Students	**Student IDs**
Max van Esso	73539
Marco Fayet	125593
Felix Gutmann	125604

1 Homework lecture 1

1.1 Excercise 1

To show: $X^T X$ is positive semi definite, where X is a $p \times q$ matrix

Starting from:

$y^T(X^T X)y \geq 0$, where y is $q \times 1$ vector

$$=y^T X^T Xy \geq 0$$
$$=(Xy)^T(Xy) \geq 0$$

(Xy) can be substituted as a $p \times 1$ vector v, which leads to:

$$= v^T v \geq 0 \qquad \rightarrow \qquad v^T v = v_1^2 + \cdots + v_p^2 \geq 0$$

In the dot product of a vector times itself each element is at least zero, because each element gets squared and therefore $X^T X$ is positive semi definite.

1.2 Excercise 2

Given the high level of positive correlation between the log of cancer volume (lcavol) and the log of capsular penetration (lcp), and given that a high level of prostate specific antigens (psa) is usually an accompanying symptom of prostate cancer, one would expect both lcavol and lcp to be positively correlated with psa.

However, the linear model presented in table 3.2 shows a negative coefficient associated with the lcp variable, and a positive one with lcavol.

From a medical standpoint, this could be due to the fact that a high level of psa can be linked to causes other than prostate cancer.

From a statistical standpoint, this is an example of Simpson's paradox, in which "a trend appears in different groups of data but disappears or reverses when these groups are combined" (wikipedia). A simple illustration can be extracted from the relationship between the number of robberies and the size of a police force: while these two events are positively correlated, the amount of property stolen would be positively correlated with the former yet negatively correlated with the latter.

This paradox demonstrates the limits of human intuition. Furthermore, it emphasizes the need for cautiousness when one tries to infer statistical causality, or attempts to extrapolate or interpret the correlation between variables.

1.3 Excercise 3

Excercise 3.1

To show: $H^T = H$

Used transpose properties:

- $(AB)^T = B^T A^T$
- $((AB)^{-1})^T = (B^T A^T)^{-1}$
- $(A^T)^T = A$

Proof:

$$
\begin{aligned}
&H = \phi(\phi^T\phi)^{-1}\phi^T \\
&H^T = (\phi(\phi^T\phi)^{-1}\phi^T)^T \\
&H^T = (\phi^T)^T((\phi^T\phi)^{-1})^T\phi^T = \phi(\phi^T\phi)^{-1}\phi^T = H
\end{aligned}
$$

Excercise 3.2

To show: $H^2 = H$

Using the definition of an Invers:

- $AA^{-1} = A^{-1}A = I$

Proof:

$$
\begin{aligned}
H &= \phi(\phi^T\phi)^{-1}\phi^T \\
H^2 &= (\phi(\phi^T\phi)^{-1}\phi^T)(\phi(\phi^T\phi)^{-1}\phi^T) \\
&= \phi(\phi^T\phi)^{-1}(\phi^T\phi)(\phi^T\phi)^{-1}\phi^T \\
&= \phi(\phi^T\phi)^{-1}\phi^T = H
\end{aligned}
$$

Excercise 3.3

To show: $tr(H) = M + 1$

Prerequisite information for proof:

- ϕ is an $N \times M + 1$ Matrix
- $(\phi^T\phi)$ is then a $M + 1 \times M + 1$ Matrix
- The trace has a cyclical property: $tr(ABC) = tr(BCA)$

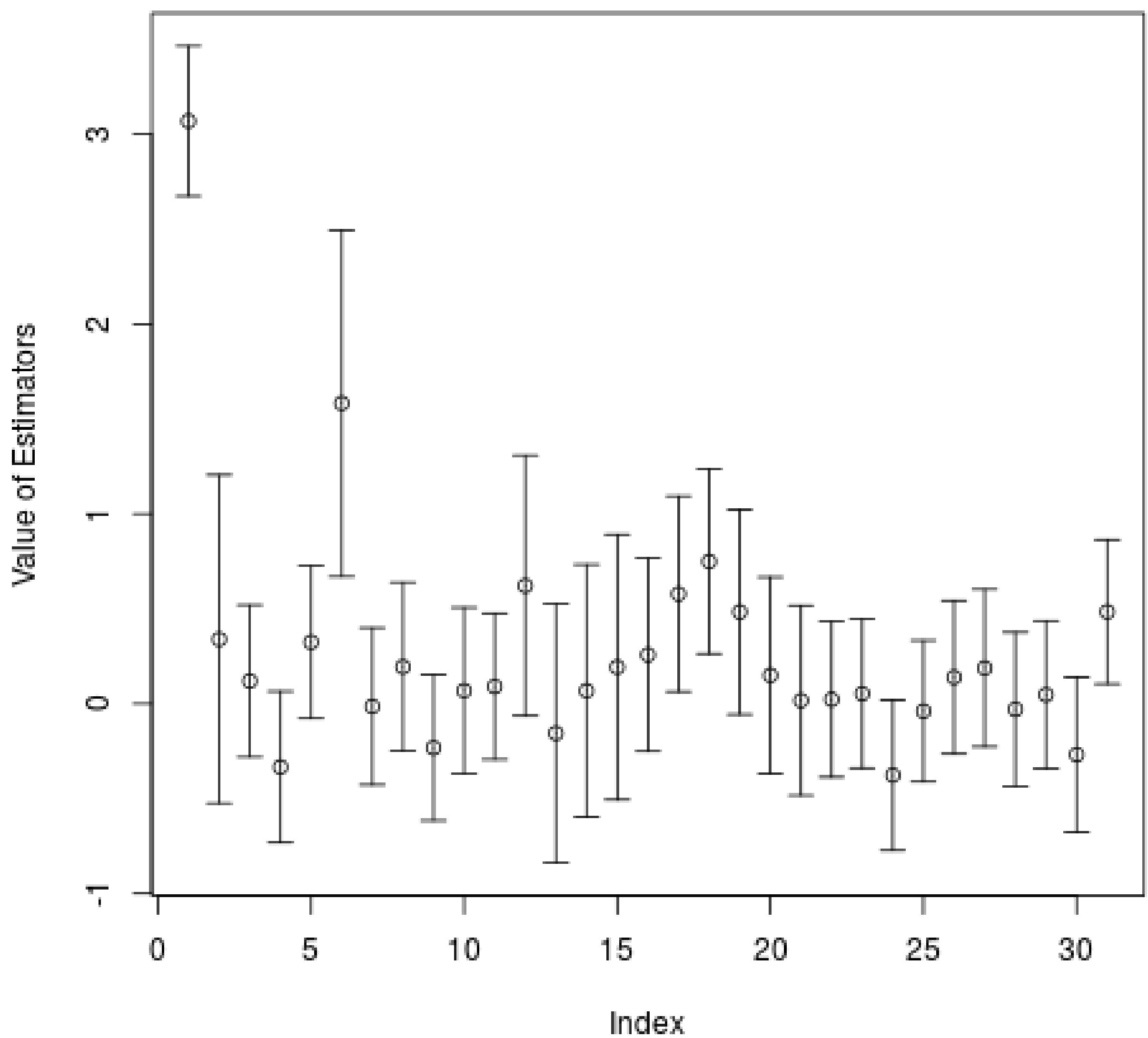
- Definition of an invertible matrix $AA^{-1} = A^{-1}A = I$
- I is the identity matrix

Proof:

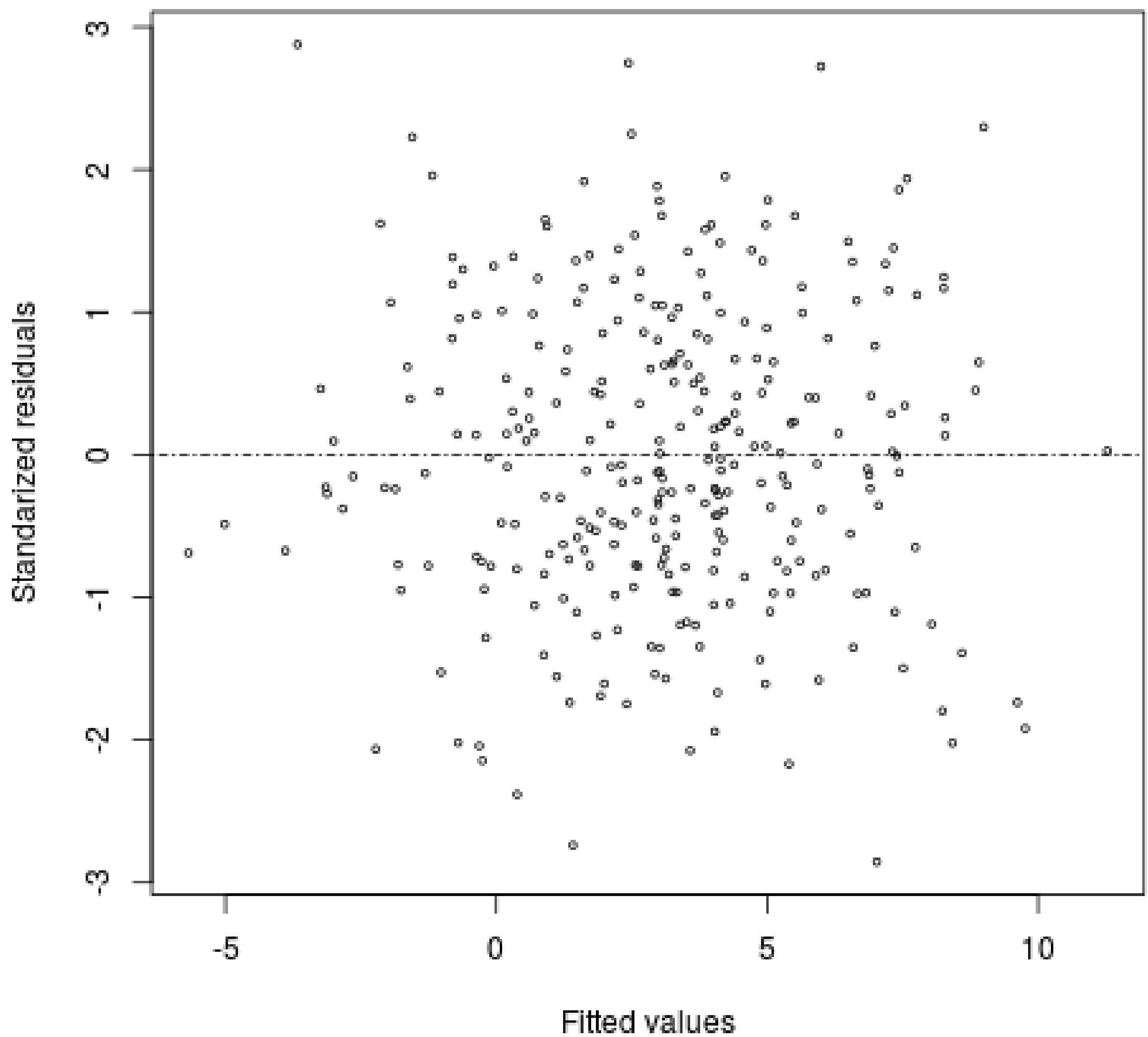
$$
\begin{aligned}
&tr(\phi(\phi^T\phi)^{-1}\phi^T) \\
=&tr((\phi^T\phi)(\phi^T\phi)^{-1}) \\
=&tr(I_{M+1\times M+1}) \\
=&\sum_{i=1}^{M+1} I_{ii} = M+1
\end{aligned}
$$

1.4 R-excercises homework 1

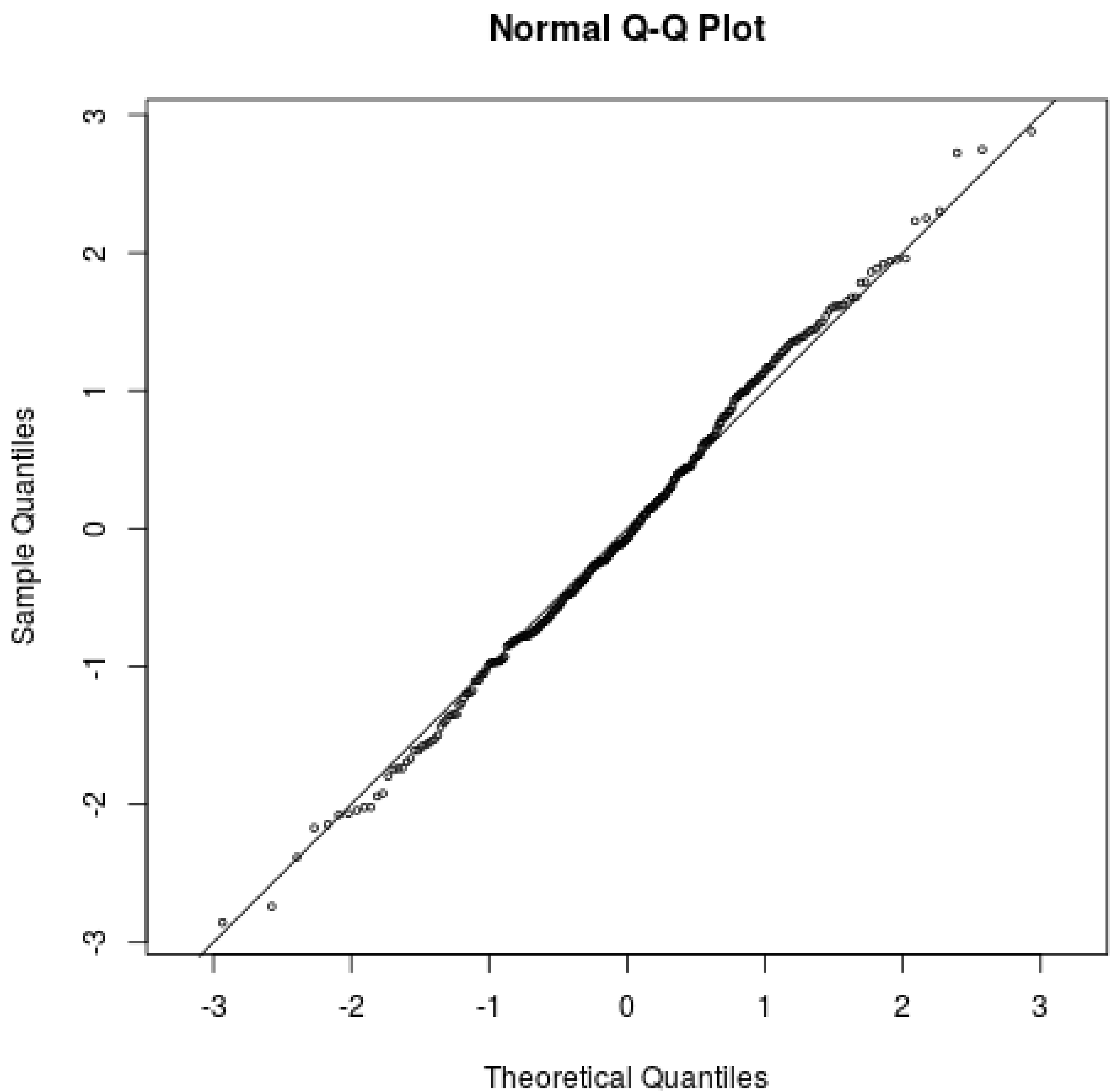
1.4.1 Estimators with 1.96 Errors



1.4.2 Standarized residuals vs. fitted values



1.4.3 Q-Q-Plot



Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

2 Homework lecture 2

2.1 Excercise 1

To Show: $V_r V_r^T$ is a projection matrix 1.1) with trace equals r 1.2)

Excersice 1.1.

In course of showing that $V_r V_r^T$ is a Projection Matrix we have to show two properties. First that it is indempotent and the transpose is euqal to itself (See also homework 1):

To Show: $(V_r V_r^T)^2 = V_r V_r^T$

Used property:

- $V_r^T V_r = I$, (columns are orthogonal)

Proof:

$$\begin{aligned}(V_r V_r^T)^2 &= (V_r V_r^T)(V_r V_r^T) \\ &= V_r (V_r^T V_r) V_r^T \\ &= V_r V_r^T\end{aligned}$$

To Show: $(V_r V_r^T)^T = V_r V_r^T$

Used property of a transpose:

- $(AB)^T = B^T A^T$

Proof:

$$\begin{aligned}(V_r V_r^T)^T &= (V_r^T)^T (V_r)^T \\ &= (V_r V_r^T)\end{aligned}$$

Excersice 1.2.

To Show: $tr(V_r V_r^T) = r$

Used properties:

- The trace has a cyclical property: $tr(ABC) = tr(BCA)$
- $V_r^T V_r$ is a $r \times r$ Matrix
- $V_r^T V_r = I$, (columns are orthogonal)

- I is the identity matrix

Proof:

$$\begin{aligned} tr(V_r V_r^T) &= tr(V_r^T V_r) \\ &= tr(I_{r\times r}) \\ &= \sum_{i=1}^{r} I_{ii} = r \end{aligned}$$

2.2 Excercise 2

To show: For $t|x \sim N(\phi w, (qD)^{-1})$ the estimator becomes $\phi^T Dt = \phi^T D\phi w$

Used property:

- D is the precision Matrix and therefore per definition symetric ($D = D^T$)

Replace qI in the former log likelihood equation by qD:

$$\begin{aligned} l &= (t-\phi w)^T qD(t-\phi w) + const \\ &= t^T Dt - t^T D\phi w - w^T \phi^T Dt + w^T \phi^T D\phi w + const \end{aligned}$$

Maximaize the log likelihood function (l) via setting the derivative to zero:

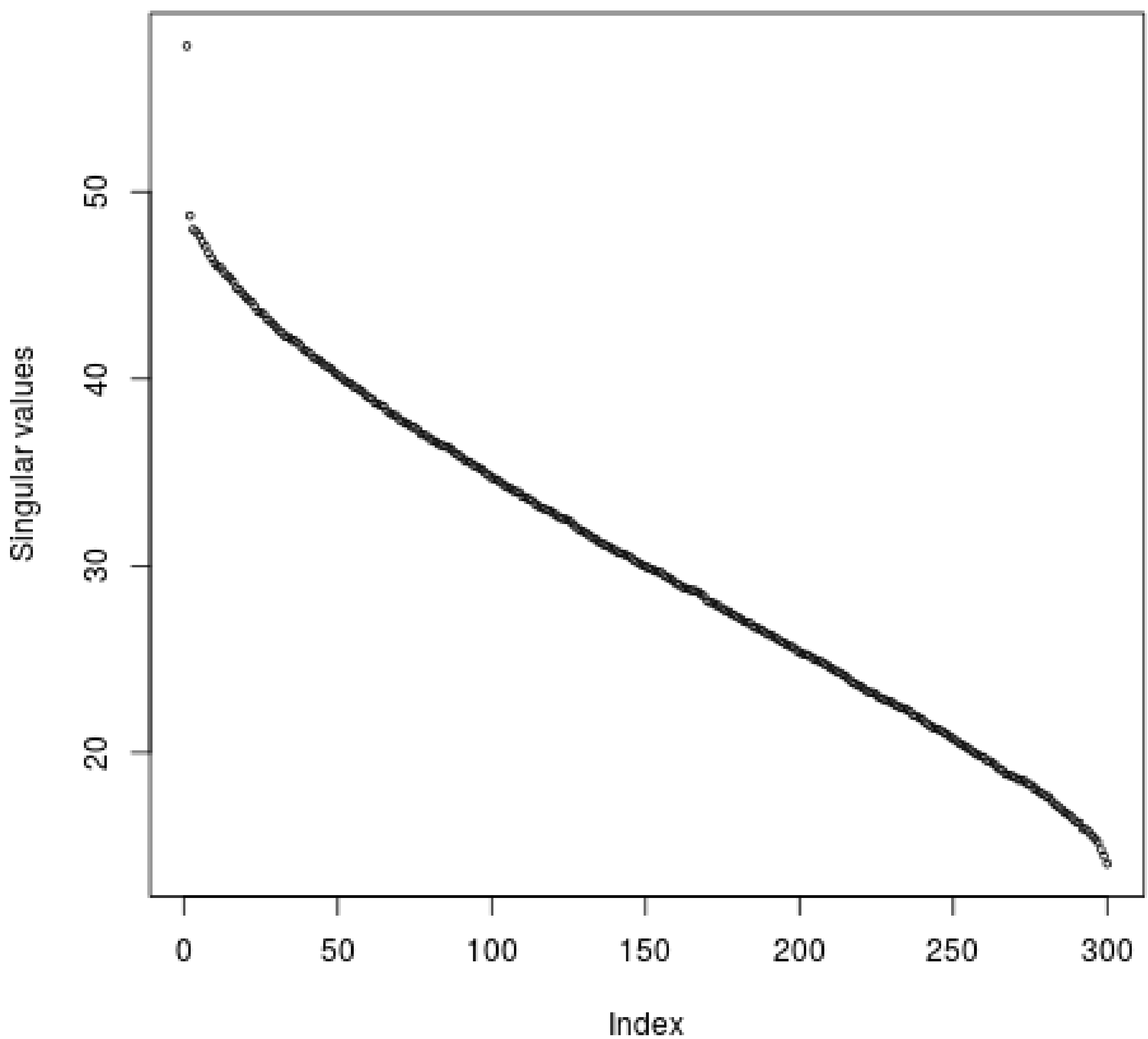
$$\begin{aligned} \frac{\partial l}{\partial w} &\stackrel{!}{=} 0 \\ &= -t^T D\phi - t^T D^T \phi + w^T(\phi^T D\phi + (\phi^T D\phi)^T) = 0 \\ 2t^T D^T \phi &= w^T(2\phi^T D\phi) \end{aligned}$$

Take the transpose of both sides:

$$\begin{aligned} (t^T D^T \phi)^T &= (w^T(\phi^T D\phi))^T \\ \phi^T Dt &= \phi^T D\phi w \end{aligned}$$

2.3 R Excercises Homework 2

2.3.1 Singular Values of ϕ



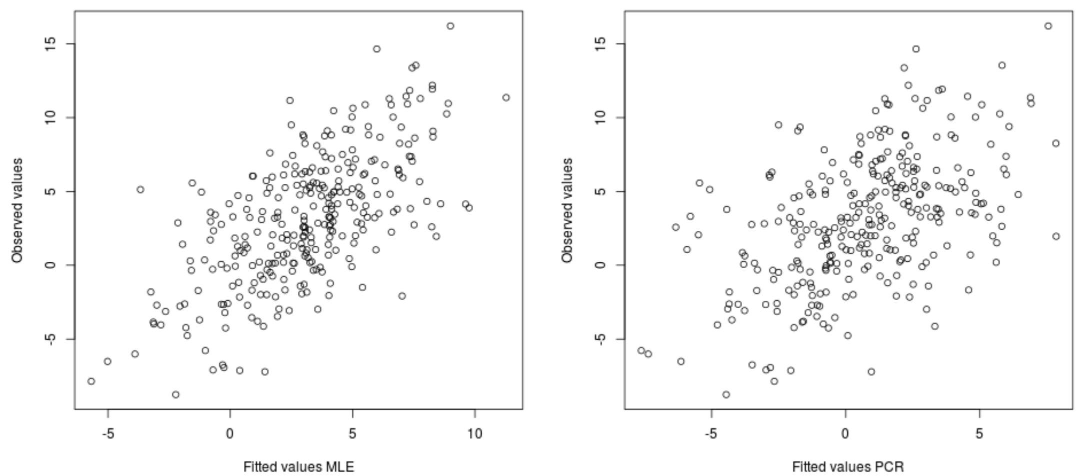
2.3.2 Rank of the input matrix

The rank of ϕ is equal to the rank of $\phi^T\phi$. We are using the matrix.rank comand from the matrixcalc package. The variable name of ϕ in our R code is "new.design.matrix".

Hence the following code provides the rank of the input matrix ϕ, which is **300**.

Code

```
matrix.rank( t(new.design.matrix) %*% new.design.matrix )
```



(a) MLE results

(b) PCR-Results

Figure 1: Compare fits

2.3.3 Compare MLE fit and PCA fit

The fitted values of the MLE regression are closer to the observed values. However, PCR is used in case of high dimensional data and selects features based on their level of variance for the whole sample. The results of the PCR estimators are more stable with a trade off with respect to precision.