Confidence Intervals in Inverse Regression

by

M. Thonnard

Supervisor:
Dr. A. Di Bucchianico

Eindhoven, January 2006
Preface

This Masters thesis is the result of 10 months of research at the Department of Statistics, Probability and Operations Research at the Eindhoven University of Technology (TU/e). With this thesis I will end my study of Industrial and Applied Mathematics, and I will receive the degree of Master of Science.

The subject of my research, Confidence Intervals in Inverse Regression, was introduced to me by dr. A. Di Bucchianico, who became my supervisor on this project.

I would like to take some time to thank everybody who, in one way or another, helped me during my 10 months of research. First of all there is my supervisor. He helped me every time I had a question or problem. And always in the most patient and precise way, in spite of his very busy schedule. I would also like to thank I. Schreur-Piet from the Chemical Engineering and Chemistry Department for the datasets that she gave me.

I would love to thank my parents who made it possible for me to begin (and finish) my study. Not only financially, but also for their support and love. For this support and love I also want to thank my sister and all my friends, Erik and Wannes in particular, who helped me with some problems, and Barbara for her support.
Summary

In most regression problems we have to determine the value of Y corresponding to a given value of x. We will consider the inverse problem, which is called inverse regression or calibration. We will only investigate the simple linear regression model, which is a model with one regressor x that has a relationship with a response Y, which is a straight line. It is not always easy to measure these variables, the regressor x or the response Y. Assume we have known values of x and their corresponding Y values, which both form a simple linear regression model and we have also an unknown value of x, such as \( x_0 \), which can not be measured and we can observe the corresponding value of Y, say \( Y_0 \). Then, \( x_0 \) can be estimated and a confidence interval for \( x_0 \) can be obtained. There are many possibilities to solve this problem.

In the introduction we give the two most used solutions to estimate the unknown \( x_0 \), the classical method and the inverse method, and we will explain them briefly.

Firstly, we will give some simple examples of the calibration problem. We will consider a physical and a chemical example, because inverse regression is frequently used in physical and chemical engineering. First, we will investigate the linearity between the two variables x and Y. Afterwards, we give the classical and inverse estimator of an unknown value \( x_0 \).

Secondly, we will consider the history of inverse regression and we will present a systematic overview.

Afterwards, we will apply Graybill's method from Graybill (1976) on the centred simple linear regression model to obtain an estimator for the unknown value of \( x_0 \) and a \( 100(1 - \alpha)\% \) confidence interval for \( x_0 \). For this model we will calculate the Least Squares estimators for the intercept, the slope and the unknown value of \( x_0 \). Because this estimator of \( x_0 \) is biased, we also present Naszódi's estimator, which is approximately corrected for bias. Then, we calculate the Maximum Likelihood estimator of \( \sigma^2 \) with the likelihood function, in which we substitute the Least Squares estimators of the intercept, the slope and the unknown value of \( x_0 \). This method is not the full Maximum Likelihood method, but a plug-in approach. We will also correct this estimator of \( \sigma^2 \) for bias. We will also obtain a confidence interval for \( x_0 \). Graybill claimed that the confidence coefficient of this interval is less than \( 100(1 - \alpha)\% \).

Then, we will consider a second method to obtain an estimator for the unknown value of \( x_0 \), Brown's profile likelihood. The profile likelihood is the log likelihood maximized over the set of four parameters (slope, intercept, \( \sigma^2 \), \( x_0 \)), which is deducted from the log likelihood maximized over the set of three parameters (slope, intercept, \( \sigma^2 \)) and fixed \( x_0 \). Brown's method is the full Maximum likelihood method in contrast with Graybill's approach. We will also give an estimator for the unknown \( x_0 \).

Finally, in Chapter 6 we will try to investigate if Graybill's claim is true. We will simulate the confidence of this interval in Mathematica for some specific examples. We also deduce \( 100(1 - \alpha)\% \) confidence intervals for the mean \( \mu \) and the variance \( \sigma^2 \) of the confidence. We will also investigate the effect of three factors, the slope, the variance and the number of observations \( k \) on the unknown \( x_0 \), with a face-centred central composite design, because we think that the values of the slope, \( \sigma^2 \) and the number of observations \( k \) on the unknown \( x_0 \) can have an influence on the confidence of the interval.
Contents

1 Introduction 7

2 Examples of the Calibration Problem 11

3 Historical Background 23
    3.1 Classical against Inverse Estimator 24
    3.2 Confidence and Discrimination Intervals for an Unknown Predictor 26
    3.3 Bayesian Solution 29
    3.4 Optimum Design 29
    3.5 Multivariate Calibration 29

4 Graybill’s Method 31
    4.1 Estimation of $\gamma_0$, $\gamma_1$ and $x_0$ 31
    4.2 Estimation of $\sigma^2$ 34
    4.3 Confidence Interval for $x_0$ for a Given $Y_0$ 42

5 Brown’s Profile Likelihood 47

6 Simulation of Graybill’s Confidence Interval 55
    6.1 Confidence Coefficient of Graybill’s Interval for Some Specific Examples 55
    6.2 Experimental Design 66
        6.2.1 Design with $\gamma_1$ Very Small 67
        6.2.2 Design with $\gamma_1$ Large Enough 68

7 Conclusion 73

A Program Codes 77
    A.1 Procedure for the Confidence Interval of Section 4.3 77
Chapter 1

Introduction

In this chapter we will discuss briefly two methods to solve the calibration problem for a simple linear regression model. In regression analysis we want to investigate the relationship between variables. Regression has many applications, which occur in many fields: engineering, economics, the physical and chemical sciences, management, biological sciences and social sciences. We only consider the simple linear regression model, which is a model with one regressor \( x \) that has a linear relationship with a response \( Y \). It is not always easy to measure the regressor \( x \) or the response \( Y \). We now consider a typical example for this problem. If \( x \) is the concentration of glucose in certain substances, then a spectrophotometric method is used to measure the absorbance. This absorbance depends on the concentration \( x \). The response \( Y \) is easy to measure with the spectrophotometric method, but the concentration on the other hand is not easy to measure. If we have \( n \) known concentrations, then the absorbance can be measured. If there is a linear relation between \( x \) and \( Y \), then a simple linear regression model can be made with these data. Suppose we have an unknown concentration, which is difficult to measure, but we can measure the absorbance of this concentration. Is it possible to estimate this concentration with the measured absorbance? This is called the calibration problem.

Suppose we have a linear model \( Y = \beta_0 + \beta_1 x + \varepsilon \) and we have an observed value of the response \( Y \), but we do not have the corresponding value of \( x \). How can we estimate this value of \( x \)? The two most important methods to estimate \( x \) are the classical method and the inverse method. We will also consider bias correction for the classical method from Naszódi (1978) and is discussed in Chapter 4. Other methods to estimate \( x_0 \) are a Bayesian solution from Hoadley (1970) or Brown (1993) proposed a method with the profile likelihood to obtain an estimator for the unknown \( x_0 \). Brown’s method is discussed in Chapter 5. For the Bayesian solution of Hoadley we will only give some references. We now discuss briefly the classical and inverse method:

- The classical method is based on the simple linear regression model

\[
Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2),
\]

where the parameters \( \beta_0 \) and \( \beta_1 \) are estimated by Least Squares as \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). At least two of the \( n \) values of \( x \) have to be distinct, otherwise we can not fit a reliable regression line. For a given value of \( x \), say \( x_0 \) (unknown), a \( Y \) value, say \( Y_0 \) (or random sample of \( k \) values of \( Y \)) is observed at the \( x_0 \) value. The problem is to estimate \( x_0 \). The classical
method uses a $Y_0$ value (or the mean of $k$ values of $Y_0$) to estimate $x_0$, which is then estimated by

$$
\hat{x}_0 = \frac{\bar{Y}_0 - \hat{\beta}_0}{\hat{\beta}_1}.
$$

(1.2)

- The inverse estimator is the simple linear regression of $x$ on $Y$. In this case, we have to fit the model

$$
x = \alpha_0 + \alpha_1 Y + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2),
$$

(1.3)

to obtain the estimator. Then the inverse estimator of $x_0$ is

$$
\hat{x}_0 = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{Y}_0,
$$

(1.4)

where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the Least Squares estimators of $\alpha_0$ and $\alpha_1$, resp.

These estimators are point estimators of $x_0$.

The goal of this Master’s thesis is to investigate interval estimators (i.e. confidence intervals) in order to get an impression of the accuracy. In Chapter 2 we give two simple examples of the calibration problem. We have a physical example, the calibration of an instrument, and a chemical example, calibration of a concentration of a solution. For both examples we will give a regression analysis and we will calculate the classical and inverse estimator to give intuition for the reader. Results on inverse regression are scattered throughout the scientific literature. In Chapter 3 we present a systematic overview. We will divide these many articles and books into 5 groups and we will link them in a diagram. Graybill presents in Graybill (1976) a method to obtain a estimator for the unknown value of $x_0$ and a $100(1 - \alpha)$% confidence interval for $x_0$ for the simple linear regression model. We consider a slight variation on Graybill’s method in Chapter 4. We have chosen for a centred simple linear regression model, because this centred model simplifies the calculations. For this model we will calculate the Least Squares estimators for the intercept, the slope and the unknown value of $x_0$. Then we calculate the Maximum Likelihood estimator of $\sigma^2$ with the likelihood function, in which we substitute the Least Squares estimators of the intercept, the slope and the unknown value of $x_0$. We will also correct this estimator for $\sigma^2$ for bias. Afterwards, we consider Graybill’s method to obtain a $100(1 - \alpha)$% confidence interval for $x_0$. We will work out the details, which are missing in Graybill (1976). Graybill claimed that the confidence coefficient of this interval is less than $1 - \alpha$. We investigate this in Chapter 6. Further on, in Chapter 5 we consider Brown’s method with the profile likelihood to obtain the estimator of $x_0$. We will maximize the log likelihood over the set of four parameters (slope, intercept, $\sigma^2$, $x_0$). But firstly $x_0$ is taken to be known to calculate the three estimators of the intercept, slope and $\sigma^2$. Afterwards, the estimator of $x_0$ is calculated from the estimator of $\sigma^2$, because the log likelihood is maximized if $\hat{\sigma}^2$ is minimized. The estimator of $x_0$ is the value which minimizes $\hat{\sigma}^2$. We supplied missing details of the extensive calculations. We also identified and corrected several mistakes. In Chapter 6 we give a simulation of the $100(1 - \alpha)$% confidence interval of Graybill’s method to investigate Graybill’s claim. We will simulate this interval 10,000 and 100,000 times for the examples given in Chapter 2 and we will count how many times the real value of $x_0$ lies in the interval. This procedure is done 30 times and afterwards a confidence interval is given for the confidence coefficient. We can conclude that Graybill’s claim is true,
but the difference is very small. The difference is even less than 0.5%. We will also investigate the effect of three factors with a face-centred central composite design, because we think that the choice of the slope, $\sigma^2$ and the number of observations $k$ on the unknown $x_0$ can have an influence on the confidence of the interval. Finally, we end in Chapter 7 with a conclusion.
Chapter 2

Examples of the Calibration Problem

In this chapter we illustrate the calibration problem with some examples. We will give a physical and a chemical example. For each example we will give the Analysis-of-Variance table for the simple linear regression model. This table investigates the linearity of the model. We will check the normality of the error with the Shapiro-Wilk test and a normal probability plot. We will also see whether there are outliers or not with the studentized residuals, Cook's D values, Dfits and Dfbetas. If the model is linear and the errors are normally distributed, then we calculate the classical and inverse estimator for an unknown $x_0$.

Inverse regression is often used in the physics to calibrate instruments. In the first example a thermocouple is calibrated.

Example 2.1 In our first example we want to calibrate a thermocouple. We have the actual temperature $x$, which is measured by a thermometer of known accuracy. The corresponding values of $Y$ are observed on the thermocouple. There are 16 observations. These values, which are taken from Montgomery et al. (2001), are shown in Table 2.1 and plotted in Figure 2.1.

Suppose we have a new observation, $Y_0 = 200^\circ C$. We now want to estimate the actual temperature, $x_0$, of $Y_0$. We now obtain the classical and inverse estimator for this example.

For the classical estimator, we have the simple linear regression model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2),$$

(2.1)

where the parameters $\beta_0$ and $\beta_1$ are estimated by Least Squares as $\hat{\beta}_0$ and $\hat{\beta}_1$. Table 2.2 is the Analysis-of-Variance table of the regression model. For this model $R^2 = 0.9993$, which does mean that 99.93% of variability in observed temperature is accounted for by the regression model. The p-value is less than 0.0001, which means that this model is highly significant. The estimators of the parameters $\beta_0$ and $\beta_1$ are given in Table 2.3. The estimator of $\sigma^2$ is 5.8835 and a 95% confidence interval for $\sigma^2$ is $(3.1559, 14.6304)$.

Thus, the simple linear regression model is

$$\hat{Y} = -6.66985 + 0.95303x.$$  

(2.2)

We now check the normality assumption of the error with a normal probability plot of the residuals (Figure 2.2). We can see that the normality assumption is satisfied, because the value of the Shapiro-Wilks statistic is 0.975915 and the p-value of the Shapiro-Wilks statistic
Table 2.1: Actual and observed temperature.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Actual temperature $x_i$ ($^\circ$C)</th>
<th>Observed temperature $Y_i$ ($^\circ$C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>88.8</td>
</tr>
<tr>
<td>2</td>
<td>120</td>
<td>108.7</td>
</tr>
<tr>
<td>3</td>
<td>140</td>
<td>129.8</td>
</tr>
<tr>
<td>4</td>
<td>160</td>
<td>146.2</td>
</tr>
<tr>
<td>5</td>
<td>180</td>
<td>161.6</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>179.9</td>
</tr>
<tr>
<td>7</td>
<td>220</td>
<td>202.4</td>
</tr>
<tr>
<td>8</td>
<td>240</td>
<td>224.5</td>
</tr>
<tr>
<td>9</td>
<td>260</td>
<td>245.1</td>
</tr>
<tr>
<td>10</td>
<td>280</td>
<td>257.7</td>
</tr>
<tr>
<td>11</td>
<td>300</td>
<td>277.0</td>
</tr>
<tr>
<td>12</td>
<td>320</td>
<td>298.1</td>
</tr>
<tr>
<td>13</td>
<td>340</td>
<td>318.8</td>
</tr>
<tr>
<td>14</td>
<td>360</td>
<td>334.6</td>
</tr>
<tr>
<td>15</td>
<td>380</td>
<td>355.2</td>
</tr>
<tr>
<td>16</td>
<td>400</td>
<td>377.0</td>
</tr>
</tbody>
</table>

Figure 2.1: Scatter plot of Table 2.1.
Table 2.2: Analysis-of-Variance table of the simple linear regression model for the classical estimator.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>123524</td>
<td>123524</td>
<td>20994.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>14</td>
<td>82.36938</td>
<td>5.88353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected Total</td>
<td>15</td>
<td>123606</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Estimates of $\beta_0$ and $\beta_1$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-6.66985</td>
<td>1.75258</td>
<td>-3.81</td>
<td>0.0019</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.95303</td>
<td>0.00658</td>
<td>144.90</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Figure 2.2: Normal probability plot of the residuals.
is 0.9229. The normal probability plot affirms also normality. The plot of the studentized residuals against the regressor is given in Figure 2.3. And a plot of the studentized residuals against the predicted values is given in Figure 2.4. Both residual plots look satisfactory.

![Residual Plot](image)

Figure 2.3: Plot of studentized residuals against the regressor.

To see whether there are points, which have an influence on the model, we can look at the studentized residuals, Cook’s D values, DFfits and DFbetas:

- **Studentized residuals**: the highest residual is 1.696 (observation 9), but it is still below the limit 2.5. There are no outliers.

- **Cook’s D**: observation 16 has the highest Cook’s D, 0.196, but it is still below the limit 1. There are also no problems.

- **DFfits**: in this case the limit is $2 \sqrt{\frac{2}{16}} \approx 0.71$. The points for which $|DFfits| > 2 \sqrt{\frac{2}{16}}$ can be outliers. In this example there is no observation with $|DFfits| > 0.71$. We do not have outliers.

- **DFbetas**: for DFbetas the absolute limit is $\frac{2}{\sqrt{16}} = 0.5$. The DFbetas of the intercept of point 3 exceeds the limit and also the DFbetas of $\beta_1$ of point 16 exceeds the limit. These two points could be outliers, but according to the other 3 methods these two points are no outliers. Thus, we keep these two points in the model.

From (1.2), the classical estimator of $x_0$ on $Y_0 = 200^\circ C$ is

$$\hat{x}_0 = \frac{Y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{200 - (-6.66985)}{0.95303} = 216.86^\circ C.$$  \hfill (2.3)
The 95% confidence interval for 216.86°C is (211.278, 222.419). The formulas to obtain this interval can be found in Chapter 4.

To obtain the inverse estimator of \( x_0 \), we have to fit the simple linear regression model of \( x \) on \( Y \):

\[
\hat{x} = 7.16051 + 1.04859Y.
\]  

(2.4)

The estimator of \( \sigma^2 \) of this model is 6.4734 and a 95% confidence interval for \( \sigma^2 \) is (3.4723, 16.0974). The inverse estimator of \( x_0 \) on \( Y_0 = 200°C \) is then

\[
\hat{x}_0 = 7.16051 + 1.04859 \times 200 = 216.88°C.
\]  

(2.5)

The 95% confidence interval for 216.88°C is (215.428, 218.327).

The second example is a typical chemical example, because in chemistry the calibration method is frequently used to solve problems.

**Example 2.2** For this example we have 10 solutions from caprolactone in the solvent tetrahydrofurhan. We also have a solution without the caprolactone. These solution are injected 3 times in a gas chromatograph. This gas chromatograph measures from each substance in the solution a surface. This surface is a size for the concentration. Now we have the concentration \( x \), which is known for the 10 solutions, and we have the surface \( Y \), which is measured 3 times for each concentration. So we have 33 observations (\( n = 33 \)), which are shown in Table 2.4 and plotted in Figure 2.5. Suppose we have a new solution, from which we do not
know the concentration of caprolactone, \( x_0 \). In this case we can use the calibration method to obtain an estimator for the concentration, because we can at measure the surface by using the gas chromatograph. We will also inject this unknown solution 3 times in the gas chromatograph, such that we will have 3 observations on the unknown \( x_0 \). These 3 observations of \( Y_0 \) are: \( Y_{01} = 1.58211 \), \( Y_{02} = 1.79325 \) and \( Y_{03} = 1.78739 \). We obtain the classical and inverse estimator for this example.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Concentration ( x_i ) ((g/l))</th>
<th>Surface ( Y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.71</td>
<td>24.276</td>
</tr>
<tr>
<td>2</td>
<td>9.71</td>
<td>24.083</td>
</tr>
<tr>
<td>3</td>
<td>9.71</td>
<td>24.276</td>
</tr>
<tr>
<td>4</td>
<td>8.52</td>
<td>20.206</td>
</tr>
<tr>
<td>5</td>
<td>8.52</td>
<td>20.199</td>
</tr>
<tr>
<td>6</td>
<td>8.52</td>
<td>20.223</td>
</tr>
<tr>
<td>7</td>
<td>7.96</td>
<td>19.773</td>
</tr>
<tr>
<td>8</td>
<td>7.96</td>
<td>19.759</td>
</tr>
<tr>
<td>9</td>
<td>7.96</td>
<td>19.765</td>
</tr>
<tr>
<td>10</td>
<td>6.82</td>
<td>16.743</td>
</tr>
<tr>
<td>11</td>
<td>6.82</td>
<td>16.587</td>
</tr>
<tr>
<td>12</td>
<td>6.82</td>
<td>16.744</td>
</tr>
<tr>
<td>13</td>
<td>5.85</td>
<td>15.081</td>
</tr>
<tr>
<td>14</td>
<td>5.85</td>
<td>15.121</td>
</tr>
<tr>
<td>15</td>
<td>5.85</td>
<td>15.274</td>
</tr>
<tr>
<td>16</td>
<td>4.95</td>
<td>12.636</td>
</tr>
<tr>
<td>17</td>
<td>4.95</td>
<td>12.641</td>
</tr>
<tr>
<td>18</td>
<td>4.95</td>
<td>12.682</td>
</tr>
<tr>
<td>19</td>
<td>3.91</td>
<td>9.869</td>
</tr>
<tr>
<td>20</td>
<td>3.91</td>
<td>9.906</td>
</tr>
<tr>
<td>21</td>
<td>3.91</td>
<td>9.883</td>
</tr>
<tr>
<td>22</td>
<td>2.98</td>
<td>7.624</td>
</tr>
<tr>
<td>23</td>
<td>2.98</td>
<td>7.592</td>
</tr>
<tr>
<td>24</td>
<td>2.98</td>
<td>7.585</td>
</tr>
<tr>
<td>25</td>
<td>2.07</td>
<td>4.638</td>
</tr>
<tr>
<td>26</td>
<td>2.07</td>
<td>4.666</td>
</tr>
<tr>
<td>27</td>
<td>2.07</td>
<td>4.649</td>
</tr>
<tr>
<td>28</td>
<td>1.02</td>
<td>2.86</td>
</tr>
<tr>
<td>29</td>
<td>1.02</td>
<td>2.859</td>
</tr>
<tr>
<td>30</td>
<td>1.02</td>
<td>2.896</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.4: Concentration and surface of caprolactone in the solution.

For the classical estimator, we have the following simple linear regression model

\[
Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2), \tag{2.6}
\]
where the parameters $\beta_0$ and $\beta_1$ are estimated by Least Squares as $\hat{\beta}_0$ and $\hat{\beta}_1$. Table 2.5 is the Analysis-of-Variance table of the regression model. For this model $R^2 = 0.9970$, which does mean that 99.70% of variability in observed temperature is accounted for by the regression model. The $p$-value is less than 0.0001, which means that this model is highly significant. The estimators of the parameters $\beta_0$ and $\beta_1$ are given in Table 2.6. The estimator of $\sigma^2$ is 0.1750 and a 95% confidence interval for $\sigma^2$ is $(0.1126, 0.3100)$.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>$F$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>1836.95638</td>
<td>1836.95638</td>
<td>10332.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>31</td>
<td>5.51130</td>
<td>0.17778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected Total</td>
<td>32</td>
<td>1842.46768</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: Analysis-of-Variance table of the simple linear regression model for the classical estimator.

Thus, the simple linear regression model is

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>$t$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.16943</td>
<td>0.13889</td>
<td>1.22</td>
<td>0.2317</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.45092</td>
<td>0.02411</td>
<td>101.65</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 2.6: Estimates of $\beta_0$ and $\beta_1$. 
\[ \hat{Y} = 0.16943 + 2.45092x. \]  

(2.7)

We now check the normality assumption of the error with a normal probability plot of the residuals (Figure 2.6). The value of the Shapiro-Wilks statistic is 0.926054 and the p-value of the Shapiro-Wilks statistic is 0.0270, which indicates that we have normality at the level of \( \alpha = 0.01 \). The normal probability plot does not look acceptable. To see what is wrong we will plot a density trace for the response (surface) (Figure 2.7). We can see that the distribution appears to have two peaks, which can indicate that the data are taken from two distributions.

To see whether there are points, which have an influence on the model, we can look at the studentized residuals, Cook’s D values, DFfits and DFbetas:

- **Studentized residuals**: the highest residual is 1.849 (observation 15), but it is still below the limit 2.5. There are no outliers.

- **Cook’s D**: observation 5 has the highest Cook’s D, 0.175, but it is still below the limit 1. There are also no problems.

- **DFfits**: in this case the limit is \( 2\sqrt{\frac{2}{33}} \approx 0.49 \). The points for which \( |DF\text{fits}| > 2\sqrt{\frac{2}{33}} \) can be outliers. In this example 3 points can be outliers: point 4 with DFfits = -0.6217, point 5 with DFfits = -0.6277 and point 6 with DFfits = -0.6072.

- **DFbetas**: for DFbetas the absolute limit is \( \frac{2}{\sqrt{33}} = 0.35 \). The DFbetas of the intercept of point 25 exceeds the limit and also the DFbetas of \( \beta_1 \) of points 4, 5 and 6 exceeds the
Figure 2.7: Density trace for the response (surface).

Because the normal probability plot of the residuals in Figure 2.6 and the density trace in Figure 2.7 are not acceptable, we have to remove some data. We remove the 3 influential points (observations 4, 5 and 6) and also the points with surface less than 3 (observations 28, 29, 30, 31, 32 and 33). The normal probability plot of the residuals of these 24 data is given in Figure 2.8 and the density trace is given in Figure 2.9. It can be seen that both figures look acceptable. We can also see that the normality assumption is satisfied, because the value of the Shapiro-Wilks statistic is 0.96324 and the p-value of the Shapiro-Wilks statistic is 0.5069. The Analysis-of-Variance table of these 24 data is given in Figure 2.7. For this model $R^2 = 0.9971$, which does mean that 99.71% of variability in observed temperature is accounted for by the regression model. The p-value is less than 0.0001, which means that this model is highly significant. The estimators of the parameters $\beta_0$ and $\beta_1$ are given in Table 2.8. The estimator of $\sigma^2$ is 0.11497 and a 95% confidence interval for $\sigma^2$ is (0.06873, 0.22994). Thus, the simple linear regression model for the 24 observations is

$$ \hat{Y} = -0.02533 + 2.50441x. \quad (2.8) $$

From (1.2), the classical estimator of $x_0$ on $\bar{Y}_0 = 1.72092$ is

$$ \bar{x}_0 = \frac{\bar{Y}_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{1.72092 - (-0.02533)}{2.50441} = 0.69727. \quad (2.9) $$
Figure 2.8: Normal probability plot of the residuals of the 24 data.

Figure 2.9: Density trace for the response (surface) of the 24 data.
Table 2.7: Analysis-of-Variance table of the simple linear regression model for the classical estimator for the 24 data.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>876.43665</td>
<td>876.43665</td>
<td>7623.16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>2.52934</td>
<td>0.11497</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected Total</td>
<td>23</td>
<td>878.96599</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.8: Estimates of $\beta_0$ and $\beta_1$ for the 24 data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-0.02533</td>
<td>0.17310</td>
<td>-0.15</td>
<td>0.8850</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.50441</td>
<td>0.02868</td>
<td>87.31</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

The 95% confidence interval for 0.69727 is (0.49149, 0.90303). The formulas to obtain this interval can be found in Chapter 4.

To obtain the inverse estimator of $x_0$, we have to fit the simple linear regression model of $x$ on $Y$:

$$\hat{x} = 0.0260012 + 0.398146Y.$$  \hfill (2.10)

The estimator of $\sigma^2$ of this model is 0.01827 and a 95% confidence interval for $\sigma^2$ is (0.01093, 0.03656). The inverse estimator of $x_0$ on $\bar{Y}_0 = 1.72092$ is then

$$\hat{x}_0 = 0.0260012 + 0.398146 \times 1.72092 = 0.71118.$$  \hfill (2.11)

The 95% confidence interval for 0.71118 is (0.58318, 0.83918).

In the next chapter we will give a historical background of inverse regression.
Chapter 2 Examples of the Calibration Problem
Chapter 3

Historical Background

In this chapter we will give a historical background of the calibration problem. In the twentieth century statisticians began to study the calibration problem. We can divide their researches into 5 different groups: classical method against inverse method, confidence and discrimination intervals for an unknown predictor, Bayesian solution, optimum design and multivariate calibration. The Bayesian solution, optimum design and multivariate calibration are discussed briefly in this report. Further on, we give a overview and also a few names of statisticians who studied some aspects of inverse regression.

Figure 3.1 gives a diagram of all the groups that are mentioned above and some statisticians who have investigated some aspects of these 5 groups.

Figure 3.1: Diagram of the 5 groups.
3.1 Classical against Inverse Estimator

Many statisticians compared the classical method with the inverse method, because both estimators are in general not the same. They wanted to investigate which one has the most advantages. Eisenhart (1939) is one of the first statisticians who discussed the problem of choosing between these methods. He compared the Analysis-of-Variance tables of both estimators. The Analysis-of-Variance table of the classical estimator is shown in Table 3.1 and the Analysis-of-Variance table of the inverse estimator is shown in Table 3.2.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>$SS_R = \hat{\beta}<em>1 \sum</em>{i=1}^{N} (x - \bar{x})(Y - \bar{Y})$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - 2$</td>
<td>$SS_{RES} = \sum_{i=1}^{N} (Y - \hat{Y})^2$</td>
</tr>
<tr>
<td>Uncorrected Total</td>
<td>$n - 1$</td>
<td>$SS_T = \sum_{i=1}^{N} (Y - \bar{Y})^2$</td>
</tr>
</tbody>
</table>

Table 3.1: Analysis-of-Variance table of the classical estimator.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>$SS_R = \hat{\alpha}<em>1 \sum</em>{i=1}^{N} (x - \bar{x})(Y - \bar{Y})$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - 2$</td>
<td>$SS_{RES} = \sum_{i=1}^{N} (x - \hat{x})^2$</td>
</tr>
<tr>
<td>Uncorrected Total</td>
<td>$n - 1$</td>
<td>$SS_T = \sum_{i=1}^{N} (x - \bar{x})^2$</td>
</tr>
</tbody>
</table>

Table 3.2: Analysis-of-Variance table of the inverse estimator.

The Analysis-of-Variance table of the inverse estimator can be misleading. In the first place the $x$ values are fixed and these values of $x$ do not depend on the observed values of $Y$. The $SS_T$ represents the variability in these chosen $x$ values and this variability results from the way they are chosen. Secondly, the $SS_R$ measures the dependence of the $x$ on the $Y$, but this is a spurious dependence because $x$ does not depend on $Y$. Finally, the $SS_{RES}$ cannot be interpreted as a measure for the error of the $x$ values because the $x$ values are fixed and do not have an error. Eisenhart concluded that if the $x$ values are selected and the corresponding $Y$ values are observed, then the classical estimator is the only correct estimator.

Krutchkoff (1967) also compared the two estimators. He used the standard mean squared error $E(\hat{X} - X)^2$ criterion. He compared the mean squared errors for both methods using the Monte Carlo procedure. In his procedure he considers the following model, $Y = \alpha + \beta x + \varepsilon$ with $\alpha = 0$ and $\beta = 0.5$. The range of $x$ is $[0, 1]$ and the standard deviation of the error is 10% of the range, $\sigma = 0.1$. A design, where the researcher does observations on the end points of the range of $x$, is called an end point design. For example, if the range of $x$ is an interval $[a, b]$, then the end points are the points $x = a$ and $x = b$. On each of these points the researcher does $n$ observations. In the Monte Carlo procedure is chosen for the end point design with three observations at each end point $(x = 0, x = 1)$. With these six observations the parameters $\beta_0$ and $\beta_1$ of (1.1) and $\alpha_0$ and $\alpha_1$ of (1.3) are estimated. The two calibration equations are also obtained, (1.2) and (1.4). With these equations, squared errors can be
found for $x = 0, 0.2, 0.4, 0.6, 0.8, 1$. Also the values $x = 1.2, 2, 5, 10$ are added to the model. The experiment is repeated 10,000 times for each value of $x$. The results were tabulated and Krutchkoff concluded that the inverse approach has a uniformly smaller mean squared error than the classical approach. He has also investigated that the results did not depend on the choice of the parameters: effect of intercept, effect of slope, effect of error variance, effect of number of observations at each design point, effect of design, effect of non-normal errors and effect of quadratic term. In Chapter 6 we give a simulation to obtain the confidence coefficient of a confidence interval of the classical estimator. A few years later Krutchkoff (1969) reconsidered the comparison of the classical method with the inverse method. In a previous article he had given results for extrapolation, but the results for $x = 5$ and $x = 10$ were incorrectly labeled. He extended the results to the points $x = 3, 4, 5, 6, 7, 8, 9, 10$. He concluded that there exist situations in extrapolation in which the classical method is better than the inverse method.

Williams (1969) criticized the papers of Krutchkoff, because Krutchkoff compared the MSE’s of two estimators, one which has finite MSE (inverse method) and one of which has infinite MSE (classical method). He also established the fact that no unbiased estimator has finite variance. He also concluded that the minimum variance or minimum MSE criterion is not suitable in the comparison of the classical method with the inverse method.

Halperin (1970) suggested another approach for comparing the two estimators. He used three criteria to compare the estimators:

- The relative “closeness” of the classical estimator ($\hat{x}_0$) and the inverse estimator ($\tilde{x}_0$) to $x_0$: $\tilde{x}_0$ is a closer estimator than $\hat{x}_0$ if, for all $x_0$,

$$Pr\{|\tilde{x}_0 - x_0| < |\hat{x}_0 - x_0|\} > \frac{1}{2}.$$  \hspace{1cm} (3.1)

- Consistency, which is only relevant if the experiment is based on large samples: $\hat{x}_0$ and $\tilde{x}_0$ converge in probability to $x_0$: $\hat{x}_0$ and $\tilde{x}_0$ are consistent estimators for $x_0$ if for every $\varepsilon > 0$,

$$\lim Pr\{|\hat{x}_0 - x_0| < \varepsilon\} = 1 \text{ and } \lim Pr\{|\tilde{x}_0 - x_0| < \varepsilon\} = 1,$$  \hspace{1cm} (3.2)

- MSE of relevant asymptotic distributions, which are finite for both estimators. The method with the smallest asymptotic MSE is the best method. This criterion to compare estimators is less meaningful than the “closeness” criterion, because this takes not into account the correlation between the estimators.

He concluded that the inverse estimator is superior if the unknown independent variable $x_0$ lies in a small interval around the mean $\bar{x}$, $|\bar{x} - x_0| < \sqrt{\frac{1}{k(\bar{x})} + 2M\sigma^2_x}$, where $\sigma^2_x = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $M$ is a variable on which the inverse estimator depends. In other cases the classical estimator seems to be superior.

Berkson (1969) considered the asymptotic MSE for both estimators and he concluded that if $x_0$ is estimated from a mean of $n > 1$ observations of $Y_0$, then there always exist an $n_0$ such that for all values of $n \geq n_0$ the asymptotic MSE of the classical estimator will be smaller than the asymptotic MSE of the inverse estimator except for $x_0 = \bar{x}$. He also prefers the classical estimator because it is a consistent estimator.
Shukla (1972) also compared the two methods. If the number of observations is small, then the inverse estimator has smaller MSE for interpolation. But when a large number of observations is used, then the classical estimator is more advantageous than the inverse estimator except in very trivial cases. If one can not take more than one observation on unknown \(x_0\) then the inverse estimator is preferable when \(x_0\) lies close to the mean of the design points. However, if one has a large sample and no prior information about the unknown \(x_0\) then a consistent estimator (classical estimator) is to be preferred.

Graybill (1976) proposed the classical method to obtain an estimator for the unknown value of \(x_0\). His technique is based on a sample of size \(n + k\)

\[
(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n), (x_0, Y_{n+1}), (x_0, Y_{n+2}), \ldots, (x_0, Y_{n+k}),
\]

(3.3)

where \(x_0\) is unknown and thus an additional parameter. With the first \(n\) observations the unknown parameters \(\beta_0\) and \(\beta_1\) are estimated by \(\hat{\beta}_0\) and \(\hat{\beta}_1\), resp. Graybill proposes the following estimator of \(x_0\)

\[
\hat{x}_0 = \frac{Y_0 - \hat{\alpha}}{\hat{\beta}}
\]

(3.4)

where \(\bar{Y}_0 = \frac{1}{k} \sum_{i=n+1}^{n+k} Y_i\). Afterwards, he gives also an estimator for \(\sigma^2\). In Chapter 4 a variation on this technique is explained.

Brown (1993) followed in his theory Eisenhart, one should minimize the errors in the direction they occur. If the \(x\) values are taken from a random sample, then one should use the inverse estimator, which is called natural or random calibration. If, on the other hand, the errors occur in the \(Y\)-direction and the \(x\) values are fixed, then one should use the classical estimator, which is called controlled calibration. Brown also considers the profile likelihood for the controlled calibration model with \(k\) observations on the unknown \(x_0\). With this method we can obtain the Maximum Likelihood estimators for the 4 parameters, which is the main difference with Graybill’s method because Graybill’s method obtains Least Squares estimators for the intercept, slope and the unknown \(x_0\). The profile likelihood is the likelihood over the four parameters \((\mu_0, \gamma_1, \sigma^2, x_0)\). If one maximizes the profile log likelihood then one can find an Maximum Likelihood estimator for \(x_0\).

Naszódi (1978) considered a bias correction, because the classical estimator of \(x_0\) is unbiased. Afterwards, he compared his corrected estimator with the classical estimator. He observed that the corrected, unbiased estimator is more efficient than the classical estimator and is also consistent. In Chapter 4 we give the derivation of the unbiased estimator of Naszódi.

Most statisticians prefer the classical method, because of its consistency and its MSE of relevant asymptotic distributions. Another theory is that one should minimize the errors in the direction they occur, which is the \(Y\)-direction. Thus, one should use the classical method. But the inverse estimator seems to be superior if the unknown \(x_0\) lies in a small interval around the mean \(\bar{x}\).

### 3.2 Confidence and Discrimination Intervals for an Unknown Predictor

Graybill (1976) also proposed a technique to obtain a confidence interval for the classical estimator of \(x_0\). He uses the standard \(t\)-test to conclude whether \(\beta_1\) is zero or not. If the test
is not rejected one could assume that $\beta_1$ is zero and that there does not exist a confidence interval. In the other case one can find a $100(1-\alpha)\%$ confidence interval. But he claimed that the confidence coefficient of this interval is less than $100(1-\alpha)\%$. In Chapter 4 a variation on this technique is explained and in Chapter 6 a simulation of the confidence coefficient is given to study to which extent Graybill’s claim is true.

Lieberman et al. (1967) considered unlimited simultaneous discrimination intervals in calibration. They used more than one unknown $x_0$ to obtain the intervals. Unlimited simultaneous discrimination intervals are based upon the classical estimated linear regression and they have the property that at least $100P$ percent of the discrimination intervals will contain the true $x$’s with confidence $1-\alpha$. They give 2 methods, Bonferroni’s method and a technique based upon a idea of Lieberman and Miller, Jr. (1963), which can be found, to obtain unlimited simultaneous discrimination intervals.

Scheffé (1973) computed also a method to obtain unlimited simultaneous discrimination intervals. He had a general technique for every kind of regression model: univariate and multivariate, linear and nonlinear. We will give this technique for the simple linear regression model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2),$$

where the parameters $\beta_0$ and $\beta_1$ are estimated by Least Squares on $(x_1,Y_1)$, $(x_2,Y_2)$, ..., $(x_n,Y_n)$ as $\hat{\beta}_0$ and $\hat{\beta}_1$. For a given value of $x$, say $x_0$ (unknown), a $Y$ value, say $Y_0$ is observed at the $x_0$ value. Afterwards, $x_0$ is estimated by the classical approach. Let $x^{(1)} = \min_i x_i$ and $x^{(2)} = \max_i x_i$, $i = 1, \ldots, n$. Scheffé called the closed interval $T = [x^{(1)}, x^{(2)}]$ the calibration interval. We have to obtain the calibration curve, $x = \frac{Y - \hat{\beta}_0}{\hat{\beta}_1}$, and the upper and lower calibration curves,

$$x_{\text{upper}} = \bar{x} + C^{-1} \left( \hat{\beta}_1 D_1 + \hat{\sigma} c_2 \left( n^{-1} C + \frac{D_1^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right)^{\frac{3}{2}},$$

$$x_{\text{lower}} = \bar{x} + C^{-1} \left( \hat{\beta}_1 D_2 - \hat{\sigma} c_2 \left( n^{-1} C + \frac{D_2^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right)^{\frac{3}{2}},$$

where

$$C = \hat{\beta}_1^2 - \frac{\hat{\sigma}^2 c_2^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$D_h = Y - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} - (-1)^h \hat{\sigma} c_1, \quad h = 1, 2,$$

$$u = \hat{\beta}_0 + \hat{\beta}_1 x - (-1)^h \hat{\sigma} \left( c_1 + c_2 \left( n^{-1} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right), \quad h = 1, 2,$$

$$c_1 = c Z_\alpha \nu^2 (\chi_{1-\delta,\nu}^2)^{-\frac{1}{2}},$$

$$c_2 = c (2F_{\delta,2,\nu}),$$

where $c$ can be found in tables and these tables can be found in Scheffé (1973) and $Z_\alpha$ is the two-tailed $\alpha$-point from the standard normal distribution. Scheffé also calculated three other
calibration intervals: \([Y^{(1)}, Y^{(2)}], [Y^{01}, Y^{I2}]\) and \([Y^{I1}, Y^{02}]\), where

\[
Y^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x^{(i)}, \quad i = 1, 2, 
\]

\[
Y^{I1} = \hat{\beta}_0 + \hat{\beta}_1 x^{(1)} + \hat{\sigma} \left( c_1 + c_2 \left( n^{-1} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)^{\frac{1}{2}} \right),
\]

\[
Y^{I2} = \hat{\beta}_0 + \hat{\beta}_1 x^{(2)} - \hat{\sigma} \left( c_1 + c_2 \left( n^{-1} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)^{\frac{1}{2}} \right),
\]

\[
Y^{01} = \hat{\beta}_0 + \hat{\beta}_1 x^{(1)} - \hat{\sigma} \left( c_1 + c_2 \left( n^{-1} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)^{\frac{1}{2}} \right),
\]

\[
Y^{02} = \hat{\beta}_0 + \hat{\beta}_1 x^{(2)} + \hat{\sigma} \left( c_1 + c_2 \left( n^{-1} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)^{\frac{1}{2}} \right).
\]

For \(Y^{(1)} \leq Y_0 \leq Y^{(2)}\), the point estimator of \(x_0\) is \(\hat{x}_0 = \frac{Y_0 - \hat{\beta}_0}{\hat{\beta}_1}\). For \(Y^{01} \leq Y_0 \leq Y^{I2}\), the upper endpoint of the interval estimator for \(x_0\) can be found by replacing \(Y\) by \(Y_0\) in (3.6) and for \(Y^{I1} \leq Y_0 \leq Y^{02}\), the lower endpoint of the interval for \(x_0\) can be found by replacing \(Y\) by \(Y_0\) in (3.7). All these points and intervals can be seen in Figure 3.2. The computation of these formulas can be found in Scheffé (1973).

For \(Y^{(1)} \leq Y_0 \leq Y^{(2)}\), the point estimator of \(x_0\) is \(\hat{x}_0 = \frac{Y_0 - \hat{\beta}_0}{\hat{\beta}_1}\). For \(Y^{01} \leq Y_0 \leq Y^{I2}\), the upper endpoint of the interval estimator for \(x_0\) can be found by replacing \(Y\) by \(Y_0\) in (3.6) and for \(Y^{I1} \leq Y_0 \leq Y^{02}\), the lower endpoint of the interval for \(x_0\) can be found by replacing \(Y\) by \(Y_0\) in (3.7). All these points and intervals can be seen in Figure 3.2. The computation of these formulas can be found in Scheffé (1973).

\[\]

Figure 3.2: Calibration chart.
3.3 Bayesian Solution

The model used by Graybill has some undesirable properties, for example, the classical estimator $\hat{x}_0$ has infinite MSE and for the $(1-\alpha)100\%$ confidence interval $S \Pr\{S = (-\infty, +\infty)\} > 0$. Because of these disadvantages, Hoadley (1970) used an other technique, a Bayesian solution.

Perng and Tong (1974) propose a two-stage sequential procedure for the construction of a fixed-width confidence interval based on the Bayesian method.

Kalotay (1971) gives also a alternative method, a structural technique, and he compared the results with a Bayesian solution.

3.4 Optimum Design

Ott and Myers (1968) considered an appropriate design for the classical approach, which is related to controlled calibration. They have shown that for the classical estimator the end-point design is optimal according to the $D$-criterium. The end-point design is the following: if the interval $(a \leq x \leq b)$ is the most interesting, then $\frac{n}{2}$ observations are done on $x = a$ and $\frac{n}{2}$ observations are done on $x = b$.

Shukla (1972) considered optimum designs for the classical and the inverse method. He concluded that for the end-point design the MSE of the inverse estimator is always smaller than the MSE of the classical estimator for interpolation when we have only one observation $Y_0$ on the unknown $x_0$ and $n \geq 8$.

3.5 Multivariate Calibration

Martens and Næs (1989) looked at the calibration problem from a chemical point of view. They consider many different calibration models, univariate against multivariate and linear against nonlinear. A regression model can contain more than one regressor variable, such that it becomes multivariate. A model is nonlinear, if the coefficients $\beta_i$ does not occur linear in the formula. They also looked at the classical and inverse estimators.
Chapter 4

Graybill’s Method

In this chapter we present a slight variation on Graybill’s solution to obtain an estimator for the unknown \(x_0\) and a confidence interval for \(x_0\). Graybill’s model is the following

\[
Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)
\]  

(4.1)

and can be found in Graybill (1976). Our model is the centred simple linear regression model, for which the regressor variable \(x\) is redefined as the deviation from its own average, \(x - \bar{x}\), such that the \(\text{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = 0\), which we will prove in (4.22). Thus, the model becomes

\[
Y = \gamma_0 + \gamma_1 (x - \bar{x}) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2),
\]  

(4.2)

where \(\gamma_0\) and \(\gamma_1\) are parameters. It is easy to show that the models are equivalent by setting \(\gamma_1 = \beta_1\) and \(\gamma_0 - \gamma_1 \bar{x} = \beta_0\).

(4.3)

To obtain an estimator of \(x_0\) with confidence bounds we observe \(k > 1\) values of \(Y\) at an unknown \(x_0\). We have a sample of size \(n + k\) denoted by

\[
(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n), (x_0, Y_{n+1}), (x_0, Y_{n+2}), \ldots, (x_0, Y_{n+k}),
\]  

(4.4)

where \(x_0\) is unknown and at least two of the controlled variables \(x_i\) have to be distinct. The \(k\) values of \(Y\), denoted by \(Y_{n+1}, Y_{n+2}, \ldots, Y_{n+k}\), come from a normal distribution with mean \(\gamma_0 + \gamma_1 (x_0 - \bar{x})\) and variance \(\sigma^2\), where \(x_0, \gamma_0, \gamma_1\) and \(\sigma^2\) are unknown.

4.1 Estimation of \(\gamma_0, \gamma_1\) and \(x_0\)

In this section, we estimate the parameters \(\gamma_0\) and \(\gamma_1\) of (4.2) and also the unknown \(x_0\). We also investigate whether the classical estimator of \(x_0\) is unbiased.

First of all, we need an important theorem, the Invariance Principle of Maximum Likelihood estimators, which can be found in Zehna (1966):

**Theorem 4.1 (Invariance Principle for Maximum Likelihood Estimators)** Let \(P\) be a distribution depending on parameters \(\theta_1, \ldots, \theta_k\) and let \(\hat{\Theta} = (\hat{\Theta}_1, \ldots, \hat{\Theta}_\ell)\) be a Maximum Likelihood estimator of \((\theta_1, \ldots, \theta_\ell)\). If \(\tau\) is an arbitrary function with domain \(\Theta \subset \mathbb{R}^\ell\), then \(\tau(\hat{\Theta})\) is a Maximum Likelihood estimator of \(\tau((\theta_1, \ldots, \theta_\ell))\). If moreover the Maximum Likelihood estimator \(\hat{\Theta}\) is unique, then \(\tau(\hat{\Theta})\) is unique too.
It is easy to show that the Least Squares estimators of the parameters $\beta_0$ and $\beta_1$ of model (1.1) are also the Maximum Likelihood estimators of these parameters. From (4.3) we know that $\gamma_0$ and $\gamma_1$ are linear transformation of $\beta_0$ and $\beta_1$, such that the estimators of $\gamma_0$ and $\gamma_1$ are the Maximum Likelihood estimators of $\gamma_0$ and $\gamma_1$, because of Theorem 4.1. The Maximum Likelihood estimators for $\gamma_0$ and $\gamma_1$, based on the first $n$ values of the data, $(x_1, Y_1)$, $(x_2, Y_2)$, \ldots, $(x_n, Y_n)$, are

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} Y_i(x_i - \bar{x}) - \bar{Y} \sum_{i=1}^{n} (x_i - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} Y_i(x_i - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \frac{\sum_{i=1}^{n} x_i - n \bar{x}}{\sum_{j=1}^{n} (x_j - \bar{x})^2}, \quad (4.5)$$

$$\hat{\gamma}_0 = \hat{\beta}_0 + \hat{\gamma}_1 \bar{x}$$

$$= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$= \bar{Y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}$$

$$= \bar{Y}, \quad (4.6)$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. Note that normality of $\varepsilon$ in (4.1) and (4.2) implies that $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are normally distributed too. The classical estimator of $x_0$, based on the first $n + k$ values of the data, $(x_1, Y_1)$, $(x_2, Y_2)$, \ldots, $(x_n, Y_n)$, $(x_0, Y_{n+1})$, $(x_0, Y_{n+2})$, \ldots, $(x_0, Y_{n+k})$

$$\hat{x}_0 = \bar{x} + \frac{Y_0 - \hat{\gamma}_0}{\hat{\gamma}_1}$$

$$= \bar{x} + \frac{Y_0 - \bar{Y}}{\hat{\gamma}_1}, \quad (4.7)$$

where $Y_0 = \frac{1}{k} \sum_{i=n+1}^{n+k} Y_i$.

To investigate whether the estimator of $x_0$ is unbiased or not we have to obtain the expectation of the estimator. We follow the method of Naszódi (1978). Naszódi expands $\hat{x}_0$ into a Taylor series about the values $\gamma_0$, $\gamma_1$ and $Y_0$. The stochastic Taylor expansion of third order of $Y = f(X_1, X_2, X_3)$ round $\mu = (\mu_1, \mu_2, \mu_3) = (E(X_1), E(X_2), E(X_3))$, which can be found in Casella and Berger (2002), is:

$$Y = f(\mu) + (X_1 - \mu_1)f_1'(\mu) + (X_2 - \mu_2)f_2'(\mu) + (X_3 - \mu_3)f_3'(\mu)$$

$$+ \frac{1}{2} ((X_1 - \mu_1)^2 f_1''(\mu) + (X_2 - \mu_2)^2 f_2''(\mu) + (X_3 - \mu_3)^2 f_3''(\mu))$$

$$+ (X_1 - \mu_1)(X_2 - \mu_2)f_{12}'(\mu) + (X_1 - \mu_1)(X_3 - \mu_3)f_{13}'(\mu) + (X_2 - \mu_2)(X_3 - \mu_3)f_{23}'(\mu)$$

$$+ \frac{1}{3} ((X_1 - \mu_1)^3 f_{111}'(\mu) + (X_2 - \mu_2)^3 f_{222}'(\mu) + (X_3 - \mu_3)^3 f_{333}'(\mu))$$

$$+ (X_1 - \mu_1)^2 (X_2 - \mu_2)f_{112}'(\mu) + (X_1 - \mu_1)(X_2 - \mu_2)^2 f_{122}'(\mu)$$

$$+ (X_1 - \mu_1)^2 (X_3 - \mu_3)f_{113}'(\mu) + (X_1 - \mu_1)(X_3 - \mu_3)^2 f_{133}'(\mu)$$

$$+ (X_2 - \mu_2)^2 (X_3 - \mu_3)f_{223}'(\mu) + (X_2 - \mu_2)(X_3 - \mu_3)^2 f_{233}'(\mu). \quad (4.8)$$
The stochastic Taylor expansion of third order of $Y = \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1}$ round $(E(Y_0), E(\hat{\gamma}_0), E(\hat{\gamma}_1)) = (E(Y_0), \gamma_0, \gamma_1)$ is:

$$f(Y_0, \hat{\gamma}_0, \hat{\gamma}_1) = \bar{x} + \frac{E(Y_0) + \gamma_0 - E(Y_0) - \gamma_0}{\gamma_1} (\Delta \hat{\gamma}_1) - \frac{1}{\gamma_1} (\Delta \hat{\gamma}_0) + \frac{1}{\gamma_1} (\Delta \hat{Y}_0)$$

$$+ \frac{E(Y_0) - \gamma_0}{\gamma_1^3} (\Delta \hat{\gamma}_1)^2 - \frac{1}{\gamma_1^2} (\Delta \hat{\gamma}_1)(\Delta \hat{Y}_0) + \frac{1}{\gamma_1^2} (\Delta \hat{\gamma}_1)(\Delta \hat{\gamma}_0)$$

$$- \frac{E(Y_0) - \gamma_0}{\gamma_1^4} (\Delta \hat{\gamma}_1)^3 + \frac{1}{\gamma_1^3} (\Delta \hat{\gamma}_1)^2 (\Delta \hat{Y}_0) - \frac{1}{\gamma_1^3} (\Delta \hat{\gamma}_1)^2 (\Delta \hat{\gamma}_0),$$

(4.9)

where

$$\Delta \hat{\gamma}_1 = \hat{\gamma}_1 - E(\hat{\gamma}_1) = \hat{\gamma}_1 - \gamma_1,$$

$$\Delta \hat{\gamma}_0 = \hat{\gamma}_0 - E(\hat{\gamma}_0) = \hat{\gamma}_0 - \gamma_0,$$

$$\Delta Y_0 = Y_0 - E(Y_0).$$

(4.10)

We will now take expectations in (4.9). Because $\hat{\gamma}_1 - E(\hat{\gamma}_1), \hat{\gamma}_0 - E(\hat{\gamma}_0)$ and $\hat{Y}_0 - E(\hat{Y}_0)$ are symmetrically distributed round 0 with expectation 0, all their moments of odd order are 0. Hence, the second, third, fourth, sixth, seventh, eighth, ninth and tenth terms of (4.9) vanish. Thus, the expected value of (4.9) is approximately equal to

$$E(\hat{x}_0) \simeq E\left( \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} (\Delta \hat{\gamma}_1)^2 \right)$$

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} E((\Delta \hat{\gamma}_1)^2)$$

(by definition)

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} \text{Var}(\hat{\gamma}_1)$$

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} \text{Var}\left( \sum_{i=1}^{n} c_i Y_i \right), \text{ with } c_i = \frac{x_i - \bar{x}}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} \sum_{i=1}^{n} c_i^2 \text{Var}(Y_i)$$

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} \sum_{i=1}^{n} c_i^2 \sigma_i^2$$

$$= \bar{x} + \frac{Y_0 - \gamma_0}{\gamma_1} + \frac{Y_0 - \gamma_0}{\gamma_1^3} \left( \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

$$= \bar{x} + (x_0 - \bar{x}) \frac{m_{12}}{\gamma_1^2},$$

(4.11)

where

$$m_{12} = \text{Var}(\hat{\gamma}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

(4.12)
Hence, (4.7) gives an estimator for \( x_0 \) which is almost biased with
\[
T(x) = \frac{(x_0 - \bar{x}) m_{12}}{\gamma_1^2}
\]  
with approximation. The following is the estimator of the bias \( T(x) \):
\[
\hat{T}(x) = \frac{(x_0 - \bar{x}) m_{12}}{\gamma_1^2}.
\]  
The new, almost unbiased estimator of \( x_0 \) is
\[
\hat{x}_0 = \hat{x}_0 - \hat{T}(\hat{x}_0).
\]  
After rearranging we obtain
\[
\hat{x}_0 = \bar{x} + \frac{Y_0 - \hat{\gamma}_0}{\hat{\gamma}_1 + \frac{m_{12}}{\gamma_1^2}}.
\]  

Naszódi claims that his estimator \( \hat{x}_0 \) is more efficient than the classical estimator and is also consistent (definition in Chapter 2). It can also easily be derived from the classical estimator. Naszódi gives also a \( T \)-optimum design for this estimator. A definition of \( T \)-optimality can be found in Chapter 9 of Pukelsheim (1993). The optimum design is the end-point design.

**Example 4.1** In this example we will obtain Naszódi’s estimator of \( x_0 \) for the data from Example 2.1. We also take \( Y_0 = 200 \degree C \).
\[
\hat{x}_0 = 250 + \frac{200 - 231.5875}{0.95303 + \frac{0.8835}{0.95303}} = 216.8574 \degree C.
\]  
The classical estimator for this example for the centred model is
\[
\hat{x}_0 = 250 + \frac{200 - 231.5875}{0.95303} = 216.8557 \degree C.
\]  

Naszódi’s estimator is almost the same as the classical estimator, because \( m_{12} \) or \( \text{Var}(\hat{\gamma}_1) \) is very small for this example.

**4.2 Estimation of \( \sigma^2 \)**

We use the likelihood function to estimate the variance \( \sigma^2 \) based on all \( n + k \) data points. The full likelihood function is based on all the \( n + k \) values of the data. The likelihood function is
\[
L(\gamma_0, \gamma_1, \sigma^2 ; x_0, y_1, x_1, \ldots, y_n, x_n ; y_{n+1}, y_{n+2}, \ldots, y_{n+k})
\]  
\[
= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (Y_i - \gamma_0 - \gamma_1(x_i - \bar{x}))^2 \right) \prod_{i=n+1}^{n+k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (Y_i - \gamma_0 - \gamma_1(x_0 - \bar{x}))^2 \right)
\]  
\[
= \frac{1}{(2\pi\sigma^2)^{n+k}} \exp \left( -\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} (Y_i - \gamma_0 - \gamma_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \gamma_0 - \gamma_1(x_0 - \bar{x}))^2 \right) \right).
\]
4.2 Estimation of $\sigma^2$

Hence, the log likelihood equals

$$- \left( \frac{n+k}{2} \right) \log 2\pi - \left( \frac{n+k}{2} \right) \log \sigma^2 - \sum_{i=1}^{n} \frac{(Y_i - \gamma_0 - \gamma_1(x_i - \bar{x}))^2}{2\sigma^2} - \sum_{i=n+1}^{n+k} \frac{(Y_i - \gamma_0 - \gamma_1(x_0 - \bar{x}))^2}{2\sigma^2}. $$

Taking the partial derivative with respect to $\sigma^2$, setting it equal to zero, substituting the estimators of $\gamma_0$, $\gamma_1$ and $x_0$, (4.5), (4.6), (4.7), we obtain an estimator of $\sigma^2$. This is not the full Maximum Likelihood function, because $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are based on $x_1, \ldots, x_n$ only. For the full Maximum Likelihood method, see the next chapter. The partial derivative with respect to $\sigma^2$ is

$$\frac{\partial \log L}{\partial \sigma^2} \bigg|_{\gamma_0, \gamma_1, \hat{x}_0, \sigma^2} = - \frac{n+k}{2\sigma^2} + \frac{1}{2\sigma^4} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2 \right).$$

Setting (4.19) to zero, we obtain the following equation for $\hat{\sigma}^2$:

$$- \frac{n+k}{2\sigma^2} + \frac{1}{2\sigma^4} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2 \right) = 0.$$

Multiplication of both sides with $\hat{\sigma}^2$ and substitution of $\hat{x}_0$ by $\bar{x} + \frac{\Sigma_0 - \bar{Y}}{\hat{\gamma}_1}$ and $\hat{\gamma}_0$ by $\bar{Y}$ yields

$$-(n+k) + \frac{1}{\sigma^2} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y})^2 \right) = 0.$$

Simplifying and taking $-(n+k)$ to the other side, we obtain

$$n + k = \frac{1}{\sigma^2} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y})^2 \right).$$

Hence,

$$\hat{\sigma}_{ML}^2 = \frac{1}{n+k} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y})^2 \right).$$

(4.20)
We now compute the expectation of $\hat{\sigma}^2_{ML}$, since ML estimators need not be unbiased. First of all, we have to calculate some covariances:

\[
\text{Cov}(Y_i, \hat{\gamma}_0) = \text{Cov}(Y_i, \bar{Y}) = \text{Cov}(Y_i, \frac{1}{n} \sum_{j=1}^{n} Y_j) = \frac{1}{n} \text{Var}(Y_i) = \frac{\sigma^2}{n}, \quad (4.21)
\]

\[
\text{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = \text{Cov} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} c_j Y_j \right), \text{ with } c_j = \frac{x_j - \bar{x}}{\sum_{l=1}^{n} (x_l - \bar{x})^2}
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{n} c_j \text{Cov}(Y_i, Y_j) = \sum_{i=1}^{n} \frac{1}{n} c_i \text{Var}(Y_i) = \frac{\sigma^2}{n} \sum_{i=1}^{n} c_i = 0, \quad (4.22)
\]

\[
\text{Cov}(Y_i, \hat{\gamma}_1) = \text{Cov}(Y_i, \sum_{j=1}^{n} c_j Y_j) = \sum_{j=1}^{n} c_j \text{Cov}(Y_i, Y_j) = c_i \text{Var}(Y_i) = \frac{\sigma^2 c_i}{\sum_{j=1}^{n} (x_j - \bar{x})^2}, \quad (4.23)
\]

\[
\text{Cov}(Y_i, \bar{Y}_0) = \text{Cov} \left( Y_i, \frac{1}{k} \sum_{i=n+1}^{n+k} Y_i \right) = \frac{1}{k} \text{Var}(Y_i) = \frac{\sigma^2}{k}. \quad (4.24)
\]

The expectation of $\hat{\sigma}^2_{ML}$ is

\[
E \left( \frac{1}{n+k} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y}_0)^2 \right) \right).
\]

We have a linear operation for the expectation: $E(aX + bY) = aE(X) + bE(Y)$. Hence,

\[
E(\hat{\sigma}^2_{ML}) = \frac{1}{n+k} \left( \sum_{i=1}^{n} E(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} E(Y_i - \bar{Y}_0)^2 \right).
\]

Applying the well-known property: $E(X^2) = \text{Var}(X) + (E(X))^2$, we rewrite this as

\[
\frac{1}{n+k} \left( \sum_{i=1}^{n} \text{Var}(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \bar{x})) + (E(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \bar{x})))^2 \right)
\]

\[
+ \frac{1}{n+k} \left( \sum_{i=n+1}^{n+k} \text{Var}(Y_i - \bar{Y}_0) + (E(Y_i - \bar{Y}_0))^2 \right).
\]

Using the addition rule for variances: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$, we further deduce that

\[
E(\hat{\sigma}^2_{ML}) = \frac{1}{n+k} \sum_{i=1}^{n} (\text{Var}(Y_i) + \text{Var}(\hat{\gamma}_0) + (x_i - \bar{x})^2 \text{Var}(\hat{\gamma}_1))
\]
4.2 Estimation of $\sigma^2$

$$+ \frac{1}{n+k} \sum_{i=1}^{n+k} (-2\text{Cov}(Y_i, \gamma_0) + 2(x_i - \bar{x})\text{Cov}(\gamma_0, \gamma_1) - 2(x_i - \bar{x})\text{Cov}(Y_i, \gamma_1))$$

$$+ \frac{1}{n+k} \sum_{i=1}^{n} \left( E(Y_i) - E(\gamma_0) - (x_i - \bar{x})E(\gamma_1) \right)^2$$

$$+ \frac{1}{n+k} \sum_{i=n+1}^{n+k} (\text{Var}(Y_i) + \text{Var}(\bar{Y}_0) - 2\text{Cov}(Y_i, \bar{Y}_0) + (E(Y_i) - E(\bar{Y}_0))^2).$$

We now substitute the expectations, variances and covariances (4.21, 4.22, 4.23 and 4.24).

The expectation of $\sigma^2_{ML}$ then becomes

$$\frac{1}{n+k} \left( \sum_{i=1}^{n} \left( \frac{\sigma^2}{n} + \frac{\sigma^2(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} - \frac{2\sigma^2}{n} \sum_{j=1}^{n}(x_i - \bar{x})^2 \right) \right) + \sum_{i=n+1}^{n+k} \left( \frac{\sigma^2}{n} \right).$$

After simplifying the expression, we obtain

$$\frac{1}{n+k} \left( \sum_{i=1}^{n} \left( \frac{\sigma^2}{n} - \frac{\sigma^2(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_i - \bar{x})^2} \right) + \sum_{i=n+1}^{n+k} \left( \frac{\sigma^2}{n} \right) \right) = \frac{1}{n+k} \left( n\sigma^2 - \sigma^2 - \sigma^2 + k\sigma^2 - \sigma^2 \right).$$

Hence,

$$E(\sigma^2_{ML}) = \frac{n+k-3}{n+k} \sigma^2.$$

Since the bias is of multiplicative nature, we can easily correct $\sigma^2_{ML}$. So, $\sigma^2_{ML}$ corrected for bias is

$$\hat{\sigma}^2 = \frac{1}{n+k-3} \left( \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y}_0)^2 \right). \quad (4.25)$$

**Remark 4.1** If $k = 1$ then the estimator of $\sigma^2$ corrected for bias reduces to

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2. \quad (4.26)$$

We need a few theorems before we can derive a confidence interval for $x_0$.

**Theorem 4.2** Let $(x_i, Y_i)$, $i=1,2,\ldots,n$, be related by the simple linear model $Y_i = \gamma_0 + \gamma_1(x_i - \bar{x}) + \epsilon_i$, $\epsilon_i$ are distributed $\text{NID}(0, \sigma^2)$. Further let $Y_{n+i}$, $i=1,2,\ldots,k$, $(k \geq 1)$ be a random sample from a normal distribution with mean $\gamma_0 + \gamma_1(x_0 - \bar{x})$ ($x_0$ unknown) and variance $\sigma^2$. Let all random variables $Y_i$ be independent. The estimators of $\gamma_1$, $\gamma_0$, $x_0$, $\sigma^2$ and also $\sigma^2$ corrected for bias are given in (4.5), (4.6), (4.7), (4.20) and (4.25).

**Theorem 4.3** If $Y_1, \ldots, Y_n$ are independent and normally distributed, then

$$\begin{pmatrix}
\sum_{i=1}^{n} \alpha_1 Y_i \\
\sum_{i=1}^{n} \alpha_2 Y_i \\
\vdots \\
\sum_{i=1}^{n} \alpha_p Y_i
\end{pmatrix} \quad (4.27)$$

has a $p$-dimensional normal distribution.
Proof The components $\sum_{i=1}^{n} \alpha_{ij}Y_i, j = 1, \ldots, p$, of the vector are normally distributed, because the $Y_i$ are normally distributed. Hence, all linear combinations of the components are normally distributed. By the theorem of Cramér-Wold (see e.g. Theorem 2.6.2 in Anderson (1984)), the vector

$$
\begin{pmatrix}
\sum_{i=1}^{n} \alpha_{i1}Y_i \\
\sum_{i=1}^{n} \alpha_{i2}Y_i \\
\vdots \\
\sum_{i=1}^{n} \alpha_{ip}Y_i
\end{pmatrix}
\tag{4.28}
$$

has a $p$-dimensional normal distribution $\blacksquare$

We now derive some distributions.

**Theorem 4.4** Let $X \sim \chi^2_n$ and $Y \sim \chi^2_p$ and $Y$ and $X$ are independent, then $X + Y \sim \chi^2_{n+p}$.

**Proof**

$$M_{X+Y}(t) = (1-2t)^{-\frac{n}{2}} \cdot (1-2t)^{-\frac{p}{2}} = (1-2t)^{-\frac{n+p}{2}},$$

which is the Moment Generating Function (MGF) of $\chi^2_{n+p}$. $\blacksquare$

**Theorem 4.5** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then

1. $U_0$ is distributed $\chi^2_n$, where $U_0 = \sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2}$;

2. $U_1$ is distributed $\chi^2_{n-1}$, where $U_1 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2}$.

**Proof** (1) Since $Y_i$ is normally distributed with mean $\mu$ and variance $\sigma^2$, $\frac{Y_i - \mu}{\sigma}$ is normally distributed with mean 0 and variance 1. By Theorem 8.5.3 in Bain and Engelhardt (1992),

$$\sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2} \sim \chi^2_n. \tag{4.29}$$

(2)

$$V_1 = \sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{Y} - \mu)^2}{\sigma^2} = V_2 + V_3, \tag{4.30}$$

where

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}. \tag{4.31}$$

By Theorem 8.5.3 in Bain and Engelhardt (1992), $\sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2} \sim \chi^2_n$ and $\frac{n(\bar{Y} - \mu)^2}{\sigma^2} \sim \chi^2_1$ and they are independent. So,

$$M_{V_2}(t) = \frac{M_{V_1}(t)}{M_{V_3}(t)} = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{n-1}{2}}, \tag{4.32}$$
which is the Moment Generating Function of $\chi^2_{n-1}$. Thus,

$$ \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y})^2}{\sigma^2} \sim \chi^2_{n-1}. \quad (4.33) $$

\[ \square \]

**Theorem 4.6** Let $Y = X'\gamma + \varepsilon$, where $\varepsilon$ is normally distributed with mean 0 and variance $\sigma^2$, be a linear model with $Y$ a $n \times 1$ observable vector of random variables, $X$ an $n \times p (n > p)$ of known fixed numbers and $X$ has full rank, $\gamma$ an $p \times 1$ vector of unknown parameters, $\varepsilon$ an $n \times 1$ unobservable vector of random variables. Then this result follows:

1. $(n-p)\hat{\sigma}^2 = U$ is distributed $\chi^2_{n-p}$.

**Proof**

$$(n-p)\hat{\sigma}^2 = SS_{Res} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= (Y - \bar{Y})' (Y - \bar{Y})$$

$$= [Y - X(X'X)^{-1}X'Y]'[Y - X(X'X)^{-1}X'Y]$$

$$= Y'[I - X(X'X)^{-1}X']Y. \quad (4.34)$$

We will now show that $[I - X(X'X)^{-1}X']$ is symmetric idempotent:

$$[I - X(X'X)^{-1}X'] [I - X(X'X)^{-1}X'] = I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X'$$

$$= I - X(X'X)^{-1}X'$$

$$= I - X(X'X)^{-1}X' \quad (4.35)$$

and

$$[I - X(X'X)^{-1}X']' = I' - (X(X'X)^{-1}X')' = I - X(X'X)^{-1}X'$$

$$= I - X(X'X)^{-1}X'. \quad (4.36)$$

If $Y \sim N(\mu, \sigma^2 I)$ and $A$ is idempotent with rank $q$, then

$$\frac{Y'A Y}{\sigma^2} \sim \chi^2_{q, \lambda}, \quad (4.37)$$

where $\lambda = \frac{\mu' A \mu}{\sigma^2}$ (noncentrality parameter) (see e.g. Appendix C in Montgomery et al. (2001)).

Now we have $Y \sim N(\mu, \sigma^2 I)$ and $[I - X(X'X)^{-1}X']$ is idempotent. Hence,

$$\frac{SS_{Res}}{\sigma^2} = \frac{1}{\sigma^2} Y'[I - X(X'X)^{-1}X']Y \quad (4.38)$$

follows a $\chi^2$ distribution. The degrees of freedom follow from the rank of $[I - X(X'X)^{-1}X']$, which is the trace of $[I - X(X'X)^{-1}X']$, because the rank of an idempotent matrix is the trace (see e.g. Appendix C in Montgomery et al. (2001)). The trace is

$$\text{trace}([I - X(X'X)^{-1}X']) = \text{trace}(I_n) - \text{trace}([X(X'X)^{-1}X'])$$

$$= n - \text{trace}(X'X(X'X)^{-1})$$

$$= n - \text{trace}(I_p) = n - p. \quad (4.39)$$
We also have $E(Y) = \mu = X\gamma$. Hence, the noncentrality parameter is
\[ \frac{1}{\sigma^2} \mu' [I - X'(X'X)^{-1} X'] \mu = \frac{1}{\sigma^2} \gamma' X' [I - X(X'X)^{-1} X'] X \gamma \]
\[ = \frac{1}{\sigma^2} \gamma' [X'X - X'X(X'X)^{-1} X'] \gamma \]
\[ = \frac{1}{\sigma^2} \gamma' [X'X - X'X] \gamma = 0. \tag{4.40} \]
Thus,
\[ \frac{SS_{Res}}{\sigma^2} = (n - p) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}. \tag{4.41} \]

We shall now prove the following theorem about the distribut ional properties of the esti-
mators in (4.5), (4.6), (4.7) and (4.25).

**Theorem 4.7** Consider the model given in Theorem 4.2 and the estimators given in (4.5), (4.6), (4.7) and (4.25).

1. $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are independent.

2. $\hat{\gamma}_0 \sim N\left(\gamma_0, \frac{\sigma^2}{n}\right)$ and $\hat{\gamma}_1 \sim N\left(\gamma_1, \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)$.

3. $U = (n + k - 3) \frac{\hat{\sigma}^2}{\sigma^2}$ is distributed as $\chi^2_{n+k-3}$ (so $E(\hat{\sigma}^2) = 0$).

4. $\hat{\sigma}^2$ is independent of $\left(\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0\right)$.

**Proof** (1) In this part we will prove the independence of $\hat{\gamma}_0$ and $\hat{\gamma}_1$. First, we prove that both $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are normally distributed. Both $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are sums of normally distributed terms ($\hat{\gamma}_0 = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} Y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sum_{i=1}^{n} c_i Y_i$), so $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are normally distributed. By Theorem 4.3 the joint distribution of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ is a 2-dimensional normal distribution. Secondly, we prove that they are uncorrelated:
\[
\text{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = \text{Cov} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} c_j Y_j \right), \text{ with } c_j = \frac{x_j - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \]
\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{n} c_j \text{Cov}(Y_i, Y_j) = \sum_{i=1}^{n} \frac{1}{n} c_i \text{Var}(Y_i) = \frac{\sigma^2}{n} \sum_{i=1}^{n} c_i = 0. \]

Thus, we can conclude that $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are uncorrelated. Since $\hat{\gamma}_0$ and $\hat{\gamma}_1$ have a joint normal distribution, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are independent.
(2) From part (1) we know that \( \hat{\gamma}_0 \) and \( \hat{\gamma}_1 \) are normally distributed. We now calculate the mean and variance for both \( \hat{\gamma}_0 \) and \( \hat{\gamma}_1 \). The mean of \( \hat{\gamma}_0 \) is

\[
E(\hat{\gamma}_0) = E(\bar{Y}) = E\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(Y_i) = 1/n \sum_{i=1}^{n} \gamma_0 = \gamma_0.
\]

The variance of \( \hat{\gamma}_0 \) is

\[
\text{Var}(\hat{\gamma}_0) = \text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) = \frac{\sigma^2}{n}.
\]

The mean of \( \hat{\gamma}_1 \) is

\[
E(\hat{\gamma}_1) = E\left(\frac{\sum_{i=1}^{n} c_i Y_i}{\sum_{j=1}^{n} (x_j - \bar{x})^2}\right) = E\left(\sum_{i=1}^{n} c_i Y_i\right), \quad \text{with} \quad c_i = \frac{x_i - \bar{x}}{\sum_{j=1}^{n} (x_j - \bar{x})^2}
\]

\[
= \sum_{i=1}^{n} c_i E(Y_i) = \sum_{i=1}^{n} c_i (\gamma_0 + \gamma_1 (x_i - \bar{x})) = \gamma_1 \sum_{i=1}^{n} c_i (x_i - \bar{x}) = \gamma_1.
\]

From (4.12) we know that

\[
\text{Var}(\hat{\gamma}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.
\]

(3) We will denote \( S_1^2 \) and \( S_2^2 \), respectively, by

\[
S_1^2 = \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 (x_i - \bar{x}))^2 \quad \text{and} \quad S_2^2 = \sum_{i=n+1}^{n+k} (Y_i - \bar{Y}_0)^2.
\]

Hence, \( \hat{\sigma}^2 = \frac{1}{n+k-3}(S_1^2 + S_2^2) \). It follows from Theorem 4.6 that \( U_1 = \frac{S_1^2}{\sigma^2} \) is distributed as \( \chi^2_{n-2} \) and from Theorem 4.5 that \( U_2 = \frac{S_2^2}{\sigma^2} \) is distributed as \( \chi^2_{k-1} \). But \( U_1 \) depends only on \( Y_1, \ldots, Y_n \), and \( U_2 \) depends only on \( Y_{n+1}, \ldots, Y_{n+k} \). Since the \( Y_i \) are mutually independent, it follows that \( U_1 \) and \( U_2 \) are independent, and hence \( S_1^2 \) and \( S_2^2 \) are independent. Thus by Theorem 4.4, \( U_1 + U_2 \) is distributed as \( \chi^2_{n+k-3} \). But \( U_1 + U_2 = \frac{S_1^2}{\sigma^2} + \frac{S_2^2}{\sigma^2} = (n+k-3)\hat{\sigma}^2 \).
In this part we will prove the fact that \( \hat{\sigma}^2 \) is independent of \((\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)\). First of all, we know from (4.21) and (4.23) that
\[
\text{Cov}(Y_i, \hat{\gamma}_0) = \frac{\sigma^2}{n}
\]
\[
\text{Cov}(Y_i, \hat{\gamma}_1) = \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2}.
\]
Hence, we can calculate the following covariances:
\[
\text{Cov}(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}), \hat{\gamma}_0) = \text{Cov}(Y_i, \hat{\gamma}_0) - \text{Var}(\hat{\gamma}_0) = \frac{\sigma^2}{n} - \text{Var}(\hat{Y})
\]
\[
= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0
\]
\[
\text{Cov}(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}), \hat{\gamma}_1) = \text{Cov}(Y_i, \hat{\gamma}_1) - \text{Cov}(\hat{\gamma}_1(x_i - \bar{x}), \hat{\gamma}_1)
\]
\[
= \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} - (x_i - \bar{x})\text{Var}(\hat{\gamma}_1)
\]
\[
= \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} - \frac{\sigma^2(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} = 0.
\]
Thus, \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \) and \( \hat{\gamma}_0 \) are uncorrelated and by Theorem 4.3 they have a 2-dimensional normal distribution, so \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \) and \( \hat{\gamma}_0 \) are independent. Because \( S_1^2 \) is a function of \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \) and \( \hat{\gamma}_0 \), we know that \( S_1^2 \) and \( \hat{\gamma}_0 \) are independent. We also know that \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \) and \( \hat{\gamma}_1 \) are uncorrelated and by Theorem 4.3 they have a 2-dimensional normal distribution, so \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \) and \( \hat{\gamma}_1 \) are independent. Because \( S_1^2 \) is a function of \( Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}) \), we know that \( S_1^2 \) and \( \hat{\gamma}_0 \) are independent. Since \( S_2^2 \) depends only on the last \( k \) \( Y_i \)'s, it is clear that the four random variables \( \hat{\gamma}_0, \hat{\gamma}_1, \hat{Y}_0, S_1^2 \) and \( S_2^2 \) are mutually independent. Thus \( \hat{\gamma}_0, \hat{\gamma}_1 \) and \( \hat{Y}_0 \) are also independent of \( S_1^2 + S_2^2 \) and hence of \( \hat{\sigma}^2 \). But \( \hat{x}_0 \) depends only on \( \hat{\gamma}_1 \) and \( \hat{Y}_0 \); hence the result of part (3).

In the next section we will obtain a \( 100(1 - \alpha)\% \) confidence interval for \( x_0 \).

### 4.3 Confidence Interval for \( x_0 \) for a Given \( Y_0 \)

With the aid of the results of the previous section, we are now ready to discuss confidence intervals for \( x_0 \) for a given \( Y_0 \) in simple linear regression models. It is intuitively clear that we do not have a useful confidence interval for \( x_0 \) if \( \gamma_1 \) is nearly to zero, because in this case the simple linear regression line is almost horizontal.

First, we need some formulas:
\[
\text{E}(\hat{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})) = \text{E}(\hat{Y}_0) - \text{E}(\hat{\gamma}_0) - \text{E}(\hat{\gamma}_1(x_0 - \bar{x}))
\]
\[
= \frac{1}{k} \sum_{i=n+1}^{n+k} (\gamma_0 + \gamma_1(x_0 - \bar{x})) - \gamma_0 - \gamma_1(x_0 - \bar{x})
\]
\[
= 0,
\]
\[
\text{Var}(\hat{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})) = \sigma^2 \left( \frac{1}{k} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right) = \sigma^2 A^2,
\]
From Theorem 4.7 we have the following inequality:

\[
\frac{\bar{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})}{\sqrt{\text{Var}(\bar{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))}} = \frac{\bar{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x})}{\sqrt{\sigma^2 A^2}} \sim N(0, 1). \tag{4.48}
\]

From Theorem 4.7 we have the following

\[
U = (n + k - 3) \frac{\hat{\sigma}^2}{\sigma^2}
= \sum_{i=1}^{n} (Y_i - \hat{\gamma}_0 - \hat{\gamma}_1(x_i - \bar{x}))^2 + \sum_{i=n+1}^{n+k} (Y_i - \bar{Y}_0)^2 \sim \chi^2_{n+k-3}. \tag{4.49}
\]

It can also be seen that \( Z \) and \( U \) are independent, because \( \sigma^2 \) is independent of \((\hat{\gamma}_0, \hat{\gamma}_1, \hat{x}_0)\) (see Theorem 4.7 part 4).

Hence, \( Z \sim N(0, 1) \) and \( U \sim \chi^2_{n+k-3} \) and \( Z \) and \( U \) are independent, which implicates that

\[
T = \frac{Z}{U^{1/2}} \sim t_{\frac{\alpha}{2}; n+k-3} \tag{4.50}
\]

and

\[
P\left(-t_{\frac{\alpha}{2}; n+k-3} \leq T \leq t_{\frac{\alpha}{2}; n+k-3}\right) = 1 - \alpha. \tag{4.51}
\]

The quantity in brackets of (4.51) is equivalent to

\[
T^2 \leq t^2_{\frac{\alpha}{2}; n+k-3}
\]

or

\[
\frac{(\bar{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2}{\hat{\sigma}^2 A^2} \leq t^2_{\frac{\alpha}{2}; n+k-3}
\]

or

\[
(\bar{Y}_0 - \hat{\gamma}_0 - \hat{\gamma}_1(x_0 - \bar{x}))^2 - \hat{\sigma}^2 A^2 t^2_{\frac{\alpha}{2}; n+k-3} \leq 0. \tag{4.52}
\]

Expanding the first term of (4.52) and substituting the definition of \( A^2 \), we obtain the following inequality:

\[
\left(\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t^2_{\frac{\alpha}{2}; n+k-3}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right) (x_0 - \bar{x})^2 - 2\hat{\gamma}_1 (\bar{Y}_0 - \hat{\gamma}_0)(x_0 - \bar{x})
+ \left(\bar{Y}_0 - \hat{\gamma}_0\right)^2 - \hat{\sigma}^2 t^2_{\frac{\alpha}{2}; n+k-3} \left(\frac{1}{k} + \frac{1}{n}\right) \leq 0.
\]

It can be seen that this is a quadratic inequality, which can be written as \( q(x_0 - \bar{x}) = a(x_0 - \bar{x})^2 + 2b(x_0 - \bar{x}) + c \leq 0 \), where \( a, b \) and \( c \) are defined above. If the values of \( x_0 \) that satisfy this inequality are an interval, then these values form a 100(1 - \( \alpha \))% confidence interval for \( x_0 \). We now examine this quadratic inequality.

If the discriminant \( b^2 - ac \) of a quadratic function is negative, then the quadratic function cannot be zero (Figure 4.1 (b) and (d)). In this case, two situations are possible,
Chapter 4 Graybill’s Method

Figure 4.1: Four possible quadratic forms

- $q(x_0 - \bar{x}) < 0$ for all $x_0$ (Figure 4.1(d)), then the confidence interval is $-\infty < x_0 < +\infty$,

- $q(x_0 - \bar{x}) > 0$ for all $x_0$ (Figure 4.1(b)), then there is no confidence interval.

If the discriminant $b^2 - ac$ is zero, then we have an unique point $z_0$, where $q(z_0 - \bar{x}) = 0$. In this case, the confidence interval for $x_0$ is $\mathbb{R}\setminus\{z_0 + \bar{x}\}$, which is not useful.

If the discriminant $b^2 - ac$ is positive, then there are also two possibilities (Figure 4.1 (a) and (c)):

- $a < 0$, then the values of $x_0$ for which $q(x_0 - \bar{x}) \leq 0$ form two infinite intervals (Figure 4.1 (c)), which is not useful,

- $a > 0$, then the values of $x_0$ for which $q(x_0 - \bar{x}) \leq 0$ form a confidence interval for $x_0$ (Figure 4.1(a)).

We can conclude that the inequality $q(x_0 - \bar{x}) \leq 0$ results in a confidence interval for $x_0$ if and only if $a > 0$ and $b^2 - ac > 0$. If we expand $b^2 - ac$, then we get

$$b^2 - ac = \gamma_1^2 (\bar{Y}_0 - \hat{\gamma}_0)^2 - \left( \gamma_1^2 - \frac{\hat{\sigma}^2 t^2 \frac{m+k-3}{2} \Sigma^{n}_{i=1} (x_i - \bar{x})^2}{\Sigma^{n}_{i=1} (x_i - \bar{x})^2} \right) \left( (\bar{Y}_0 - \hat{\gamma}_0)^2 - \hat{\sigma}^2 t^2 \frac{m+k-3}{2} \left( \frac{1}{k} + \frac{1}{n} \right) \right)$$

$$= \gamma_1^2 \hat{\sigma}^2 t^2 \frac{m+k-3}{2} \left( \frac{1}{k} + \frac{1}{n} \right) + \frac{\hat{\sigma}^2 t^2 \frac{m+k-3}{2} \Sigma^{n}_{i=1} (x_i - \bar{x})^2}{\Sigma^{n}_{i=1} (x_i - \bar{x})^2} (\bar{Y}_0 - \hat{\gamma}_0)^2 - \frac{\hat{\sigma}^4 t^4 \frac{m+k-3}{2} \Sigma^{n}_{i=1} (x_i - \bar{x})^2}{\Sigma^{n}_{i=1} (x_i - \bar{x})^2} \left( \frac{1}{k} + \frac{1}{n} \right)$$
4.3 Confidence Interval for $x_0$ for a Given $Y_0$

\[ \hat{\gamma}_1 = \frac{1}{k} + \frac{1}{n} \]

\[ + \left( \frac{(\bar{Y}_0 - \hat{\gamma}_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} - \frac{\hat{\sigma}^2 t_{n+k-3}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \left( \frac{1}{k} + \frac{1}{n} \right) \right) \]

\[ = \hat{\sigma}^2 t_{n+k-3}^2 \left( \frac{1}{k} + \frac{1}{n} \right) a + \left( \frac{(\bar{Y}_0 - \hat{\gamma}_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) . \]

Hence, if $a \geq 0$, then the discriminant $b^2 - ac \geq 0$. So (4.3) yields a confidence interval for $x_0$ if and only if $a \geq 0$, or, in other words, if and only if $\hat{\gamma}_1^2 - \frac{\hat{\sigma}^2 t_{n+k-3}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \geq 0$. If we simplify this, we get

\[ \frac{\hat{\gamma}_1}{\hat{\sigma}^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \geq t_{n+k-3}^2 = F_{\alpha, n+k-3} \]

which is a size $\alpha$ test of $H_0 : \gamma_1 = 0$ versus $H_a : \gamma_1 \neq 0$. If $\gamma_1$ is near to zero, then the linear function is almost horizontal, which gives a very wide confidence interval, which is not very useful.

We can now give a reasonable procedure for determining the value of $x_0$ for a given value of $\bar{y}_0$ and to obtain a confidence interval for $x_0$. The procedure is as follows:

1. Use the statistics in (4.5), (4.6), (4.7), (4.25) and obtain the estimator of $x_0$, which is $\bar{x} + \frac{\bar{Y}_0 - \bar{Y}}{\hat{\gamma}_1}$;

2. Test $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$ with a size $\alpha$ test; reject $H_0$ if and only if $\frac{\hat{\gamma}_1}{\hat{\sigma}^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \geq t_{n+k-3}^2 = F_{\alpha, n+k-3}$;

3. If $H_0$ is not rejected, assume that the model is $y_i = \gamma_0 + \varepsilon$, and hence assume that $x_0$ is not in the model, so no confidence interval exists;

4. If $H_0$ is rejected, set a $100(1 - \alpha)\%$ confidence interval on $x_0$. This confidence interval always exists if $H_0$ is rejected, and the limits are

\[ \bar{x} + \frac{\hat{\gamma}_1 (\bar{Y}_0 - \bar{Y})}{a} - \frac{t_{n+k-3} \hat{\sigma}}{a} \sqrt{a \left( \frac{1}{n} + \frac{1}{k} \right) + \frac{(\bar{Y}_0 - \bar{Y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \]

and

\[ \bar{x} + \frac{\hat{\gamma}_1 (\bar{Y}_0 - \bar{Y})}{a} + \frac{t_{n+k-3} \hat{\sigma}}{a} \sqrt{a \left( \frac{1}{n} + \frac{1}{k} \right) + \frac{(\bar{Y}_0 - \bar{Y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}. \]

where

\[ a = \frac{\hat{\gamma}_1^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \]

Graybill claimed that this is not a $100(1 - \alpha)\%$ confidence interval for $x_0$ but it has confidence coefficient less than $100(1 - \alpha)\%$, because the probability that $x_0$ satisfies the inequality in (4.51) is equal to $1 - \alpha$. In Chapter 6 we will simulate this procedure for some explicit examples.
Chapter 5

Brown’s Profile Likelihood

Brown (1993) considers a second method to obtain the estimator of the unknown \( x_0 \). He obtains the profile likelihood for the simple linear regression model with \( k \) observations at the unknown value of \( x_0 \) in the controlled case. The profile likelihood method presented in Brown (1993) is derived from Brown and Sundberg (1987) and Brown and Sundberg (1989). The profile log likelihood for the \( (n + k) \) observations is the log likelihood maximized over the complete set of four parameters (intercept, slope, \( \sigma^2 \), \( x_0 \)). Because we can not solve this algebraically, we will first obtain the log likelihood maximized over three parameters (intercept, slope, \( \sigma^2 \)) for a fixed \( x_0 \). Then the maximizing values of the intercept, the slope and \( \sigma^2 \) will be functions of \( x_0 \). It can be seen that the log likelihood is maximized, when \( \sigma^2 \) is minimized, such that we can obtain the estimator of \( x_0 \) algebraically. We can also plot the log likelihood as a function of \( x_0 \) and obtain a maximum, which is the estimator of \( x_0 \). The main difference with Graybill’s method is that Brown uses the full maximum likelihood approach. Graybill plugs in the Least Squares estimators of the intercept, slope and \( x_0 \) in the log likelihood to obtain the estimator of \( \sigma^2 \).

First, we will obtain the estimator of \( x_0 \) algebraically. Brown assumes that \( x \) and \( Y \) are centred. For \( Y \) this has to be done after the values are observed. If we have the following observations of \( Y \), \( Y_1, Y_2, \ldots, Y_n \), then we have to redefine these values as the deviation of its own average, \( y_i = Y_i - \bar{Y} \), where \( \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \). We have now the following assumptions

\[
\sum_{i=1}^{n} x_i = 0, \\
\sum_{i=1}^{n} y_i = 0. \quad (5.1)
\]

We have the following model

\[
y_i = \mu + \gamma_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n, \quad (5.2)
\]

where \( \mu = \gamma_0 + \gamma_1 \bar{x} \), the mean of the values of \( Y \) with the old \( x \)-scale and the Least Squares estimator \( \hat{\mu} = 0 \). The value of \( x_0 \) is measured on the same scale as the centred \( x \). The \( k \) observations of \( Y \) on the unknown \( x_0 \) are the original values (they are not centred). The model for the \( k \) observations on \( x_0 \) is

\[
Y_{0\ell} = \mu + \gamma_1 x_0 + \varepsilon_\ell, \quad \varepsilon_\ell \sim N(0, \sigma^2), \quad \ell = n + 1, \ldots, n + k. \quad (5.3)
\]
The models in (5.2) and (5.3) are combined as a vector model

\[ V_0 = \mathbf{1}\mu_0 + \gamma_1 w_0 + \varepsilon_0, \quad \varepsilon_0 \sim N(0, \sigma^2), \]  

where

\[ V_0 = (y_1, y_2, \ldots, y_n, Y_{n+1}, \ldots, Y_{n+k}), \]  

\[ w_0 = (x_1 - a, x_2 - a, \ldots, x_n - a, x_0 - a, \ldots, x_0 - a), \]  

\[ \mathbf{1} = (1, 1, \ldots, 1), \]  

with constraints \( \sum_{i=1}^{n} x_i = 0 \) and \( \sum_{i=1}^{n} y_i = 0 \). We want that the components of \( w_0 \) sum up to zero:

\[ \sum_{i=1}^{n} (x_i - a) + \sum_{\ell=1}^{k} (x_0 - a) = 0 \]  

\[ \sum_{i=1}^{n} x_i - na + k(x_0 - a) = 0 \]  

\[ kx_0 - (n + k)a = 0. \]

Thus,

\[ a = \frac{kx_0}{n + k} \]  

and hence

\[ \mu_0 = \mu + \frac{k\gamma_1 x_0}{n + k}. \]

We will now take the profile likelihood for the \( n+k \) observations over the set of four parameters \((\mu_0, \gamma_1, \sigma^2, x_0)\):

\[
\text{PL}(\mu_0, \gamma_1, \sigma^2, x_0 : y_1, x_1, \ldots, y_n, x_n; Y_{n+1}, Y_{n+2}, \ldots, Y_{n+k}) = \prod_{i=1}^{n+k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2\right).
\]

Hence, the log likelihood equals

\[
-\frac{n + k}{2} \log(2\pi\sigma^2) - \sum_{i=1}^{n+k} \frac{1}{2\sigma^2}(V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2.
\]

The maximum likelihood estimators \( \mu_0, \gamma_1 \) and \( \sigma^2 \), when we keep \( x_0 \) fixed, must satisfy

\[
\frac{\partial \log L}{\partial \mu_0}|_{\mu_0, \gamma_1, \hat{x}_0, \hat{\sigma}^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i}) = 0
\]

\[
\frac{\partial \log L}{\partial \gamma_1}|_{\mu_0, \hat{x}_0, \hat{\sigma}^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i}) w_{0i} = 0
\]

\[
\frac{\partial \log L}{\partial \sigma^2}|_{\mu_0, \gamma_1, \hat{x}_0, \hat{\sigma}^2} = -\frac{n + k}{2\sigma^2} + \frac{1}{2\sigma^4} \left( \sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2 \right) = 0.
\]
Thus, the Maximum Likelihood estimators, when we assume that $x_0$ is known, are

$$
\hat{\mu}_0 = \bar{V}_0 \quad (5.15)
$$

$$
\hat{\gamma}_1 = \frac{\sum_{i=1}^{n+k} w_{0i} V_{0i}}{\sum_{j=1}^{n+k} w_{0j}^2} \quad (5.16)
$$

$$
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2}{n + k} \quad (5.17)
$$

If we replace in (5.13) $\mu_0$, $\gamma_1$ and $\sigma^2$ by $\hat{\mu}_0$, $\hat{\gamma}_1$ and $\hat{\sigma}^2$, resp., then

$$
- \frac{n + k}{2} \log(2\pi\hat{\sigma}^2) - \frac{\sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2}{2 \sum_{i=1}^{n+k} (V_{0i} - \mu_0 - \gamma_1 w_{0i})^2} = - \frac{n + k}{2} \log(2\pi\hat{\sigma}^2) - \frac{n + k}{2} \quad (5.18)
$$

is the maximized value. It can be seen that the log likelihood in (5.13) is maximized if the estimator of $\sigma^2$ in (5.17), which is a function of $x_0$, is minimized. To see for which estimator of $x_0$ (5.17) is minimized, we have to simplify the formula in (5.17):

$$
(n + k)\hat{\sigma}^2 = \sum_{i=1}^{n+k} (V_{0i} - \hat{\mu}_0 - \hat{\gamma}_1 w_{0i})^2 \\
= \sum_{i=1}^{n+k} ((V_{0i} - \bar{V}_0)^2 - 2(V_{0i} - \bar{V}_0)\hat{\gamma}_1 w_{0i} + \hat{\gamma}_1^2 w_{0i}^2) \\
= \sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0)^2 - 2\hat{\gamma}_1 \sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0) w_{0i} + \hat{\gamma}_1^2 \sum_{i=1}^{n+k} w_{0i}^2 \\
= \sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0)^2 - 2 \left( \frac{\sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0) w_{0i}}{\sum_{i=1}^{n+k} w_{0i}^2} \right)^2 + \left( \frac{\sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0) w_{0i}}{\sum_{i=1}^{n+k} w_{0i}^2} \right)^2 \\
= \sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0)^2 - \left( \frac{\sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0) w_{0i}}{\sum_{i=1}^{n+k} w_{0i}^2} \right)^2 \quad (5.19)
$$
The formulas of the right-hand side of (5.19) can be simplified:

\[
\sum_{i=1}^{n+k} \left( V_{0i} - \bar{V}_0 \right)^2 = \sum_{i=1}^{n+k} V_{0i}^2 - 2\bar{V}_0 \sum_{i=1}^{n+k} V_{0i} + (n+k)\bar{V}_0^2 \\
= \sum_{i=1}^{n} y_i^2 + \sum_{\ell=n+1}^{n+k} Y_{0\ell}^2 - 2\bar{V}_0 \sum_{\ell=n+1}^{n+k} Y_{0\ell} + (n+k)\bar{V}_0^2 \\
= \sum_{i=1}^{n} y_i^2 + \sum_{\ell=n+1}^{n+k} Y_{0\ell}^2 - \frac{2\sum_{i=1}^{n+k} V_{0i}}{n+k} \sum_{\ell=n+1}^{n+k} Y_{0\ell} + \left( \frac{\sum_{i=1}^{n+k} V_{0i}}{n+k} \right)^2 \\
= \sum_{i=1}^{n} y_i^2 + \sum_{\ell=n+1}^{n+k} Y_{0\ell}^2 - \frac{\sum_{i=1}^{n+k} V_{0i} \sum_{\ell=n+1}^{n+k} Y_{0\ell}}{n+k} \\
= \sum_{i=1}^{n} y_i^2 + \sum_{\ell=n+1}^{n+k} Y_{0\ell}^2 - \frac{\left( \sum_{\ell=n+1}^{n+k} Y_{0\ell} \right)^2}{n+k} \\
= \sum_{i=1}^{n} y_i^2 + \sum_{\ell=n+1}^{n+k} Y_{0\ell}^2 - \frac{k^2\bar{Y}_0^2}{n+k} \quad (5.20)
\]

\[
\sum_{i=1}^{n+k} (V_{0i} - \bar{V}_0)w_{0i} = \sum_{i=1}^{n+k} V_{0i}w_{0i} - \bar{V}_0 \sum_{i=1}^{n+k} w_{0i} = \sum_{i=1}^{n+k} V_{0i}w_{0i} \\
= \sum_{i=1}^{n} y_i(x_i - a) + \sum_{\ell=n+1}^{n+k} Y_{0\ell}(x_0 - a) \\
= \sum_{i=1}^{n} y_i x_i + x_0 k\bar{Y}_0 - ak\bar{Y}_0 \\
= \sum_{i=1}^{n} y_i x_i + x_0 k\bar{Y}_0 - \frac{k^2 x_0 \bar{Y}_0}{n+k} \\
= \sum_{i=1}^{n} y_i x_i + k x_0 \bar{Y}_0 \left( 1 - \frac{k}{n+k} \right) \\
= \sum_{i=1}^{n} y_i x_i + \frac{nk x_0 \bar{Y}_0}{n+k}. \quad (5.21)
\]
We can also simplify the following term:

\[
\sum_{i=1}^{n+k} w_{0i}^2 = \sum_{i=1}^{n} (x_i - a)^2 + \sum_{\ell=n+1}^{n+k} (x_0 - a)^2 \\
= \sum_{i=1}^{n} (x_i^2 - 2ax_i + a^2) + k(x_0^2 - 2ax_0 + a^2) \\
= \sum_{i=1}^{n} x_i^2 + na^2 + kx_0^2 - 2kax_0 + ka^2 \\
= \sum_{i=1}^{n} x_i^2 + \frac{nk^2 x_0^2}{(n+k)^2} + kx_0^2 - \frac{2k^2 x_0^2}{n+k} + \frac{k^3 x_0^2}{(n+k)^2} \\
= \sum_{i=1}^{n} x_i^2 + kx_0^2 \left( \frac{nk}{(n+k)^2} + 1 - \frac{2k}{n+k} + \frac{k^2}{(n+k)^2} \right) \\
= \sum_{i=1}^{n} x_i^2 + kx_0^2 \left( \frac{nk + (n+k)^2 - 2k(n+k) + k^2}{(n+k)^2} \right) \\
= \sum_{i=1}^{n} x_i^2 + kx_0^2 \left( \frac{nk + n^2}{(n+k)^2} \right) \\
= \sum_{i=1}^{n} x_i^2 + \frac{nkx_0^2}{n+k}. \quad (5.22)
\]

After replacing the second term of the right-hand side of (5.19) by (5.21) and (5.22), it equals

\[
- \left( \frac{\sum_{i=1}^{n} y_ix_i + \frac{nkx_0y_0}{n+k}}{\sum_{i=1}^{n} x_i^2 + \frac{nkx_0^2}{n+k}} \right)^2 \quad (5.23)
\]

We will now multiply both nominator and denominator by \( \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} \):

\[
- \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} \left( \frac{\sum_{i=1}^{n} y_ix_i + \frac{nkx_0y_0}{n+k}}{\sum_{i=1}^{n} x_i^2 + \frac{nkx_0^2}{n+k}} \right)^2 = - \frac{\sum_{i=1}^{n} y_ix_i + \frac{nkx_0y_0}{(n+k)\sum_{i=1}^{n} x_i^2}}{\sum_{i=1}^{n} x_i^2 \left( 1 + \frac{nkx_0^2}{(n+k)\sum_{i=1}^{n} x_i^2} \right)} \left( \frac{\hat{\beta} + \frac{nkx_0y_0}{(n+k)^2}}{gr(x_0)} \right)^2, \quad (5.24)
\]

where

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} y_ix_i}{\sum_{i=1}^{n} x_i^2}, \quad (5.25)
\]

\[
g = \frac{1}{\sum_{i=1}^{n} x_i^2}, \quad (5.26)
\]

\[
r(x_0) = 1 + \frac{nkx_0^2}{n+k}. \quad (5.27)
\]
In particular, we have

\[
\frac{n^2 k g x_0^2}{n + k} = r(x_0) - 1. \tag{5.28}
\]

With further algebra,

\[
\frac{\left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2}{g r(x_0)} = - \frac{\left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2}{\sum_{i=1}^{n} x_i^2 \left[ 1 + \frac{n k x_0^2}{(n+k) \sum_{i=1}^{n} x_i^2} \right]} = - \frac{\sum_{i=1}^{n} x_i^2 \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2}{\left[ 1 + \frac{n k x_0^2}{(n+k) \sum_{i=1}^{n} x_i^2} \right] \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2} = \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2 \left( \frac{n k x_0^2}{(n+k) \left[ 1 + \frac{n k x_0^2}{(n+k) \sum_{i=1}^{n} x_i^2} \right]} - \sum_{i=1}^{n} x_i^2 \right) = \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2 \left( \frac{n k x_0^2}{(n+k) \left[ 1 + \frac{n k x_0^2}{(n+k) \sum_{i=1}^{n} x_i^2} \right]} - \frac{1}{g} \right) = \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \left( \hat{\beta} + \frac{n k x_0 \bar{Y}_0 g}{(n+k)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \left( \frac{\hat{\beta} x_0 + \bar{Y}_0 (r(x_0) - 1))^2}{(n+k) r(x_0)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \left( \frac{\hat{\beta} x_0 + \bar{Y}_0 (r(x_0))^2}{(n+k) r(x_0)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \frac{\left( \hat{\beta} x_0 - \bar{Y}_0 r(x_0)^2 n k}{(n+k) r(x_0)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \frac{\left( \hat{\beta} x_0 - \bar{Y}_0 r(x_0)^2 n k}{(n+k) r(x_0)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \frac{\left( \hat{\beta} x_0 - \bar{Y}_0 r(x_0) + \bar{Y}_0 r(x_0) + \bar{Y}_0 r(x_0)^2 n k}{(n+k) r(x_0)} \right)^2 \frac{g}{\left( n + k \right) r(x_0)} = \frac{\left( \hat{\beta}^2 + 2\hat{\beta} \frac{n k x_0 \bar{Y}_0 g}{(n+k)} + n^2 k x_0^2 \bar{Y}_0^2 g^2 \right)^2}{(n+k) r(x_0)} = \frac{g}{\left( n + k \right) r(x_0)} \right)^2
\[
\frac{(\hat{\beta}x_0 - \bar{Y}_0)^2nk}{(n+k)r(x_0)} + \frac{2(\hat{\beta}x_0 - \bar{Y}_0)\bar{Y}_0 nkrnk}{(n+k)r(x_0)} + \frac{\bar{Y}_0^2 r(x_0)^2nk}{(n+k)r(x_0)} - \frac{\hat{\beta}^2}{g} - \frac{2\hat{\beta}n^2k^2\bar{Y}_0^2}{(n+k)^2} \]
\[
\frac{\hat{\beta}^2}{g} - \frac{2\hat{\beta}nkx_0\bar{Y}_0}{n+k} - \frac{n^2k^2x_0^2\bar{Y}_0^2}{(n+k)^2} \]
\[
\frac{(\hat{\beta}x_0 - \bar{Y}_0)^2nk}{(n+k)r(x_0)} + \frac{2\hat{\beta}x_0\bar{Y}_0 nknk}{n+k} + \frac{2\bar{Y}_0^2 nknk}{(n+k)} - \frac{\hat{\beta}^2}{g} - \frac{n^2k^2\bar{Y}_0^2}{(n+k)^2} \]
\[
\frac{\hat{\beta}^2}{g} - \frac{n^2k^2x_0^2\bar{Y}_0^2}{(n+k)^2} \]
\[
\frac{\hat{\beta}^2}{g} - \frac{(\hat{\beta}x_0 - \bar{Y}_0)^2nk}{(n+k)r(x_0)} + \frac{\bar{Y}_0^2 nknk}{(n+k)} - \frac{\hat{\beta}^2}{g} \]
\[
\frac{\hat{\beta}^2}{g} - \frac{n^2k^2x_0^2\bar{Y}_0^2}{(n+k)^2} \]
\[
\frac{(\hat{\beta}x_0 - \bar{Y}_0)^2nk}{(n+k)r(x_0)} + 2\bar{Y}_0^2 nknk \quad \text{(5.29)}
\]

Now we can substitute (5.29) back into (5.19) and we will also use (5.20):

\[
(n + k)\bar{\sigma}^2 = \sum_{i=1}^{n} g_i^2 + \sum_{i=n+1}^{n+k} Y_{i+}^2 - \frac{k^2\bar{Y}_0^2}{n+k} + \frac{(\hat{\beta}x_0 - \bar{Y}_0)^2nk}{(n+k)r(x_0)} - \frac{\bar{Y}_0^2 nknk}{(n+k)r(x_0)} - \frac{\hat{\beta}^2}{g} \quad \text{(5.30)}
\]

where \(s_+^2\) are all terms without \(x_0\). Thus \(s_+^2\) is constant.

The log likelihood in (5.13) is maximized when (5.30) is minimized. Thus we have to minimize the second term of (5.30), because the first term \(s_+^2\) is constant. This is at

\[
\hat{x}_0 = \frac{\bar{Y}_0}{\hat{\beta}}. \quad \text{(5.31)}
\]

**Example 5.1** In this example we will obtain the estimator, which is derived from the profile likelihood, for the new data of Example 2.2. The new data are the data without the removed data. The profile likelihood estimator is

\[
\hat{x}_0 = \frac{\bar{Y}_0}{\hat{\beta}} = \frac{1.72092}{0.96539} = 1.78156. \quad \text{(5.32)}
\]
This estimator differs a bit from the classical estimator for these data, which is 0.69727.

A second approach to obtain the profile likelihood estimator of $x_0$ is to write the profile log likelihood of (5.13) as a function of $x_0$. We also have to substitute the maximizing values of $\mu_0$, $\gamma_1$ and $\sigma^2$, which can be found in (5.15), (5.16) and (5.17), resp. We will also use the new data of Example 2.2 to plot the profile log likelihood in Mathematica. A plot of this function can be seen in Figure 5.1. We can obtain the $x$-value of the maximum with Mathematica,

\[
\hat{x}_0 = 0.687154, \tag{5.33}
\]

which is approximately the same as the estimator of $x_0$ in (5.32) of the algebraically approach.
Chapter 6

Simulation of Graybill’s Confidence Interval

In this chapter we will simulate the confidence interval of Graybill’s method, which can be found in Chapter 4. We have chosen for Graybill’s method, because the other methods to obtain an estimator of $x_0$ do not differ much from the estimator of Graybill’s method. In Section 6.1 we will simulate the confidence coefficient for the specific examples of Chapter 2. Afterwards, in Section 6.2 we will present an experimental design to investigate whether some factors have an influence on the confidence coefficient.

6.1 Confidence Coefficient of Graybill’s Interval for Some Specific Examples

In this section we will simulate the procedure for the confidence interval of Section 4.3 to investigate to which extent Graybill’s claim is true. Graybill claimed that the confidence interval of Section 4.3 is not a $100(1 - \alpha)\%$ confidence interval, but has confidence less than $100(1 - \alpha)\%$. We will now simulate the confidence of the interval. This simulation will give the appropriate confidence for the interval.

First of all, we will describe the set up of the simulation, which is given in Appendix A.1. Afterwards, we will give some results.

We will use the centred simple linear regression model, which is

$$Y = \gamma_0 + \gamma_1(x - \bar{x}) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

(6.1)

To execute the simulation some variables has to be given. These variables are:

- The true value of $x_0$;
- The true value of $\gamma_0$;
- The true value of $\gamma_1$;
- The true value of $\sigma^2$;
- The number of observations $n$;
- The number of observations $k$ on $x_0$;
• The $n$ values of $x$: $x_1, \ldots, x_n$;

• The confidence coefficient $\alpha$;

• The number of times $T$ the loop, which is explained beneath, is repeated (at least 10,000 times, otherwise the confidence coefficient is not reliable enough).

We will now explain the loop in the simulation. First of all, we have to pick the $n$ values of $Y$, $Y_i$, $i = 1, \ldots, n$, randomly from the normal distribution with mean $\gamma_0 + \gamma_1(x_i - \bar{x})$ and standard deviation $\sigma$. Secondly, the $k$ observations on $x_0$ has to be chosen from a normal distribution with mean $\gamma_0 + \gamma_1(x_0 - \bar{x})$ and standard deviation $\sigma$. Further on, the estimates of $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\sigma}^2$ and $\hat{x}_0$ will be calculated using (4.6), (4.5), (4.25) and (4.7), resp. Afterwards, the confidence interval of (4.53) and (4.54) is calculated. There are two tests included in the simulation. The first test is the following: $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$. The hypothesis $H_0$ is rejected if and only if $\hat{\gamma}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 \geq \frac{t^2_{\frac{\alpha}{2} n+k-3}}{\hat{\sigma}^2} = F_{\alpha;1,n+k-3}$ (see Chapter 4). If $H_0$ is rejected, then the simulation gives a 0 as outcome, otherwise a 1. The other test checks whether the true value of $x_0$ lies in the calculated confidence interval. If $x_0$ lies in the interval, then the simulation returns a 1 and a 0 otherwise. This loop is repeated $T$ times. Finally, the simulation has 2 results: the number of times of $T$ the true value of $x_0$ lies in the confidence interval and the number of times of $T H_0$ is accepted. Now, the confidence of this interval is the number of times the true value of $x_0$ lies in the confidence interval minus the number of times $H_0$ is accepted.

We will now simulate the two examples of Chapter 2. For each example we will repeat 30 times the simulation with $T = 10,000$ and 30 times the simulation with $T = 100,000$. We have chosen to simulate 30 times, such that we can investigate the normality of these values.

**Example 6.1** In this example we will simulate Example 2.1 with $T = 10,000$. For Example 2.1 we take the following values for the variables:

• The true value of $x_0$: 216.86;

• The true value of $\gamma_0$: -6.67;

• The true value of $\gamma_1$: 0.953;

• The true value of $\sigma^2$: 2.420743687;

• The number of observations $n$: 16;

• The number of observations $k$ on $x_0$: 10;

• The 16 values of $x$: $x_1 = 100$, $x_2 = 120$, $x_3 = 140$, $x_4 = 160$, $x_5 = 180$, $x_6 = 200$, $x_7 = 220$, $x_8 = 240$, $x_9 = 260$, $x_{10} = 280$, $x_{11} = 300$, $x_{12} = 320$, $x_{13} = 340$, $x_{14} = 360$, $x_{15} = 380$, $x_{16} = 400$;

• The confidence coefficient $\alpha$: 0.05;

• The number of times $T$ the loop, which is explained beneath, is repeated: 10,000.
With these values we have simulated the confidence 30 times. These 30 values are given in Table 6.1. Figure 6.1 gives a histogram of these 30 values. We will now test whether these 30 values are normally distributed. This can be tested with the Shapiro-Wilks test and a normal probability plot of the 30 values. We can see that the 30 values are normally distributed, because the value of the Shapiro-Wilks statistic is 0.95918 and the p-value is 0.2951. The normal probability plot in Figure 6.1 also implicates normality. Thus, the 30 values $X_i$, $i = 1, \ldots, 30$, are normally distributed with mean $\mu$ and variance $\sigma^2$ unknown. We can obtain a $100(1 - \alpha)\%$ confidence interval for $\mu$ and for $\sigma^2$. We know the following distributional properties:

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$  \hspace{1cm} (6.2)

and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$ \hspace{1cm} (6.3)

We also know from the simulation that $\bar{X} = 94.935$ and $S = 0.240398$. The $100(1 - \alpha)\%$ confidence interval for $\mu$ is

$$\left( \bar{X} - t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}} \right) = (94.8069, 95.0631)$$ \hspace{1cm} (6.4)

and the $100(1 - \alpha)\%$ confidence interval for $\sigma^2$ is

$$\left( \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}} \right) = (0.031522, 0.138477).$$ \hspace{1cm} (6.5)

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.00</td>
<td>95.10</td>
</tr>
<tr>
<td>95.12</td>
<td>95.36</td>
</tr>
<tr>
<td>95.08</td>
<td>94.92</td>
</tr>
<tr>
<td>94.57</td>
<td>95.19</td>
</tr>
<tr>
<td>94.93</td>
<td>94.96</td>
</tr>
<tr>
<td>95.03</td>
<td>94.85</td>
</tr>
<tr>
<td>94.54</td>
<td>94.80</td>
</tr>
<tr>
<td>95.04</td>
<td>94.96</td>
</tr>
<tr>
<td>94.43</td>
<td>95.03</td>
</tr>
<tr>
<td>94.69</td>
<td>95.05</td>
</tr>
<tr>
<td>94.82</td>
<td>94.69</td>
</tr>
<tr>
<td>95.04</td>
<td>95.27</td>
</tr>
<tr>
<td>94.84</td>
<td>95.11</td>
</tr>
<tr>
<td>95.35</td>
<td>94.44</td>
</tr>
<tr>
<td>94.83</td>
<td>95.01</td>
</tr>
</tbody>
</table>
Figure 6.1: Histogram of Table 6.1.

Figure 6.2: Normal probability plot of Table 6.1.
Example 6.2 In this example we will simulate Example 2.1 with $T = 100,000$. For Example 2.1 we take the following values for the variables:

- The true value of $x_0$: 216.86;
- The true value of $\gamma_0$: −6.67;
- The true value of $\gamma_1$: 0.953;
- The true value of $\sigma^2$: 2.420743687;
- The number of observations $n$: 16;
- The number of observations $k$ on $x_0$: 10;
- The 16 values of $x$: $x_1 = 100$, $x_2 = 120$, $x_3 = 140$, $x_4 = 160$, $x_5 = 180$, $x_6 = 200$, $x_7 = 220$, $x_8 = 240$, $x_9 = 260$, $x_{10} = 280$, $x_{11} = 300$, $x_{12} = 320$, $x_{13} = 340$, $x_{14} = 360$, $x_{15} = 380$, $x_{16} = 400$;
- The confidence coefficient $\alpha$: 0.05;
- The number of times $T$ the loop, which is explained beneath, is repeated: 100,000.

With these values we have simulated the confidence 30 times. These 30 values are given in Table 6.2. Figure 6.2 gives a histogram of these 30 values. We will now test whether these 30 values are normally distributed. This can be tested with the Shapiro-Wilks test and a normal probability plot of the 30 values. We can see that the 30 values are normally distributed, because the value of the Shapiro-Wilks statistic is 0.946295 and the $p$-value is 0.1344. The normal probability plot in Figure 6.2 also implicates normality. Thus, the 30 values $X_i$, $i = 1, \ldots, 30$, are normally distributed with mean $\mu$ and variance $\sigma^2$ unknown. We can obtain

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>94.927</td>
<td>94.972</td>
</tr>
<tr>
<td>94.968</td>
<td>94.996</td>
</tr>
<tr>
<td>95.046</td>
<td>94.990</td>
</tr>
<tr>
<td>94.935</td>
<td>94.891</td>
</tr>
<tr>
<td>95.016</td>
<td>95.109</td>
</tr>
<tr>
<td>94.915</td>
<td>94.954</td>
</tr>
<tr>
<td>94.932</td>
<td>95.000</td>
</tr>
<tr>
<td>94.924</td>
<td>95.025</td>
</tr>
<tr>
<td>95.109</td>
<td>94.991</td>
</tr>
<tr>
<td>95.052</td>
<td>94.959</td>
</tr>
<tr>
<td>95.087</td>
<td>94.977</td>
</tr>
<tr>
<td>94.966</td>
<td>95.043</td>
</tr>
<tr>
<td>94.960</td>
<td>95.088</td>
</tr>
<tr>
<td>95.129</td>
<td>94.950</td>
</tr>
<tr>
<td>94.961</td>
<td>95.000</td>
</tr>
</tbody>
</table>

Table 6.2: The confidence of the interval of (4.53) and (4.54).
Figure 6.3: Histogram of Table 6.2.

Figure 6.4: Normal probability plot of Table 6.2.
a 100(1 − α)% confidence interval for μ and for σ²*. We know the following distributional properties:

\[
\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}
\]

(6.6)

and

\[
\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2_{n-1}.
\]

(6.7)

We also know from the simulation that ̅\(\bar{X} = 94.9957\) and ̅\(\bar{S} = 0.0629416\). The 100(1 − α)% confidence interval for μ is

\[
\left( \bar{X} - t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}} \right) = (94.9622, 95.0292)
\]

(6.8)

and the 100(1 − α)% confidence interval for σ²* is

\[
\left( \frac{(n - 1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \frac{(n - 1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}} \right) = (0.002161, 0.009493).
\]

(6.9)

**Example 6.3** In this example we will simulate Example 2.2 with ̅\(\bar{T} = 10,000\). For Example 2.2 we take the following values for the variables:

- The true value of \(x_0\): 0.69727;
- The true value of \(γ_0\): 13.8272;
- The true value of \(γ_1\): 2.50441;
- The true value of \(σ^2\): 0.106594;
- The number of observations \(n\): 24;
- The number of observations \(k\) on \(x_0\): 10;
- The 33 values of \(x\): \(x_1 = 9.71\), \(x_2 = 9.71\), \(x_3 = 9.71\), \(x_4 = 7.96\), \(x_5 = 7.96\), \(x_6 = 7.96\), \(x_7 = 6.82\), \(x_8 = 6.82\), \(x_9 = 6.82\), \(x_{10} = 5.85\), \(x_{11} = 5.85\), \(x_{12} = 5.85\), \(x_{13} = 4.95\), \(x_{14} = 4.95\), \(x_{15} = 4.95\), \(x_{16} = 3.91\), \(x_{17} = 3.91\), \(x_{18} = 3.91\), \(x_{19} = 2.98\), \(x_{20} = 2.98\), \(x_{21} = 2.98\), \(x_{22} = 2.07\), \(x_{23} = 2.07\), \(x_{24} = 2.07\);
- The confidence coefficient \(α\): 0.05;
- The number of times \(T\) the loop, which is explained beneath, is repeated: 10,000.

With these values we have simulated the confidence 30 times. These 30 values are given in Table 6.3. Figure 6.3 gives a histogram of these 30 values. We will now test whether these 30 values are normally distributed. This can be tested with the Shapiro-Wilks test and a normal probability plot of the 30 values. We can see that the 30 values are normally distributed, because the value of the Shapiro-Wilks statistic is 0.948083 and the \(p\)-value is 0.1502. The normal probability plot in Figure 6.3 also implicates normality. Thus, the 30 values \(X_i\),
Table 6.3: The confidence of the interval of (4.53) and (4.54).

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.03</td>
<td>95.20</td>
</tr>
<tr>
<td>95.08</td>
<td>95.02</td>
</tr>
<tr>
<td>94.67</td>
<td>94.82</td>
</tr>
<tr>
<td>94.67</td>
<td>95.17</td>
</tr>
<tr>
<td>95.06</td>
<td>94.88</td>
</tr>
<tr>
<td>94.93</td>
<td>94.90</td>
</tr>
<tr>
<td>95.02</td>
<td>94.84</td>
</tr>
<tr>
<td>94.87</td>
<td>95.09</td>
</tr>
<tr>
<td>95.26</td>
<td>94.94</td>
</tr>
<tr>
<td>94.94</td>
<td>94.57</td>
</tr>
<tr>
<td>95.13</td>
<td>95.13</td>
</tr>
<tr>
<td>94.69</td>
<td>95.12</td>
</tr>
<tr>
<td>95.07</td>
<td>94.88</td>
</tr>
<tr>
<td>94.63</td>
<td>94.70</td>
</tr>
<tr>
<td>95.07</td>
<td>95.10</td>
</tr>
</tbody>
</table>

Figure 6.5: Histogram of Table 6.3.
$i = 1, \ldots, 30$, are normally distributed with mean $\mu$ and variance $\sigma^2$ unknown. We can obtain a $100(1 - \alpha)\%$ confidence interval for $\mu$ and for $\sigma^2$. We know the following distributional properties:

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

(6.10)

and

$$\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

(6.11)

We also know from the simulation that $\bar{X} = 94.9493$ and $S = 0.185583$. The $100(1 - \alpha)\%$ confidence interval for $\mu$ is

$$\left( \bar{X} - t_{1 - \frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1 - \frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}} \right) = (94.8709, 95.0277)$$

(6.12)

and the $100(1 - \alpha)\%$ confidence interval for $\sigma^2$ is

$$\left( \frac{(n - 1)S^2}{\chi^2_{n-1}}, \frac{(n - 1)S^2}{\chi^1_{2,n-1}} \right) = (0.020791, 0.067705).$$

(6.13)

**Example 6.4** In this example we will simulate Example 2.2 with $T = 100,000$. For Example 2.2 we take the following values for the variables:
• The true value of $x_0$: 0.69727;
• The true value of $\gamma_0$: 13.8272;
• The true value of $\gamma_1$: 2.50441;
• The true value of $\sigma^2$: 0.106594;
• The number of observations $n$: 24;
• The number of observations $k$ on $x_0$: 10;
• The $33$ values of $x$: $x_1 = 9.71$, $x_2 = 9.71$, $x_3 = 9.71$, $x_4 = 7.96$, $x_5 = 7.96$, $x_6 = 7.96$, $x_7 = 6.82$, $x_8 = 6.82$, $x_9 = 6.82$, $x_{10} = 5.85$, $x_{11} = 5.85$, $x_{12} = 5.85$, $x_{13} = 4.95$, $x_{14} = 4.95$, $x_{15} = 4.95$, $x_{16} = 3.91$, $x_{17} = 3.91$, $x_{18} = 3.91$, $x_{19} = 2.98$, $x_{20} = 2.98$, $x_{21} = 2.98$, $x_{22} = 2.07$, $x_{23} = 2.07$, $x_{24} = 2.07$;
• The confidence coefficient $\alpha$: 0.05;
• The number of times $T$ the loop, which is explained beneath, is repeated: 10,000.

With these values we have simulated the confidence 30 times. These 30 values are given in Table 6.4. Figure 6.4 gives a histogram of these 30 values. We will now test whether these 30 values are normally distributed. This can be tested with the Shapiro-Wilks test and a normal probability plot of the 30 values. We can see that the 30 values are normally distributed, because the value of the Shapiro-Wilks statistic is 0.935034 and the p-value is 0.0669. The normal probability plot in Figure 6.4 also implicates normality. Thus, the 30 values $X_i$,

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>95.026</td>
<td>95.026</td>
</tr>
<tr>
<td>94.914</td>
<td>94.970</td>
</tr>
<tr>
<td>95.051</td>
<td>95.011</td>
</tr>
<tr>
<td>94.895</td>
<td>95.054</td>
</tr>
<tr>
<td>94.988</td>
<td>95.125</td>
</tr>
<tr>
<td>95.025</td>
<td>94.894</td>
</tr>
<tr>
<td>95.014</td>
<td>94.892</td>
</tr>
<tr>
<td>94.923</td>
<td>95.036</td>
</tr>
<tr>
<td>95.000</td>
<td>95.015</td>
</tr>
<tr>
<td>95.030</td>
<td>95.026</td>
</tr>
<tr>
<td>95.006</td>
<td>94.903</td>
</tr>
<tr>
<td>94.934</td>
<td>94.963</td>
</tr>
<tr>
<td>95.018</td>
<td>95.041</td>
</tr>
<tr>
<td>94.852</td>
<td>94.784</td>
</tr>
<tr>
<td>95.054</td>
<td>94.957</td>
</tr>
</tbody>
</table>

Table 6.4: The confidence of the interval of (4.53) and (4.54).

$i = 1, \ldots, 30$, are normally distributed with mean $\mu$ and variance $\sigma^2$ unknown. We can obtain
6.1 Confidence Coefficient of Graybill’s Interval for Some Specific Examples

Figure 6.7: Histogram of Table 6.4.

Figure 6.8: Normal probability plot of Table 6.4.
a 100(1 − α)% confidence interval for μ and for σ^2. We know the following distributional properties:

\[
\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \tag{6.14}
\]

and

\[
\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{6.15}
\]

We also know from the simulation that \(\bar{X} = 94.9809\) and \(S = 0.0729626\). The 100(1 − α)% confidence interval for μ is

\[
\left(\bar{X} - t_{1-\frac{\alpha}{2}:n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}:n-1} \frac{S}{\sqrt{n}}\right) = (94.9501, 95.0117) \tag{6.16}
\]

and the 100(1 − α)% confidence interval for σ^2 is

\[
\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}:n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}:n-1}}\right) = (0.003214, 0.010465). \tag{6.17}
\]

We can conclude that in all the four examples the mean value of the confidence coefficient is less than 100(1 − α)%. For many examples Graybill’s claim is true, but the difference is very small, such that we can conclude that this interval is very reliable.

### 6.2 Experimental Design

In this section we will systematically investigate the effects of 3 factors of the centred simple linear regression model with a face-centred central composite design. A face-centred central composite design is a 2^3 factorial design with star points, which are face-centred. Face-centred implicates that the star points are located on the centres of the faces of the cube, which can be seen in Figure 6.9. The 3 factors, which we want to investigate, are:

- The true value of the slope \(\gamma_1\);
- The true value of \(\sigma^2\);
- The number of observations \(k\) on the unknown \(x_0\);

because we think that these factors can have an influence on the confidence coefficient. We take each factor at two levels, low and high. We will distinguish two designs. For the first design we choose \(\gamma_1\) very small, such that in some cases the standard \(t\)-test, which tests \(H_0 : \gamma_1 = 0\) against \(H_1 : \gamma_1 \neq 0\), does not reject \(H_0\) (see Chapter 4). This implicates that we do not have a confidence interval for \(x_0\). For the second design we take \(\gamma_1\) large enough, so that \(H_0\) is always rejected.
6.2 Experimental Design

Figure 6.9: A face-centred central composite design for $k = 3$.

6.2.1 Design with $\gamma_1$ Very Small

For the first design we choose for the following levels:

- $\gamma_{1\text{Low}} = \gamma_{1L} = 0.1$ and $\gamma_{1\text{High}} = \gamma_{1H} = 10$;
- $\sigma_{L\text{Low}}^2 = \sigma_{L}^2 = 0.1$ and $\sigma_{H\text{High}}^2 = \sigma_{H}^2 = 1$;
- $k_{\text{Low}} = k_{L} = 2$ and $k_{\text{High}} = k_{H} = 10$.

The other factors are the following such as in Example 2.2:

- The true value of $x_0$: 0.69727;
- The true value of $\gamma_0$: 13.8272;
- The number of observations $n$: 24;
- The 33 values of $x$: $x_1 = 9.71$, $x_2 = 9.71$, $x_3 = 9.71$, $x_4 = 7.96$, $x_5 = 7.96$, $x_6 = 7.96$, $x_7 = 6.82$, $x_8 = 6.82$, $x_9 = 6.82$, $x_{10} = 5.85$, $x_{11} = 5.85$, $x_{12} = 5.85$, $x_{13} = 4.95$, $x_{14} = 4.95$, $x_{15} = 4.95$, $x_{16} = 3.91$, $x_{17} = 3.91$, $x_{18} = 3.91$, $x_{19} = 2.98$, $x_{20} = 2.98$, $x_{21} = 2.98$, $x_{22} = 2.07$, $x_{23} = 2.07$, $x_{24} = 2.07$;
- The confidence coefficient $\alpha$: 0.05;
- The number of times $T$ the loop is repeated: 10,000.
The response is the confidence of the interval of Section 4.3, which is the result of the simulation in Section 6.1. We will also consider 4 centrepoints and we will replicate this design 5 times. We will first simulate the 18 points 5 times in Mathematica. These 18 runs are tabulated in Table 6.5.

<table>
<thead>
<tr>
<th>$\gamma_1$</th>
<th>$\sigma^2$</th>
<th>$k$</th>
<th>$\gamma_1$</th>
<th>$\sigma^2$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>2</td>
<td>10</td>
<td>0.55</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>2</td>
<td>5.05</td>
<td>0.1</td>
<td>6</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>2</td>
<td>5.05</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2</td>
<td>5.05</td>
<td>0.55</td>
<td>2</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>10</td>
<td>5.05</td>
<td>0.55</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>10</td>
<td>5.05</td>
<td>0.55</td>
<td>6</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>10</td>
<td>5.05</td>
<td>0.55</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>5.05</td>
<td>0.55</td>
<td>6</td>
</tr>
<tr>
<td>0.1</td>
<td>0.55</td>
<td>6</td>
<td>5.05</td>
<td>0.55</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 6.5: The 18 runs.

This design is analyzed in Statgraphics and the Analysis-of-Variance table is given in Table 6.6. The value of $R^2_{adj}$ is 93.1977. It can be seen that the three main effects $A : \gamma_1$, $B : \sigma^2$ and $C : $ observations are highly significant. Also the $AA$, $BB$ and $CC$ effects and the $AB$ and $AC$ interactions are highly significant. The $BC$ interaction is also significant. It can also be seen that the lack-of-fit test is significant, which implicates that the regression function is not linear. The estimates of the effects are given in Table 6.7. A standardized Pareto chart of the effects is given in Figure 6.10. We can conclude that the factors $\gamma_1$, $\sigma^2$ and observations are significant and that the confidence coefficient depends on them.

### 6.2.2 Design with $\gamma_1$ Large Enough

For the second design we choose for the following levels:
6.2 Experimental Design

Table 6.7: The estimated effects.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A : \gamma_1$</td>
<td>-45.0704</td>
</tr>
<tr>
<td>$B : \sigma^2$</td>
<td>-28.8136</td>
</tr>
<tr>
<td>$C : \text{observations}$</td>
<td>0.2784</td>
</tr>
<tr>
<td>$AA$</td>
<td>-54.9774</td>
</tr>
<tr>
<td>$AB$</td>
<td>36.162</td>
</tr>
<tr>
<td>$AC$</td>
<td>-0.401</td>
</tr>
<tr>
<td>$BB$</td>
<td>8.98657</td>
</tr>
<tr>
<td>$BC$</td>
<td>-0.216</td>
</tr>
<tr>
<td>$CC$</td>
<td>8.96657</td>
</tr>
</tbody>
</table>

Figure 6.10: Standardized Pareto chart of the effects.
Chapter 6 Simulation of Graybill’s Confidence Interval

- \( \gamma_{\text{Low}} = \gamma_1 L = 1 \) and \( \gamma_{\text{High}} = \gamma_1 H = 10 \);
- \( \sigma^2_{\text{Low}} = \sigma^2_L = 0.1 \) and \( \sigma^2_{\text{High}} = \sigma^2_H = 1 \);
- \( k_{\text{Low}} = k_L = 2 \) and \( k_{\text{High}} = k_H = 10 \).

The other factors are the following such as in Example 2.2:

- The true value of \( x_0 \): 0.69727;
- The true value of \( \gamma_0 \): 13.8272;
- The number of observations \( n \): 24;
- The 33 values of \( x \): \( x_1 = 9.71, x_2 = 9.71, x_3 = 9.71, x_4 = 7.96, x_5 = 7.96, x_6 = 7.96, x_7 = 6.82, x_8 = 6.82, x_9 = 6.82, x_{10} = 5.85, x_{11} = 5.85, x_{12} = 5.85, x_{13} = 4.95, x_{14} = 4.95, x_{15} = 4.95, x_{16} = 3.91, x_{17} = 3.91, x_{18} = 3.91, x_{19} = 2.98, x_{20} = 2.98, x_{21} = 2.98, x_{22} = 2.07, x_{23} = 2.07, x_{24} = 2.07 \);
- The confidence coefficient \( \alpha \): 0.05;
- The number of times \( T \) the loop is repeated: 10,000.

The response is the confidence of the interval of Section 4.3, which is the result of the simulation in Section 6.1. We will also consider 4 centrepoints and we will replicate this design 5 times. We will first simulate the 18 points 5 times in Mathematica. These 18 runs are tabulated in Table 6.8. This design is analyzed in Statgraphics and the Analysis-of-Variance table is given in Table 6.9. The value of \( R_{adj}^2 \) is 15.9208, which is very small. It can be seen that the three main effects \( A : \gamma_1, B : \sigma^2 \) and \( C : \text{observations} \) are not significant. Also the \( CC \) effects and the \( AC \) and \( AB \) interactions are not significant. The \( AA \) and \( BB \) effects and the \( AB \) interaction are significant. It can also be seen that the lack-of-fit test is not significant, which implicates that the regression function is linear. The estimations of the effects are given in Table 6.10. A standardized Pareto chart of the effects is given in Figure 6.11. We can conclude that the values of \( \gamma_1, \sigma^2 \) and the number of observations on the unknown \( x_0 \) is not very important, as long as \( \gamma_1 \) is large enough, such that Graybill’s confidence coefficient is approximately for 95% reliable.

\[
\begin{array}{cccccc}
\gamma_1 & \sigma^2 & k & \gamma_1 & \sigma^2 & k \\
1 & 0.1 & 2 & 10 & 0.55 & 6 \\
10 & 0.1 & 2 & 5.5 & 0.1 & 6 \\
1 & 1 & 2 & 5.5 & 1 & 6 \\
10 & 1 & 2 & 5.5 & 0.55 & 2 \\
1 & 0.1 & 10 & 5.5 & 0.55 & 10 \\
10 & 0.1 & 10 & 5.5 & 0.55 & 6 \\
1 & 1 & 10 & 5.5 & 0.55 & 6 \\
10 & 1 & 10 & 5.5 & 0.55 & 6 \\
1 & 0.55 & 6 & 5.5 & 0.55 & 6 \\
\end{array}
\]

Table 6.8: The 18 runs.
### Table 6.9: Analysis-of-Variance table.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A : $\gamma_1$</td>
<td>0.13005</td>
<td>1</td>
<td>0.13005</td>
<td>2.80</td>
<td>0.0982</td>
</tr>
<tr>
<td>B : $\sigma^2$</td>
<td>0.095922</td>
<td>1</td>
<td>0.095922</td>
<td>2.07</td>
<td>0.1545</td>
</tr>
<tr>
<td>C : observations</td>
<td>0.023762</td>
<td>1</td>
<td>0.023762</td>
<td>0.51</td>
<td>0.4763</td>
</tr>
<tr>
<td>A $\times$ A</td>
<td>0.193783</td>
<td>1</td>
<td>0.193783</td>
<td>4.18</td>
<td>0.0444</td>
</tr>
<tr>
<td>A $\times$ B</td>
<td>0.24336</td>
<td>1</td>
<td>0.24336</td>
<td>5.25</td>
<td>0.0248</td>
</tr>
<tr>
<td>A $\times$ C</td>
<td>0.01936</td>
<td>1</td>
<td>0.01936</td>
<td>0.42</td>
<td>0.5202</td>
</tr>
<tr>
<td>B $\times$ B</td>
<td>0.412102</td>
<td>1</td>
<td>0.412102</td>
<td>8.89</td>
<td>0.0039</td>
</tr>
<tr>
<td>B $\times$ C</td>
<td>0.10816</td>
<td>1</td>
<td>0.10816</td>
<td>2.33</td>
<td>0.1309</td>
</tr>
<tr>
<td>C $\times$ C</td>
<td>0.145401</td>
<td>1</td>
<td>0.145401</td>
<td>3.14</td>
<td>0.0807</td>
</tr>
<tr>
<td>Lack-of-fit</td>
<td>0.248384</td>
<td>5</td>
<td>0.0496767</td>
<td>1.07</td>
<td>0.3832</td>
</tr>
<tr>
<td>Pure error</td>
<td>3.47805</td>
<td>75</td>
<td>0.0463741</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.10: The estimated effects.

<table>
<thead>
<tr>
<th>A : $\gamma_1$</th>
<th>-0.102</th>
</tr>
</thead>
<tbody>
<tr>
<td>B : $\sigma^2$</td>
<td>-0.0876</td>
</tr>
<tr>
<td>C : observations</td>
<td>-0.0436</td>
</tr>
<tr>
<td>A $\times$ A</td>
<td>-0.23919</td>
</tr>
<tr>
<td>A $\times$ B</td>
<td>-0.156</td>
</tr>
<tr>
<td>A $\times$ C</td>
<td>0.044</td>
</tr>
<tr>
<td>B $\times$ B</td>
<td>0.34881</td>
</tr>
<tr>
<td>B $\times$ C</td>
<td>0.104</td>
</tr>
<tr>
<td>C $\times$ C</td>
<td>-0.20719</td>
</tr>
</tbody>
</table>
Figure 6.11: Standardized Pareto chart of the effects.
Chapter 7

Conclusion

We have discussed the calibration problem for the simple linear regression model. If we have the following model $Y = \beta_0 + \beta_1 x + \varepsilon$ and an observed value (or a random sample of $k$ values) of the response $Y$, $Y_0$, then we can estimate the corresponding value of the independent variable, $x_0$. The two most used methods are the classical method and the inverse method. These two approaches are introduced in Chapter 1.

In Chapter 2 these two methods are applied to two datasets, which originate from a physical and a chemical example, respectively. In Chapter 3 we give a systematic literature overview of the results on inverse regression. We distinguish five groups: classical against inverse estimator, confidence and discrimination intervals for an unknown predictor, Bayesian solution, optimum design and multivariate calibration.

In Chapter 4 we have presented a slight variation on Graybill’s method, which is not the full Maximum Likelihood approach, to obtain an estimator for the unknown value of $x_0$ and a $100(1 - \alpha)$% confidence interval for $x_0$ for the simple linear regression model. We have investigated the centred simple linear regression model, which is the following:

$$ Y = \gamma_0 + \gamma_1 (x - \bar{x}) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \quad (7.1) $$

This model is based on $n + k$ observations

$$ (x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n), (x_0, Y_{n+1}), (x_0, Y_{n+2}), \ldots, (x_0, Y_{n+k}). \quad (7.2) $$

Firstly, the Least Squares estimators of $\gamma_0$ and $\gamma_1$, based on the first $n$ observations, are obtained. The estimator of $x_0$, which is based on all the $n + k$ observations, is

$$ \hat{x}_0 = \bar{x} + \frac{\bar{Y}_0 - \bar{Y}}{\hat{\gamma}_1}. \quad (7.3) $$

Because this estimator of $x_0$ is biased, we also present Naszódi’s estimator, which is approximately corrected for bias.

Afterwards, we calculate the estimator for $\sigma^2$ by plugging in the Least Squares estimates of the intercept $\gamma_0$, the slope $\gamma_1$ and $x_0$ in the Maximum Likelihood function. After maximizing the log likelihood function we found the Maximum Likelihood estimator of $\sigma^2$. We have also obtained the unbiased estimator of $\sigma^2$. If we want to obtain a confidence interval for $x_0$, then we first have to test the following hypothesis $H_0 : \gamma_1 = 0$ against $H_a : \gamma_1 \neq 0$ with the standard $t$-test. Only if $H_0$ is rejected we have a $100(1 - \alpha)$% confidence interval for $x_0$. 
Graybill claimed that this is not a $100(1 - \alpha)\%$ confidence interval for $x_0$ but it has confidence coefficient less than $1 - \alpha$, which we have investigated in Chapter 6.

In Chapter 5 we considered Brown’s method with the profile likelihood to obtain the estimator of $x_0$. This method applies the full Maximum Likelihood method on the $n + k$ observations to obtain the estimators of the intercept, the slope, $\sigma^2$ and the unknown $x_0$. First, we obtained the log likelihood maximized over three parameters (intercept, slope, $\sigma^2$) for a fixed $x_0$. Then, the maximizing values of the intercept, the slope and $\sigma^2$ are functions of $x_0$. The log likelihood is maximized, when $\hat{\sigma}^2$ is minimized, such that we can obtain the estimator of $x_0$ algebraically. The main difference with Graybill’s method is that Brown uses the full maximum likelihood approach. Graybill plugs in the Least Squares estimators of the intercept, slope and $x_0$ in the log likelihood to obtain the estimator of $\sigma^2$. We supplied missing details of the extensive calculations. We also identified and corrected several mistakes. We applied this estimator, which we derived algebraically on the data of Example 2.2 of Chapter 2. We have also presented a graphical method to obtain an estimator of $x_0$, which we applied on the data of Example 2.2.

In Chapter 6 we have given a simulation of the $100(1 - \alpha)\%$ confidence interval of Graybill’s method to investigate Graybill’s claim. We have only investigated 4 specific examples. From these results we can conclude that Graybill’s claim is true, but the difference is very small. It is even less than $0.5\%$. We have also investigated the effect of three factors with a face-centred central composite design. These factors are the slope $\gamma_1$, the variance $\sigma^2$ and the number of observations $k$ on the unknown $x_0$. We have distinguished two designs. For the first design we have chosen $\gamma_{1,Low}$ very small, such that in some cases the standard $t$-test, which tests $H_0 : \gamma_1 = 0$ against $H_1 : \gamma_1 \neq 0$, does not reject $H_0$. We can conclude that the values of the values of the factors $\gamma_1$, $\sigma^2$ and observations have an influence on the confidence coefficient. For the second design we take $\gamma_1$ large enough, such that $H_0$ is always rejected. Now we can see that the values of $\gamma_1$, $\sigma^2$ and the number of observations are not very important, as long as $\gamma_1$ is high enough, such that Graybill’s confidence coefficient is approximately for 95\% reliable.

We can conclude that Graybill’s method to obtain an estimator of an unknown $x_0$ and a $100(1 - \alpha)\%$ confidence interval is reliable. Naszódi’s estimator and Brown’s profile likelihood estimator do not differ much from Graybill’s method, such that they are also reliable.
Bibliography


Appendix A

Program Codes

A.1 Procedure for the Confidence Interval of Section 4.3

```mathematica
f2[x_, y_, C_, \[Alpha]_: 0.05] :=
Module[{i, j, k = Length[C], n = Length[x], xbar, ygem, y0bar, gamma1,
gamma0, x0, sigma2, a, midden, afwijking},
If[Length[x] == Length[y], xbar = Mean[x]; ygem = Mean[y];
y0bar = Mean[C];
gamma1 = Sum[(y[[i]] - ygem) (x[[i]] - xbar), {i, 1, n}]/Sum[(x[[j]] - xbar)^2, {j, 1, n}], {i, 1, n}];
\[IndentingNewLine]
\[IndentingNewLine]gamma0 = ygem; \[IndentingNewLine]x0 = xbar + \ y0bar - ygem)/gamma1;
\[IndentingNewLine]sigma2 = (1/(n + k - 3) * (Sum[((y[[i]]) - gamma0 - gamma1 (x[[i]] - xbar))^2, {i, 1, n}] + Sum[(C[[i]] - y0bar)^2, {i, 1, k}]))/n); \[IndentingNewLine]
\[IndentingNewLine]Q = Quantile[StudentTDistribution[n + k - 3],
1 - \[Alpha]/2]^2; \[IndentingNewLine]a =
gamma1^2 - \[SquareRoot][sigma2] * ((1/(n + k - 3)) * (Sum[((y[[i]]) - gamma0 - gamma1 (x[[i]] - xbar))^2, {i, 1, n}] + Sum[(C[[i]] - y0bar)^2, {i, 1, k}]])/n); \[IndentingNewLine]
\[IndentingNewLine]midden = xbar + gamma1 ((y0bar - ygem))/a;
\[IndentingNewLine]afwijking = \[SquareRoot][Q] * Sqrt[ sigma2]/a)
\[IndentingNewLine]
If[gamma1^2 * Sum[(x[[i]] - xbar)^2, {i, 1, n}] / sigma2 >= Q,
Return[{-Infinity, Infinity, x0, 1}]]; \[IndentingNewLine]
Return[{midden - afwijking, midden + afwijking, x0, 0}]]]]](\[IndentingNewLine])
```

\[
Sim[aantal_, xnullcht_:0.633021, \[Alpha]_:0.05, delta0_:12.1544,
delta1_:2.45092, xgem_:4.89, sigmaecht_:0.167885, k_:10] :=
\]
Module[{teller = 0, reject = 0, y, CC, Int},
  Do[y = Table[
    Random[NormalDistribution[delta0 + delta1 (x[[i]] - xgem),
      Sqrt[sigmaeucht]]], {i, 1, Length[x]}];
  CC =
    Table[Random[
      NormalDistribution[delta0 + delta1 (xnulecht - xgem),
        Sqrt[sigmaeucht]]], {i, 1, k}]; Int = f2[x, y, CC, \[Alpha]];
  If[Int[[4]] == 1, reject = reject + 1];
  If[xnulecht <= Int[[2]] && xnulecht >= Int[[1]],
    teller = teller + 1], {j, 1, aantal}]; Return[{teller, reject}]]