#Relationship between BMI and Skin Thickness

library(readr)  
diabetes <- read\_csv("diabetes.csv")

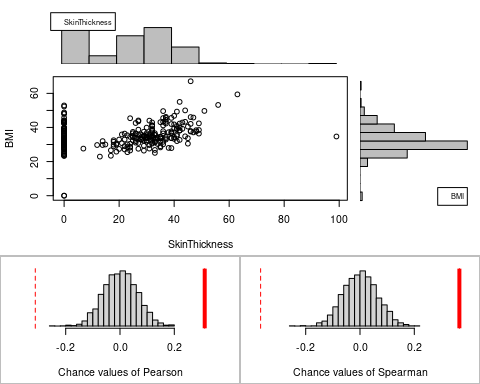
## Rows: 768 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (9): Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, D...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

options(width=90)  
library(regclass)

## Loading required package: bestglm  
## Loading required package: leaps  
## Loading required package: VGAM  
## Loading required package: stats4  
## Loading required package: splines  
## Loading required package: rpart  
## Loading required package: randomForest  
## randomForest 4.7-1  
## Type rfNews() to see new features/changes/bug fixes.  
## Important regclass change from 1.3:  
## All functions that had a . in the name now have an \_  
## all.correlations -> all\_correlations, cor.demo -> cor\_demo, etc.

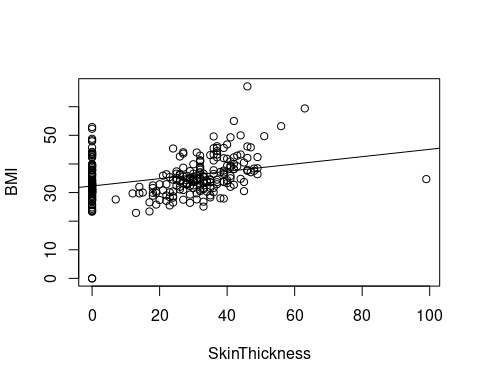
Diabetes\_Data <- subset(diabetes, Outcome > 0)  
associate(BMI~SkinThickness, data = Diabetes\_Data, permutations = 2000)

## Association between SkinThickness (numerical) and BMI (numerical)  
## using 268 complete cases



## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.3120583 0  
## Spearman's rank correlation 0.3660637 0  
## With 2000 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0 and 0.002   
## the p-value of Spearman's rank correlation is between 0 and 0.002   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

X<-lm(BMI~SkinThickness, data = Diabetes\_Data)  
plot(BMI~SkinThickness, data = Diabetes\_Data)  
abline(X)



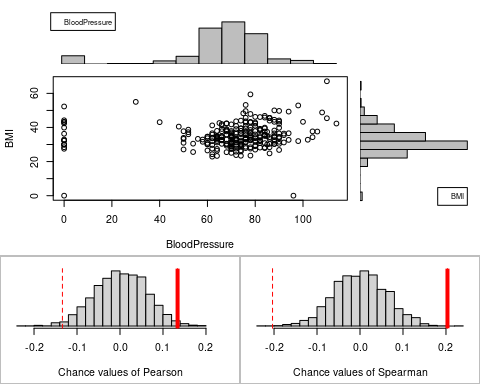
summary(X)

##   
## Call:  
## lm(formula = BMI ~ SkinThickness, data = Diabetes\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32.301 -3.927 -0.948 3.712 28.902   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 32.30118 0.67798 47.643 < 2e-16 \*\*\*  
## SkinThickness 0.12820 0.02393 5.357 1.83e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.913 on 266 degrees of freedom  
## Multiple R-squared: 0.09738, Adjusted R-squared: 0.09399   
## F-statistic: 28.7 on 1 and 266 DF, p-value: 1.833e-07

#Relationship between BMI and BloodPressure

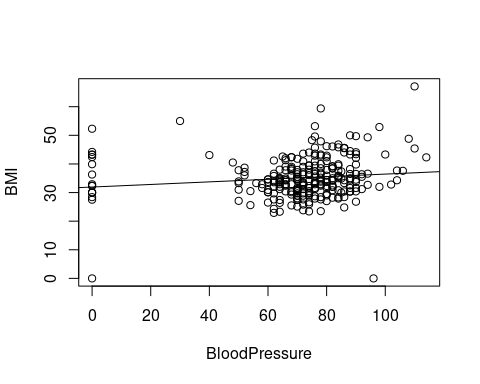
Diabetes\_Data <- subset(diabetes, Outcome > 0)  
associate(BMI~BloodPressure, data = Diabetes\_Data, permutations = 2000)

## Association between BloodPressure (numerical) and BMI (numerical)  
## using 268 complete cases



## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.1339510 0.0245  
## Spearman's rank correlation 0.2034747 0.0010  
## With 2000 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.018 and 0.032   
## the p-value of Spearman's rank correlation is between 0 and 0.004   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

BMI\_BloodPressure<-lm(BMI~BloodPressure, data = Diabetes\_Data)  
plot(BMI~BloodPressure, data = Diabetes\_Data)  
abline(BMI\_BloodPressure)



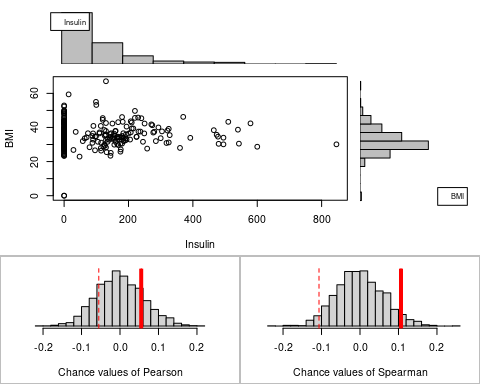
summary(BMI\_BloodPressure)

##   
## Call:  
## lm(formula = BMI ~ BloodPressure, data = Diabetes\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.282 -4.234 -0.892 3.615 30.184   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.93648 1.51954 21.017 <2e-16 \*\*\*  
## BloodPressure 0.04527 0.02053 2.205 0.0283 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.211 on 266 degrees of freedom  
## Multiple R-squared: 0.01794, Adjusted R-squared: 0.01425   
## F-statistic: 4.86 on 1 and 266 DF, p-value: 0.02834

#Relationship between BMI and Glucose

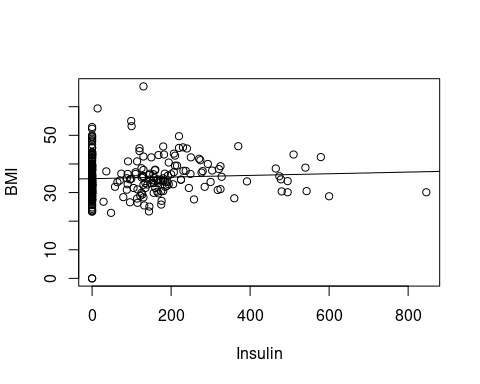
Diabetes\_Data <- subset(diabetes, Outcome > 0)  
associate(BMI~Insulin, data = Diabetes\_Data, permutations = 2000)

## Association between Insulin (numerical) and BMI (numerical)  
## using 268 complete cases



## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r 0.05511211 0.3605  
## Spearman's rank correlation 0.10654443 0.0785  
## With 2000 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.339 and 0.382   
## the p-value of Spearman's rank correlation is between 0.067 and 0.091   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

BMI\_Insulin<-lm(BMI~Insulin, data = Diabetes\_Data)  
plot(BMI~Insulin, data = Diabetes\_Data)  
abline(BMI\_Insulin)



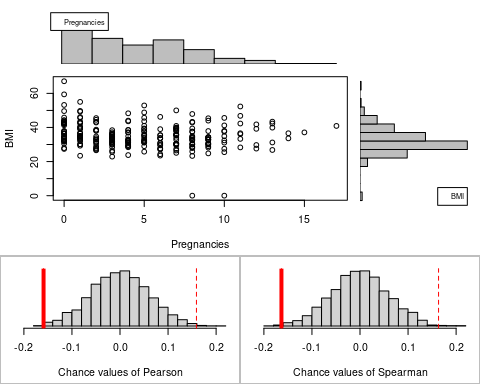
summary(BMI\_Insulin)

##   
## Call:  
## lm(formula = BMI ~ Insulin, data = Diabetes\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.853 -4.353 -1.002 3.647 31.872   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 34.852953 0.548133 63.59 <2e-16 \*\*\*  
## Insulin 0.002886 0.003206 0.90 0.369   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.266 on 266 degrees of freedom  
## Multiple R-squared: 0.003037, Adjusted R-squared: -0.0007106   
## F-statistic: 0.8104 on 1 and 266 DF, p-value: 0.3688

#Relationship between BMI and Pregnancies

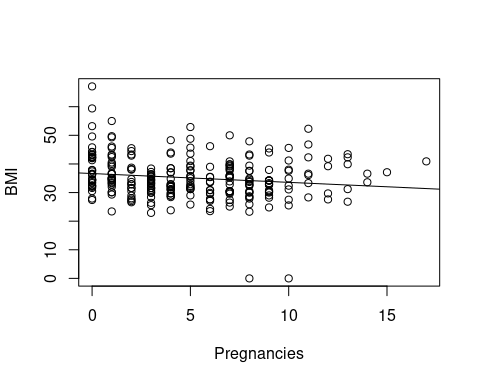
Diabetes\_Data <- subset(diabetes, Outcome > 0)  
associate(BMI~Pregnancies, data = Diabetes\_Data, permutations = 2000)

## Association between Pregnancies (numerical) and BMI (numerical)  
## using 268 complete cases



## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r -0.1590709 0.005  
## Spearman's rank correlation -0.1634791 0.006  
## With 2000 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.002 and 0.009   
## the p-value of Spearman's rank correlation is between 0.003 and 0.01   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

BMI\_Pregnancies<-lm(BMI~Pregnancies, data = Diabetes\_Data)  
plot(BMI~Pregnancies, data = Diabetes\_Data)  
abline(BMI\_Pregnancies)



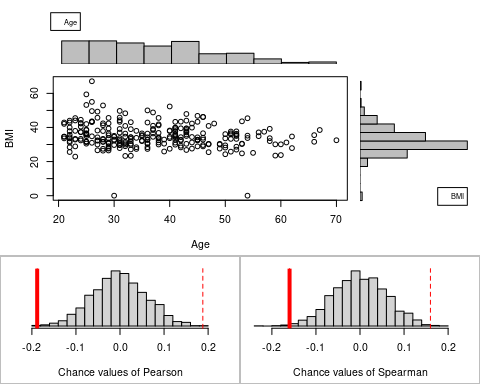
summary(BMI\_Pregnancies)

##   
## Call:  
## lm(formula = BMI ~ Pregnancies, data = Diabetes\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.175 -4.152 -1.097 4.036 30.455   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.6451 0.7208 50.841 < 2e-16 \*\*\*  
## Pregnancies -0.3088 0.1175 -2.628 0.00909 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.184 on 266 degrees of freedom  
## Multiple R-squared: 0.0253, Adjusted R-squared: 0.02164   
## F-statistic: 6.905 on 1 and 266 DF, p-value: 0.009092

#Relationship between BMI and Age

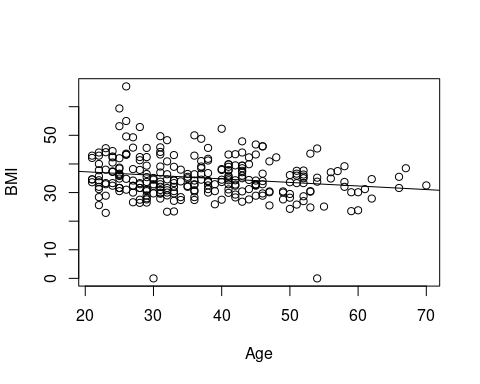
Diabetes\_Data <- subset(diabetes, Outcome > 0)  
associate(BMI~Age, data = Diabetes\_Data, permutations = 2000)

## Association between Age (numerical) and BMI (numerical)  
## using 268 complete cases



## Permutation procedure:  
## Value Estimated p-value  
## Pearson's r -0.1880117 0.0025  
## Spearman's rank correlation -0.1598639 0.0060  
## With 2000 permutations, we are 95% confident that:  
## the p-value of Pearson's correlation (r) is between 0.001 and 0.006   
## the p-value of Spearman's rank correlation is between 0.003 and 0.01   
## Note: If 0.05 is in this range, increase the permutations= argument.  
##   
##   
##   
## Advice: If stream of points is well described by an ellipse, use Pearson's r.  
## Otherwise, as long as stream is monotonic, use Spearman's rank correlation  
## or try logs, e.g. associate( log10(y)~log10(x) )

BMI\_Age<-lm(BMI~Age, data = Diabetes\_Data)  
plot(BMI~Age, data = Diabetes\_Data)  
abline(BMI\_Age)



summary(BMI\_Age)

##   
## Call:  
## lm(formula = BMI ~ Age, data = Diabetes\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.022 -4.378 -0.884 3.994 30.580   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 39.75732 1.54124 25.796 < 2e-16 \*\*\*  
## Age -0.12450 0.03988 -3.122 0.00199 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.147 on 266 degrees of freedom  
## Multiple R-squared: 0.03535, Adjusted R-squared: 0.03172   
## F-statistic: 9.747 on 1 and 266 DF, p-value: 0.001994