kunna beräkna cellernas tillväxthastighet, delningsintensitet. Enkla linjära regressionsmodeller används generellt för att skatta tillväxthastigheterna.

I uppsatsen ämnar vi även att besvara frågan om det är möjligt att kvantifiera effekten av olika doser av för cellen giftiga substanser på dess tillväxthastighet med hjälp av de givna experimentsuppställningarna. Multipla regressionsmodeller används i denna del av analysen, eftersom vi tar i beaktande flera oberoende variabler såsom positions-, brunn- och experimenteffekter.

För att kunna studera hur cellerna blir påverkade av giftet behöver vi förstå hur deras tillväxt och celledelning ser ut i dess normala tillstånd. Tyngdpunkten för uppsatsen ligger därför i att studera tillväxthastigheter för kontrollgruppen i de tre experimenten. Kontrollgruppen kommer även fungera som en referens vid jämförelse av tillväxten vid de andra doserna. Samtliga analyser och tester av datamaterialet har utförts i statistikprogrammet R.

## 2 Teori

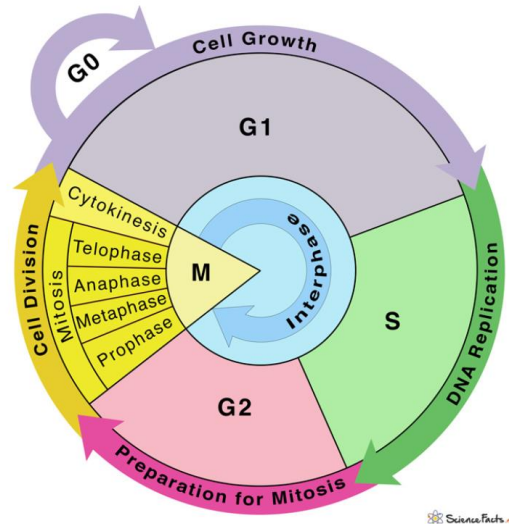
Teorin som ligger till grund för denna uppsats presenteras i två huvudavsnitt. Cellbiologin går igenom i avsnitt 2.1 och statistiken i avsnitt 2.2.

### 2.1 Celltillväxt ur ett biologiskt perspektiv

Kunskap om cellcykeln och dess olika faser är en av de mest centrala delarna inom cellbiologin. Att förstå hur celldelningen går till och vilka händelser en cell genomgår vid delningsprocessen är väsentligt för forskningen. Det är alltså nödvändigt att ha en förståelse för hur celldelningen förlöper och hur tillväxthastigheten ska beräknas för att kunna genomföra en analys. I detta avsnitt kommer vi att presentera principerna för cellcykeln, de olika faser en cell genomgår under sin tillväxt - själva celldelningsprocessen.

#### 2.1.1 Cellcykeln

Erlanson-Albertsson och Gullberg (2007) beskriver cellcykeln som en väl kontrollerad process som cellen går igenom vid sin delning. Den karaktäriseras av två huvudsakliga faser: mitos (M-fas) och interfase. Interfasen består i sin tur av tre olika faser som betecknas Gap 1 (G1), Syntes (S) och Gap 2 (G2). Under cellcykeln finns även ett antal kontrollstationer mellan interfasen och M-fasen, där det säkerställs att cellen successivt har genomgått faserna, att varje fas är korrekt utförd och att alla villkor för celldelning är uppfyllda. Om inte alla kriterier är uppfyllda backas celldelningsprocessen och återupptas efter åtgärd, en sorts kontrollmekanism (Källa 1). Det är också viktigt att förstå G0-fasen som en viktig fas (Källa 1). Under G0-fasen, den så kallade vilofasen, slutar cellerna att dela sig och förblir inaktiva under lång tid.



Figur 2: Cellcykelns faser, sciencefacts.net

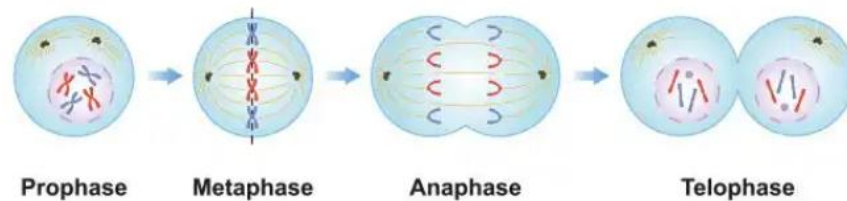
### *Interfasen*

Under interfasen sker cellens metabola aktiviteter, proteiner, lipider och kolhydrater tillverkas. Cellen ökar även i storlek under denna fas (Erlanson-Albertsson & Gullberg, 2007). Under den första fasen G1 sker det inte särskilt mycket förutom att en tillväxt sker av cellen, den blir större, mängden proteiner och kolhydrater ökar och den samlar på sig tillräckligt med energi för nästkommande faser. Nästa fas är S-fasen, där cellens DNA replikeras så att varje dottercell kan få ett komplett set kromosomer efter celldelningens uppsättning. S-fasen övergår sedan till G2-fasen där det sker en förberedelse för delning av cellen, cellen växer ytterligare och en ökning av mängden protein sker återigen. Här säkerställs även att cellen är i rätt skick och är korrekt innan själva celldelningen startar. (Källa 1); (Henriksson, 2019).

### *Mitos-fasen*

Under M-fasen, även kallad för mitos, sker själva celldelningen och den består av fem delfaser: Profas, Prometafas, Metafas, Anafas, Telofas (Wiklund, 2014). Enkelt beskrivet är det under dessa faser som kromosomerna i cellkärnan kopieras och fördubblas. Kromosomerna sitter sedan i par i formen av ett X. Kärnmembranet, en dubbelskiktad membranstruktur som omger cellens cellkärna och fungerar som en skyddande barriär som reglerar vad som går in och ut ur kärnan, löses upp och kromosomparen dras isär åt varsin pol inne i cellen.

Cellmembranet drar ihop sig och kärnmembranet återbildas runt kromosomerna med hjälp av speciella proteiner i cellkärnan. Nu har två nya dotterceller bildats som har exakt lika många kromosomer, med samma uppsättning DNA som modercellen (Erlanson-Albertsson & Gullberg, 2007).



*Figur 3: Mitos-fasens delar, turito.com*

*Figuren visar mitosfasens 4 huvuddelar: (1) Profas: Kärnmembranet bryts ner för att frigöra kromosomerna. (2) Metafas: Kromosomerna är i linje i cellens mitt och fästa vid spindelmikrotubuli (3) Anafas: Kromosomerna går sönder och dras till motsatta sidor av cellen. (4) Telofas: Två nya kompletta kärnor bildas och cellen kan dela sig i två dotterceller. Kärnmembranet återbildas.*

Efter mitosen går de båda dottercellerna in i antingen G0 fasen, cellerna lämnar cellcykeln och slutar dela sig, eller så fortsätter de till G1 fasen och cellcykeln upprepar sig. Cellen kan vara i G0 fasen för alltid eller endast tillfälligt. Många av kroppens celler ligger i G0 fasen men om tex en skada sker så kan de påbörja cellcykeln åter igen (Källa 1); (Henriksson, 2019).

Enligt Kielberg et al. (1994) fördubblas cellerna med fasta intervall om de befinner sig under goda tillväxtförhållanden med avseende på bl.a. temperatur, celltäthet, syre, näringsämnen. Det betyder att vid celledningen ombildas en cell till två celler som i sin tur blir 4 celler, 8 celler osv. Den tid det tar för cellerna att fördubblas i antal kallas fördubblingstiden och den varierar för olika celltyper.

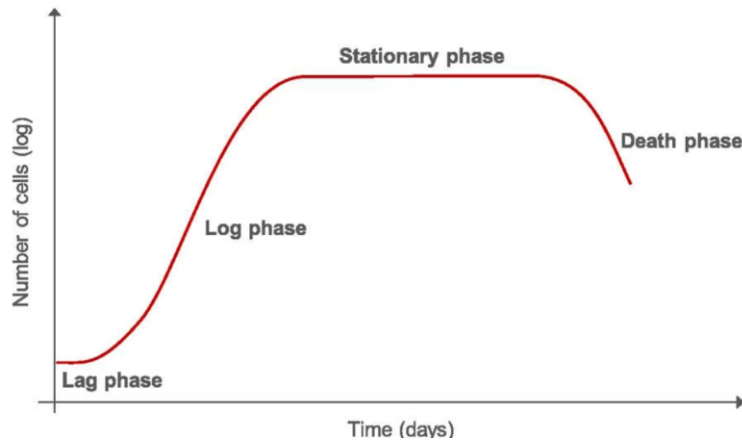


### 2.1.2 Tillväxtfaser

Tillväxtkurvor visar förändringen i antalet celler som funktion av tiden och är enligt Kielberg et al. (1994) det bästa sättet att bestämma tillväxthastigheten hos en cell. Vid cellodlingen går cellen igenom fyra tillväxtfaser, (se Figur 4 för den principiella indelningen av tillväxtkurvan i sina olika faser). Enligt Wiklund (2014) är det väsentligt att veta hur många celler som finns från början för att kunna göra en relevant jämförelse mot slutet av tillväxten. Vid cellodling behöver man ta hänsyn till flera viktiga faktorer som kan påverka cellodlingen och därmed resultatet. Det är av vikt att beakta cellcykeln och cellens metabolism, samt att kontrollera pH-värde, temperatur, näringstillgång, tillväxtfaktorer och osmos (jonkoncentration) för att cellen ska kunna föröka sig så optimalt som möjligt.

Första tillväxtfasen heter *lag-fasen* och enligt Freshney (1994) är det den period då cellen anpassar sig, det sker ingen eller väldigt långsam tillväxt eftersom cellerna just har kommit in i den nya miljön och behöver tid för att vänja sig. När cellerna har anpassat sig till miljön börjar de föröka sig och hamnar då i *log-fasen* där tillväxttakten är hög och kurvan går stadigt uppåt. Cellerna mår som bäst i denna fas och cellantalet fördubblas med bestämda tidsintervall, tillväxttakten ligger på ungefär 90-100% (Freshney, 1994). Därför kallas denna fas även för den exponentiella tillväxtfasen. Det är de celler som befinner sig i denna fas som är bäst lämpade för odlingsförsök.

Om denna kurva är rätlinjig påstår Kielberg et al. (1994) att cellerna växer bra eftersom deras tillväxtförhållanden är gynnsamma. Enligt Freshney (1994) finns det ett antal faktorer som påverkar hur länge denna fas pågår, bl.a. nämner han såddensitet och cellens tillväxttakt. När tillväxten sedan avstannar hamnar cellerna i vilostadiet och tillväxtkurvan planar ut. Denna fas kallas *platåfasen* och kan orsakas av att cellerna uppnått konfluens. Med konfluens menas att cellerna täcker botten i cellodlingsflaskan eller petriskålen. Tillväxttakten ligger mellan 0%-10% i denna fas (Freshney, 1994). En annan orsak till konfluens kan vara brist av tillväxtfaktorer och näring eller att något tillväxthämmande ämne är tillfört. Wiklund (2014) förklarar detta som att cellerna hamnar i denna fas när man slutar ”mata” dem med mer näring, då dör det lika många celler som det skapas. Till sist hamnar cellerna i dödsfasen så kallad *declin-fas*. Inga nya celler bildas och cellerna dör efterhand. Tillväxtkurvan pekar därefter nedåt.



Figur 4: Tillväxtkurvor, leica-microsystems.com

## 2.2 Regressionsanalys

### 2.2.1 Enkel linjär regression

För att studera sambandet mellan variablerna i uppsatsen använder vi oss av enkel och multipel linjär regressionsanalys. Regressionsanalys används för att finna om det finns ett linjärt samband mellan två eller flera variabler. Modellen för enkel linjär regression beskriver sambandet mellan en beroende variabel ( $Y$ ) och en oberoende variabel ( $x_i$ ). Ekvationen för den enkla linjära regressionsmodellen skrivs enligt följande:

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad (1)$$

där  $\beta_0$  representerar interceptet,  $\beta_1$  representerar värdet för linjens lutning och  $e_i$  representerar feltermerna.

När vi studerar en variabel ( $y$ ) som beror linjärt av en variabel ( $x$ ) analyseras en slumpmässig störning eller avvikelse. En enkel linjär regressionsmodell bygger på fyra grundläggande antagande enligt (Sheather, 2009, 21):

1. Linjära modellen beskrivs som:  $Y_i = \beta_0 + \beta_1 x_i + e_i$
2. Feltermerna är oberoende, dvs  $\text{Cov}(e_i, e_j | X_i X_j) = 0$  om  $i \neq j$ .
3. Feltermerna har samma varians, dvs  $\text{Var}(e_i | X_i) = \sigma^2$
4. Feltermerna är normalfördelade, dvs  $e_i | X_i \sim N(0, \sigma^2)$  med ett väntevärde på 0 och variansen  $\sigma^2$ .

Dessa fundamentala förutsättningar måste vara uppfyllda för att modellen ska vara giltig. Om det uppstår/finns någon avvikelse från modellantagandena kommer vi att få någon typ av fel i bland annat prediktionsintervall, konfidensintervall eller signifikans (Sheather, 2009). För att undersöka om dessa villkor är uppfyllda utgör residualanalysen ett mer eller mindre oundvikligt verktyg och steg i veriferingen av den linjära regressionsmodellen.

### 2.2.2. *Multipel Regressionsmodell*

En multipel regressionsanalys bygger på samma antaganden som den enkla linjära regressionsmodellen (Sheather, 2009, 21). Skillnaden är att en multipel linjär regression innehåller flera oberoende - förklarande - variabler ( $x_{ji}$ ) som tillsammans förklarar variationen i den beroende variabeln ( $Y_i$ ).

Ekvationen för den multipla regressionsmodellen skrivs enligt följande:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad (2)$$

Enligt Cohen et al. (2013) är en förklaringsvariabel nominell om det finns flera olika kodningssystem och om alla kodningssystemen använder  $g - 1$  antal kodade variabler för att representera de  $g$ -grupperna, där  $g$  anger antal grupper/kategorier som finns hos denna variabel. I denna uppsats kommer ett dummy variabel kodning system att tillämpas. Detta innebär att en referensgrupp väljs ut och tilldelas värdet 0 för varje kodad variabel. För de andra  $g-1$  kategorierna/grupperna får var och en koden 1 medan de återstående kategorierna är kodade till 0.

Det finns tre aspekter som man behöver ta hänsyn till vid valet av referensgrupp. För det första bör referensgruppen fungera som en användbar jämförelse. För det andra bör referensgruppen vara väldefinierad och inte klassas som en grupp som innehåller värden som inte kan kategoriseras. För det tredje bör referensgruppens storlek inte vara mycket mindre i jämförelse med de andra grupperna.

Valet av referensgrupp tycks inte vara väsentligt vid de experiment som vi har analyserat eftersom alla positioner i experimentuppställningarna som kan användas som förklarande variabel förekommer lika ofta i alla experiment och ingen position verkar ha någon större betydelse än andra positioner.

## 2.3 Minsta kvadratmetoden

Inom regressionsanalys används minsta kvadratmetoden för att ta fram parameterskattningarna i den valda modellen. Skattningarna beräknas genom att minimera RSS som står för ”sum of squared residuals” och mäter nivån av variansen för feltermerna. Vi kommer att presentera minsta kvadratmetoden enligt (Sheater, 2009, 15-20).

Formeln för RSS ges enligt:

$$RSS = \sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

För att RSS ska kunna minimeras för  $\hat{\beta}_0$  och  $\hat{\beta}_1$  gäller att:

$$\frac{\partial RSS}{\partial b_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Och

$$\frac{\partial RSS}{\partial b_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Vilket ger oss formlerna nedan som beskriver normalekvationen:

$$\sum_{i=1}^n y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^n x_i$$

och

$$\sum_{i=1}^n x_i y_i = \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2$$

Till sist ges minsta kvadratskattningen av interceptet som representeras enligt:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Och av lutningen enligt:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{SXY}{SXX} \end{aligned}$$

## 2.4 Analys av modellerna

### 2.4.1 Multikolinjäritet

När en modell består av fler oberoende variabler är det viktigt att undersöka om det finns multikolinjäritet som visar på ett starkt linjärt beroende mellan de förklarande variablerna (x). Enligt Cohen et al. (2013) uppstår multikolinjäritet när en eller flera oberoende variabler är starkt korrelerade med de andra oberoende variablerna i regressionskvationen. Detta leder till att en skattning av regressionskoefficienten för den korrelerade prediktorn kommer att vara oerhört otillförlitlig eftersom det innehåller för lite unik information för att uppskatta dess värde. Skattning av regressionskoefficienten får ett stort medelfel och blir ofta svår att tolka. Då den starka korrelationen enligt (Sheather, 2009, 195-203) bland annat leder till varierande parameterskattningar i olika stickprov, breda konfidensintervall och missvisande resultat anses multikolinjäritet inte vara ett önskvärt tillstånd. För att mäta multikolinjäritet kan man gå tillväga på flera olika sätt, till exempel genom att undersöka korrelationsmatrisen, använda ett F-test, undersöka VIF-värdet, chi-två test med mera. Vi har valt att utföra det sistnämnda som presenteras i kapitel 4 eftersom våra variabler är nominella variabler.

### 2.4.2 Justerad förklaringsgrad, $R_{adj}^2$

Förklaringsgraden  $R^2$  kan alltid undersökas vid jämförelser av modeller, måttet är inte avtagande utan ökar ju fler förklarande variabler man använder, men tar inte hänsyn till komplexitet i modellen. Därför används istället justerad förklaringsgrad;  $R_{adj}^2$ , där frihetsgraderna i modellen används för att ta hänsyn till modellens komplexitet.

Formeln för  $R_{adj}^2$  ges enligt:

$$R_{adj}^2 = 1 - \frac{RSS/(n - p - 1)}{SST/(n - 1)}$$

där RSS (residual sum of squares) mäter nivån av variansen för feltermerna, SST mäter den totala variationen som finns i den beroende variabeln, p är antalet prediktorer i modellen (Sheather, 2009, s.228).

## 2.5 Residualanalys

Som vi har nämnt tidigare är residualanalys ett oundvikligt steg inom regressionsanalys. I detta avsnitt kommer vi att gå in på teorin och beskriva de olika måtten som används vid en sådan analys. Residualanalysen som har gjorts presenteras sedan i resultatavsnittet.

### 2.5.1 Leverage och Outliers

Leverage mått används för att identifiera inflytelserika fall, det vill säga extrema värden som kan påverka resultatet när de inkluderas eller utesluts från analysen. Leverage mäter inflytande av  $x_i$  värde på  $\hat{y}_i$  värde, dvs det används för att identifiera inflytelserika observationer som kan vara extrema värden och påverka modellen. Ett sätt att upptäcka outlier är att använda standardiserade residualer ( $r$ ). Enligt en vanlig tumregel är en outlier en punkt som har ett  $r$ -värde som ligger utanför intervallet mellan  $-2$  och  $2$ . (Sheather, 2009, s.60).

Sheater (2009, s.64) förklarar att ett leverage-värde kan vara både bra och dåligt. Observationer med högt leverage-värde som inte följer mönstret har ett mer betydande inflytande på  $\hat{y}$ -värdet och bedöms vara en ”bad-outlier”. Det vill säga om observationens leverage-värde är högt samtidigt som den är en outlier är det dåligt, om däremot leverage-värdet är högt men observationen inte är en ”outlier” anses det istället vara en inflytelserik observation. Skillnaden mellan dessa kan även påvisas genom att en outlier avviker i  $y$ -led (höjdlid från regressionslinjen), medan en inflytelserik observation har ett avvikande  $x$ -värde.

*Cooks distance* är ett mått som har utvecklats för att kunna mäta om en observation har stort inflytande. En enkel tumregel är att måttet för leveragegränsvärdet beräknas genom att dividera 4 med stickprovsstorleken. Tumregeln är att de observationer som har leverage-värde ( $h_{ii}$ ) som är större än  $\frac{4}{n}$  anses ha högt leverage-värde (Sheather, 2009, s.51-52). Formeln för regeln ges enligt följande:

$$h_{ii} > 2 \times \text{average}(h_{ii}) = 2 \times \frac{2}{n} = \frac{4}{n}$$

### 2.5.2 Konstant varians

Konstant varians även benämnt *homoskedasticitet* är önskvärt för att feltermerna ska uppfylla kravet för stabilitet. En horisontell linje med jämnt fördelade punkter är att föredra. Om modellen inte följer en konstant varians, om variansen för residualerna till exempel ökar med de anpassade värdena (fitted values) kan det finnas heteroskedasticitet. Det finns olika transformationer som kan användas, bland annat kan en logaritmering av den beroende variabeln i modellen leda till homoskedasticitet, (Sheather, 2009, s.71).

### 2.5.3 Normalfördelning

Ett av kraven för att kunna använda en regressionsmodell är att feltermerna är normalfördelade. En QQ-plot (Quantile-Quantile plot) är ett verktyg som används för att jämföra fördelningen av en datamängd med en specifik teoretisk fördelning, oftast en normalfördelning. Denna är särskilt användbar för att bedöma om feltermerna i en regressionsmodell är normalfördelade. I QQ-ploten ritas de teoretiska kvantilerna på x-axeln och de standardiserade feltermerna på y-axeln. Om punkterna ligger nära den teoretiska styrlinjen tyder det på att feltermerna är normalfördelade. Om punkterna däremot avviker från denna raka linje kan detta indikera att residualerna inte är normalfördelade, vilket kan påverka giltigheten av regressionsresultatet (Sheather, 2009, s.70).

## 2.6 Chi-två test

När fler än en oberoende variabel är involverade i en regressionsmodell är det nödvändigt att analysera om dessa oberoende variabler korrelerar med varandra. I vår analys kommer vi att behöva kontrollera om variabel brunn, position och dos associerar med varandra. Oberoendet av två nominella variabler kan undersökas med hjälp av ett chi-två test (Cortinhas & Black, 2012, 697-699). Nollhypotesen om att två variabler är oberoende av varandra testas mot mothypotesen att det finns ett beroende. Formel för chi-två testet för oberoende ser ut enligt nedan:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

där  $O_{ij}$  är den observerade frekvensen i cell  $i, j$  i den  $i$ :te raden och  $j$ :te kolumnen.  $E_{ij}$  är den förväntade frekvensen för motsvarande observation. Antal frihetsgrader bestäms av försökssituationen och är beräknad som  $(r-1)(k-1)$ , där  $r$  står för antal rader och  $k$  står för antal kolumner i tabellen. Nollhypotesen förkastas om testvariabeln överstiger motsvarande kritiska värde.

## 2.7 Konfidensintervall för skattad parameter

Julious (2004) beskriver i sin artikel att det finns ett enkelt visuellt sätt att avgöra om det finns någon statistiskt signifikant skillnad mellan olika gruppers medelvärden. Genom att rita upp/plotta deras konfidensintervall tillsammans i en figur och kontrollera om dessa överlappar varandra eller ej. Om konfidensintervallen inte överlappar betyder det att deras medelvärden är signifikant skilda åt. I denna uppsats kommer vi att tillämpa detta tillvägagångssätt för att undersöka om olika skattade parametrar  $\widehat{\beta}_1$  är signifikant skilda åt. Ett konfidensintervall för  $\beta_1$  beräknas med formeln:

$$\widehat{\beta}_1 \pm t \sqrt{\left(\frac{s_e^2}{\sum(x_i - \bar{x})^2}\right)}$$

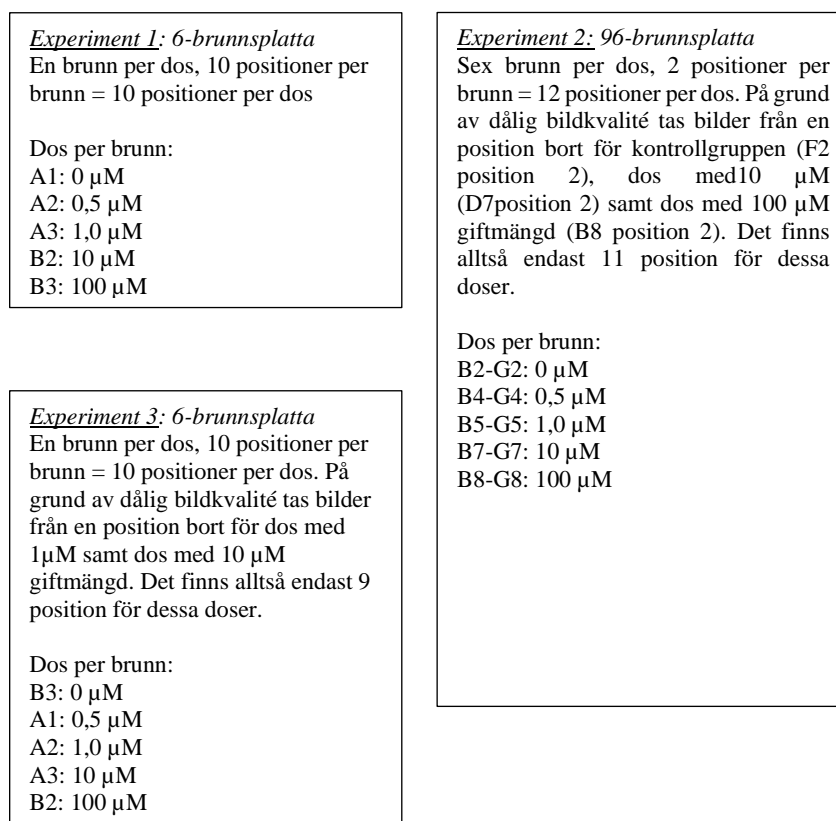
där  $t$  bestäms av antalet frihetsgrader  $f$  och konfidensgraden  $1 - \alpha$ . Eftersom alla experiment innehåller stor mängd data med många observationer kommer  $t$ -fördelningen att konvergera mot den standardiserade normalfördelningen, vilket ger ett  $t$ -värde på 1,96 för  $p$ -värdet 5%. Både  $\widehat{\beta}_1$  och medelfelet för skattningen av  $\beta_1$  kan läsas ut ur resultatet från regressionsanalysen. Ovanstående formel för konfidensintervall kan därför skrivas om till:

$$\widehat{\beta}_1 \pm 1,96 \cdot s_b$$



### 3 Experimentell uppställning

Datamaterialet för uppsatsen består av tre experiment. Cellerna odlas i 48 timmar i labbmiljö i plattor med antingen sex eller 96 brunnar. Varje platta innehåller samma medium för att säkerställa lika förutsättningar för att cellerna ska få optimala tillväxtförhållanden. En brunn i en sex-brunnsplatta är större än en i en 96-brunnsplatta. I figur 5 visas en övergripande bild av de tre experimenten och hur brunnarna på de olika plattorna har använts vid respektive experiment.



Figur 5: Bild över experimenten

Den optiska läsaren genererar stora mängder data med olika variabler. Vi har valt att begränsa denna analys till fem variabler: *reviderade antal celler*, *löpande tider i sekunder*, *tidsindex*, *positioner och brunn*. 'Reviderade antal celler' refererar till antalet celler som har justerats post-bildtagning för att korrigera cytometerns eventuella felbedömningar av cellidentifikation, baserat på cellens area och volym. Ett tidsindex motsvarar en mättidpunkt eller en bildsekvens. Eftersom vi har 97 bildsekvenser för hela mätserien får

vi alltså 97 tidsindex. Tidsindexen används främst för att plotta tillväxtkurvor för alla positioner i samma experiment som redovisas i en och samma plot. Positioner anger den platsen som cytometern ska ta bilder av. Varje position har unika koordinater och betraktas därför som ett oberoende stickprov i våra analyser. Hur många positioner det finns för varje dos beror på vilket experiment den tillhör. Vad gäller variabeln brunn så är den endast relevant i experiment två där cellodlingen görs i sex olika brunnar per dos, alltså sex olika replikat per dos. Vår utgångspunkt är att cellens tillväxthastighet är konstant för en given celltyp, vilket innebär att tiden för en cell att fördubblas bör vara enhetlig.

För varje experiment undersöks hur preparat påverkar celltillväxten vid fem olika doser. Första dosen är en kontrollgrupp där ingen gift tillsätts. Övriga doser innehåller samma preparat men i varierande koncentrationer: 0,5  $\mu\text{M}$ ; 1,0  $\mu\text{M}$ ; 10  $\mu\text{M}$  respektive 100  $\mu\text{M}$ . En mer utförlig beskrivning av förutsättningarna för respektive experiment redovisas nedan.

### 3.1 Experiment 1

I experiment ett odlas celler i en sex-brunnspatta och endast fem brunnar används. I respektive brunn väljs tio fasta positioner för bildtagning. Valet av positioner sker oftast manuellt för att garantera närvaro av celler, vilket kan resultera i en överestimation av cellantal men anses inte problematiskt då fokus ligger på proliferation, det vill säga delningstakten av cellerna som typiskt är densamma och inte beror av cellantalet. Totalt tas 970 bilder under 48 timmar, med en bild var 28–30:e minut. Eftersom endast en bild i taget kan fotograferas medför det en fördröjning på sex till nio sekunder mellan de tio positionerna i samma bildsekvens. Denna tidsskillnad anses försumbar då cellerna inte har hunnit förändra sig speciellt mycket under ett så kort tidsintervall.

### 3.2 Experiment 2

I experiment två används en platta med 96 brunnar, därav sex brunnar per dos men endast två positioner per brunn. Detta innebär att endast 30 av 96 brunnar utnyttjas och vi får 12 positioner per dos. I detta fall behöver vi ta hänsyn till brunnseffekten i vår analys. På grund av dålig bildkvalitet elimineras bilder från en position i kontrollgruppen samt i två

högre doser, giftmängd 10  $\mu\text{M}$  samt 100  $\mu\text{M}$ . Dessa tre doser får alltså 1067 bilder medan de två andra doserna får 1164 bilder för alla 12 positionerna.

### 3.3 Experiment 3

Det tredje experimentet är ett upprepade försök av experiment ett. Cellodlingen sker i en sex-brunn platta, alltså en dos per brunn och 10 positioner per brunn. I detta experiment försvinner en mätserie av en position för doserna med giftmängd 1,0  $\mu\text{M}$  respektive 100  $\mu\text{M}$  på grund av dålig bildkvalitet. Vi får alltså bilder för endast nio positioner, vilket resulterar i 873 bilder för dessa två doser.

### 3.4 Positionernas betydelse

Enligt tidigare beskrivningen betyder position den plats där cellerna ska mätas och följas upp. Idealiskt sett bör positionerna i brunnarna väljas slumpmässigt genom likformig systematiskt randomiserad sampling för att korrekt representera innehållet i hela brunnen. I praktiken medför detta svårigheter vid analys av experimentet, eftersom slumpmässigt utvalda positioner inte alltid innehåller celler. Därför tenderar man att manuellt välja positioner där celler är synliga. Detta innebär att mätningarna av tillväxthastigheter är beroende av att det faktiskt förekommer tillväxt. Använder vi dessa bilder för att uppskatta det totala antalet celler i brunnen kan det leda till ett systematiskt fel, en överskattning, eftersom tomma positioner exkluderas från beräkningen. Om man istället fokuserar på proliferation, alltså celledelningen och hur ofta den sker och inte antalet celler, antar vi att det inte ställer till några problem. Vi utgår från att cellernas tillväxthastighet är konstant för en given celltyp, vilket innebär att tiden för en cell att fördubblas bör vara enhetlig. Om brunn, position samt dos inte har någon effekt på proliferationen bör man kunna se det i cellernas tillväxthastighet. I denna uppsats vill vi undersöka om detta antagande stämmer ur statistisk synpunkt.

### 3.4 Komplexitet i experimenten

Efter att ha granskat experimenten och diskuterat med företaget konstaterades det att experimentens komplexitet är ett faktum. När cytometern har gjort sina mätningar görs en manuell uppskattning av resultatet. Det finns brister i denna kontroll eftersom det inte med säkerhet går att påvisa vad som är en cell och inte, det är även oklart om mätning

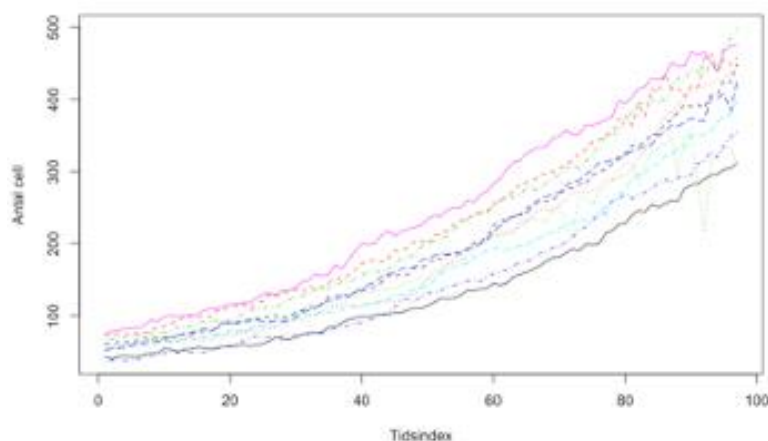
görs i början eller i slutet av en celldelning. På bilderna som cytometern tar är det omöjligt att veta om bilden visar en tjock cell eller flera celler på varandra, dvs massan av cellen är inte säkerställd. Celler är även svåra att föra in i brunnarna eftersom de är genomskinliga och lätta att förstöra. De kan även förflytta sig och därför röra sig mellan olika positioner, vilket påverkar indatan. Vi har också förstått att det finns svårigheter med att förstå om cellerna har delat sig eller endast blivit större, det är även svårare att räkna antalet celler ifall dessa har krympt. Trots att cellerna är i samma platta får de inte likvärdiga förhållanden såsom värme och syre bland annat, vilket innebär att valet av brunn påverkar villkoren för tillväxt, som i sin tur har inverkan på cellernas tillväxthastighet. Komplexiteten inom experimenten leder till vissa nödvändiga inskränkningar och förenklingar för att kunna modellera celltillväxten och analysera experimenten. Populationseffekt, täthetseffekt och bruseffekt exkluderas därför i analysen.

## 4 Statistisk modellering av celltillväxten

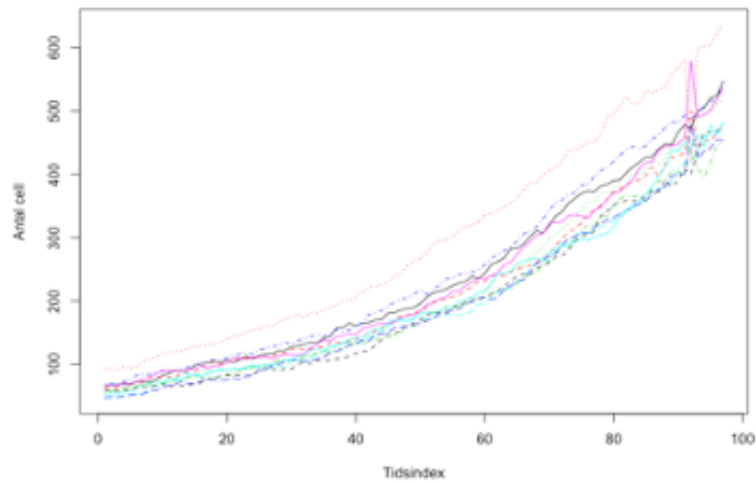
Kapitel fyra fokuserar på analys av cellernas tillväxthastighet. I de två huvudavsnitten presenteras olika tillvägagångssätt och statistiska modeller. I avsnitt 4.1 utförs en detaljerad analys av kontrollgrupperna i de tre experimenten. För kontrollgrupperna tillämpas både enkla och multipla linjära regressionsmodeller för att skatta tillväxthastigheten hos cellerna. Resultaten från kontrollgrupperna ger en grund för jämförelser med de andra behandlingarna. I avsnitt 4.2 undersöks celltillväxten vid olika doser. Två strategier implementeras: först studeras varje dos separat, och därefter byggs en gemensam regressionsmodell för alla doser. Slutligen används konfidensintervall för att avgöra om det finns någon statistiskt signifikant doseffekt.

### 4.1 Kontrollgrupp

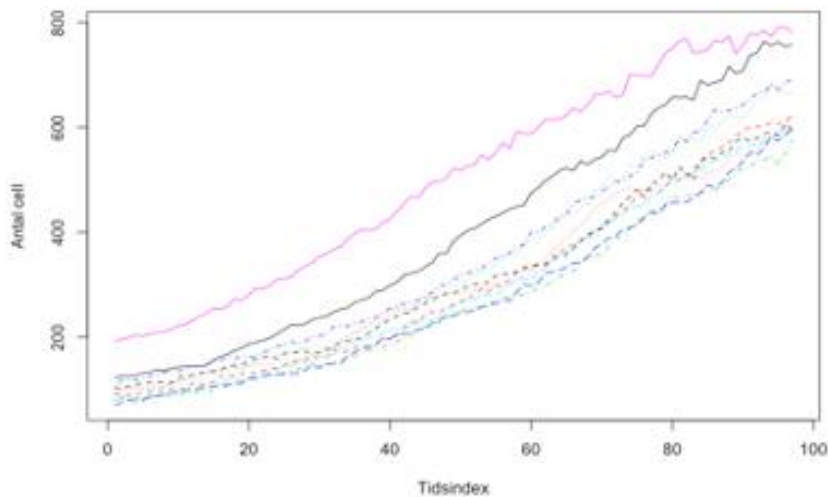
Som tidigare nämnts ligger uppsatsens fokus på att skatta cellernas tillväxthastighet och hur denna påverkas av behandlingar. Första steget är att skatta cellernas tillväxthastighet hos kontrollgruppen för att kunna förstå hur cellerna beter sig under normala tillväxtbetingelser. Kontrollgruppen kommer att fungera som en referens när vi analyserar effekten av doserna med olika mängd av tillväxthämmande preparat. I figur 6, 7 och 8 nedan visas hur antalet celler växer över tid för kontrollgrupperna i de tre experimenten. Celltillväxten för respektive position presenteras i form av en linje där y-axeln anger antal celler och x-axeln visar tidsindexet.



*Figur 6: Celltillväxten i de 10 positionerna för kontrollgruppen i experiment 1*



*Figur 7: Celltillväxten i de 11 positionerna för kontrollgruppen i experiment 2*



*Figur 8: Celltillväxten i de 10 positionerna för kontrollgruppen i experiment 3*

Tillväxtdiagrammen visar tydligt att cellerna befinner sig i sin log-fas, deras tillväxtkurvor visar en exponentiell tillväxt (Kielberg, Brunner, & Briand, 1994). Det framgår av diagrammen att experiment tre verkar ha kommit längre in i log-fasen, vilket reflekteras i de brantare kurvorna jämfört med de relativt raka linjerna i experiment ett och två. Figur 7 illustrerar en avvikelse i en av tillväxtkurvorna där cellantalet inte följer den förväntade kurvan vid tidsindex 92, vilket möjligtvis kan förklaras av mätproblem vid denna tidpunkt.

### 4.1.1 Enkel linjär regression

För att analysera tillväxthastigheten tillämpas en enkel linjär regressionsmodell med löpande tider i sekund som förklarande variabel. Eftersom antalet celler ökar exponentiellt med tider (Kielberg, Brunner, & Briand, 1994) kan följande modell användas för att beskriva tillväxten:

$$Y_t = k_0 \cdot e^{rt} \cdot \varepsilon, \text{ där } \begin{cases} Y_t: \text{antalet celler} \\ k_0: \text{antalet initialceller} \\ r: \text{cellens tillväxthastighet} \\ t: \text{tiden} \\ \varepsilon_t: \text{felterm} \end{cases} \quad (3)$$

Då varje position i en brunn innehåller olika antal initiala celler kommer skillnaden i antal celler i början och i slutet av mätningen jämfört mellan positionerna att kunna bli avsevärd. Med hjälp av modell (3) kan vi konstatera att antalet initialceller, alltså antalet celler i början av mätningen, har en betydande effekt på antalet celler i slutet, vilket kan leda till misstolkningar av tillväxthastigheten. Genom logaritmering skapas en linjär regressionsmodell:

$$\ln(Y_t) = \ln(k_0) + rt + \ln(\varepsilon_t) \quad (4)$$

Tillvägagångssättet vid experimenten är att positionerna väljs ut manuellt. Detta leder till ett intressant övervägande om det är bättre att analysera utvalda positioner eller använda all tillgänglig data för att uppnå större precision. För att kunna utreda detta använder vi oss av fyra metoder för att skatta fram cellernas tillväxthastighet och motsvarande medelfel.

Den första metoden är *min-punkt-metoden*, där den minsta cellmängden per mättpunkt identifieras och en linjär regressionslinje anpassas. Min-punkterna kan komma från en eller flera positioner. Vi använder oss av medelvärdet av de tidpunkter för det aktuella tidsindexet som vår oberoende variabel vid regressionsmodelleringen av tillväxten. Den andra metoden som används heter *max-punkt-metoden*. Tillvägagångssättet är detsamma som för min-punkt-metoden, men här används cellantalet från positionerna som har störst antal celler. Tredje metoden som tillämpas är *medelvärdesmetoden* där medelvärdet av antalet celler i positionerna från varje mätpunkt beräknas. Vi har nu fått 97 medelvärden av antalet celler och medeltider i sekunder för varje mätning som kommer

att ligga till grund för regressionsmodelleringen. I den fjärde metoden görs ett antagande om att cellerna inte förflyttar sig mellan positionerna och skattningen av tillväxthastigheten,  $r$ , är normalfördelad. Detta innebär att varje mätning av tillväxthastigheten i en position betraktas som ett oberoende stickprov. Tillväxthastigheten och dess standardavvikelse för varje position skattas fram och sedan beräknas dess medelvärde. Resultaten från dessa analyser, inklusive skattade tillväxthastigheter och deras medelfel, presenteras i tabeller nedan för varje experiment och metod. Dessa resultat ger insikter i cellernas beteende under olika experimentella förhållanden och stöder vidare analys av hur behandlingarna påverkar celltillväxten.

Tabell 1: Resultat från enkla regressionsmodeller via fyra metoder i Experiment 1

Metod		Koefficient	Medelfel
1. Min-punkt	Intercept	3,614	$8,56e^{-3}$
	$\hat{r}$	$1,26e^{-5}$	$8,561e^{-8}$
2. Max-punkt	Intercept	4,392	$1,28e^{-2}$
	$\hat{r}$	$1,126e^{-5}$	$1,279 \cdot e^{-7}$
3. Medelvärden av cellantal	Intercept	4,042	$6,564e^{-3}$
	$\hat{r}$	$1,196e^{-5}$	$6,562e^{-8}$
4. Medelvärden av $\hat{r}$	Intercept $\overline{\beta_0}$	4,0159	$1,11e^{-2}$
	$\bar{r}$	$1,203e^{-5}$	$1,123e^{-7}$

Tabell 2: Resultat från enkla regressionsmodeller via fyra metoder i Experiment 2

Metod		Koefficient	Medelfel
1. Min-punkt	Intercept	3,891	$8,011e^{-3}$
	$\hat{r}$	$1,315e^{-5}$	$8,008e^{-8}$
2. Max-punkt	Intercept	4,551	$7,33e^{-3}$
	$\hat{r}$	$1,145e^{-5}$	$7,336e^{-8}$
3. Medelvärden av cellantal	Intercept	4,146	$4,91e^{-3}$
	$\hat{r}$	$1,232e^{-5}$	$4,91e^{-8}$
4. Medelvärden av $\hat{r}$	Intercept $\overline{\beta_0}$	4,131	$9,306e^{-3}$
	$\bar{r}$	$1,238e^{-5}$	$9,903e^{-8}$



Tabell 3: Resultat från enkla regressionsmodeller via fyra metoder i Experiment 3

Metod		Koefficient	Medelfel
1. Min-punkt	Intercept	4,34	$9,95e^{-3}$
	$\hat{r}$	$1,229e^{-5}$	$9,95e^{-8}$
2. Max-punkt	Intercept	5,368	$1,835e^{-2}$
	$\hat{r}$	$8,726e^{-6}$	$1,835e^{-7}$
3. Medelvärden av cellantal	Intercept	4,688	$1,07e^{-2}$
	$\hat{r}$	$1,123e^{-5}$	$1,065e^{-7}$
4. Medelvärden av $\hat{r}$	Intercept $\overline{\beta_0}$	4,645	$1,227e^{-2}$
	$\hat{r}$	$1,143e^{-5}$	$1,227e^{-7}$

Förklaringsgraden  $R^2$ , vilket är ett mått på hur stor del av variationen i de logaritmerade cellantal som kan förklaras av tidsvariationen, ligger stabilt mellan 95%-99% för alla modeller. Detta höga värde visar att tid är en stark prediktor för celltillväxt under experimentella förhållanden. P-värden för alla regressioner och även för intercept och riktningskoefficienter är lägre än  $2,2 \cdot 10^{-16}$ , vilket tydligt indikerar att alla dessa är statistiskt signifikanta.

Residualanalys som presenteras i bilagan i figur 9, 10 och 11 för max-punkt, min-punkt och medelvärdesmetoden indikerar att antagandet om normalfördelning uppfyller för de flesta av modellerna. QQ-plottarna som genererats för varje experiment visar att residualerna, eller feltermerna, är normalfördelade vilket är en förutsättning för många statistiska tester (Sheather, 2009). Även om det förekommer vissa avvikelser i början och slutet av dataserien, visar residualernas spridning inte någon uppenbar skevhet eller kurvatur, vilket bekräftar modellernas lämplighet. Diagrammen som jämför residualer med anpassade värden (residuals vs. fitted plots) visar ett svagt mönster för min- och max-punkt-metoden, vilket kan antyda en mild icke-linjäritet. För medelvärdesmetoden verkar residualerna vara mer slumpmässigt spridda. Dessutom visar scale-location diagrammen, som analyserar spridningen av residualerna, på god konstans i variansen, vilket är viktig indikation för homoskedasticitet. För maxpunktsmodellen i experiment två observerades dock en ökning av residualernas varians med de anpassade värdena, ett fenomen kallat heteroskedasticitet, vilket kan behöva ytterligare undersökning för att säkerställa robustheten i modellens antaganden (White, 1980). Leverage-plottar har också

använts för att identifiera outliers som potentiellt kan påverka regressionen negativt. Det finns några punkter med högre Cook's avstånd för min-punkt och max-punkt-modellen i experiment ett, vilket indikerar potentiellt inflytserika datapunkter. För de andra modellerna finns det inga extremt höga värden på Cook's avstånd (Sheather, 2009).

Sammanfattningsvis är medelvärdesmetoden den mest robusta för att uppskatta celltillväxt i dessa experiment. Denna metod visar genomgående mindre problem med residualernas fördelning, normalitet och homoskedasticitet. Avvikelserna som observerats i residualplottarna kan bero på flera faktorer.

En logaritmisk transformation av cellantalet,  $\log(Y)$ , är särskilt lämplig för att fånga den exponentiella tillväxtfasen (log-fasen) hos en cellkultur. Denna fas kännetecknas av en konstant och maximal tillväxthastighet. När cellerna befinner sig i lag-fasen (initial anpassningsperiod) och stationär fas (platå på grund av resursbegränsningar), som har olika tillväxtdynamik kan inte en enkel linjär regressionsmodellen fånga upp exakt vad som sker under celltillväxten. Samtidigt kan cellerna fortfarande dela sig vid ungefär samma tidpunkter, vilket leder till periodiska mönster i residualplottarna. Vi väljer därför att fortsätta med enkla linjära regressionsmodeller i de fortsatta studierna. Modellen är enkel att använda och ger en rimlig approximation av tillväxtprocessen under log-fasen.

Det framgår att de skattade tillväxthastigheterna alltid är störst i min-punkt-metoden och lägst i max-punkt-metoden. Detta beror troligen på att högre celltäthet som man har i max-punkt-metoden skapar större konkurrens om resurser som näring och utrymme, vilket hämmar celltillväxten. Celler i min-punkten har däremot tillgång till fler resurser per ytenhet, vilket gör att de kan växa snabbare (Freshney, 1994). Skattade värdena ur regressionsmodellerna enligt de fyra metoderna ligger väldigt nära varandra. De varierar mellan  $1,20 \cdot 10^{-5}$  och  $1,23 \cdot 10^{-5}$ . Cellernas fördubblingstid varierar därmed från 15,63 till 16,1 timmar, vilket innebär att två till tre celcykler kan observeras under en 48-timmarsperiod. Spridningen är dock lägst i medelvärdesmetoden i experiment ett och två, medan min-punkt-metoden ger lägst medelfel i experiment tre. Baserad på residualanalysen kan vi konstatera att medelvärdesmetoden verkar vara den bästa metoden att skatta cellens tillväxthastighet, eftersom den verkar uppfylla modellens antaganden bäst.

Vid granskning av tillväxtdiagrammen för kontrollgrupperna framgår det att tillväxtkurvorna i experiment tre i figur 8 är brantare än i experiment ett och två i figur 6 och 7. Detta indikerar att cellerna i experiment tre har kommit längre in i den exponentiella tillväxtfasen och sannolikt hunnit genomgå flera celldelningar innan experimentet inleddes. Som ett resultat närmar sig cellerna i experiment tre slutet av denna fas, vilket är synligt genom att celltillväxten planar ut och stabiliserar sig när de närmar sig platåfasen där cellerna minskar i antal eller upphör att dela sig, vilket ofta ses i slutet av cellcykelns livscykel (Erlanson-Albertsson & Gullberg, 2007).

#### 4.1.2 Multipel linjär regression

Det tidigare nämnda sambandet mellan initialt och slutligt cellantal bekräftas av resultaten från både min- och max-metoden. Positioner med färre initialceller uppvisar högre tillväxthastigheter än de med fler celler från start. Detta väcker frågan om position och brunn påverkar resultaten, särskilt på grund av olika storlekar och fördelningar av brunnar mellan experimenten. För kontrollgruppen är brunnseffekten endast identifierbar i experiment två eftersom där tillförs celler av samma dos i sex brunnar medan experiment ett och tre endast i en brunn.

Jämförelser mellan kontrollgrupperna i de olika experimenten kan även avslöja en experimenteffekt, speciellt för experiment ett och tre eftersom cellerna odlas i samma 6-brunnsplatta men under olika tillfällen.

För att undersöka position- och brunnseffekter, används en multipel regressionsmodell:

$$\ln(Y_t) = \beta_0 + \beta_1 X_1 + \beta_{2,i} X_{2,i} + \beta_{3,i} X_{3,i} + \ln(\varepsilon_t) \quad (4)$$

$$\text{där } \left\{ \begin{array}{l} \ln(Y_t): \text{logaritm av antalet celler} \\ \beta_0: \text{intercept} \\ \beta_1: \text{tillväxthastighet} \\ X_1: \text{tiden i sekund} \\ \beta_{2,i}: \text{positionseffekt} \\ X_{2,i}: \text{position} \\ \beta_{3,i}: \text{brunnseffekt} \\ X_{3,i}: \text{brunn} \\ \ln(\varepsilon_t): \text{logaritm av felterm} \end{array} \right.$$

Utifrån ekvation (4) har den multipla linjära regressionsmodellen tre oberoende variabler: *tiden i sekunder, position och brunn*. När en modell består av fler oberoende variabler är det viktigt att kontrollera om det finns multikolinjäritet, ett problem som uppstår när oberoende variabler är starkt korrelerade enligt Cohen et al. (2013). Detta medför att skattningen av regressionskoefficienten för den korrelerade oberoende variabeln kommer att vara otillförlitlig därför att den innehåller för lite unik information. Regressionskoefficienters skattning kommer att ha mycket stort medelfel och blir ofta svår att tolka. Chi-två test används för att undersöka om det finns något beroende mellan brunn och position i experiment två eftersom dessa två är kategoriska variabler med nominalskala (Cortinhas & Black, 2012). Resultatet visar ett observerat värde av Chi-två på 5 335 med 50 frihetsgrader, motsvarande ett p-värde på mindre än som är signifikant. Detta indikerar att brunn och position är korrelerade, vilket leder till att brunnsvariabeln utesluts från den multipla regressionsmodellen. Vi vill ta med position i modellen för att positioner finns i tre experimenten medan brunn är endast aktuell i experiment två.

Den uppdaterade multipla regressionen består av följande:

$$\ln(Y_t) = \beta_0 + \beta_1 X_1 + \beta_{2,1} X_2 + \beta_{2,2} X_2 + \dots + \beta_{2,i} X_2 + \ln(\varepsilon_t) \quad (4)$$

$$\text{där } \left\{ \begin{array}{l} \ln(Y_t): \text{logaritm av antalet celler} \\ \beta_0: \text{intercept} \\ \beta_1: \text{tillväxthastighet} \\ X_1: \text{tiden i sekund} \\ \beta_{2,i}: \text{positionseffekt} \\ X_2: \text{position} \\ \ln(\varepsilon_t): \text{logariment av felterm} \end{array} \right.$$

Position är en kvalitativ variabel på nominalskala, vilket innebär att en cell endast kan finnas i en position och alla celler kategoriseras i någon av positionerna. Position ett används som referensgrupp för alla tre experiment. I experiment ett och tre får vi nio kodvariabler för de tio positionerna medan experiment två har tio kodvariabler för de elva positionerna. Resultaten för de tre experimenten sammanfattas i bilagan som tabell 4,5 och 6.

Enligt p-värdena i dessa tre tabeller är tiden och positionerna signifikanta variabler och bör inkluderas i modellerna. Förklaringsgraden för experiment ett är 98,99%, experiment två 99,33% och experiment tre 98,31%. När antalet oberoende variabler justeras ökar den

justerade förklaringsgraden till 98,98%, 99,32% respektive 98,29%. Därför visar de multipla regressionsmodellerna en förbättring i att förklara variationen jämfört med de enkla regressionerna i avsnitt 4.1.1. P-värden för hela regressionen är under  $2,2 \cdot 10^{-16}$ , vilket betyder att hela regressionen är signifikant. De multipla linjära regressionsmodellerna ger följande skattade tillväxthastigheter:  $1,2 \cdot 10^{-5}$ ;  $1,24 \cdot 10^{-5}$  respektive  $1,14 \cdot 10^{-5}$ . Dessa är likvärdiga med de bästa enkla regressionerna som är  $1,196 \cdot 10^{-5}$ ;  $1,232 \cdot 10^{-5}$  respektive  $1,229 \cdot 10^{-5}$ , men med något lägre medelfel:  $4,13 \cdot 10^{-8}$ ;  $3,2 \cdot 10^{-8}$  respektive  $5,13 \cdot 10^{-8}$  jämför med  $6,562 \cdot 10^{-8}$ ;  $4,91 \cdot 10^{-8}$  respektive  $9,95 \cdot 10^{-8}$ .

Residualplottar för de multipla regressionsmodellerna finns i figur 12, 13 och 14 för respektive experiment. Både residual vs Fitted och Scale-location figureerna visar en spridning av residualerna med några outliers. Residualerna är inte heller perfekt normalfördelade, vilket framgår av avvikelserna från den röda linjen i Q-Q-plottarna. Leverage-plottarna visar dessutom några observationer med hög leverage.

Liksom för den enkla linjära regressionsmodellen verkar de multipla linjära regressionsmodellerna för kontrollgrupperna inte uppfylla alla antaganden fullt ut. Med hänsyn till modellens prediktiva förmåga och komplexitet hos celltillväxten, som beskrivs i residualanalysen för den enkla regressionsmodellen kommer vi att använda den multipla linjära regressionsmodellen med *tid* och *position* som förklarande variabler i den fortsatta analysen. Detta är på grund av denna modell ger oss bättre skattningar av tillväxthastigheterna än de enkla linjära regressionsmodellerna.

## 4.2 Analys av doser

En viktig del av uppsatsens syfte är att undersöka möjligheten till en kvantifiering av dosernas effekt på cellernas tillväxthastighet. För att kunna besvara denna fråga har vi använt oss av två olika metoder.

I den första metoden analyseras varje enskild dos för sig, en regressionsmodell anpassas separat för varje dos. I den andra metoden skapas en gemensam regressionsmodell som inkluderar alla doserna vid ett experiment. Modellerna beskrivs mer i detalj i avsnitt 4.2.1

och avsnitt 4.2.2. I appendixet återfinns tillväxtdiagram i figur 15, 17, 19 och 21 för de fyra doserna i de tre experimenten.

Baserat på resultaten av regressionsanalysen för kontrollgrupperna kommer den beroende variabeln utgöras av logaritmen av cellantalet medan *tiden* (i sekunder) och *position* blir förklarande variabler. Den multipla regressionsmodellen för doserna kommer att ha samma struktur som den multipla regressionsmodellen för kontrollgruppen.

#### 4.2.1 Separata regressionsmodeller för var sin dos

De fyra doserna består av samma preparat med olika koncentrationer av gift: 0,5  $\mu\text{M}$ ; 1,0  $\mu\text{M}$ ; 10  $\mu\text{M}$  respektive 100  $\mu\text{M}$ , därmed kommer vi att få fyra regressionsmodeller för varje experiment. I experiment ett och tre används en brunn per dos, där varje brunn har tio positioner, vilket ger nio kategorier för variabeln position för varje dos. På grund av dålig bildkvalitet saknas en position för dos två och tre i experiment tre, vilket innebär att det finns åtta kategorier för dessa doser. I experiment två tillsätts varje giftidos till sex brunnar, sex replikat per dos, och cellantalet räknas i två positioner per brunn. Totalt 12 positioner för varje dos och därmed elva kategorier för variabeln position. Då det saknas bilder i en position för dos tre och fyra kommer variabeln position endast ha 10 kategorier för dessa doser. Resultaten för alla doser och experiment återfinns i tabell 7-10 i bilagan.

I tabell 7 finns resultaten av de skattade multipla regressionsmodellerna för dos ett (0,5  $\mu\text{M}$ ) i de tre experimenten. De skattade tillväxthastigheterna i experiment ett, två och tre blir  $1,005 \cdot 10^{-5}$ ;  $1,013 \cdot 10^{-5}$  respektive  $1,046 \cdot 10^{-5}$ , vilket är något lägre än de skattade tillväxthastigheterna för kontrollgrupperna som är  $1,196 \cdot 10^{-5}$ ;  $1,232 \cdot 10^{-5}$  respektive  $1,229 \cdot 10^{-5}$ . Medelfel blir  $4,439 \cdot 10^{-8}$ ;  $3,796 \cdot 10^{-8}$  respektive  $4,459 \cdot 10^{-8}$ . Den lägre tillväxten för dos ett kan även utläsas från figur 15, då linjerna i denna figur är något mindre kurviga än dem i figur 6, 7 och 8 för kontrollgrupperna. Utifrån tabell 7 kan vi konstatera att effekten av tid samt för position i allmänhet är signifikanta på väldigt hög nivå, ( $p < 2,2 \cdot 10^{-16}$ ). Positionerna fem och åtta är inte signifikanta för experiment ett detsamma gäller position 11 i experiment två, dessa kan därför plockas bort från respektive modell. Förklaringsgraden  $R^2$  för modellen i experiment ett blev 98,4%, för experiment två 98,52% samt för experiment tre 98,74%. De justerade förklaringsgraderna  $R^2_{adj}$ , blev i sin tur 98,39%, 98,5% respektive 98,73%. Alla tre regressionerna har ett p-

värde som understiger  $2,2 \cdot 10^{-16}$ . Förklaringsgraderna och regressionernas signifikans anses vara tillräckligt höga för att kunna beskriva celltillväxten på ett adekvat sätt och med hög precision vid denna dos.

Resultatet av den multipla regressionsmodelleringen för dos två ( $1,0 \mu\text{M}$ ) finns i tabell 8. Alla parametrar, intercept och koefficienter är signifikanta och ska därmed ingå i regressionsmodellen. Förklaringsgraden  $R^2$  för experiment ett är 96,51%, för experiment två 98,15% samt för experiment tre 98,71%. Den justerade förklaringsgraden  $R_{adj}^2$  blir i sin tur 96,47%, 98,18% respektive 98,69%. Alla tre regressioner är signifikanta med ett p-värde som understiger  $2,2 \cdot 10^{-16}$ . De skattade tillväxthastigheterna blir  $7,477 \cdot 10^{-6}$ ;  $8,66 \cdot 10^{-6}$  respektive  $9,193 \cdot 10^{-6}$ , vilka är lägre jämfört med kontrollgrupperna. Linjerna i figur 17 är betydligt brantare än de för kontrollgrupperna i figur 6, 7 och 8, vilket indikerar en kraftigt reducerad tillväxttakt på grund av en starkare inhiberande effekt. Medelfel på dessa skattningar blir  $4,805 \cdot 10^{-8}$ ;  $3,67 \cdot 10^{-8}$  respektive  $4,139 \cdot 10^{-8}$ .

Tabell 9 redovisar resultaten för dos tre ( $10,0 \mu\text{M}$ ). Parameterskattningarna för tid och position är signifikanta för experiment ett och två med förklaringsgrader på 95,77% respektive 83,54%, och justerade förklaringsgrader på 95,73% respektive 83,37%. De skattade tillväxthastigheterna blir  $1,083 \cdot 10^{-7}$ ;  $5,827 \cdot 10^{-7}$  respektive  $1,272 \cdot 10^{-8}$ , med medelfel på  $2,484 \cdot 10^{-8}$ ;  $3,839 \cdot 10^{-8}$  respektive  $4,875 \cdot 10^{-8}$ . P-värdet för hela regressionerna är lägre än  $2,2 \cdot 10^{-16}$ , vilket innebär att regressionerna är signifikanta. Anmärkningsvärt för experiment tre är tiden som inte är signifikant, med ett högt p-värde på 0,79417. Likaså är riktningskoefficienternas värden vid denna dos avsevärt lägre än för de mindre doserna ett och två, åtminstone drygt en faktor 100 lägre i tillväxthastighet. Detta betyder att tiden inte har någon större effekt på tillväxten av cellantalet, det vill säga tillväxthastigheten och även doseffekten är inte mätbara i detta experiment. Spridningsdiagrammet för dos tre i experiment tre (se figur 19) visar att cellantalet förblir stabilt under hela mätningen. Med andra ord räknas förmodligen samma celler hela tiden och skillnaden i antal celler som uppstår verkar vara brus eller slumpfel vid mätningen eller identifieringen av celler. En förklaring till skillnaden mellan experiment tre och de två övriga är att cellerna redan har nått platåfasen och därmed inte längre reagerar på toxinet (Weinberg, 2013). Observera att för samtliga experiment är den skattade

tillväxthastigheten betydligt lägre än för kontrollgrupperna och för experimenten vid dos ett och två. Detta visas tydligt i figur 19 då celltillväxten för dos tre verkar endast öka något över tid, där linjerna är betydligt plattare jämfört med linjerna för kontrollgrupperna och de andra doserna.

Utifrån tabell 10 kan vi konstatera att alla positioner och tiden är signifikanta hos experiment ett och tre för dos fyra. Position 9 och 10 är dock inte signifikant skilda från noll hos experiment två och kommer därför att plockas bort från regressionsmodellen. Förklaringsgraden  $R^2$  är 99,21% för experiment ett, 82,18% för experiment två samt 97,92% för experiment tre. Den justerade förklaringsgraden  $R_{adj}^2$  är 99,2%, 81,99% respektive 97,9%. Alla tre regressioner är signifikanta. Däremot får den skattade tillväxthastigheten negativa värden i alla tre experimenten  $-2,358 \cdot 10^{-7}$ ;  $-7,738 \cdot 10^{-7}$  respektive  $-2,803 \cdot 10^{-7}$ , vilket betyder att cellantalet minskar över tid. Denna minskning i cellantalet visas tydligt i figur 21 då linjerna är mycket plattare och även lutar nedåt. En möjlig förklaring kan vara att den mängd gift som tillsätts påverkar cellernas delningstakt negativt, det vill säga cellerna dör snabbare eller så hinner de inte alls att dela sig.

Baserat på analysen av residualplottarna för doserna, som visas i figur 16, 18, 20 och 22, avviker dos två och dos tre mest från modellens antaganden. För dos två visar residualerna större spridning, fler outliers och avvikelser från normalfördelningen i alla tre experiment. Även dos tre har större spridning av residualerna och avvikelser från normalfördelningen, särskilt i experiment två och tre.

#### *4.2.2 Gemensam regressionsmodell för alla doser*

För att kunna jämföra resultaten i de tre experimenten på ett enkelt sätt rakt över, anpassas en gemensam multipel regressionsmodell som inkluderar både doser och kontrollgrupp för respektive experiment. Anledningen till att ta med både dos och kontrollgrupper i samma modell och inte använda parvisa jämförelser av tillväxtresultaten är för att undvika masssignifikansproblem, risken att förkasta en korrekt nollhypotes felaktigt vid hypotesprövningen, (Körner & Wahlgren, 2006).



Baserat på resultatet från de multipla regressionsmodellerna för kontrollgrupper i avsnitt 4.1.2 bör de multipla regressionsmodellerna för doserna bestå av *logaritmerat antal celler* som beroende variabel medan *tid* i sekunder, *position* och *dos* ska vara förklarande variabler. *Logaritmerat antal celler* och *tid* är kvantitativ data med kvotskala medan *position* och *dos* är nominella variabler. Experiment ett innehåller 4 850 observationer, 30 positioner och fem doser. Experiment två omfattar 5 529 observationer, 57 positioner och fem doser. Experiment tre består av 4 656 observationer, 48 positioner och fem doser. Eftersom vi har två förklarande variabler *position* och *dos* behöver vi undersöka om dessa två variabler korrelerar med varandra, alltså om multikolinjäritet finns i modellen. Vi har därför utfört ett Chi-två test för experiment ett, det observerade värdet av chi-två beräknas till 19 400 med 196 frihetsgrader, vilket motsvarar ett p-värde på mindre än  $2,2 \cdot 10^{-16}$ . Detta betyder att nollhypotes om oberoende mellan variablerna *position* och *dos* förkastas, eller med andra ord, *position* och *dos* är korrelerade. Vi tar därför bort variabel **position** och använder endast *tid* och *dos* som förklarande variabler i modellerna. Resultatet redovisas i tabell 11, 12 och 13 som återfinns i bilaga.

Utifrån resultat av regressionsanalysen (tabell 11, 12 och 13) kan vi påstå att alla förklarande variabler är signifikanta. Förklaringsgraden  $R^2$  hos experiment ett är 64,14%, hos experiment två 74,35% samt hos experiment tre 67,08%. Den justerade förklaringsgraden  $R_{adj}^2$  blir 64,1%, 74,33% respektive 67,04%. Parameterskattningarna i regressionsmodellerna har mycket låga p-värden, de är i allmänhet mindre än  $2,2 \cdot 10^{-16}$ , vilket betyder att regressionsmodellerna är höggradigt signifikanta. Det som är anmärkningsvärt för dessa tre regressioner är att medelfelet för *tid* är mycket större än de medelfel vi får från de separata multipla regressionerna för kontrollgrupperna. Vi kan utläsa ur tabell 11 att medelfelet för *tid* är  $1,02 \cdot 10^{-7}$ , medan medelfelet ur tabell 4 är  $4,13 \cdot 10^{-8}$ , vilket betyder att det första medelfelet är 246% större än det andra medelfelet. För experiment två är denna skillnad 251% och för experiment tre 259%.

Vi får därför en sämre skattning av tillväxthastigheter genom att slå ihop alla doser samt kontrollgruppen i en och samma modell. En möjlig förklaring till detta är att de stora skillnaderna i giftmängden som gör att celltillväxten har blivit påverkad olika mycket men även kanske på olika sätt. För de stora doserna, alltså dos tre och fyra, verkar det som att samma celler räknas hela tiden med några få avvikelser vilket resulterar i samma

antal celler i början som i slutet av odlingen. Det vi ser är antagligen inget annat än brus - slumpfelet vid mätningarna. Detta betyder att vi inte kan identifiera - kvantifiera effekten för dessa två stora doser. En annan förklaring kan vara att cellerna inte får samma tillväxtförhållande i form av värme och syre i de olika brunnarna, trots att de står i samma platta.

Residualplottar i figur 22, 23, 24 visar en tydligare avvikelser av dessa modeller jämfört med de regressionsmodeller för enskilda doser. För alla experiment visar residualerna en större spridning och flera outliers när alla doser analyseras tillsammans. Detta indikerar att det finns betydande variabilitet som inte fångas upp av modellen. Normal Q-Q-plottarna visar större avvikelser från den teoretiska normalfördelningen, vilket indikerar att residualerna är icke-normalfördelade. Den ökande variationen i Scale-location plotten tyder på att modellens antaganden om homogen varians (homoscedasticity) inte håller. I Residuals vs Leverage plotten finns det flera observationer med hög leverage och några med extrem Cook's distance, vilket indikerar att vissa datapunkter har en oproportionerligt stor inverkan på modellen.

Allt tyder på att vi inte kan slå ihop doserna i en och samma modell utan får analysera dem var för sig. Vi kommer därför att använda oss av de skattade tillväxthastigheterna som vi får från de multipla regressionerna för varje dos och använder dess konfidensintervall för att avgöra om det finns en doseffekt eller ej.

#### 4.2.3 Konfidensintervall för tillväxthastigheten - doseffekt

Varje dos analyseras separat och en multipel regressionsmodell används för att skatta tillväxthastigheten -- intensiteten. De skattade intensiteterna för alla doser i samma experiment jämförs därefter med varandra för att bestämma om de skiljer sig ifrån varandra eller ej. Ett sätt att utföra jämförelserna är att räkna ut konfidensintervall för de skattade tillväxthastigheterna och sedan undersöka om konfidensintervallen överlappar varandra. Om det inte finns överlappning mellan konfidensintervallen innebär det att de skattade parametrarna skiljer sig från varandra (Julious, 2004). Konfidensintervall för alla doser i alla tre experimenten visualiseras i figur 23, 24 och 25 i bilaga. Dessa är även uträknade och presenteras i tabell 14 nedan.

Tabell 14: Konfidensintervall för kontrollgrupper och alla doser

	Undregräns	Punktskattning	Övregräns	Medelfel
<b>Kontroll Exp 1</b>	1,196e-5	1,204e-5	1,212e-5	4,13-8
<b>Dos 1 Exp 1</b>	9,958e-6	1,005e-5	1,013e-5	4,439-8
<b>Dos 2 Exp 1</b>	7,383e-6	7,477e-6	7,571e-6	4,805-8
<b>Dos 3 Exp 1</b>	5,963e-8	1,083e-7	1,570e-7	2,484-8
<b>Dos 4 Exp 1</b>	-2,817e-7	-2,358e-7	-1,899e-7	2,343-8
<b>Kontroll Exp 2</b>	1,231e-5	1,238e-5	1,244e-5	3,203e-8
<b>Dos 1 Exp 2</b>	1,006e-5	1,013e-5	1,021e-5	3,796e-8
<b>Dos 2 Exp 2</b>	8,589e-6	8,66e-6	8,732e-6	3,67e-8
<b>Dos 3 Exp 2</b>	5,074e-7	5,827e-7	6,579e-7	3,839e-8
<b>Dos 4 Exp 2</b>	-8,406e-7	-7,738e-7	-7,071e-7	3,407e-8
<b>Kontroll Exp 3</b>	1,133e-5	1,143e-5	1,153e-5	5,134e-8
<b>Dos 1 Exp 3</b>	1,0368e-5	1,046e-5	1,0543e-5	4,459e-8
<b>Dos 2 Exp 3</b>	9,112e-6	9,193e-6	9,274e-6	4,139e-8
<b>Dos 3 Exp 3</b>	-8,282e-8	1,272e-8	1,083e-7	4,875e-8
<b>Dos 4 Exp 3</b>	-3,254e-7	-2,803e-7	-2,353e-7	2,297e-8

I dessa figurer har vi fem konfidensintervall för kontrollgruppen och de fyra doserna för alla tre experiment. Lägg märke till dos tre i figur 25 där dess konfidensintervall täcker nollpunkten, vilket bekräftar vårt resultat i avsnitt 4.2.1 om att tiden och även cellens tillväxthastighet inte är signifikant skild ifrån noll. Det som är gemensamt för alla tre experimenten är att ingen av konfidensintervallen överlappar varandra, vilket betyder att de skattade tillväxthastigheterna är signifikant skilda ifrån varandra. Vi kan då konstatera att det finns en doseffekt som påverkar cellernas tillväxthastighet för alla doser i alla experimenten.

Det är även av vikt att undersöka om det finns experimentseffekt. Detta gör vi genom att jämföra konfidensintervall för de skattade tillväxthastigheterna för kontrollgruppen från respektive experiment.

I figur 26 kan vi utläsa att den skattade tillväxthastigheten är lägst hos experiment tre och högst hos experiment två. De tre konfidensintervallen överlappar inte varandra, vilket

resulterar i att de skattade tillväxthastigheterna för kontrollgrupper kan anses signifikant skilda åt. Det finns en tydlig experimentseffekt som påverkar cellernas tillväxthastighet. Vi har i avsnitt 4.1 plottat fram ett spridningsdiagram för hur antalet celler hos kontrollgrupper i experiment ett, två och tre förändras över tiden och konstaterat att tillväxtkurvan hos experiment tre ser annorlunda ut än tillväxtkurvan hos de två resterande experimenten. Denna skillnad kan uppstå på grund av att cellerna i experiment tre har kommit längre i tillväxtfasen och därmed hunnit med fler delningar än i experiment ett och två. Cellerna närmar sig alltså platåfasen där de slutar att dela sig. En annan tänkbar förklaring är att cellerna inte har helt identiska experimentella förhållanden i form av till exempel värme, ph-värde, syre med mera, där små variationer i miljöförhållanden kan signifikant påverka cellens tillväxtsförmåga (Calabrese et al., 2007). Denna problematik har lyfts i avsnittet om experimentens komplexitet, där de utmaningar som uppstår på grund av variationer och begränsningar kan ha en inverkan på experimenten.

Att den skattade celltillväxthastigheten skiljer sig mellan experiment ett och två är även intressant. Detta kan tyda på att variationer mellan brunnarna kan påverka cellernas tillväxt. Som tidigare nämnts odlas cellerna i en 6-brunnspatta i experiment ett och en 96-brunnspatta i experiment två. I en 96-brunnspatta är både volym och ytan per brunn mindre jämfört med en 6-brunnspatta, vilket kan leda till snabbare konsumtion av näringsämnen och påverka cellernas tillväxttillstånd. Mindre brunnar kan ha högre evaporation, vilket kan leda till koncentrationsförändringar i mediet (Sirenko et al., 2015).

## 5 Diskussion

Syftet och den övergripande målsättningen med uppsatsen har varit att utveckla en kvantitativ statistisk modell som kan användas för att undersöka effekterna av olika behandlingar på celltillväxt. Genom att använda cytometer kan vi i detalj följa celledningen under hela cellodlingsperioden. Ett viktigt mål har varit att identifiera en nyckelvariabel som enkelt kan användas för att avgöra signifikansen av olika behandlingar och jämföra deras effekter mellan olika experiment. Den data som har använts har tillhandahållits i form av tre olika experiment. Frågeställningarna och modellansatsen har tagits fram i diskussion med företaget.

Ett av de första resultaten av den statistiska analysen av celltillväxten i de olika försöken visade att det inte är möjligt eller rekommenderat att bygga en kvantitativ analys av experimentresultaten genom att jämföra cellantalet i början respektive slutet av cellodlingsperioden. Eftersom celledningsprocessen är av exponentiell karaktär och det är praktiskt taget omöjligt att exakt kontrollera antalet celler som varje odlingsförsök startar med, går det inte att använda en statistiska som grundar sig på skillnaden i antalet celler. Icke signifikanta skillnader i cellantal vid experimentets början kan resultera i stora och potentiellt missvisande skillnader vid slutet av experimenten, enbart på grund av de initiala variationerna. Därför är denna metod inte lämplig för relevant statistisk analys.

Slutsatsen blev istället att det är mer intressant att fokusera på tillväxthastigheten (den intensitet som cellerna delar sig under experimenten) som en möjlig statistisk nyckelvariabel. Vid olika behandlingar såsom gifter, tillväxthämmande eller tillväxtstimulerande preparat förväntas dessa påverka just förutsättningarna för en celledning, vilket gör celledningshastigheten mer pålitlig och informativ för analys.

En viktig observation har bland annat varit kontrollgruppens centrala roll i varje experiment. Kontrollgruppen fungerade som en återkommande referens, vilket är i linje med tidigare forskning som betonar vikten av robusta kontrollgrupper (James, Witten, Hastie, & Tibshirani, 2022). Trots konstant referens till kontrollgruppen stötte vi på utmaningar med brunnseffekt som visar på experimentella variationer. Denna effekt var

dock svår att kvantifiera på grund av varierande förutsättningar och odlingsmiljöer mellan experimenten. Detta pekar på behovet av förbättrade mätmetoder eller analysstrategier.

Resultaten av vår studie bekräftar att initialt antal celler samt deras placering i brunnarna betydligt påverkar resultaten. När celler manuellt fördelas i brunnarna, påverkar detta cellernas tillväxthastigheter på grund av positionseffekten. Trots detta problem kan repetition av experimenten under liknande förhållanden ge konsistenta resultat. Det är viktigt att mäta celltillväxten när cellerna befinner sig i samma tillväxtfas för att undvika felaktiga tolkningar av tillväxtkurvornas variation mellan experimenten.

Våra observationer tyder på att kombinationer av positions- och doseffekter samt brunnseffekter är betydande och som nämnt påvisade studien ett samband mellan initialt och slutligt antal celler, där positioner med färre initialceller tenderar att ha högre tillväxthastigheter. Denna observation väckte frågan om huruvida position och brunn kan påverka resultaten av experimenten. För att utforska detta valdes de oberoende variablerna *tid i sekunder*, *brunn* och *position* för den multipla regressionsmodellen. Det visade sig att position och brunn var korrelerade och brunn uteslöts för att förbättra modellens precision. Den gemensamma regressionsmodellen för dos och kontroll indikerade en korrelation mellan position och dos. Resultatet blev en mindre tillfredsställande skattning av tillväxthastigheterna.

Genom att genomföra flera experiment kan vi möjligen uppnå mer generaliserade resultat om celltillväxthastigheten, särskilt genom att inkludera mätdata från olika stadier i cellernas tillväxtfas. En potentiell lösning för att säkerställa en mer enhetlig datamängd är att exkludera de första och sista mätpunkterna där celltillväxten ännu inte stabiliserats.

## Möjligheter och vidare forskning

Under uppsatsens gång har vi stött på oklarheter och möjliga osäkerhetsfaktorer vid experimenteringen vilket påverkar resultattolkningens tillförlitlighet. Till exempel är det svårt att avgöra cellantal på grund av osäkerhet kring om bilderna visar en enda cell eller flera överlappande celler. Dessutom kan inkubatorns värmefördelning påverka resultaten beroende på brunnarnas position på plattan, vilket kan leda till varierande tillväxtförhållanden.

Vi står också konfunderade över syftet med tillsättningen av giftet. Om syftet är att kontrollera tillväxten visar resultaten att detta kan uppnås redan vid lägre doser. Vid högre doser verkar celledning och tillväxt helt upphöra. Det är också oklart om det finns en systematisk randomisering vid tillsättning av kontrollprover, vilket kan påverka resultatens tillförlitlighet.

För vidare forskning och förbättring av modellens anpassning är ett förslag att använda polynomisk regression, styckvis linjär regression eller generaliserade linjära modeller. Alternativt kan olika transformationer av den beroende variabeln (t ex. kvadratroten eller Box-Cox transformationer) stabilisera variansen och normalisera residualerna. Dessa metoder kan ge en förbättrad representation av hela tillväxtprocessen, vilket säkerställer att modellen optimalt tar hänsyn till den inneboende variabiliteten och de olika tillväxtfaserna.

Ytterligare förslag är att utforska effekten av lägre doser av gift för att bestämma minimivåer för kontroll av celltillväxthastighet, vilket kan leda till mer kostnadseffektiv och mindre invasiv forskning. Det är även önskvärt att undersöka omfattningen av experimenteffekten och utforska potentialen i alternativa statistiska transformationer för att stärka modellernas prediktiva kraft. Vi har identifierat en experimentseffekt som kan vara värd att studera, men på grund av begränsningar i datamaterialet har vi valt att inte fördjupa oss i detta. Frågan lämnas öppen inför framtida forskning och nya experiment.

## 6. Litteraturförteckning

- Calabrese, E., Bachmann, K., Bailer, A., Bolger, P., Borak, J., Cai, L., . . . Cherian, M. (den 1 July 2007). Biological stress response terminology: Integrating the concepts of adaptive response and preconditioning stress within a hormetic dose-response framework. *Toxicology and Applied Pharmacology*, ss. 122-128.
- Cohen, J., Cohen, P., Aiken, L., & West, S. (2013). *Applied multiple regression/correlation for the behavioral sciences*. Routledge.
- Cortinhas, C., & Black, K. (2012). *Statistics for Business and Economics*. John Wiley & Sons Ltd.
- Erlanson-Albertsson, C., & Urban, G. (2007). *Cellbiologi*. Studentlitteratur AB.
- Freshney, R. (1994). *Culture of Animal Cells: A Manual of Basic Technique*. Wiley-Liss.
- [https://biomedicinskanalytiker.org/2014/01/22/cellodling/Magdalena Wiklund 2014](https://biomedicinskanalytiker.org/2014/01/22/cellodling/Magdalena%20Wiklund%202014). (u.d.).
- <https://cancer.se/cellcykeln-nar-en-cell-blir-tva/>. (u.d.).
- <https://www.microscope.com/education-center/microscopes-101/history-of-microscopes>. (u.d.).
- <https://www.news-medical.net/life-sciences/Brief-History-of-Microscopy.aspx>. (u.d.).
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2022). *An Introduction to Statistical Learning*. Springer.
- Julious, S. A. (den 14 September 2004). Using confidence intervals around individual means to assess statistical significance between two means. *Pharmaceutical Statistics*, ss. 217-222.
- Körner, S., & Wahlgren, L. (2006). *Statistisk dataanalys*. Studentlitteratur AB.
- Kielberg, V., Brunner, N., & Briand, P. (1994). *Cellodling - En praktisk handbok i odling av mammalieceller*. Lund: AB Labassco förlag.
- <https://www.leica-microsystems.com/science-lab/life-science/how-to-do-a-proper-cell-culture-quick-check/> Straube T, Muller C, Perestrelo A.R. 2023
- Montgomery, D., Peck, E., & G. Geoffrey, V. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- <https://www.sciencefacts.net/cell-cycle.html> (u.d).
- Sheather, S. (2009). *A Modern Approach to Regression with R*. Springer.



Sirenko, O., Mitlo, T., Hesley, J., Luke, S., Owens, W., & Cromwell, E. (den 28 August 2015). High-Content Assays for Characterizing the Viability and Morphology of 3D Cancer Spheroid Cultures. *ASSAY and Drug Development Technologies*, ss. 402-414.

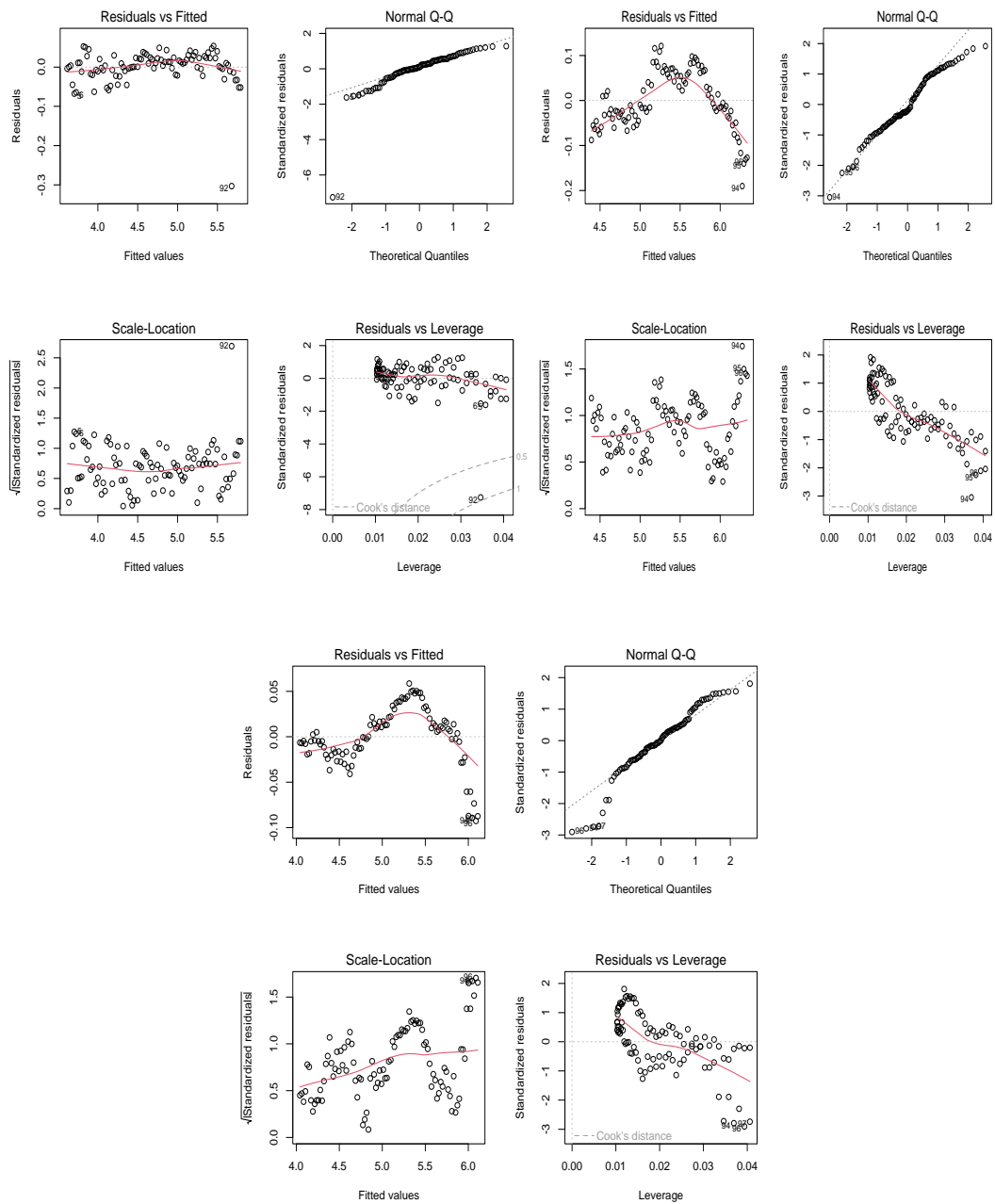
<https://www.turito.com/blog/biology/phases-of-mitosis> (2022)

Weinberg, R. (2013). *The Biology of Cancer*. Taylor & Francis Ltd.

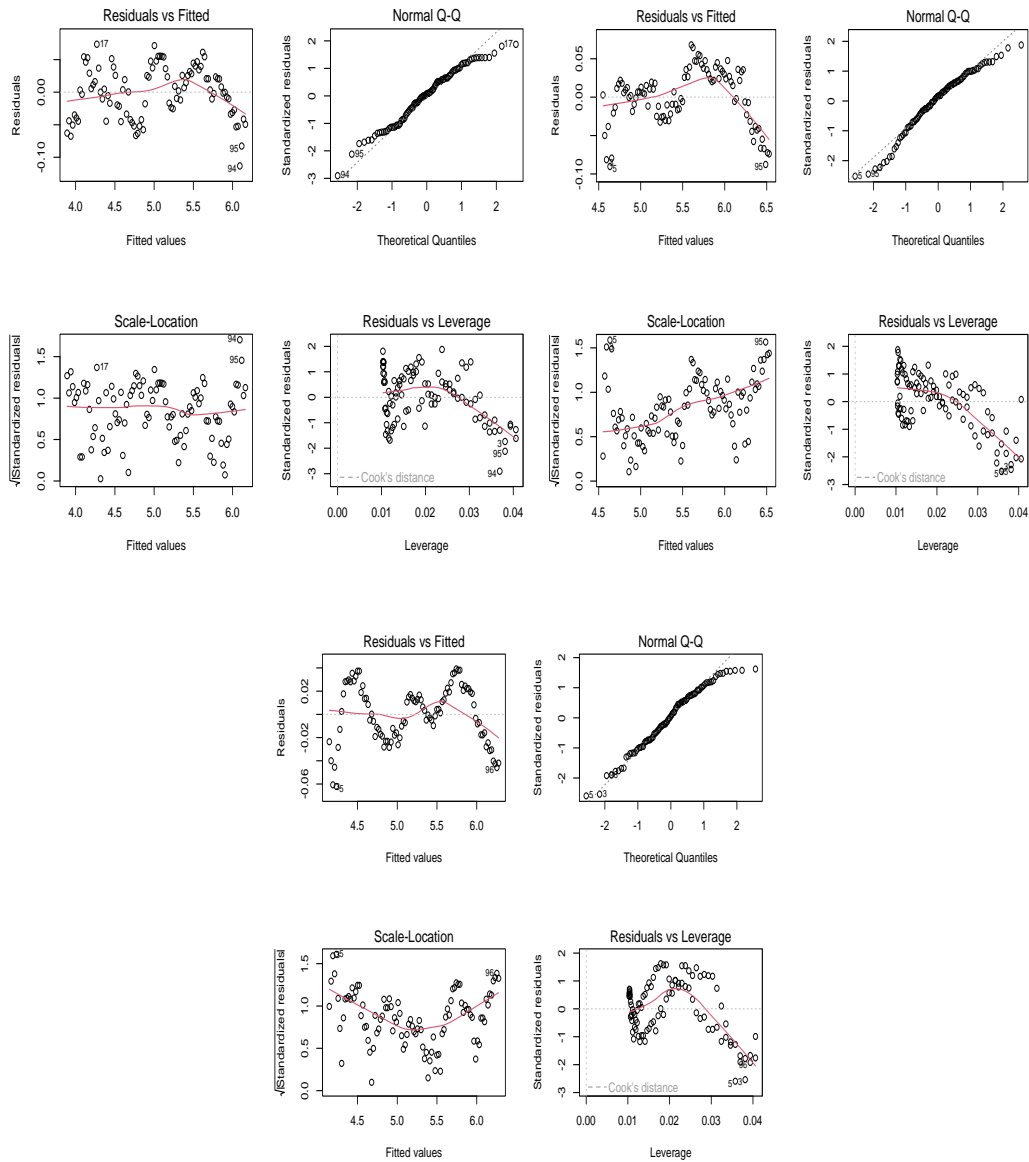
Wexler, P. (2014). *Encyclopedia of Toxicology*. Elsevier Science .

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *The Econometric Society*, ss. 817-838.

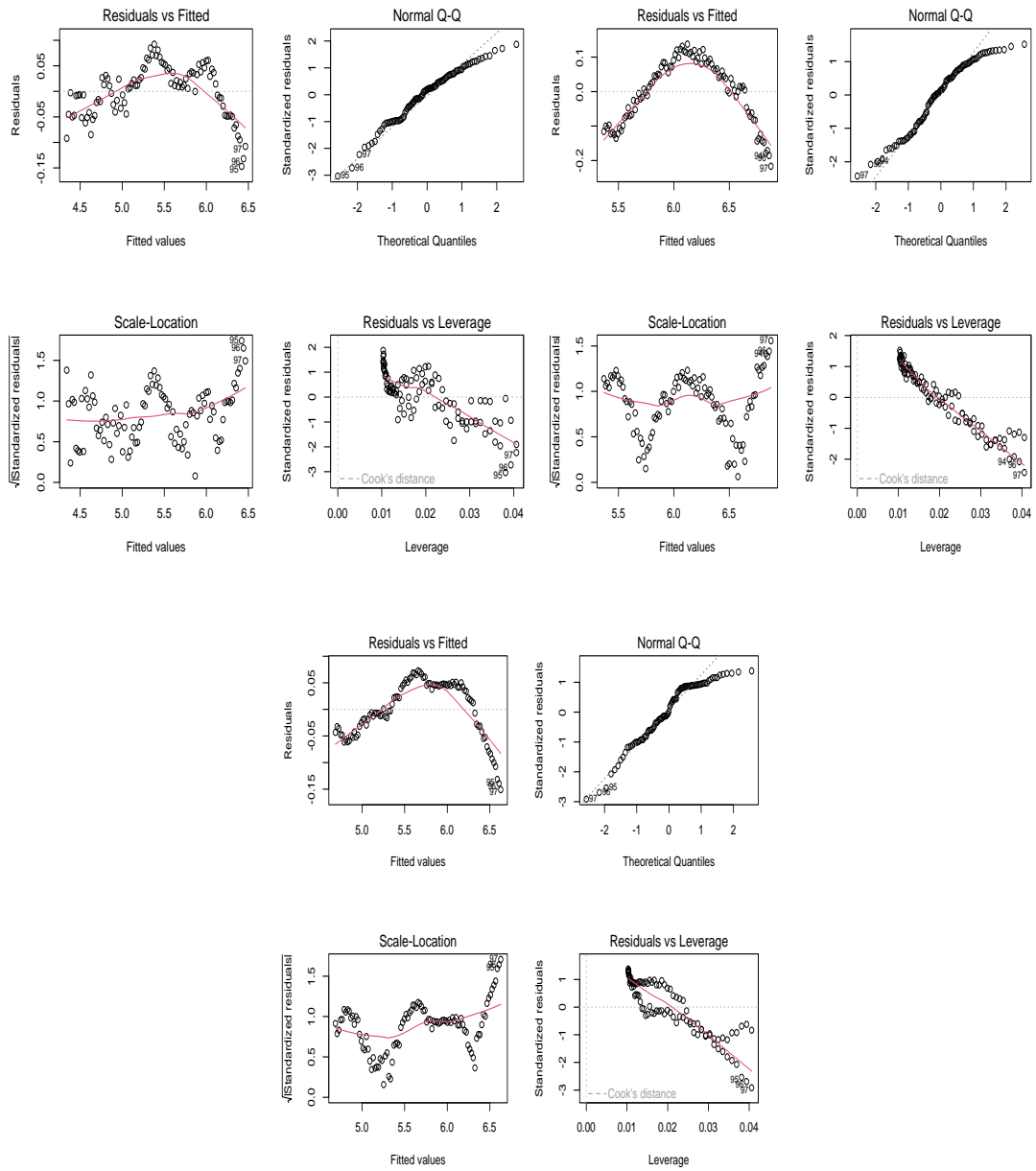
# 7 Bilagor



Figur 8: Residualanalys för Min-vänster, Max-punkt (höger) och Medelvärdesmetoden (mitten) experiment 1



*Figur 9: Residualanalys för Min-(vänster), Max-punkt (höger) och Medelvärdesmetoden (mitten) experimt 2*



Figur 10: Residualanalys för Min-(vänster), Max-punkt (höger) och Medelvärdesmetoden (mitten) experimt 3

Tabell 4: Resultat multipel regression över kontrollgrupp i experiment 1

	Koefficient	Medelfel	Pr(> t )
<b>Intercept</b>	3,687	7,49E-03	<2e-16 ***
<b>Tid</b>	1,20E-05	4,13E-08	<2e-16 ***
<b>Pos 2</b>	0,551	9,31E-03	<2e-16 ***
<b>Pos 3</b>	0,281	9,31E-04	<2e-16 ***
<b>Pos4</b>	0,0466	9,31E-05	6,79e-07 ***
<b>Pos 5</b>	0,235	9,31E-06	<2e-16 ***
<b>Pos 6</b>	0,635	9,31E-07	<2e-16 ***
<b>Pos 7</b>	0,379	9,31E-08	<2e-16 ***
<b>Pos 8</b>	0,289	9,31E-09	<2e-16 ***
<b>Pos 9</b>	0,501	9,31E-10	<2e-16 ***
<b>Pos 10</b>	0,363	9,31E-11	<2e-16 ***

Utifrån resultatet i tabell 4 blir regressionsmodellen för kontrollgruppen i experiment ett enligt följande:

- För position ett, där denna position är referensgrupp:

$$\log(\hat{Y}_t) = 3,687 + 1,2 \cdot 10^{-5} \cdot t + \ln(\hat{\varepsilon}_t)$$

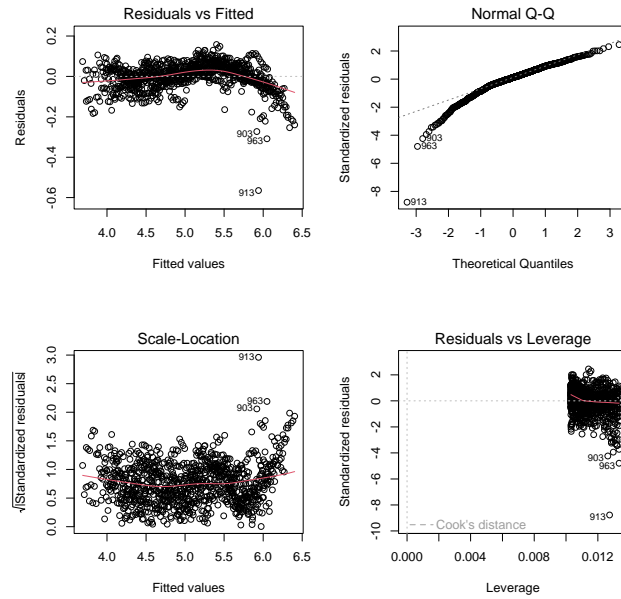
- För position två:

$$\log(\hat{Y}_t) = 3,687 + 1,2 \cdot 10^{-5} \cdot t + 0,551 + \ln(\hat{\varepsilon}_t) = 4,238 + 1,2 \cdot 10^{-5} \cdot t + \ln(\hat{\varepsilon}_t)$$

- För position tre:

$$\log(\hat{Y}_t) = 3,687 + 1,2 \cdot 10^{-5} \cdot t + 0,281 + \ln(\hat{\varepsilon}_t) = 3,968 + 1,2 \cdot 10^{-5} \cdot t + \ln(\hat{\varepsilon}_t)$$

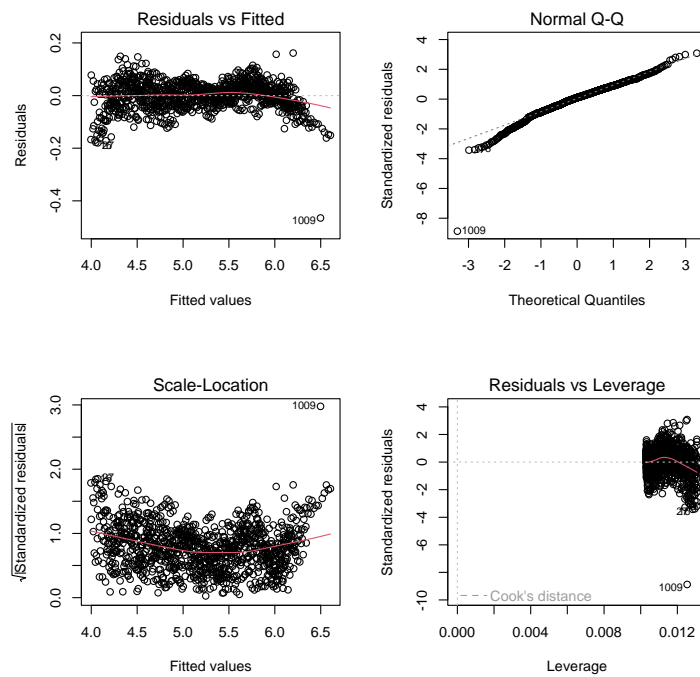
o.s.v



Figur 11: Residualanalys multipel linjär regression för kontrollgruppen experiment 1

Tabell 5: Resultat multipel regression över kontrollgrupp i experiment 2

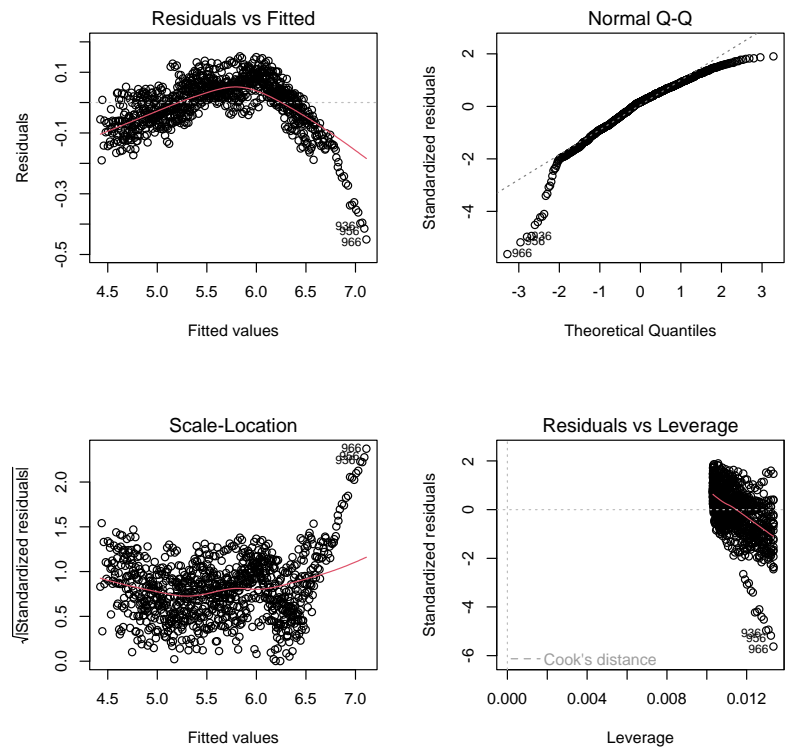
	Koefficient	Medelfel	Pr(> t )
<b>Intercept</b>	4,204	6,03E-03	<2e-16 ***
<b>Tid</b>	1,24E-05	3,20E-08	<2e-16 ***
<b>Pos 2</b>	-0,0756	7,57E-03	<2e-16 ***
<b>Pos 3</b>	-0,117	7,57E-04	<2e-16 ***
<b>Pos4</b>	0,0487	7,57E-05	1,87e-10 ***
<b>Pos 5</b>	-0,180	7,57E-06	<2e-16 ***
<b>Pos 6</b>	-0,0365	7,57E-07	1,68e-06 ***
<b>Pos 7</b>	-0,209	7,57E-08	1,68e-06 ***
<b>Pos 8</b>	0,259	7,57E-09	<2e-16 ***
<b>Pos 9</b>	-0,178	7,57E-10	<2e-16 ***
<b>Pos 10</b>	-0,216	7,57E-11	<2e-16 ***
<b>Pos 11</b>	-0,143	7,57E-12	<2e-16 ***



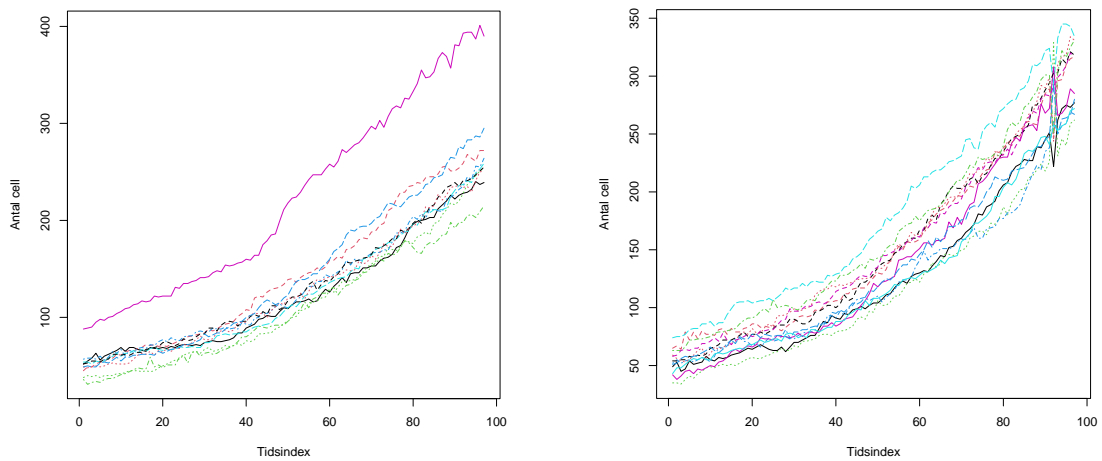
Figur 12: Residualanalys multipel linjär regression för kontrollgruppen experiment 2

Tabell 6: Resultat multipel regression över kontrollgrupp i experiment 3

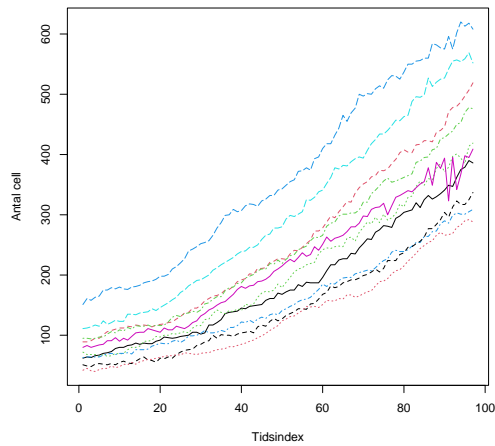
	Koefficient	Medelfel	Pr(> t )
<b>Intercept</b>	4,851	9,31E-03	<2e-16 ***
<b>Tid</b>	1,14E-05	5,13E-08	<2e-16 ***
<b>Pos 2</b>	-0,319	1,16E-02	<2e-16 ***
<b>Pos 3</b>	-0,197	1,16E-02	<2e-16 ***
<b>Pos 4</b>	-0,134	1,16E-02	<2e-16 ***
<b>Pos 5</b>	-0,35	1,16E-02	<2e-16 ***
<b>Pos 6</b>	0,282	1,16E-02	<2e-16 ***
<b>Pos 7</b>	-0,25	1,16E-02	<2e-16 ***
<b>Pos 8</b>	-0,258	1,16E-02	<2e-16 ***
<b>Pos 9</b>	-0,427	1,16E-02	<2e-16 ***
<b>Pos 10</b>	-0,4,14	1,16E-02	<2e-16 ***



Figur 13: Residualanalys multipel linjär regression för kontrollgruppen experiment 3







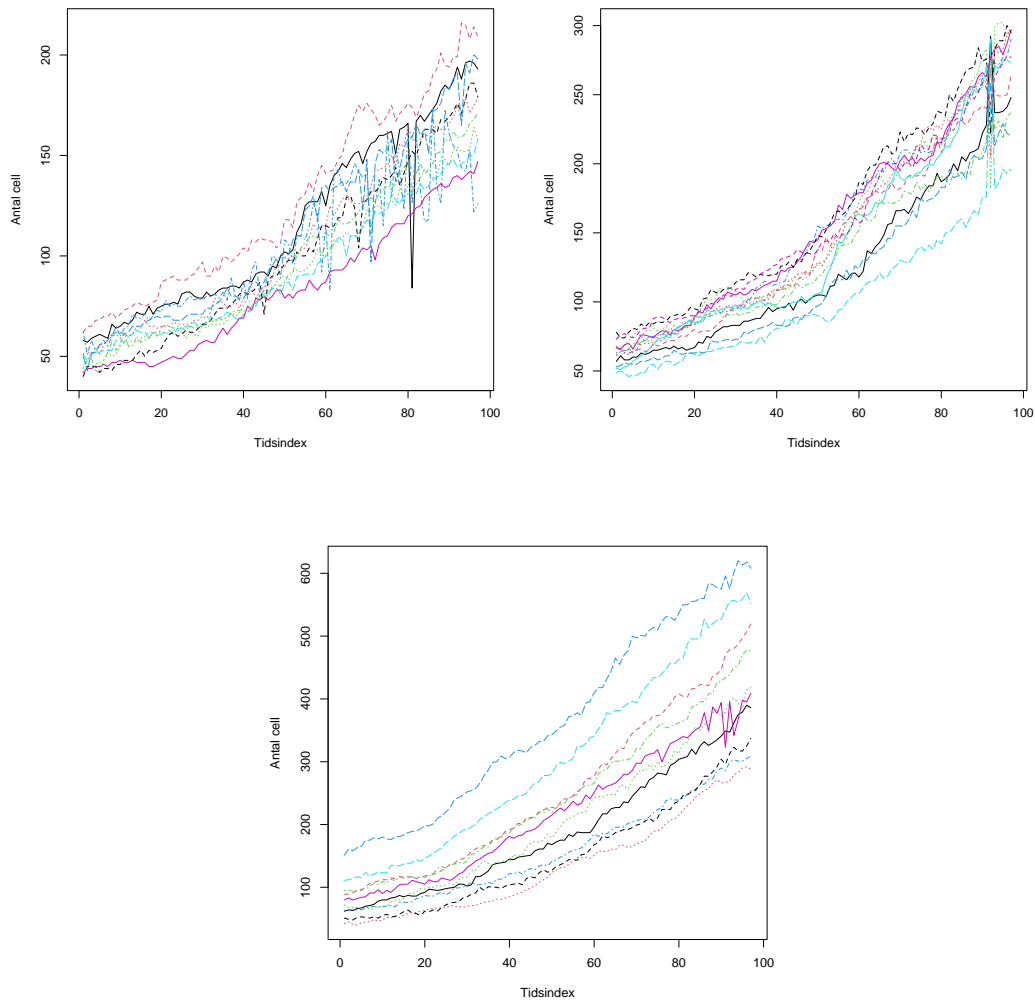
Figur 14: Celltillväxten för dos 1 i experiment 1 (vänster), experiment 2 (höger), experiment 3 (mitten)

Tabell 7: Resultat multipla regressioner över dos 1 i tre experiment

		Koefficient	Medelfel	Pr(> t )
<b>Experiment 1</b>	<b>Intercept</b>	3,834	8,05E-03	<2e-16 ***
	<b>Tid</b>	1,01E-05	4,44E-08	<2e-16 ***
	<b>Pos 2</b>	0,131	1,00E-02	<2e-16 ***
	<b>Pos 3</b>	-0,143	1,00E-03	<2e-16 ***
	<b>Pos4</b>	0,0598	1,00E-04	3,29e-09***
	<b>Pos 5</b>	0,003,23	1,00E-05	0,747
	<b>Pos 6</b>	0,593	1,00E-06	<2e-16 ***
	<b>Pos 7</b>	0,0534	1,00E-07	1,17e-07***
	<b>Pos 8</b>	0,0123	1,00E-08	0,221
	<b>Pos 9</b>	-0,172	1,00E-09	<2e-16 ***
<b>Pos 10</b>	0,1013	1,00E-10	<2e-16 ***	
<b>Experiment 2</b>	<b>Intercept</b>	3,813	7,39E-03	<2e-16 ***
	<b>Tid</b>	1,01E-05	3,80E-08	<2e-16 ***
	<b>Pos 2</b>	0,217	9,37E-03	<2e-16 ***
	<b>Pos 3</b>	-0,098	9,37E-04	<2e-16 ***
	<b>Pos4</b>	0,0205	9,37E-05	0,029 *
	<b>Pos 5</b>	0,399	9,37E-06	<2e-16 ***

	<b>Pos 6</b>	0,0516	9,37E-07	4,63e-08***
	<b>Pos 7</b>	0,183	9,37E-08	<2e-16 ***
	<b>Pos 8</b>	0,209	9,37E-09	<2e-16 ***
	<b>Pos 9</b>	0,284	9,37E-10	<2e-16 ***
	<b>Pos 10</b>	0,0821	9,37E-11	<2e-16 ***
	<b>Pos 11</b>	0,00575	9,37E-12	0,54
	<b>Pos 12</b>	0,206	9,37E-13	<2e-16 ***
<b>Experiment 3</b>	<b>Intercept</b>	4,187	8,08E-03	<2e-16 ***
	<b>Tid</b>	1,05E-05	4,46E-08	<2e-16 ***
	<b>Pos 2</b>	0,307	1,01E-02	<2e-16 ***
	<b>Pos 3</b>	0,0672	1,01E-03	3,97e-11***
	<b>Pos4</b>	-0,146	1,01E-04	<2e-16 ***
	<b>Pos 5</b>	0,497	1,01E-05	<2e-16 ***
	<b>Pos 6</b>	0,168	1,01E-06	<2e-16 ***
	<b>Pos 7</b>	-0,262	1,01E-07	<2e-16 ***
	<b>Pos 8</b>	-0,371	1,01E-08	<2e-16 ***
	<b>Pos 9</b>	0,273	1,01E-09	<2e-16 ***
	<b>Pos 10</b>	0,715	1,01E-10	<2e-16 ***



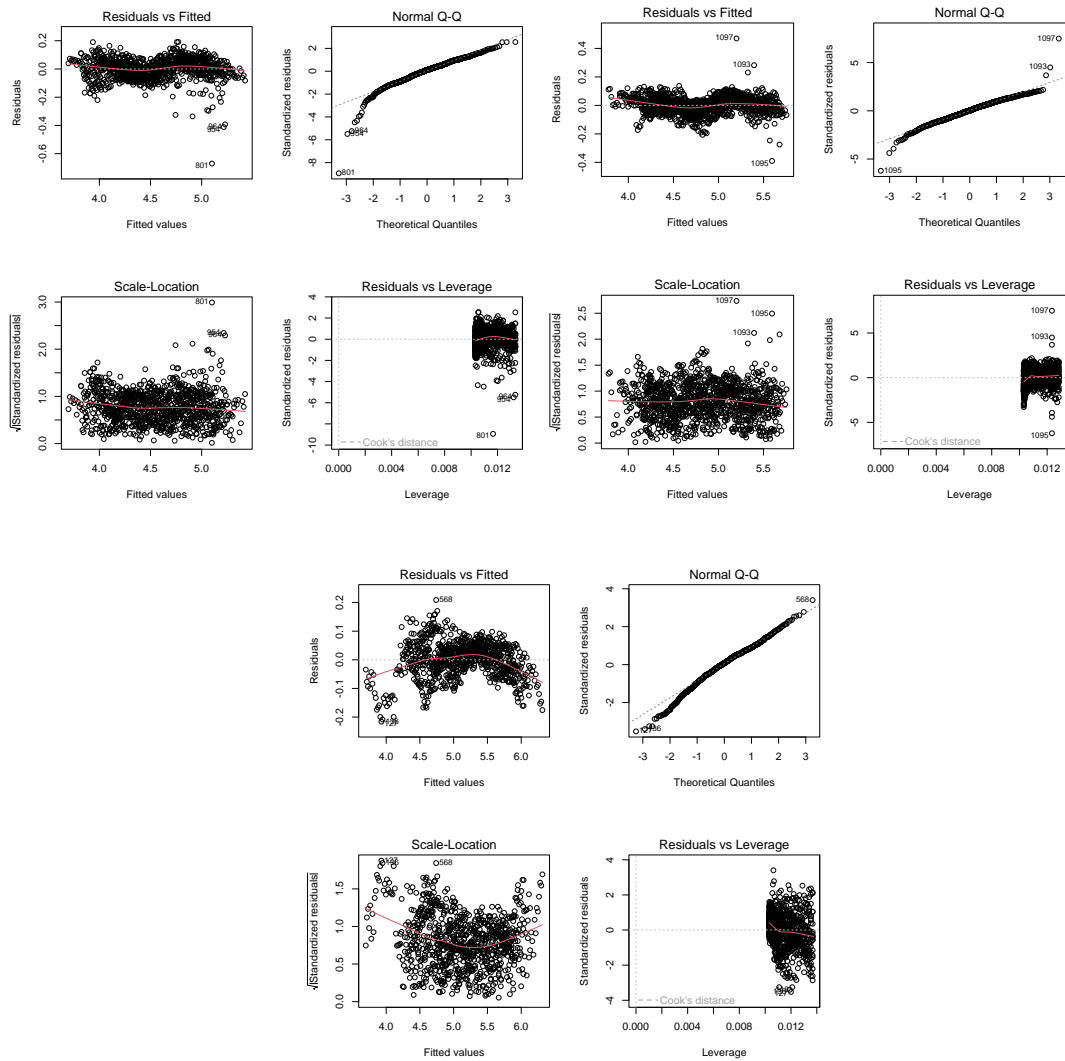


Figur 16: Celltillväxten för dos 2 i experiment 1 (vänster), experiment 2 (höger), experiment 3 (mitten)

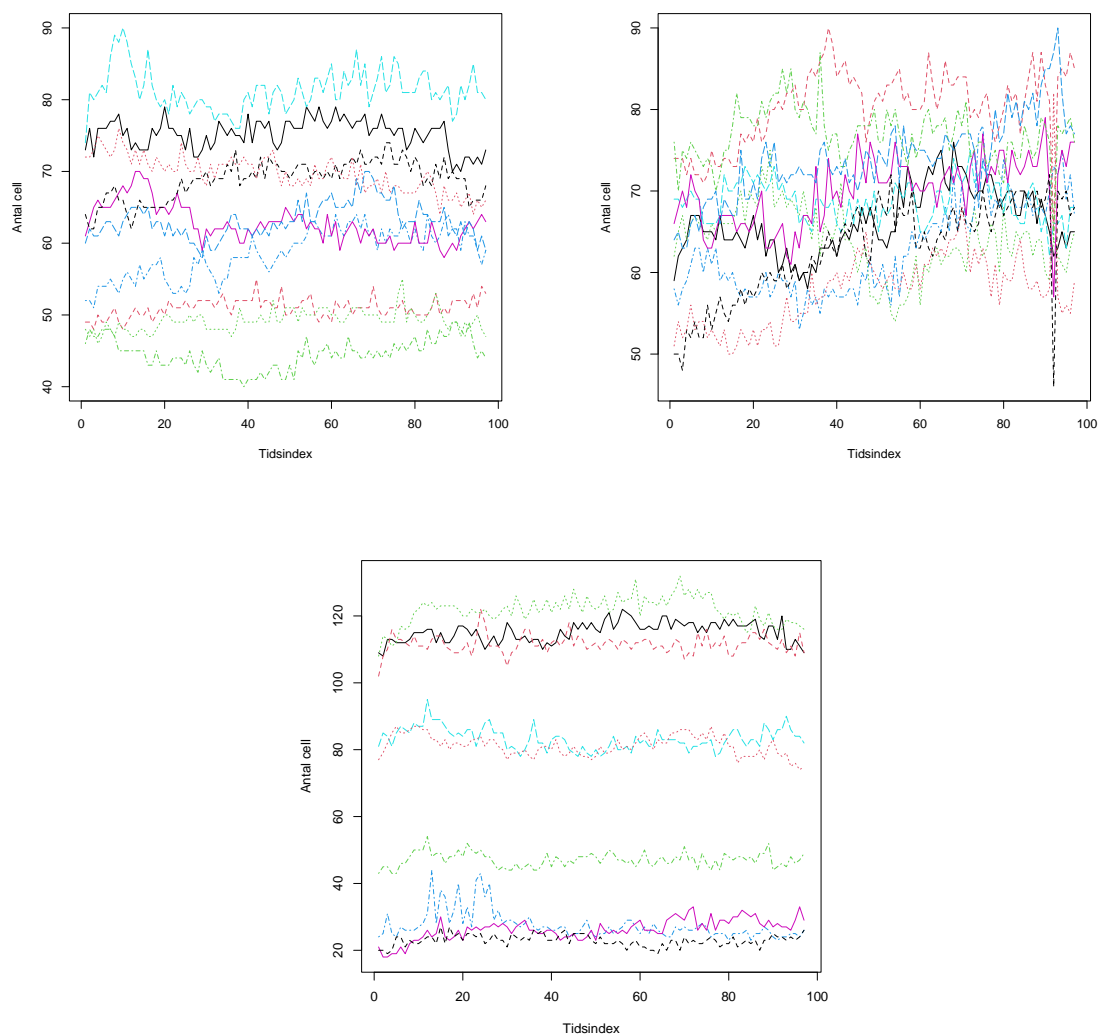
Tabell 8: Resultat multipla regressioner över dos 2 i tre experiment

		Koefficient	Medelfel	Pr(> t )
<b>Experiment 1</b>	<b>Intercept</b>	4,024	8,71E-03	<2e-16 ***
	<b>Tid</b>	7,48E-06	4,81E-08	<2e-16 ***
	<b>Pos 2</b>	0,109	1,08E-02	<2e-16 ***
	<b>Pos 3</b>	-0,203	1,08E-02	<2e-16 ***
	<b>Pos4</b>	-0,0874	1,08E-03	2,16e-15 ***
	<b>Pos 5</b>	-0,205	1,08E-04	<2e-16 ***
	<b>Pos 6</b>	-0,327	1,08E-05	<2e-16 ***

	<b>Pos 7</b>	-0,168	1,08E-06	<2e-16 ***
	<b>Pos 8</b>	-0,0987	1,08E-07	<2e-16 ***
	<b>Pos 9</b>	-0,169	1,08E-08	<2e-16 ***
	<b>Pos 10</b>	-0,0561	1,08E-09	2,73e-07 ***
<b>Experiment</b>	<b>Intercept</b>	3,975	7,15E-03	<2e-16 ***
<b>2</b>	<b>Tid</b>	8,66E-06	3,67E-08	<2e-16 ***
	<b>Pos 2</b>	0,135	9,07E-03	<2e-16 ***
	<b>Pos 3</b>	0,198	9,07E-04	<2e-16 ***
	<b>Pos4</b>	0,204	9,07E-05	<2e-16 ***
	<b>Pos 5</b>	-0,197	9,07E-06	<2e-16 ***
	<b>Pos 6</b>	0,223	9,07E-07	<2e-16 ***
	<b>Pos 7</b>	0,279	9,07E-08	<2e-16 ***
	<b>Pos 8</b>	0,171	9,07E-09	<2e-16 ***
	<b>Pos 9</b>	0,0843	9,07E-10	<2e-16 ***
	<b>Pos 10</b>	-0,0735	9,07E-11	1,27e-15 ***
	<b>Pos 11</b>	0,108	9,07E-12	<2e-16 ***
	<b>Pos 12</b>	0,228	9,07E-13	<2e-16 ***
<b>Experiment</b>	<b>Intercept</b>	3,698	7,21E-03	<2e-16 ***
<b>3</b>	<b>Tid</b>	9,19E-06	4,14E-08	<2e-16 ***
	<b>Pos 2</b>	0,619	8,85E-03	<2e-16 ***
	<b>Pos 3</b>	0,758	8,85E-04	<2e-16 ***
	<b>Pos4</b>	0,498	8,85E-05	<2e-16 ***
	<b>Pos 5</b>	1,027	8,85E-06	<2e-16 ***
	<b>Pos 6</b>	0,66	8,85E-07	<2e-16 ***
	<b>Pos 7</b>	0,595	8,85E-08	<2e-16 ***
	<b>Pos 8</b>	0,748	8,85E-09	<2e-16 ***
	<b>Pos 9</b>	0,844	8,85E-10	<2e-16 ***



Figur 17: Residualanalys multipel regression för dos 2 i Experiment 1 (vänster), Experiment 2 (höger) och Experiment 3 (mitten)



Figur 18: Celltillväxten för dos 3 i experiment 1 (vänster), experiment 2 (höger), experiment 3 (mitten)

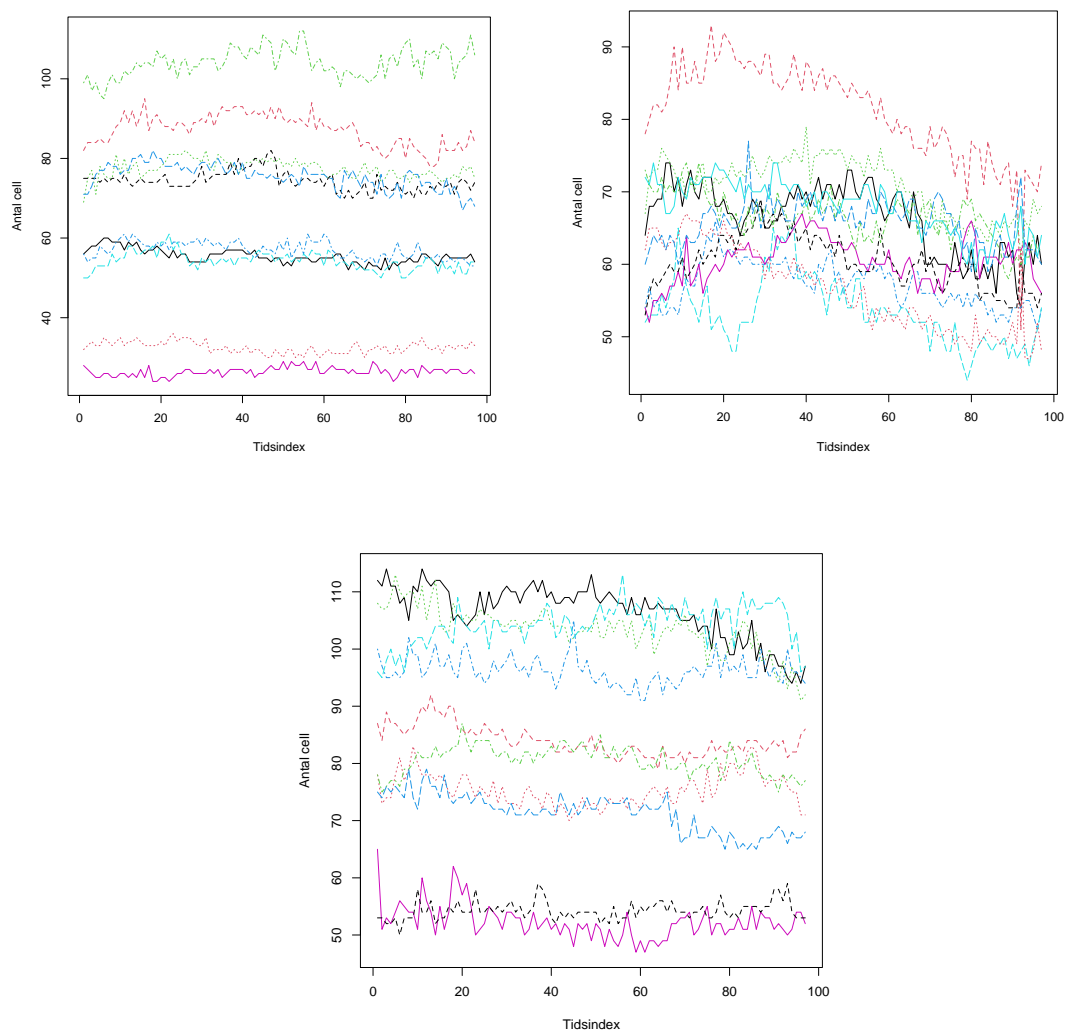
Tabell 9: Resultat multipla regressioner över dos 3 i tre experiment

		Koefficient	Medelfel	Pr(> t )
<b>Experiment 1</b>	<b>Intercept</b>	4,311	4,50E-03	<2e-16 ***
	<b>Tid</b>	1,08E-07	2,48E-08	1,43e-05 ***
	<b>Pos 2</b>	-0,389	5,60E-03	<2e-16 ***
	<b>Pos 3</b>	-0,424	5,60E-04	<2e-16 ***
	<b>Pos4</b>	-0,252	5,60E-05	<2e-16 ***

	<b>Pos 5</b>	0,0775	5,60E-06	<2e-16 ***
	<b>Pos 6</b>	-0,182	5,60E-07	<2e-16 ***
	<b>Pos 7</b>	-0,0898	5,60E-08	<2e-16 ***
	<b>Pos 8</b>	-0,0761	5,60E-09	<2e-16 ***
	<b>Pos 9</b>	-0,526	5,60E-10	<2e-16 ***
	<b>Pos 10</b>	-0,175	5,60E-11	<2e-16 ***
<b>Experiment 2</b>	<b>Intercept</b>	4,141	7,23E-03	<2e-16 ***
	<b>Tid</b>	5,83E-07	3,84E-08	<2e-16 ***
	<b>Pos 2</b>	0,193	9,08E-03	<2e-16 ***
	<b>Pos 3</b>	-0,0361	9,08E-04	7,61e-05 ***
	<b>Pos4</b>	-0,0564	9,08E-05	7,50e-10 ***
	<b>Pos 5</b>	0,0381	9,08E-06	3,00e-05 ***
	<b>Pos 6</b>	0,0539	9,08E-07	3,86e-09 ***
	<b>Pos 7</b>	-0,0567	9,08E-08	6,20e-10 ***
	<b>Pos 8</b>	-0,138	9,08E-09	<2e-16 ***
	<b>Pos 9</b>	-0,33	9,08E-10	<2e-16 ***
	<b>Pos 10</b>	0,15	9,08E-11	<2e-16 ***
	<b>Pos 11</b>	0,109	9,08E-12	<2e-16 ***
<b>Experiment 3</b>	<b>Intercept</b>	4,745	8,49E-03	<2e-16 ***
	<b>Tid</b>	1,27E-08	4,88E-08	0,79417
	<b>Pos 2</b>	-0,0317	1,04E-02	0,00244 **
	<b>Pos 3</b>	0,0543	1,04E-03	2,42e-07 ***
	<b>Pos4</b>	-1,442	1,04E-04	<2e-16 ***
	<b>Pos 5</b>	-0,325	1,04E-05	<2e-16 ***
	<b>Pos 6</b>	-1,485	1,04E-06	<2e-16 ***
	<b>Pos 7</b>	-1,616	1,04E-07	<2e-16 ***
	<b>Pos 8</b>	-0,35	1,04E-08	<2e-16 ***





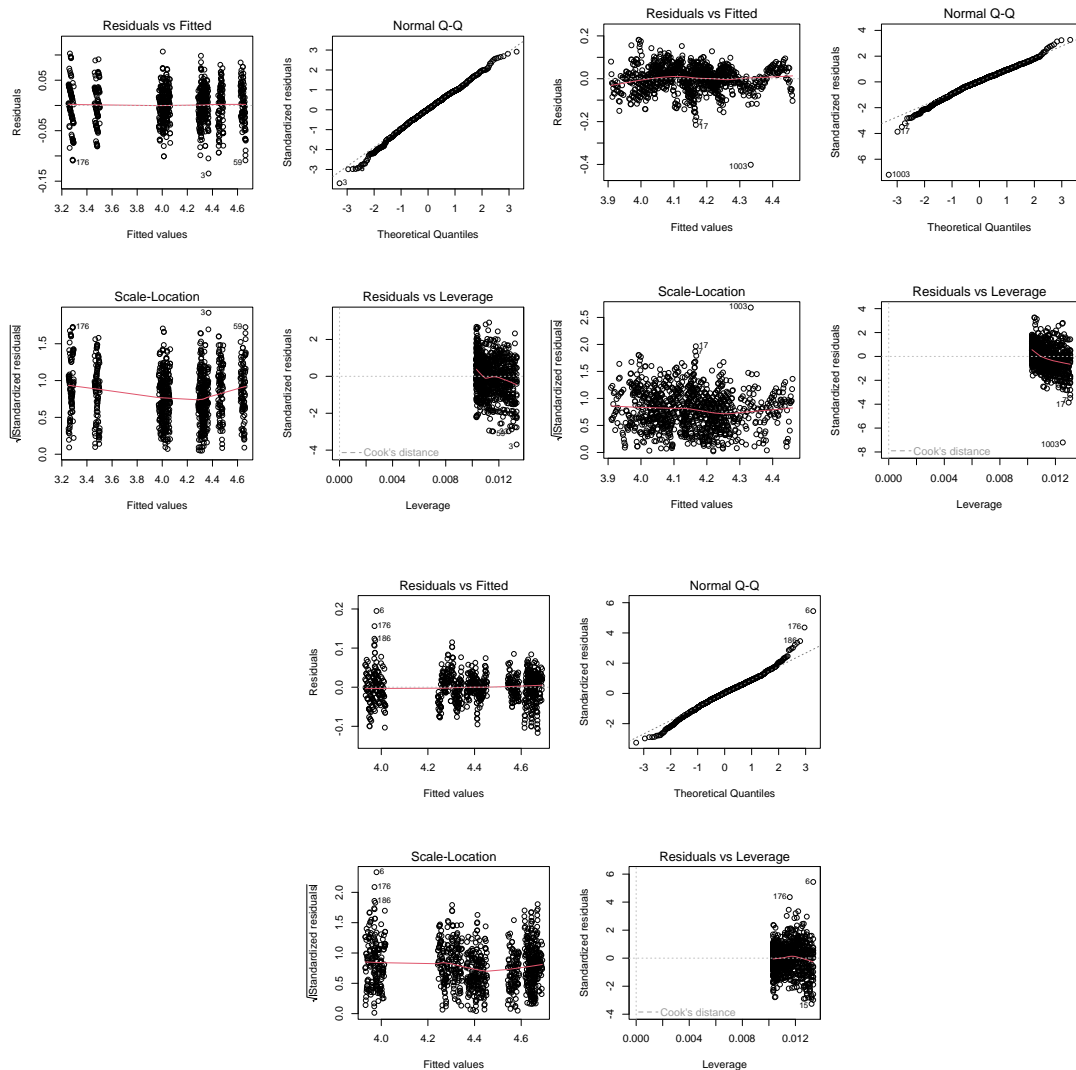


Figur 20: Celltillväxten för dos 4 i experiment 1 (vänster), experiment 2 (höger), experiment 3 (mitten)

Tabell 10: Resultat multipla regressioner över dos 4 i tre experiment

		Koefficient	Medelfel	Pr(> t )
<b>Experiment 1</b>	<b>Intercept</b>	4,035	4,25E-03	<2e-16 ***
	<b>Tid</b>	-2,36E-07	2,34E-08	<2e-16 ***
	<b>Pos 2</b>	0,451	5,28E-03	<2e-16 ***
	<b>Pos 3</b>	0,0334	5,28E-04	<2e-16 ***
	<b>Pos4</b>	0,0285	5,28E-05	8,90e-08 ***
	<b>Pos 5</b>	-0,0221	5,28E-06	3,08e-05 ***

	<b>Pos 6</b>	-0,741	5,28E-07	<2e-16 ***
	<b>Pos 7</b>	0,293	5,28E-08	<2e-16 ***
	<b>Pos 8</b>	-0,534	5,28E-09	<2e-16 ***
	<b>Pos 9</b>	0,63	5,28E-10	<2e-16 ***
	<b>Pos 10</b>	0,308	5,28E-11	<2e-16 ***
<b>Experiment 2</b>	<b>Intercept</b>	4,259	6,41E-03	<2e-16 ***
	<b>Tid</b>	-7,74E-07	3,41E-08	<2e-16 ***
	<b>Pos 2</b>	0,202	8,05E-03	<2e-16 ***
	<b>Pos 3</b>	0,0672	8,05E-04	2,23e-16 ***
	<b>Pos4</b>	-0,145	8,05E-05	<2e-16 ***
	<b>Pos 5</b>	-0,216	8,05E-06	<2e-16 ***
	<b>Pos 6</b>	-0,0904	8,05E-07	<2e-16 ***
	<b>Pos 7</b>	-0,0934	8,05E-08	<2e-16 ***
	<b>Pos 8</b>	-0,16	8,05E-09	<2e-16 ***
	<b>Pos 9</b>	0,00727	8,05E-10	0,366893
	<b>Pos 10</b>	-0,00765	8,05E-11	0,342691
	<b>Pos 11</b>	0,0277	8,05E-12	0,000611 ***
<b>Experiment 3</b>	<b>Intercept</b>	4,691	4,17E-03	<2e-16 ***
	<b>Tid</b>	-2,80E-07	2,30E-08	<2e-16 ***
	<b>Pos 2</b>	-1,2371	5,18E-03	<2e-16 ***
	<b>Pos 3</b>	-0,0282	5,18E-04	6,28e-08 ***
	<b>Pos4</b>	-0,0998	5,18E-05	<2e-16 ***
	<b>Pos 5</b>	-0,0196	5,18E-06	<2e-16 ***
	<b>Pos 6</b>	-0,711	5,18E-07	0,000168 ***
	<b>Pos 7</b>	-0,673	5,18E-08	<2e-16 ***
	<b>Pos 8</b>	-0,345	5,18E-09	<2e-16 ***
	<b>Pos 9</b>	-0,278	5,18E-10	<2e-16 ***



Figur 21: Residualanalys multipel regression för dos 4 i Experiment 1 (vänster), Experiment 2 (höger) och Experiment 3 (mitten)

Tabell 11: Resultat multipel regression över alla doser i Experiment 1

	Koefficient	Medelfel	Pr(> t )
<b>Intercept</b>	4,547	1,44E-02	<2e-16 ***
<b>Tid</b>	5,89E-06	1,02E-07	<2e-16 ***
<b>Dos 1</b>	-0,29	1,62E-02	<2e-16 ***
<b>Dos 2</b>	-0,506	1,62E-03	<2e-16 ***
<b>Dos 3</b>	-0,94	1,62E-04	<2e-16 ***
<b>Dos 4</b>	-0,967	1,62E-05	<2e-16 ***

Utifrån resultatet i tabell 11 blir regressionsmodellen för alla doser hos experiment ett enligt följande:

- För kontrollgrupp, där kontrollgruppen är referensgrupp:

$$\begin{aligned}\log(\hat{Y}_t) &= 4,547 + 5,89 \cdot 10^{-6} \cdot t - 0,29 \cdot 0 - 0,506 \cdot 0 - 0,94 \cdot 0 - 0,967 \cdot 0 \\ &\quad + \ln(\hat{\varepsilon}_t) \\ &= 4,547 + 5,89 \cdot 10^{-6} \cdot t + \ln(\hat{\varepsilon}_t)\end{aligned}$$

- För dos 1:

$$\begin{aligned}\log(\hat{Y}_t) &= 4,547 + 5,89 \cdot 10^{-6} \cdot t - 0,29 \cdot 1 - 0,506 \cdot 0 - 0,94 \cdot 0 - 0,967 \cdot 0 \\ &\quad + \ln(\hat{\varepsilon}_t) \\ &= 4,257 + 5,89 \cdot 10^{-6} \cdot t + \ln(\hat{\varepsilon}_t)\end{aligned}$$

- För dos 2:

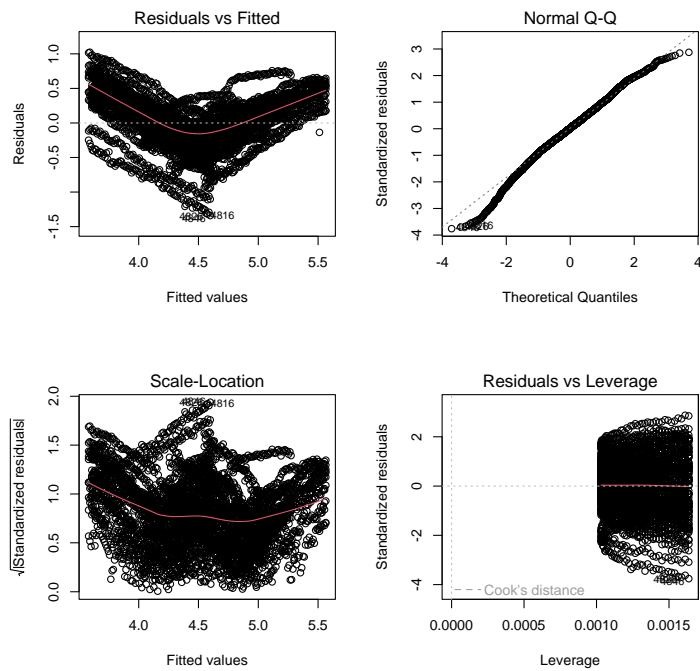
$$\begin{aligned}\log(\hat{Y}_t) &= 4,547 + 5,89 \cdot 10^{-6} \cdot t - 0,29 \cdot 0 - 0,506 \cdot 1 - 0,94 \cdot 0 - 0,967 \cdot 0 \\ &\quad + \ln(\hat{\varepsilon}_t) \\ &= 4,041 + 5,89 \cdot 10^{-6} \cdot t + \ln(\hat{\varepsilon}_t)\end{aligned}$$

- För dos 3:

$$\begin{aligned}\log(\hat{Y}_t) &= 4,547 + 5,89 \cdot 10^{-6} \cdot t - 0,29 \cdot 0 - 0,506 \cdot 0 - 0,94 \cdot 1 - 0,967 \cdot 0 + \\ \ln(\hat{\varepsilon}_t) &= 3,607 + 5,89 \cdot 10^{-6} \cdot t + \ln(\hat{\varepsilon}_t)\end{aligned}$$

- För dos 4:

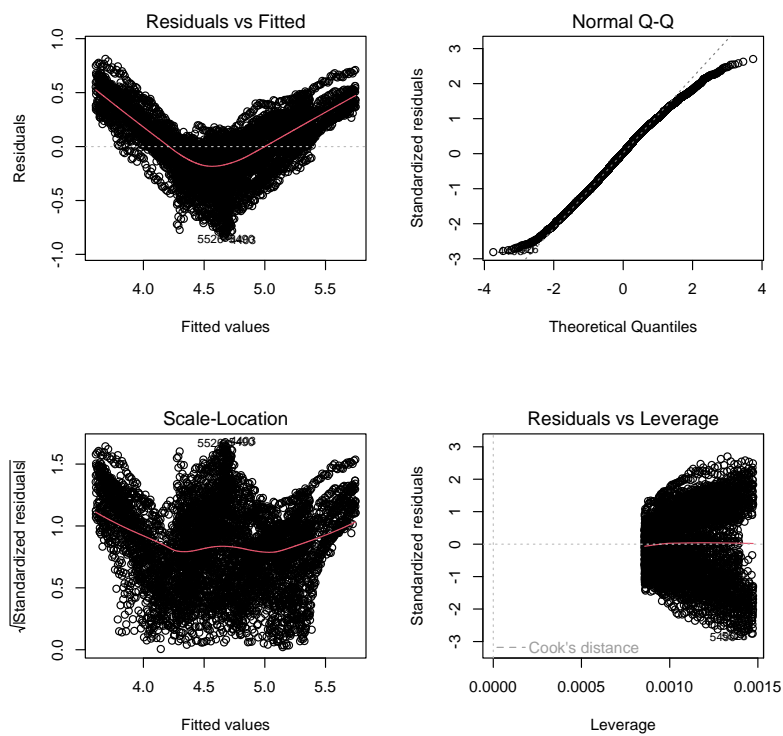
$$\begin{aligned}\log(\hat{Y}_t) &= 4,547 + 5,89 \cdot 10^{-6} \cdot t - 0,29 \cdot 0 - 0,506 \cdot 0 - 0,94 \cdot 0 - 0,967 \cdot 1 \\ &\quad + \ln(\hat{\varepsilon}_t) \\ &= 3,58 + 5,89 \cdot 10^{-6} \cdot t + \ln(\hat{\varepsilon}_t)\end{aligned}$$



Figur 22: : Residualanalys multipel regression för alla doser i Experiment 1

Tabell 12: Resultat multipel regression över alla doser i experiment 2

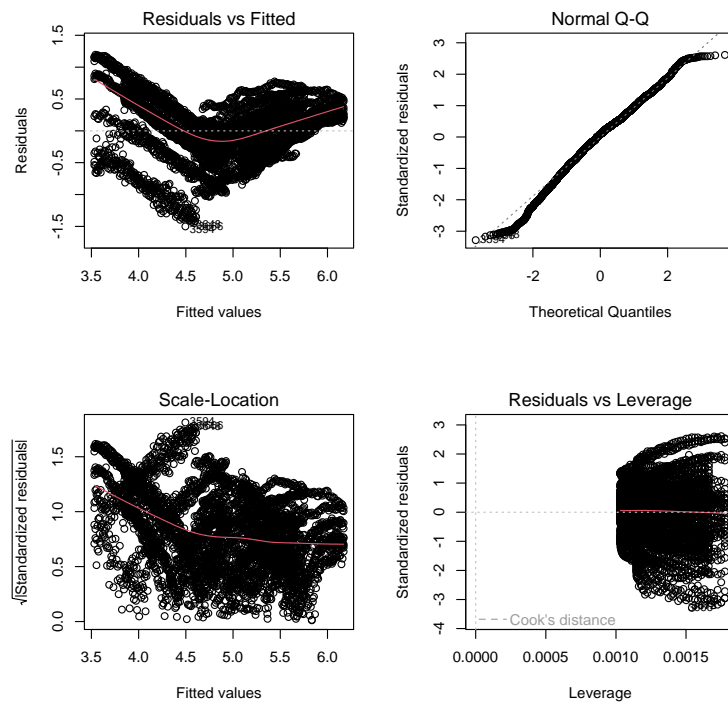
	<b>Koefficient</b>	<b>Medelfel</b>	<b>Pr(&gt; t )</b>
<b>Intercept</b>	4,653	1,16E-02	<2e-16 ***
<b>Tid</b>	6,31E-06	8,03E-08	<2e-16 ***
<b>Dos 1</b>	-0,378	1,28E-02	<2e-16 ***
<b>Dos 2</b>	-0,361	1,28E-02	<2e-16 ***
<b>Dos 3</b>	-1,015	1,30E-02	<2e-16 ***
<b>Dos 4</b>	-1,046	1,30E-02	<2e-16 ***



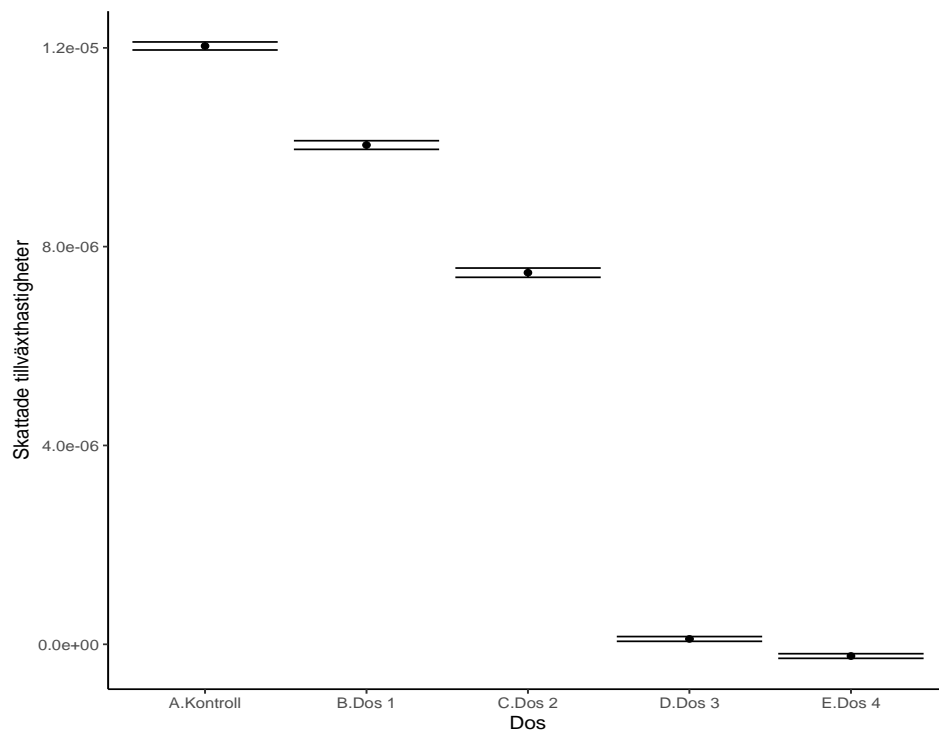
Figur 23: : Residualanalys multipel regression för alla doser i Experiment 2

Tabell 13: Resultat multipel regression över alla doser i experiment 3

	Koefficient	Medelfel	Pr(> t )
<b>Intercept</b>	5,095	1,86E-02	<2e-16 ***
<b>Tid</b>	6,23E-06	1,33E-07	<2e-16 ***
<b>Dos 1</b>	-0,417	2,08E-02	<2e-16 ***
<b>Dos 2</b>	-0,502	2,13E-02	<2e-16 ***
<b>Dos 3</b>	-1,56	2,13E-02	<2e-16 ***
<b>Dos 4</b>	-1,25	2,08E-02	<2e-16 ***

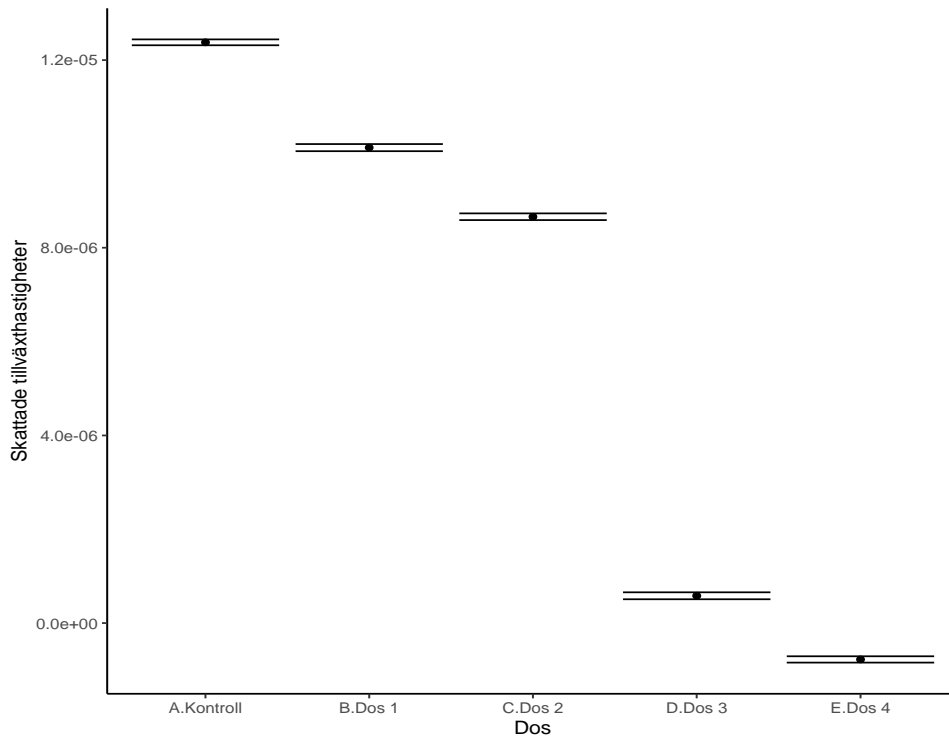


Figur 24: : Residualanalys multipel regression för alla doser i Experiment 3

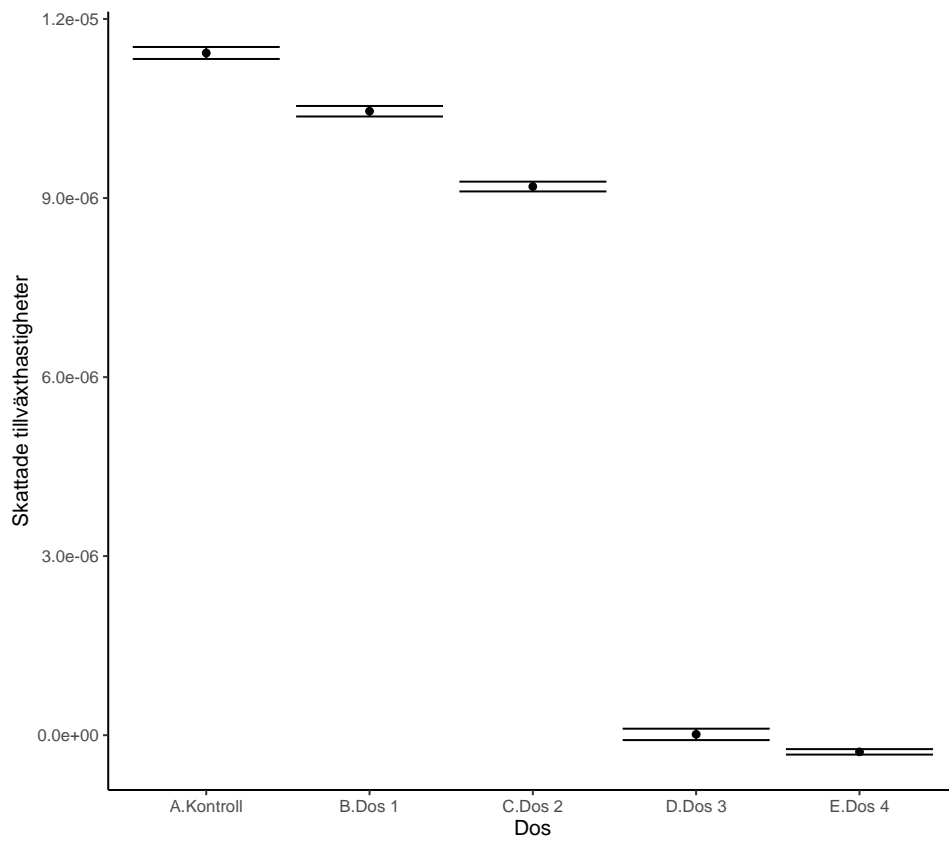


Figur 25: Konfidsintervall för alla doser i Experiment 1

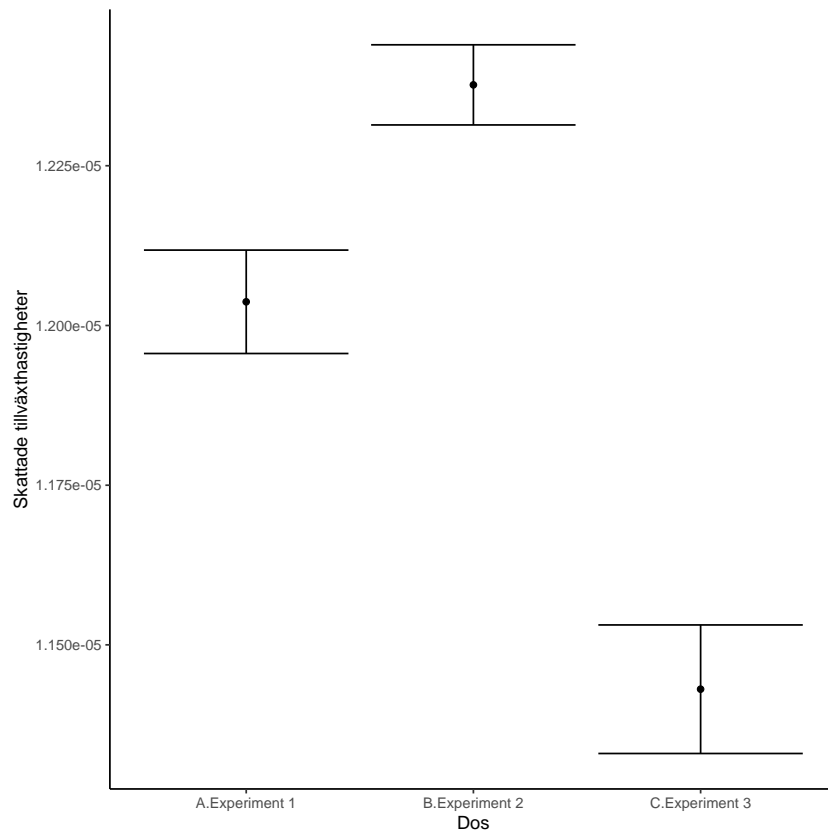




Figur 26: Konfidensintervall för alla doser i Experiment 2



Figur 27: Konfidensintervall för alla doser i Experiment 3



Figur 28: Konfidensintervall för kontrollgrupperna i 3 experiment