SQ2

###Results of SQ2: Do farmers with more intensive systems have relatively more forest reserves? This chapter models the effect of SPS, stocking rate, distance, milk yield per hectare, labour hours per hectare, total factor productivity and distance on forest reserves in a multiple linear regression.

The necessary libraries and the dataframe si.df are loaded into Rstudio. The library “psych” is used to visualize the variables in a convenient way. It displays histograms, scatterplot and correlations of forest reserves, stocking rates, annual labour hours per hectare, milk yield per hectare and distance.

library(tidyverse)

## -- Attaching packages ------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

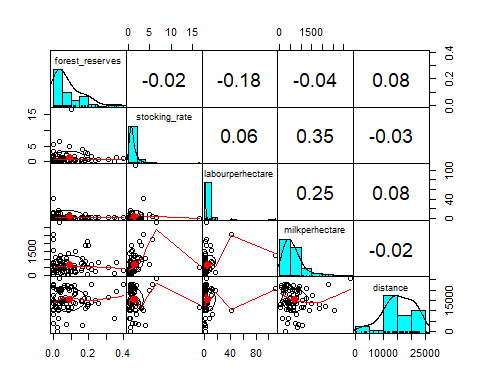
## -- Conflicts ---------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

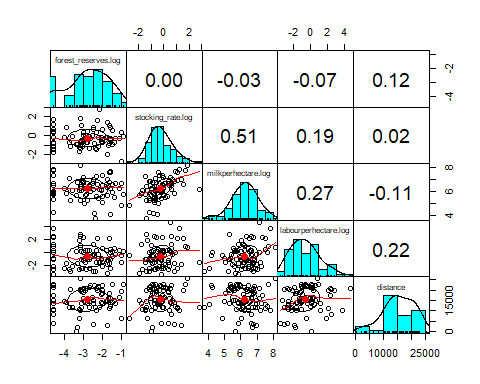
library(ggplot2)  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

load("Dataframes/si.df.rda")  
  
model.var <- c(12,32, 36, 37, 14) # forest\_reserves,stocking rate, milkperhectare, labourperhectare, distance  
pairs.panels(si.df[,model.var],  
 gap=0,  
 bg=c("red")[si.df$SPS],  
 pch = 21)

 Forest reserves, stocking rates, annual milk yield and annual labour days per hectare are log transformed to reach more normally distributed variables. Forest reserves are transformed using log(x+00.1) to avoid infinitive values. The code is seen below. The other variables were already transformed for SQ1. Scatterplots and correlations are visualized thereafter.



**Removeing Outliers using the Mahalanobis distance**

To detect and remove outliers the Mahalanobis distance is calculated. The Mahalanobis distance calculates the standard Euclidean distance of each point to the mean. It is unitless, scale-invariant and includes correlations (Wikipedia!! Scite correctly)

si.df<-si.df[!is.na(si.df$forest\_reserves.log),] # remove NA in forest\_reserves.log  
si.mhlnbs\_outl <- mahalanobis(si.df[,log.var],colMeans(si.df[,log.var]),cov(si.df[,log.var]))  
# Calculated Mahalanobis distances

The Mahalanobis distance is added to the dataframe si.df and a logical variable created. Outliers are defined as mhlnbs > 11 in accordance with the visualization thereafter.

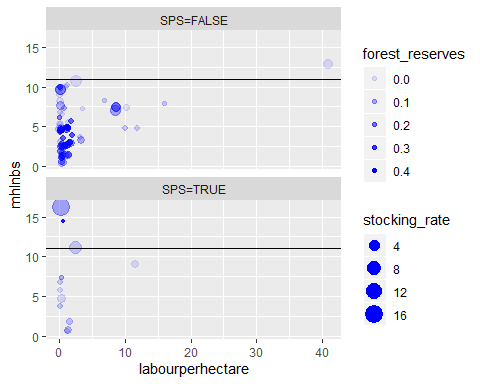
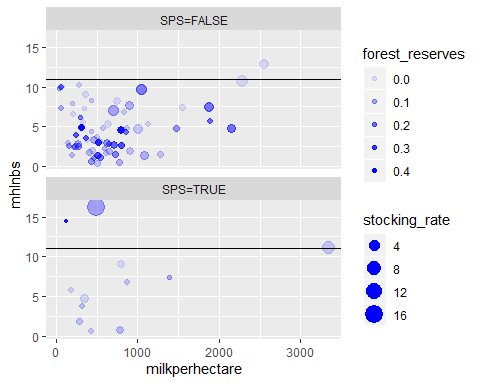
si.df$mhlnbs <- round(si.mhlnbs\_outl,3)  
summary(si.df$mhlnbs)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.369 2.364 4.497 4.938 7.318 16.272

outlier <- 11.0  
si.df$outlier\_mhlnbs <- ifelse(si.df$mhlnbs > outlier, TRUE, FALSE)  
summary(si.df$outlier\_mhlnbs)

## Mode FALSE TRUE   
## logical 76 4

The scatterplots show the Mahalanobis distance on the y axis. The horizontal line displays the boundary of mhlnbs = 11. The x axis show annual milk yield on the first and annual labour days on the second figure. Conventional farmers are displayed on the upper half and SPS farmers on the lower hald of each figure. Four outliers are indentified of which 3 are SPS farmers. The conventional farmer has a exeptionally high amount of annual labour days. Two SPS farmers have a high stocking rate. One of them has a very high milk yield per hectare and the other one very high forest reserves. Subsequently the outliers are removed from the dataframe.



si.df <- filter(si.df, outlier\_mhlnbs==FALSE) #remove outliers from dataframe

**Multiple linear regression model** A multiple linear regression model is built to test if the log-stocking rate, the log-annual milk yield per hectarem the log-annual labour days per hectare, the distance to the highway and the dummy SPS have a combined affect on the forest reserves of farmers. The model statistics are displayed below.

fr\_sr\_sps\_distance\_reg <- lm(forest\_reserves.log ~ stocking\_rate.log+milkperhectare.log+labourperhectare.log+SPS+distance, data = si.df)  
summary(fr\_sr\_sps\_distance\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ stocking\_rate.log + milkperhectare.log +   
## labourperhectare.log + SPS + distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.4744 -0.5667 0.1171 0.7556 1.5952   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.019e+00 1.290e+00 -3.116 0.00266 \*\*  
## stocking\_rate.log 1.048e-01 1.733e-01 0.604 0.54755   
## milkperhectare.log 1.091e-01 1.791e-01 0.609 0.54434   
## labourperhectare.log -6.591e-02 8.308e-02 -0.793 0.43027   
## SPSTRUE -5.855e-01 3.747e-01 -1.562 0.12269   
## distance 4.047e-05 2.379e-05 1.701 0.09334 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.051 on 70 degrees of freedom  
## Multiple R-squared: 0.07543, Adjusted R-squared: 0.009388   
## F-statistic: 1.142 on 5 and 70 DF, p-value: 0.3465

step(fr\_sr\_sps\_distance\_reg)

## Start: AIC=13.27  
## forest\_reserves.log ~ stocking\_rate.log + milkperhectare.log +   
## labourperhectare.log + SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - stocking\_rate.log 1 0.4032 77.679 11.661  
## - milkperhectare.log 1 0.4097 77.686 11.667  
## - labourperhectare.log 1 0.6948 77.971 11.945  
## <none> 77.276 13.265  
## - SPS 1 2.6950 79.971 13.871  
## - distance 1 3.1950 80.471 14.344  
##   
## Step: AIC=11.66  
## forest\_reserves.log ~ milkperhectare.log + labourperhectare.log +   
## SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - labourperhectare.log 1 0.6184 78.297 10.263  
## - milkperhectare.log 1 1.0858 78.765 10.716  
## <none> 77.679 11.661  
## - SPS 1 2.7439 80.423 12.299  
## - distance 1 3.2633 80.942 12.788  
##   
## Step: AIC=10.26  
## forest\_reserves.log ~ milkperhectare.log + SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## - milkperhectare.log 1 0.77375 79.071 9.0108  
## <none> 78.297 10.2634  
## - SPS 1 2.56504 80.863 10.7133  
## - distance 1 2.71083 81.008 10.8502  
##   
## Step: AIC=9.01  
## forest\_reserves.log ~ SPS + distance  
##   
## Df Sum of Sq RSS AIC  
## <none> 79.071 9.0108  
## - distance 1 2.2300 81.301 9.1245  
## - SPS 1 2.4961 81.567 9.3729

##   
## Call:  
## lm(formula = forest\_reserves.log ~ SPS + distance, data = si.df)  
##   
## Coefficients:  
## (Intercept) SPSTRUE distance   
## -3.1994742 -0.5615511 0.0000315

fr\_sps\_distance\_reg <- lm(forest\_reserves.log ~ SPS + distance, data = si.df)  
summary(fr\_sps\_distance\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ SPS + distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.1310 -0.5596 0.1583 0.8116 1.5957   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.199e+00 3.557e-01 -8.996 1.9e-13 \*\*\*  
## SPSTRUE -5.616e-01 3.699e-01 -1.518 0.133   
## distance 3.150e-05 2.195e-05 1.435 0.156   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.041 on 73 degrees of freedom  
## Multiple R-squared: 0.05395, Adjusted R-squared: 0.02803   
## F-statistic: 2.081 on 2 and 73 DF, p-value: 0.1321

The result shows no significance for the F-statistics. It therefore is not possible to say that the exogenous variables have no combined effect on the forest reserves. The adjusted r-square value is extremely low, therefore the explanatory power of the model is low as well. The single variables are not significant at the 95 percent level. Merely distance is significant at the 90 percent level. The signs of the estimates suggest however that with an increasing stocking rate forest reserves slightly increase. The same holds for annual milk yields. More labour hours per hectare reduce forest reserves. SPS reduces the forest reserves and with higher distance the forest reserves grow. As the significance of the model is not given however, no clear absolute statement can be given.

**T-test of mean log forest reserves of SPS and conventional farmers**  
The former model gave hints that SPS reduces the forest reserves of farmers.Subsequently a one sided t-test is executed to test whether conventional farmers indeed have more forest reserves than SPS farmers.

#si.df <- si.df[is.finite(si.df$forest\_reserves),] #remove non-finite values of forest\_reserves  
fr\_ttest <- t.test(x=si.df$forest\_reserves.log[si.df$SPS==FALSE],  
 y=si.df$forest\_reserves.log[si.df$SPS==TRUE], alternative = "greater"  
 )  
fr\_ttest

##   
## Welch Two Sample t-test  
##   
## data: si.df$forest\_reserves.log[si.df$SPS == FALSE] and si.df$forest\_reserves.log[si.df$SPS == TRUE]  
## t = 2.0254, df = 14.047, p-value = 0.03113  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.07000143 Inf  
## sample estimates:  
## mean of x mean of y   
## -2.722864 -3.258802

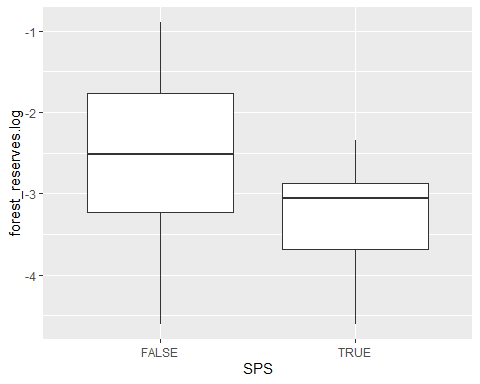
print("Log-transformation:")

## [1] "Log-transformation:"

exp(fr\_ttest$estimate)-0.01

## mean of x mean of y   
## 0.05568636 0.02843440

ggplot(data = si.df, mapping=aes(y=forest\_reserves.log, x=SPS))+  
 geom\_boxplot()

 The test results show that conventional farmers have significantly more forest reserves at the 95 percent level. While conventional farmers on average have 5.57 percent of their total area in forest the SPS farmers have merely 2.84 percent of their area in forest.