Deforestation

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.

Load dataframe

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)

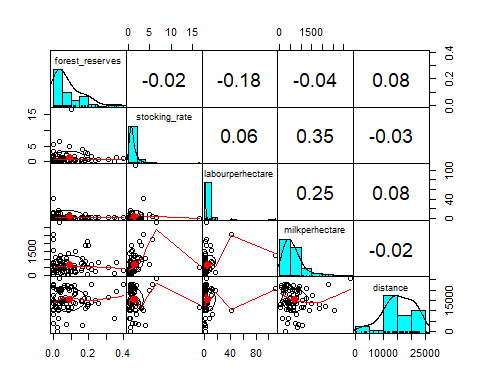
##   
## Attaching package: 'psych'

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

load("Dataframes/si.df.rda")

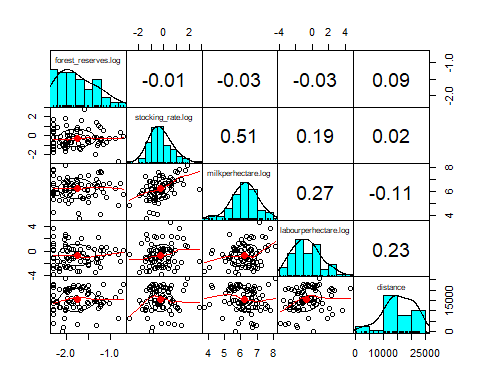
Scatterplot and correlations

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", "blue", "yellow")[si.df$SPS],  
 pch = 21)

 Forest reserves, stocking rates, annual milk yield annual labour days per hectare are log transformed

Scatterplot and correlations

log.var <- c(44,41:43,14) # forest\_reserves.log,stocking\_rate.log , milkperhectare.log, labourperhectare.log, distance  
pairs.panels(si.df[,log.var],  
 gap=0,  
 bg=c("red", "blue", "yellow")[si.df$SPS],  
 pch = 21)



## Removeing Outliers using the Mahalanobis distance

Find Outliers using Mahalanobis distance of “distance” ,“stocking\_rate”, “forest\_reserves”, “labourperhectare”, “milkperhectare”. Mahalanobis distance calculates the standard Euclidean distance of each point to the mean. It is unitless, scale-invariant and includes correlations (Wikipedia!! Scite correctly)

summary(si.df[log.var])

## forest\_reserves.log stocking\_rate.log milkperhectare.log  
## Min. :-2.3026 Min. :-2.7568 Min. :3.785   
## 1st Qu.:-2.0551 1st Qu.:-0.8476 1st Qu.:5.732   
## Median :-1.8445 Median :-0.3245 Median :6.236   
## Mean :-1.7617 Mean :-0.2541 Mean :6.218   
## 3rd Qu.:-1.4917 3rd Qu.: 0.3576 3rd Qu.:6.743   
## Max. :-0.6931 Max. : 2.8168 Max. :8.112   
## NA's :3   
## labourperhectare.log distance   
## Min. :-3.8067 Min. : 161.8   
## 1st Qu.:-1.7748 1st Qu.:11776.1   
## Median :-0.9634 Median :14876.6   
## Mean :-0.6133 Mean :15365.9   
## 3rd Qu.: 0.2886 3rd Qu.:20299.4   
## Max. : 4.6965 Max. :25536.8   
##

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

Add Mahalanobis distance to dataframe si.df

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

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.379 2.245 4.529 4.938 7.032 16.635

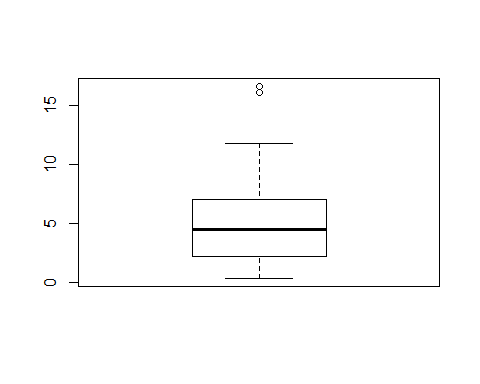
#save(si.df, file = "Dataframes/si.df.rda")

Boxplot Statistics of Mahalanobis distance

boxplot.stats(si.df$mhlnbs)

## $stats  
## [1] 0.3790 2.1995 4.5295 7.0320 11.8230  
##   
## $n  
## [1] 80  
##   
## $conf  
## [1] 3.675842 5.383158  
##   
## $out  
## [1] 16.635 16.113

boxplot(si.df$mhlnbs)



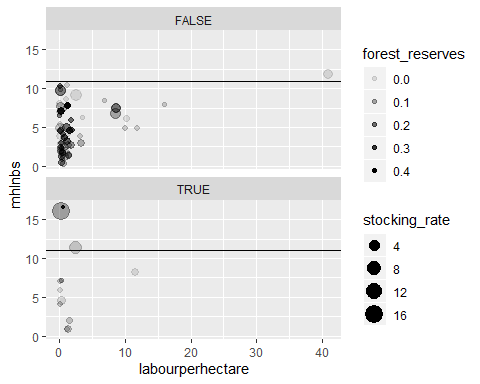
The boxplot identifies a Mahalanobis distance of bigger than 16 as outlier.

Create binomial variable of Mahalanobis Outliers, level 10 is chosen

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

2 outliers are identified

Visualize Outliers - Point clouds, above are SPS=FALSE, below SPS=TRUE. y axis displays Mahalanobis distance, the horizontal line shows the 14 boundary. 

Build subset without Mahalanobis Outliers

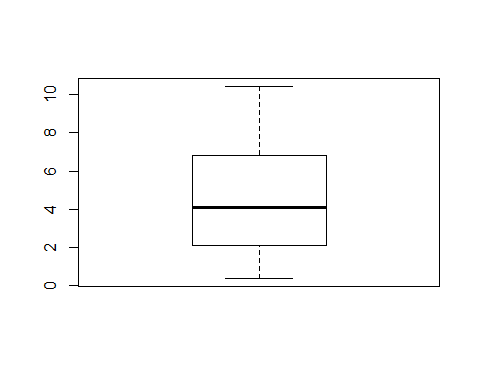
si.df <- filter(si.df, outlier\_mhlnbs==FALSE)

Boxplot Statistics of Mahalanobis distance

boxplot.stats(si.df$mhlnbs)

## $stats  
## [1] 0.3790 2.0805 4.0820 6.8310 10.4410  
##   
## $n  
## [1] 76  
##   
## $conf  
## [1] 3.221027 4.942973  
##   
## $out  
## numeric(0)

boxplot(si.df$mhlnbs)

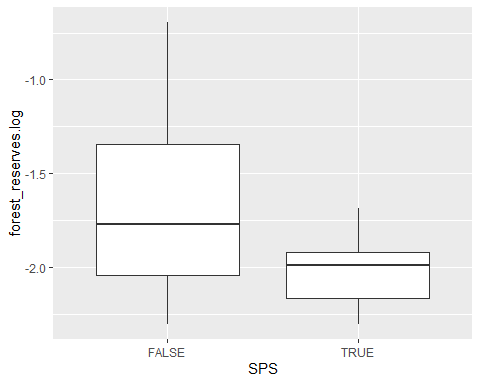


T-test of mean log forest reserves of SPS and conventional farmers

#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]  
 )  
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 = 3.4048, df = 22.279, p-value = 0.00251  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.1091206 0.4485692  
## sample estimates:  
## mean of x mean of y   
## -1.732999 -2.011843

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

 SPS farmers have significantly less forest reserves than conventional farmers

Transfroming back into level results in geometric means.

exp(fr\_ttest$estimate)-0.1

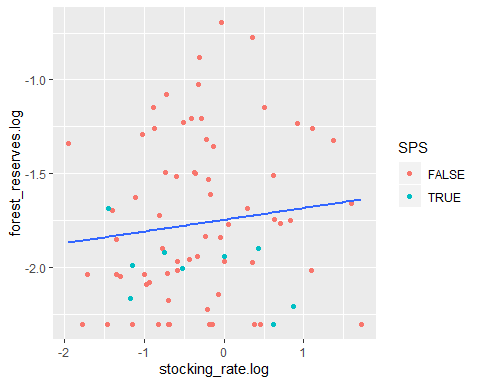
## mean of x mean of y   
## 0.07675362 0.03374191

Simple linear regression of stocking rate as independent and forest reserves as dependent variable

fr\_stock\_reg <- lm(forest\_reserves.log ~ stocking\_rate.log, data = si.df)  
summary(fr\_stock\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ stocking\_rate.log, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.66397 -0.27614 -0.09075 0.28346 1.05505   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.74638 0.05195 -33.616 <2e-16 \*\*\*  
## stocking\_rate.log 0.06256 0.06079 1.029 0.307   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4212 on 74 degrees of freedom  
## Multiple R-squared: 0.01411, Adjusted R-squared: 0.0007832   
## F-statistic: 1.059 on 1 and 74 DF, p-value: 0.3068

ggplot(data = si.df, aes(x = stocking\_rate.log, y = forest\_reserves.log)) +   
 geom\_point(mapping = aes(color=SPS)) +  
 geom\_smooth(method = "lm", se = FALSE)

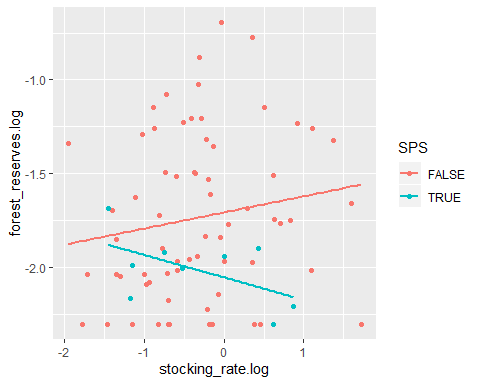


Regression of stocking rate as independent, SPS as dummy and forest reserves as dependent variable

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

##   
## Call:  
## lm(formula = forest\_reserves.log ~ stocking\_rate.log + SPS, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.6931 -0.2744 -0.0632 0.2776 1.0228   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.71420 0.05388 -31.818 <2e-16 \*\*\*  
## stocking\_rate.log 0.06078 0.05979 1.017 0.3127   
## SPSTRUE -0.27649 0.14707 -1.880 0.0641 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4142 on 73 degrees of freedom  
## Multiple R-squared: 0.05964, Adjusted R-squared: 0.03387   
## F-statistic: 2.315 on 2 and 73 DF, p-value: 0.106

ggplot(data = si.df, aes(x = stocking\_rate.log, y = forest\_reserves.log, color=SPS)) +   
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)



Including Distance

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

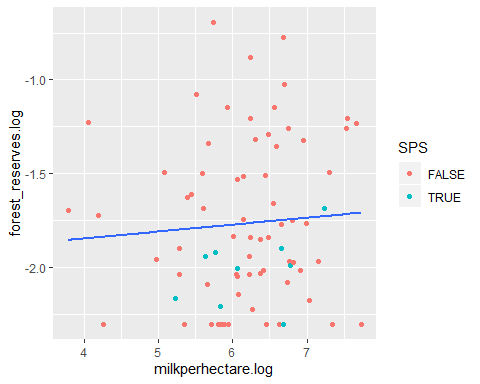
##   
## Call:  
## lm(formula = forest\_reserves.log ~ labourperhectare.log + stocking\_rate.log +   
## SPS + distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.80083 -0.26386 -0.06651 0.27591 0.92383   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.944e+00 1.518e-01 -12.811 <2e-16 \*\*\*  
## labourperhectare.log -1.739e-02 3.103e-02 -0.560 0.5769   
## stocking\_rate.log 7.196e-02 6.064e-02 1.187 0.2393   
## SPSTRUE -2.927e-01 1.469e-01 -1.992 0.0502 .   
## distance 1.464e-05 9.052e-06 1.618 0.1101   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4124 on 71 degrees of freedom  
## Multiple R-squared: 0.09326, Adjusted R-squared: 0.04218   
## F-statistic: 1.826 on 4 and 71 DF, p-value: 0.1334

Simple linear regression of forest reserves and milk yield

fr\_my\_reg <- lm(forest\_reserves.log ~ milkperhectare.log, data = si.df)  
summary(fr\_my\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ milkperhectare.log, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.59396 -0.28339 -0.08767 0.29749 1.08891   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.99442 0.38000 -5.248 1.41e-06 \*\*\*  
## milkperhectare.log 0.03696 0.06099 0.606 0.546   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4232 on 74 degrees of freedom  
## Multiple R-squared: 0.004938, Adjusted R-squared: -0.008508   
## F-statistic: 0.3673 on 1 and 74 DF, p-value: 0.5464

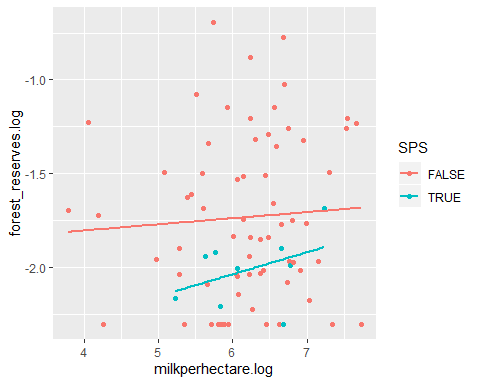
ggplot(data = si.df, aes(x = milkperhectare.log, y = forest\_reserves.log)) +   
 geom\_point(mapping = aes(color=SPS)) +  
 geom\_smooth(method = "lm", se = FALSE)

 Inlcuding SPS as dummy

fr\_my\_SPS\_reg <- lm(forest\_reserves.log ~ milkperhectare.log+SPS, data = si.df)  
summary(fr\_my\_SPS\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ milkperhectare.log + SPS,   
## data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.62967 -0.30610 -0.04734 0.28486 1.05644   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.97133 0.37370 -5.275 1.3e-06 \*\*\*  
## milkperhectare.log 0.03859 0.05995 0.644 0.5218   
## SPSTRUE -0.28021 0.14768 -1.897 0.0617 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4159 on 73 degrees of freedom  
## Multiple R-squared: 0.0517, Adjusted R-squared: 0.02572   
## F-statistic: 1.99 on 2 and 73 DF, p-value: 0.144

ggplot(data = si.df, aes(x = milkperhectare.log, y = forest\_reserves.log, color=SPS)) +   
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)

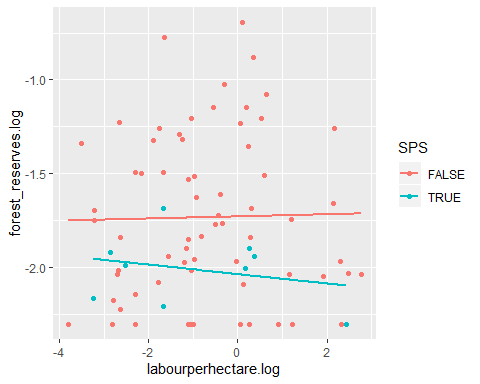


Regression of labour hours and forest reserves

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

##   
## Call:  
## lm(formula = forest\_reserves.log ~ labourperhectare.log + SPS,   
## data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.57319 -0.30275 -0.02162 0.26472 1.03891   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.732173 0.054966 -31.513 <2e-16 \*\*\*  
## labourperhectare.log 0.001195 0.029808 0.040 0.9681   
## SPSTRUE -0.278517 0.148310 -1.878 0.0644 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4171 on 73 degrees of freedom  
## Multiple R-squared: 0.04634, Adjusted R-squared: 0.02021   
## F-statistic: 1.774 on 2 and 73 DF, p-value: 0.1769

ggplot(data = si.df, aes(x = labourperhectare.log, y = forest\_reserves.log, color=SPS)) +   
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE)



Regression of milk yield, labour hours on forest reserves

fr\_my\_lh\_pfp\_reg <- lm(forest\_reserves.log ~ milkperhectare.log+labourperhectare.log+stocking\_rate.log+SPS, data = si.df)  
summary(fr\_my\_lh\_pfp\_reg)

##   
## Call:  
## lm(formula = forest\_reserves.log ~ milkperhectare.log + labourperhectare.log +   
## stocking\_rate.log + SPS, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.69690 -0.27542 -0.06277 0.28200 1.03357   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.802870 0.443743 -4.063 0.000123 \*\*\*  
## milkperhectare.log 0.013609 0.069268 0.196 0.844804   
## labourperhectare.log -0.004438 0.030592 -0.145 0.885064   
## stocking\_rate.log 0.055736 0.069230 0.805 0.423457   
## SPSTRUE -0.278385 0.149350 -1.864 0.066461 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4198 on 71 degrees of freedom  
## Multiple R-squared: 0.06035, Adjusted R-squared: 0.007413   
## F-statistic: 1.14 on 4 and 71 DF, p-value: 0.3448

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

##   
## Call:  
## lm(formula = forest\_reserves.log ~ distance, data = si.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.6295 -0.3066 -0.1270 0.2589 0.9849   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.947e+00 1.428e-01 -13.638 <2e-16 \*\*\*  
## distance 1.191e-05 8.831e-06 1.349 0.181   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4191 on 74 degrees of freedom  
## Multiple R-squared: 0.024, Adjusted R-squared: 0.01081   
## F-statistic: 1.82 on 1 and 74 DF, p-value: 0.1814