MARKDOWN – Artigo Conceitos x Rendimento Discente

## Análise

setwd("C:\\Users\\ contabilidade")  
library(MASS)  
library(reshape2)  
library(ggplot2)  
library(plyr)  
library(tidyverse)

## -- Attaching packages ---------------------------------------------- tidyverse 1.2.1 --

## v tibble 1.4.2 v purrr 0.2.5  
## v tidyr 0.8.2 v dplyr 0.7.8  
## v readr 1.2.1 v stringr 1.3.1  
## v tibble 1.4.2 v forcats 0.3.0

## -- Conflicts ------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::arrange() masks plyr::arrange()  
## x purrr::compact() masks plyr::compact()  
## x dplyr::count() masks plyr::count()  
## x dplyr::failwith() masks plyr::failwith()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::id() masks plyr::id()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::mutate() masks plyr::mutate()  
## x dplyr::rename() masks plyr::rename()  
## x dplyr::select() masks MASS::select()  
## x dplyr::summarise() masks plyr::summarise()  
## x dplyr::summarize() masks plyr::summarize()

library(corrplot)

## Warning: package 'corrplot' was built under R version 3.5.3

## corrplot 0.84 loaded

library(rpart)

## Warning: package 'rpart' was built under R version 3.5.3

library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 3.5.3

library(pscl)

## Warning: package 'pscl' was built under R version 3.5.3

## Classes and Methods for R developed in the  
## Political Science Computational Laboratory  
## Department of Political Science  
## Stanford University  
## Simon Jackman  
## hurdle and zeroinfl functions by Achim Zeileis

library(ROCR)

## Warning: package 'ROCR' was built under R version 3.5.3

## Loading required package: gplots

## Warning: package 'gplots' was built under R version 3.5.3

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.5.2

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

options(digits=4,OutDec=',')  
###################################################################################  
dados = read.table("dados.csv",sep=';',dec=',',header = T,encoding="UTF-8",stringsAsFactors=FALSE)  
names(dados)[1] = "ID"  
#  
mconceito = melt(dados[,c(1,3,5,7,9,11,13)],id="ID",variable.name = "Conceito",value.name = "Nota")  
#str(mconceito)  
#boxplot(Nota~Conceito,data=mconceito)  
round(tapply(mconceito$Nota,mconceito$Conceito, mean),2)

## conceito1 conceito2 conceito3 conceito4 conceito5 conceito6   
## 5,53 8,38 8,79 7,28 6,82 7,48

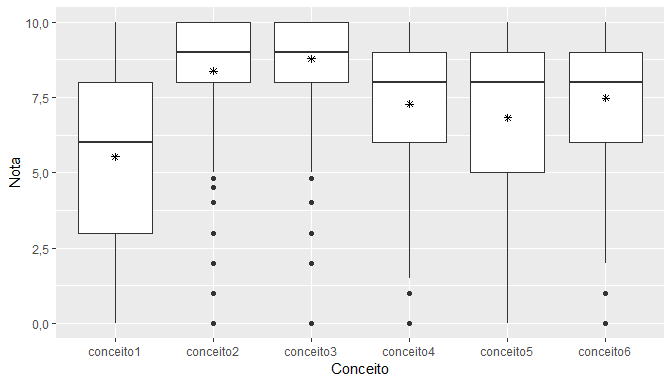
tapply(mconceito$Nota,mconceito$Conceito, median)

## conceito1 conceito2 conceito3 conceito4 conceito5 conceito6   
## 6 9 9 8 8 8

tapply(mconceito$Nota,mconceito$Conceito, summary)

## $conceito1  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 3,00 6,00 5,53 8,00 10,00   
##   
## $conceito2  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 8,00 9,00 8,38 10,00 10,00   
##   
## $conceito3  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 8,00 9,00 8,79 10,00 10,00   
##   
## $conceito4  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 6,00 8,00 7,28 9,00 10,00   
##   
## $conceito5  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 5,00 8,00 6,82 9,00 10,00   
##   
## $conceito6  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0,00 6,00 8,00 7,48 9,00 10,00

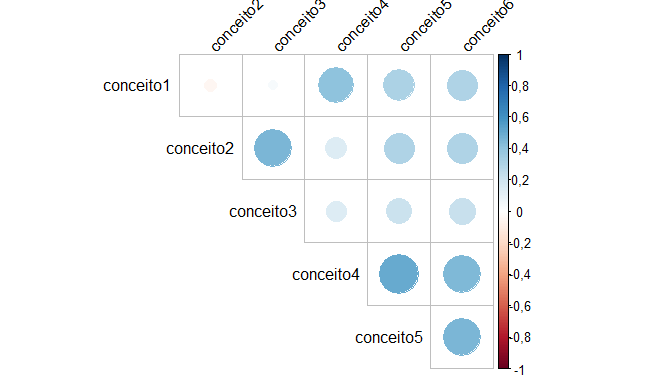
#  
p<-qplot(Conceito,Nota,data=mconceito,geom='boxplot')  
p<-p+stat\_summary(fun.y=mean,shape=8,col=1,geom='point')  
print(p)



#  
corMat = round(cor(dados[,c(3,5,7,9,11,13)]),2)  
corMat

## conceito1 conceito2 conceito3 conceito4 conceito5 conceito6  
## conceito1 1,00 -0,05 0,03 0,40 0,31 0,30  
## conceito2 -0,05 1,00 0,45 0,15 0,30 0,30  
## conceito3 0,03 0,45 1,00 0,14 0,21 0,22  
## conceito4 0,40 0,15 0,14 1,00 0,50 0,44  
## conceito5 0,31 0,30 0,21 0,50 1,00 0,45  
## conceito6 0,30 0,30 0,22 0,44 0,45 1,00

#library("PerformanceAnalytics")  
#chart.Correlation(dados[,c(3,5,7,9,11,13)], histogram=F, pch=19)  
#  
corrplot(corMat, type = "upper", order = "original", diag=F,  
 tl.col = "black", tl.srt = 45)



###### aceitação dos conceitos  
aceit\_conc<-tapply(mconceito$Nota,mconceito$Conceito,function(x) as.numeric(table(x>=5)))%>%  
 unlist() %>%  
 matrix(ncol=2,nrow=6,byrow=T)   
colnames(aceit\_conc)=c("<5",">=5")  
rownames(aceit\_conc)=paste0("Conceito",1:6)  
aceit\_conc

## <5 >=5  
## Conceito1 221 429  
## Conceito2 36 614  
## Conceito3 15 635  
## Conceito4 75 575  
## Conceito5 125 525  
## Conceito6 70 580

##  
mconceito$aceit = ifelse(mconceito$Nota<5,0,1)  
round(tapply(mconceito$aceit,mconceito$Conceito,mean),2)

## conceito1 conceito2 conceito3 conceito4 conceito5 conceito6   
## 0,66 0,94 0,98 0,88 0,81 0,89

##########################################################

##########################################################  
#parte 1: conceitos vs qualificações  
##########################################################  
dados$public\_agrp <- as.factor(ifelse(dados$public\_atual==0,"0",  
 ifelse(dados$public\_atual<=2,"1-2",  
 ifelse(dados$public\_atual<=5,"3-5","6+"))))  
dados$experiencia\_mercado = ifelse(dados$tempo\_mercado==0,0,1)  
dados$certif\_atual = ifelse((dados$certificac\_atual==1)|(dados$certif\_crcatual==1),1,0)  
dados$mercado\_atual= ifelse(dados$mercado\_atual==7,8,dados$mercado\_atual)  
###  
#####################################################################  
vars\_num=c("doutor\_tempo\_concl","mestr\_tempo\_concl","docencia\_tempo","tempo\_mercado")  
corMat <- dados %>% select(vars\_num) %>% cor(use = "complete.obs") %>% round(2)  
corMat = round(cor(dados[,c(3,5,7,9,11,13)]),2)  
corMat

## conceito1 conceito2 conceito3 conceito4 conceito5 conceito6  
## conceito1 1,00 -0,05 0,03 0,40 0,31 0,30  
## conceito2 -0,05 1,00 0,45 0,15 0,30 0,30  
## conceito3 0,03 0,45 1,00 0,14 0,21 0,22  
## conceito4 0,40 0,15 0,14 1,00 0,50 0,44  
## conceito5 0,31 0,30 0,21 0,50 1,00 0,45  
## conceito6 0,30 0,30 0,22 0,44 0,45 1,00

vars\_cat = c("se\_doutor\_cont\_1s","se\_mestre\_contabeis\_1s","public\_agrp",  
 "certif\_atual","mercado\_atual","for\_pedag","genero\_1M",  
 "regime\_atual","cat\_adm\_atual","regiao\_geografica","Organ\_acad\_atual",  
 "experiencia\_mercado")  
dados <- dados %>% mutate\_each(funs(factor(.)),vars\_cat)

## `mutate\_each()` is deprecated.  
## Use `mutate\_all()`, `mutate\_at()` or `mutate\_if()` instead.  
## To map `funs` over a selection of variables, use `mutate\_at()`

#usar interação experiencia\_mercado:tempo\_mercado  
  
### análise bivariada  
extras = c("area\_societaria","area\_custos","area\_gerencial","area\_publica","area\_teoria",  
 "area\_etica","area\_auditoria","area\_pericia","area\_tributaria")  
dados <- dados %>% mutate\_each(funs(factor(.)),extras)

## `mutate\_each()` is deprecated.  
## Use `mutate\_all()`, `mutate\_at()` or `mutate\_if()` instead.  
## To map `funs` over a selection of variables, use `mutate\_at()`

conc13 = melt(dados[,c(1,3,7)],id="ID",variable.name = "Conceito",value.name = "Nota")  
conc13\_1s = melt(dados[,c(1,2,6)],id="ID",variable.name = "Conceito",value.name = "Aceitacao")  
dconc13 = merge(dados[,c("ID",vars\_num,vars\_cat,extras)],conc13,by="ID",all.y=T)  
dconc13\_1s = merge(dados[,c("ID",vars\_num,vars\_cat,extras)],conc13\_1s,by="ID",all.y=T)  
#  
#se\_doutor\_cont\_1s  
g1<-ggplot(subset(dconc13,!is.na(se\_doutor\_cont\_1s)), aes(y=Nota, x=se\_doutor\_cont\_1s)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
table(dados$se\_doutor\_cont\_1s)

##   
## 0 1   
## 411 118

ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$se\_doutor\_cont\_1s)

## 0 1  
##   
## conc1\_s1 0 140 40  
## 1 271 78  
## conc3\_s1 0 6 7  
## 1 405 111

chisq.test(table(dados$se\_doutor\_cont\_1s,dados$conc1\_s1)) #não rejeita #se\_mestre\_contabeis\_1s

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$se\_doutor\_cont\_1s, dados$conc1\_s1)  
## X-squared = 2,7e-29, df = 1, p-value = 1

g2<-ggplot(subset(dconc13,!is.na(se\_mestre\_contabeis\_1s)), aes(y=Nota, x=se\_mestre\_contabeis\_1s)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
table(dados$se\_mestre\_contabeis\_1s)

##   
## 0 1   
## 214 315

ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$se\_mestre\_contabeis\_1s)

## 0 1  
##   
## conc1\_s1 0 69 111  
## 1 145 204  
## conc3\_s1 0 5 8  
## 1 209 307

chisq.test(table(dados$se\_mestre\_contabeis\_1s,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$se\_mestre\_contabeis\_1s, dados$conc1\_s1)  
## X-squared = 0,38, df = 1, p-value = 0,5

# null hypothesis that the joint distribution of the cell counts in a 2-dimensional contingency table is the product of the row and column marginals.  
  
#public\_atual --> public\_agrp  
table(dados$public\_atual)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 14 15 20 25   
## 192 99 56 54 27 25 15 1 5 3 8 4 4 2 1 3 1

table(dados$public\_agrp)

##   
## 0 1-2 3-5 6+   
## 192 155 106 47

g3<-ggplot(subset(dconc13,!is.na(public\_agrp)), aes(y=Nota, x=public\_agrp)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$public\_agrp)

## 0 1-2 3-5 6+  
##   
## conc1\_s1 0 60 56 37 18  
## 1 132 99 69 29  
## conc3\_s1 0 2 4 6 0  
## 1 190 151 100 47

chisq.test(table(dados$public\_agrp,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$public\_agrp, dados$conc1\_s1)  
## X-squared = 1,4, df = 3, p-value = 0,7

#  
table(dados$experiencia\_mercado)

##   
## 0 1   
## 87 413

g4<-ggplot(subset(dconc13,!is.na(experiencia\_mercado)), aes(y=Nota, x=experiencia\_mercado)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$experiencia\_mercado)

## 0 1  
##   
## conc1\_s1 0 27 144  
## 1 60 269  
## conc3\_s1 0 4 8  
## 1 83 405

chisq.test(table(dados$experiencia\_mercado,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$experiencia\_mercado, dados$conc1\_s1)  
## X-squared = 0,31, df = 1, p-value = 0,6

#certif\_atual  
table(dados$certif\_crcatual,dados$certificac\_atual) #grande maioria tem as duas

##   
## 0 1  
## 0 124 9  
## 1 0 367

table(dados$certif\_outratual) # pouca gente

##   
## 0 1   
## 465 35

table(dados$certif\_atual) # =1 se crc ou cac

##   
## 0 1   
## 124 376

g5<-ggplot(subset(dconc13,!is.na(certif\_atual)), aes(y=Nota, x=certif\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$certif\_atual)

## 0 1  
##   
## conc1\_s1 0 32 139  
## 1 92 237  
## conc3\_s1 0 3 9  
## 1 121 367

chisq.test(table(dados$certif\_atual,dados$conc1\_s1)) #rejeita independencia

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$certif\_atual, dados$conc1\_s1)  
## X-squared = 4,7, df = 1, p-value = 0,03

#### diferença significativa para conceito 1  
tapply(dados$conceito1,dados$certif\_atual,mean)

## 0 1   
## 6,231 5,359

tapply(dados$conceito1,dados$certif\_atual,sd)

## 0 1   
## 3,099 3,470

tapply(dados$conceito1,dados$certif\_atual,median)

## 0 1   
## 7 6

wilcox.test(conceito1~certif\_atual,data=dados)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: conceito1 by certif\_atual  
## W = 26000, p-value = 0,02  
## alternative hypothesis: true location shift is not equal to 0

### locação signific diferente  
  
#mercado\_atual  
table(dados$mercado\_atual) #agrupar 7 e 8

##   
## 1 2 3 4 5 6 8   
## 282 40 31 50 37 36 23

g6<-ggplot(subset(dconc13,!is.na(mercado\_atual)), aes(y=Nota, x=mercado\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$mercado\_atual)

## 1 2 3 4 5 6 8  
##   
## conc1\_s1 0 101 19 11 14 8 10 8  
## 1 181 21 20 36 29 26 15  
## conc3\_s1 0 7 0 0 3 1 1 0  
## 1 275 40 31 47 36 35 23

chisq.test(table(dados$mercado\_atual,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$mercado\_atual, dados$conc1\_s1)  
## X-squared = 7,6, df = 6, p-value = 0,3

(means <-tapply(dados$conceito1,dados$mercado\_atual,mean))

## 1 2 3 4 5 6 8   
## 5,562 4,438 5,313 5,880 6,919 5,583 5,261

tapply(dados$conceito1,dados$mercado\_atual,sd)

## 1 2 3 4 5 6 8   
## 3,395 3,503 3,633 3,280 3,345 2,999 3,570

tapply(dados$conceito1,dados$mercado\_atual,median)

## 1 2 3 4 5 6 8   
## 6 5 7 7 8 6 5

#for\_pedag  
table(dados$for\_pedag) #pouca variacao

##   
## 0 1   
## 82 418

g7<-ggplot(subset(dconc13,!is.na(for\_pedag)), aes(y=Nota, x=for\_pedag)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$for\_pedag)

## 0 1  
##   
## conc1\_s1 0 25 146  
## 1 57 272  
## conc3\_s1 0 0 12  
## 1 82 406

chisq.test(table(dados$for\_pedag,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$for\_pedag, dados$conc1\_s1)  
## X-squared = 0,42, df = 1, p-value = 0,5

#genero\_1M  
table(dados$genero\_1M)

##   
## 0 1   
## 188 312

g8<-ggplot(subset(dconc13,!is.na(genero\_1M)), aes(y=Nota, x=genero\_1M)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$genero\_1M)

## 0 1  
##   
## conc1\_s1 0 72 99  
## 1 116 213  
## conc3\_s1 0 3 9  
## 1 185 303

chisq.test(table(dados$genero\_1M,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(dados$genero\_1M, dados$conc1\_s1)  
## X-squared = 2, df = 1, p-value = 0,2

#regime\_atual  
table(dados$regime\_atual)

##   
## 1 2 3   
## 271 96 133

g9<-ggplot(subset(dconc13,!is.na(regime\_atual)), aes(y=Nota, x=regime\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$regime\_atual)

## 1 2 3  
##   
## conc1\_s1 0 101 29 41  
## 1 170 67 92  
## conc3\_s1 0 9 3 0  
## 1 262 93 133

chisq.test(table(dados$regime\_atual,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$regime\_atual, dados$conc1\_s1)  
## X-squared = 2,5, df = 2, p-value = 0,3

#cat\_adm\_atual  
table(dados$cat\_adm\_atual)

##   
## 1 2 3   
## 182 88 154

g10<-ggplot(subset(dconc13,!is.na(cat\_adm\_atual)), aes(y=Nota, x=cat\_adm\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$cat\_adm\_atual)

## 1 2 3  
##   
## conc1\_s1 0 55 32 56  
## 1 127 56 98  
## conc3\_s1 0 6 1 1  
## 1 176 87 153

chisq.test(table(dados$cat\_adm\_atual,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$cat\_adm\_atual, dados$conc1\_s1)  
## X-squared = 1,8, df = 2, p-value = 0,4

#regiao\_geografica  
table(dados$regiao\_geografica)

##   
## 1 2 3 4 5   
## 230 140 20 68 42

g11<-ggplot(subset(dconc13,!is.na(regiao\_geografica)), aes(y=Nota, x=regiao\_geografica)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$regiao\_geografica)

## 1 2 3 4 5  
##   
## conc1\_s1 0 63 60 7 27 14  
## 1 167 80 13 41 28  
## conc3\_s1 0 2 4 1 5 0  
## 1 228 136 19 63 42

chisq.test(table(dados$regiao\_geografica,dados$conc1\_s1)) #rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$regiao\_geografica, dados$conc1\_s1)  
## X-squared = 10, df = 4, p-value = 0,04

prop.table(table(dados$regiao\_geografica,dados$conc1\_s1),1)

##   
## 0 1  
## 1 0,2739 0,7261  
## 2 0,4286 0,5714  
## 3 0,3500 0,6500  
## 4 0,3971 0,6029  
## 5 0,3333 0,6667

(means <-tapply(dados$conceito1,dados$regiao\_geografica,mean))

## 1 2 3 4 5   
## 6,060 4,911 5,700 5,316 5,498

tapply(dados$conceito1,dados$regiao\_geografica,sd)

## 1 2 3 4 5   
## 3,303 3,502 2,755 3,700 3,048

tapply(dados$conceito1,dados$regiao\_geografica,median)

## 1 2 3 4 5   
## 7,0 5,0 5,5 6,0 5,0

kruskal.test(dados$conceito1 ~dados$regiao\_geografica)

##   
## Kruskal-Wallis rank sum test  
##   
## data: dados$conceito1 by dados$regiao\_geografica  
## Kruskal-Wallis chi-squared = 11, df = 4, p-value = 0,03

pairwise.wilcox.test(dados$conceito1,dados$regiao\_geografica,p.adjust.method = "BH")

## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot  
## compute exact p-value with ties

##   
## Pairwise comparisons using Wilcoxon rank sum test   
##   
## data: dados$conceito1 and dados$regiao\_geografica   
##   
## 1 2 3 4   
## 2 0,02 - - -   
## 3 0,56 0,56 - -   
## 4 0,56 0,56 0,99 -   
## 5 0,56 0,56 0,99 0,99  
##   
## P value adjustment method: BH

#null that the location parameters of the distribution of x are the same in each group (sample).  
#The alternative is that they differ in at least one  
#http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r  
  
#  
table(dados$Organ\_acad\_atual)

##   
## 1 2 3   
## 307 52 63

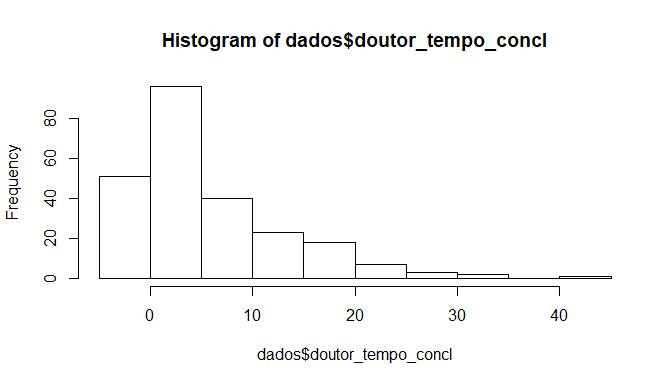
g12<-ggplot(subset(dconc13,!is.na(Organ\_acad\_atual)), aes(y=Nota, x=Organ\_acad\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)  
ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,dconc13\_1s$Organ\_acad\_atual)

## 1 2 3  
##   
## conc1\_s1 0 102 18 22  
## 1 205 34 41  
## conc3\_s1 0 7 1 0  
## 1 300 51 63

chisq.test(table(dados$Organ\_acad\_atual,dados$conc1\_s1)) #não rejeita independencia

##   
## Pearson's Chi-squared test  
##   
## data: table(dados$Organ\_acad\_atual, dados$conc1\_s1)  
## X-squared = 0,092, df = 2, p-value = 1

#doutor\_tempo\_concl  
hist(dados$doutor\_tempo\_concl)



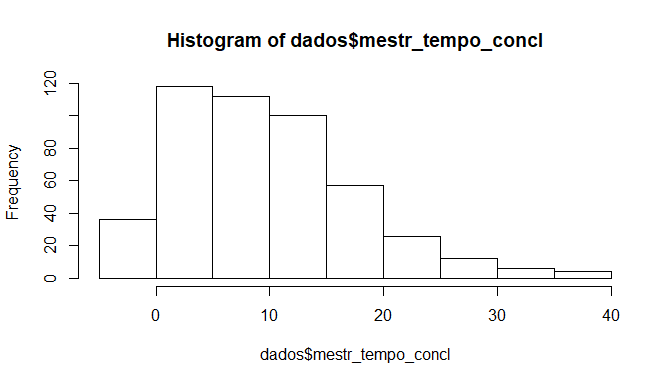
c1<-ggplot(dconc13, aes(y=Nota, x=doutor\_tempo\_concl)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point()   
cor(dados$conceito1,dados$doutor\_tempo\_concl,use="complete.obs")

## [1] 0,03141

cor(dados$conceito3,dados$doutor\_tempo\_concl,use="complete.obs")

## [1] -0,03277

#  
c2<-ggplot(dconc13\_1s, aes(y=doutor\_tempo\_concl, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação")  
  
#mestr\_tempo\_concl  
hist(dados$mestr\_tempo\_concl)



cor(dados$doutor\_tempo\_concl,dados$mestr\_tempo\_concl,use="complete.obs")

## [1] 0,8407

c3<-ggplot(dconc13, aes(y=Nota, x=mestr\_tempo\_concl)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point()   
cor(dados$conceito1,dados$mestr\_tempo\_concl,use="complete.obs")

## [1] -0,009361

cor(dados$conceito3,dados$mestr\_tempo\_concl,use="complete.obs")

## [1] -0,0887

#  
c4<-ggplot(dconc13\_1s, aes(y=mestr\_tempo\_concl, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação")  
  
#tempo\_mercado  
table(dados$tempo\_mercado) #muitos 0's, criar dummy e usar interação

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24   
## 87 6 24 10 20 26 12 7 19 10 37 4 16 9 10 24 9 4 15 2 31 4 6 6 4   
## 25 26 27 28 29 30 31 32 33 34 35 36 37 40 41 43 45 47 48 50 56   
## 17 2 3 3 3 25 5 3 1 5 8 4 1 7 1 3 2 1 1 2 1

c5<-ggplot(dconc13, aes(y=Nota, x=tempo\_mercado)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point()   
cor(dados$conceito1,dados$tempo\_mercado,use="complete.obs")

## [1] -0,09519

cor(dados$conceito3,dados$tempo\_mercado,use="complete.obs")

## [1] 0,06989

#  
c6<-ggplot(dconc13\_1s, aes(y=tempo\_mercado, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação")  
  
#docencia\_tempo  
table(dados$docencia\_tempo)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24   
## 16 29 15 25 24 29 18 20 21 9 40 22 12 12 16 31 8 14 20 5 36 5 7 17 3   
## 25 26 27 28 29 30 32 34 35 37 38 40 41 50   
## 16 5 4 14 2 12 2 3 10 1 2 2 1 1

c7<-ggplot(dconc13, aes(y=Nota, x=docencia\_tempo)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point()   
cor(dados$conceito1,dados$docencia\_tempo,use="complete.obs")

## [1] -0,03501

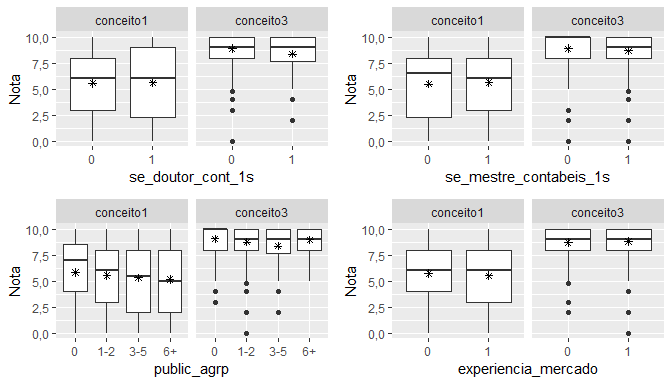
cor(dados$conceito3,dados$docencia\_tempo,use="complete.obs")

## [1] -0,006017

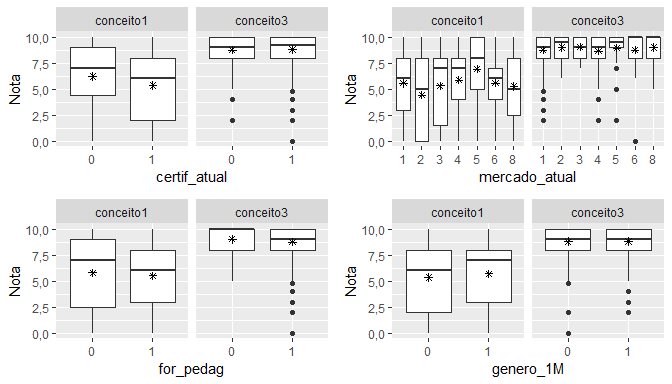
#  
c8<-ggplot(dconc13\_1s, aes(y=docencia\_tempo, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação")  
  
#table(dados$area\_societaria,dados$area\_custos)  
#area de atuação - mais de uma resposta, ver se tem alguma de interesse  
plot\_extra = function (i) {  
 ggplot(subset(dconc13,!is.na(eval(parse(text = i)))),   
 aes(y=Nota, x=eval(parse(text = i)))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab(i)  
 #tab <-ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,eval(parse(text = paste0("dconc13\_1s$",i))))  
 #pval = chisq.test(table(eval(parse(text = paste0("dados$",i))),dados$conc1\_s1))$p.value  
 #return(list(p,tab,pval))  
 }  
  
myplots <- lapply(extras, plot\_extra)  
#myplots[[4]]  
  
for (i in extras) {  
 print(i)  
 print(table(eval(parse(text = paste0("dados$",i)))))  
 print(ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,eval(parse(text = paste0("dconc13\_1s$",i)))))  
 print(chisq.test(table(eval(parse(text = paste0("dados$",i))),dados$conc1\_s1))$p.value)  
}

## [1] "area\_societaria"  
##   
## 0 1   
## 200 329   
## 0 1  
##   
## conc1\_s1 0 67 113  
## 1 133 216  
## conc3\_s1 0 6 7  
## 1 194 322  
## [1] 0,9167  
## [1] "area\_custos"  
##   
## 0 1   
## 344 185   
## 0 1  
##   
## conc1\_s1 0 122 58  
## 1 222 127  
## conc3\_s1 0 10 3  
## 1 334 182  
## [1] 0,3919  
## [1] "area\_gerencial"  
##   
## 0 1   
## 300 229   
## 0 1  
##   
## conc1\_s1 0 118 62  
## 1 182 167  
## conc3\_s1 0 9 4  
## 1 291 225  
## [1] 0,00429  
## [1] "area\_publica"  
##   
## 0 1   
## 451 78   
## 0 1  
##   
## conc1\_s1 0 149 31  
## 1 302 47  
## conc3\_s1 0 11 2  
## 1 440 76  
## [1] 0,3055  
## [1] "area\_teoria"  
##   
## 0 1   
## 388 141   
## 0 1  
##   
## conc1\_s1 0 129 51  
## 1 259 90  
## conc3\_s1 0 6 7  
## 1 382 134  
## [1] 0,6006  
## [1] "area\_etica"  
##   
## 0 1   
## 479 50   
## 0 1  
##   
## conc1\_s1 0 164 16  
## 1 315 34  
## conc3\_s1 0 13 0  
## 1 466 50  
## [1] 0,8721  
## [1] "area\_auditoria"  
##   
## 0 1   
## 453 76   
## 0 1  
##   
## conc1\_s1 0 153 27  
## 1 300 49  
## conc3\_s1 0 12 1  
## 1 441 75  
## [1] 0,867  
## [1] "area\_pericia"  
##   
## 0 1   
## 456 73   
## 0 1  
##   
## conc1\_s1 0 153 27  
## 1 303 46  
## conc3\_s1 0 11 2  
## 1 445 71  
## [1] 0,6586  
## [1] "area\_tributaria"  
##   
## 0 1   
## 489 40   
## 0 1  
##   
## conc1\_s1 0 168 12  
## 1 321 28  
## conc3\_s1 0 13 0  
## 1 476 40  
## [1] 0,6999

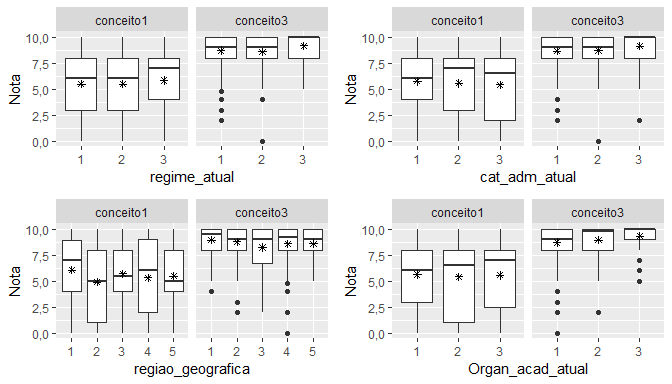
####gráficos  
grid.arrange(g1,g2,g3,g4,ncol=2)



grid.arrange(g5,g6,g7,g8,ncol=2)



grid.arrange(g9,g10,g11,g12,ncol=2)



grid.arrange(c1,c2,c3,c4,ncol=2)

## Warning: Removed 818 rows containing missing values (geom\_point).

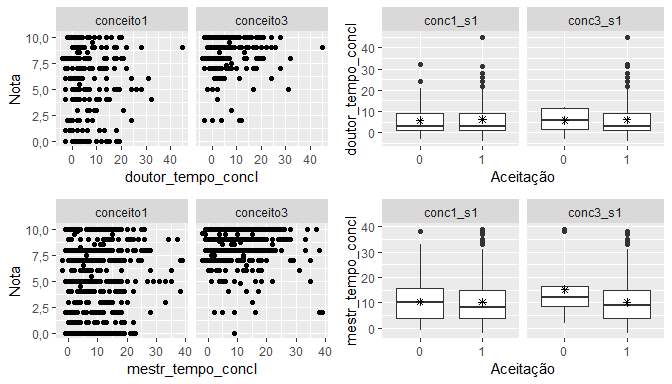
## Warning: Removed 818 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 818 rows containing non-finite values (stat\_summary).

## Warning: Removed 358 rows containing missing values (geom\_point).

## Warning: Removed 358 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 358 rows containing non-finite values (stat\_summary).



grid.arrange(c5,c6,c7,c8,ncol=2)

## Warning: Removed 300 rows containing missing values (geom\_point).

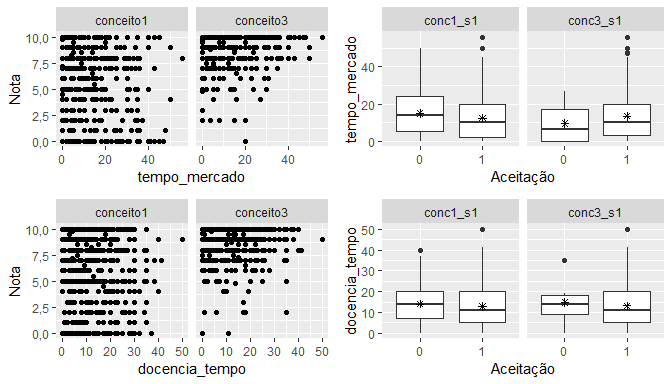
## Warning: Removed 300 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 300 rows containing non-finite values (stat\_summary).

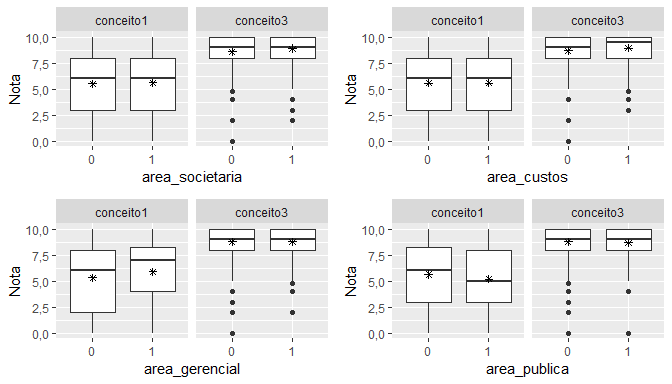
## Warning: Removed 242 rows containing missing values (geom\_point).

## Warning: Removed 242 rows containing non-finite values (stat\_boxplot).

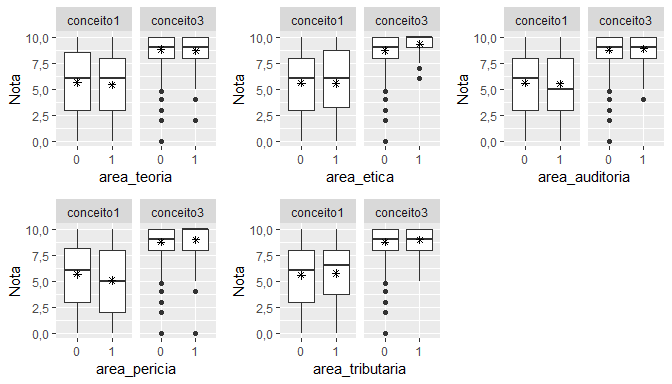
## Warning: Removed 242 rows containing non-finite values (stat\_summary).



grid.arrange(myplots[[1]],myplots[[2]],myplots[[3]],myplots[[4]],ncol=2)



grid.arrange(myplots[[5]],myplots[[6]],myplots[[7]],myplots[[8]],myplots[[9]],ncol=3)



########################################################  
explicativas = c(vars\_num[-4],vars\_cat,"experiencia\_mercado:tempo\_mercado","area\_gerencial")  
model <- glm(reformulate(explicativas,"conc1\_s1"),family=binomial(link='logit'),data=dados,na.action = "na.omit")  
anova(model,test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: conc1\_s1  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev  
## NULL 201 254  
## doutor\_tempo\_concl 1 0,11 200 254  
## mestr\_tempo\_concl 1 0,00 199 254  
## docencia\_tempo 1 1,87 198 252  
## se\_doutor\_cont\_1s 1 1,77 197 250  
## se\_mestre\_contabeis\_1s 1 4,19 196 246  
## public\_agrp 3 6,03 193 240  
## certif\_atual 1 1,16 192 239  
## mercado\_atual 6 10,15 186 228  
## for\_pedag 1 0,00 185 228  
## genero\_1M 1 0,00 184 228  
## regime\_atual 2 1,00 182 228  
## cat\_adm\_atual 2 0,23 180 227  
## regiao\_geografica 4 9,77 176 218  
## Organ\_acad\_atual 2 3,32 174 214  
## experiencia\_mercado 1 0,52 173 214  
## area\_gerencial 1 3,41 172 210  
## experiencia\_mercado:tempo\_mercado 1 1,26 171 209  
## Pr(>Chi)   
## NULL   
## doutor\_tempo\_concl 0,739   
## mestr\_tempo\_concl 0,944   
## docencia\_tempo 0,171   
## se\_doutor\_cont\_1s 0,183   
## se\_mestre\_contabeis\_1s 0,041 \*  
## public\_agrp 0,110   
## certif\_atual 0,281   
## mercado\_atual 0,119   
## for\_pedag 0,985   
## genero\_1M 0,978   
## regime\_atual 0,606   
## cat\_adm\_atual 0,894   
## regiao\_geografica 0,044 \*  
## Organ\_acad\_atual 0,190   
## experiencia\_mercado 0,472   
## area\_gerencial 0,065 .  
## experiencia\_mercado:tempo\_mercado 0,263   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

#### usando tudo, não encontra nada  
# usando as variaveis que apresentaram diferença na análise bivariada  
explicativas2 <- explicativas[c(5,7,13,17)]  
model2<- glm(reformulate(explicativas2,"conc1\_s1"),family=binomial(link='logit'),data=dados,na.action = "na.omit")  
summary(model2)

##   
## Call:  
## glm(formula = reformulate(explicativas2, "conc1\_s1"), family = binomial(link = "logit"),   
## data = dados, na.action = "na.omit")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2,005 -1,240 0,760 0,964 1,245   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1,300 0,287 4,53 5,8e-06 \*\*\*  
## se\_mestre\_contabeis\_1s1 -0,238 0,202 -1,18 0,2386   
## certif\_atual1 -0,536 0,239 -2,24 0,0250 \*   
## regiao\_geografica2 -0,678 0,230 -2,95 0,0032 \*\*   
## regiao\_geografica3 -0,185 0,500 -0,37 0,7114   
## regiao\_geografica4 -0,683 0,296 -2,30 0,0213 \*   
## regiao\_geografica5 -0,200 0,367 -0,55 0,5852   
## area\_gerencial1 0,567 0,201 2,83 0,0047 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 642,35 on 499 degrees of freedom  
## Residual deviance: 616,84 on 492 degrees of freedom  
## (150 observations deleted due to missingness)  
## AIC: 632,8  
##   
## Number of Fisher Scoring iterations: 4

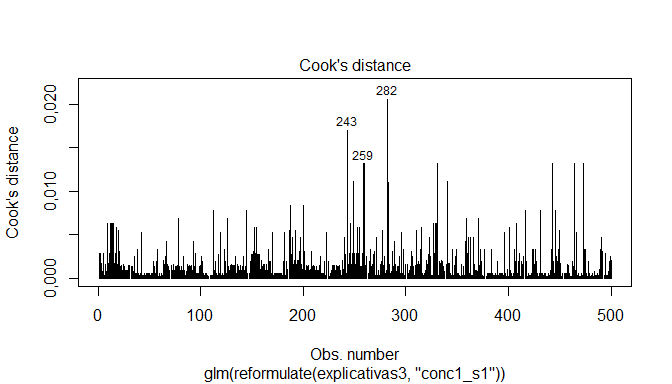
anova(model2,test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: conc1\_s1  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 499 642   
## se\_mestre\_contabeis\_1s 1 1,52 498 641 0,2174   
## certif\_atual 1 5,19 497 636 0,0228 \*   
## regiao\_geografica 4 10,64 493 625 0,0309 \*   
## area\_gerencial 1 8,16 492 617 0,0043 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

#  
explicativas3 <- explicativas[c(7,13,17)]  
model3<- glm(reformulate(explicativas3,"conc1\_s1"),family=binomial(link='logit'),data=dados,na.action = "na.omit")  
summary(model3)

##   
## Call:  
## glm(formula = reformulate(explicativas3, "conc1\_s1"), family = binomial(link = "logit"),   
## data = dados, na.action = "na.omit")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1,944 -1,204 0,730 0,961 1,204   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1,138 0,250 4,55 5,3e-06 \*\*\*  
## certif\_atual1 -0,539 0,239 -2,26 0,0240 \*   
## regiao\_geografica2 -0,661 0,229 -2,88 0,0040 \*\*   
## regiao\_geografica3 -0,193 0,499 -0,39 0,6988   
## regiao\_geografica4 -0,652 0,295 -2,21 0,0269 \*   
## regiao\_geografica5 -0,202 0,366 -0,55 0,5812   
## area\_gerencial1 0,587 0,200 2,94 0,0033 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 642,35 on 499 degrees of freedom  
## Residual deviance: 618,24 on 493 degrees of freedom  
## (150 observations deleted due to missingness)  
## AIC: 632,2  
##   
## Number of Fisher Scoring iterations: 4

plot(model3, which = 4, id.n = 3)



anova(model3,test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: conc1\_s1  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 499 642   
## certif\_atual 1 5,33 498 637 0,021 \*   
## regiao\_geografica 4 9,96 494 627 0,041 \*   
## area\_gerencial 1 8,83 493 618 0,003 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

pR2(model3)

## llh llhNull G2 McFadden r2ML r2CU   
## -309,1183 -416,6731 215,1094 0,2581 0,3496 0,4310

table(ifelse(model3$fitted.values>.5,1,0),model3$y)

##   
## 0 1  
## 0 46 36  
## 1 125 293

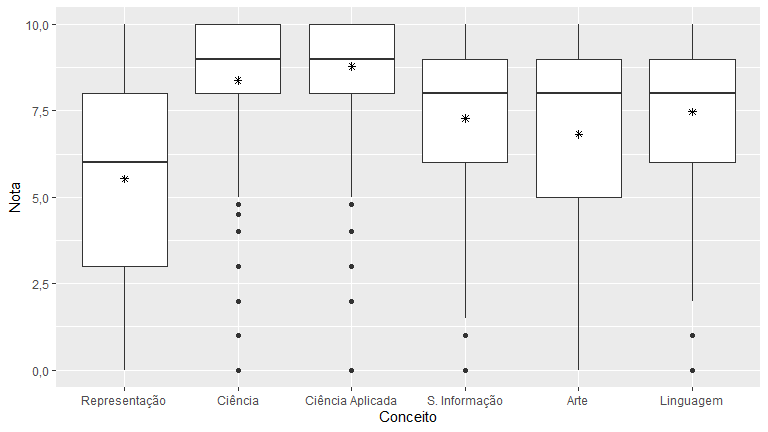
mean(ifelse(model3$fitted.values>.5,1,0)==model3$y)

## [1] 0,678

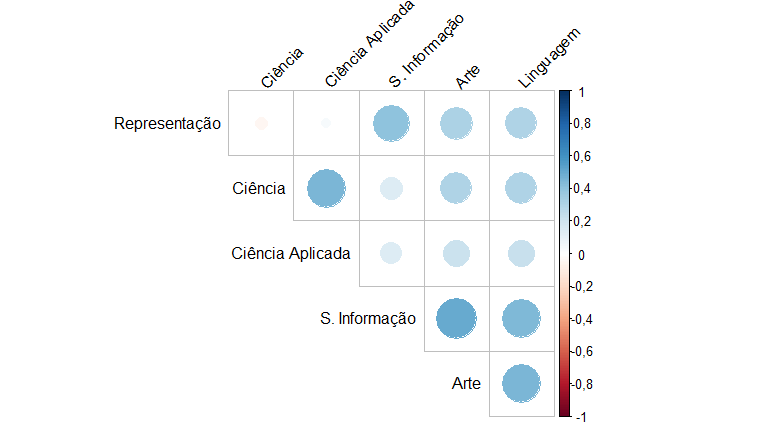
#  
p <- predict(model3, type="response")  
pr <- prediction(p, model3$y)  
prf <- performance(pr, measure = "tpr", x.measure = "fpr")  
plot(prf,xlab="Taxa de falsos positivos",ylab="Taxa de positivos verdadeiros")  
abline(a = 0, b = 1,lty=2)

Gráficos com nomenclatura ajustada

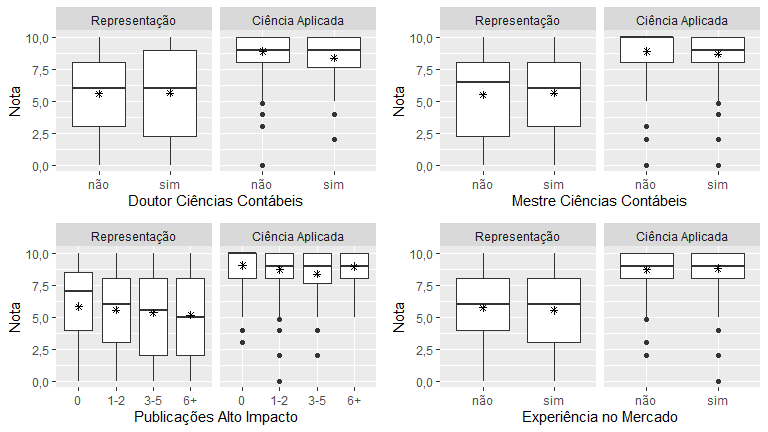
setwd("C:\\Users\\ contabilidade")  
library(MASS)  
library(reshape2)  
library(ggplot2)  
library(plyr)  
library(tidyverse)  
library(corrplot)  
library(rpart)  
library(rpart.plot)  
library(pscl)  
library(ROCR)  
library(gridExtra)  
options(digits=4,OutDec=',')  
###################################################################################  
dados = read.table("dados2.csv",sep=';',dec=',',header = T,encoding="UTF-8",stringsAsFactors=FALSE)  
dados$regiao\_geografica[dados$regiao\_geografica==25]=1  
names(dados)[1] = "ID"  
#str(dados)  
#  
mconceito = melt(dados[,c(1,3,5,7,9,11,13)],id="ID",variable.name = "Conceito",value.name = "Nota")  
#str(mconceito)  
mconceito$Conceito=mapvalues(mconceito$Conceito, from = c("conceito1\_modelo", "conceito2\_franco", "conceito3\_usp", "conceito4\_sist\_In", "conceito5\_arte", "conceito6\_linguag"), to = c("Representação", "Ciência","Ciência Aplicada","S. Informação","Arte","Linguagem"))  
  
p<-qplot(Conceito,Nota,data=mconceito,geom='boxplot')  
p<-p+stat\_summary(fun.y=mean,shape=8,col=1,geom='point')  
print(p)



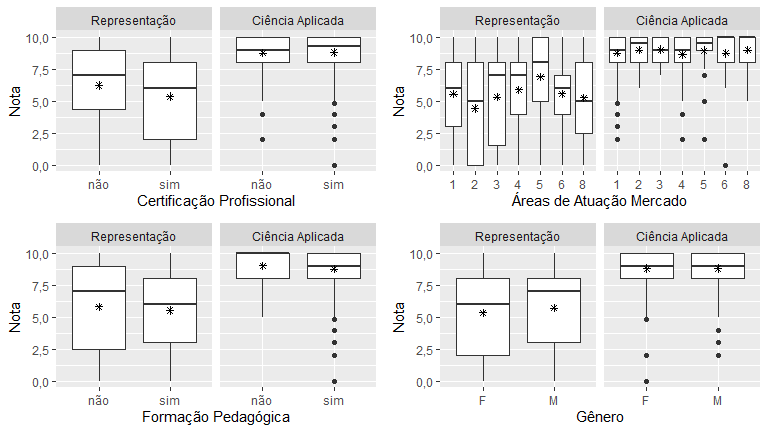
#  
dat1<-dcast(data = mconceito,formula = ID~Conceito,fun.aggregate = sum,value.var = "Nota")  
corMat = round(cor(dat1[,-1]),2)  
corrplot(corMat, type = "upper", order = "original", diag=F,  
 tl.col = "black", tl.srt = 45)



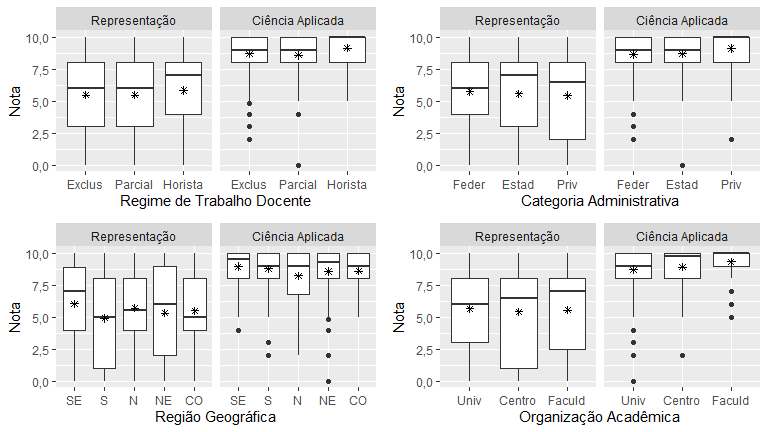
##########################################################  
#parte 1: conceitos vs qualificações  
##########################################################  
dados$public\_agrp <- as.factor(ifelse(dados$public\_atual==0,"0",  
 ifelse(dados$public\_atual<=2,"1-2",  
 ifelse(dados$public\_atual<=5,"3-5","6+"))))  
dados$experiencia\_mercado = ifelse(dados$tempo\_mercado==0,0,1)  
dados$certif\_atual = ifelse((dados$certificac\_atual==1)|(dados$certif\_crcatual==1),1,0)  
dados$mercado\_atual= ifelse(dados$mercado\_atual==7,8,dados$mercado\_atual)  
###  
#####################################################################  
vars\_num=c("doutor\_tempo\_concl","mestr\_tempo\_concl","docencia\_tempo","tempo\_mercado")  
#corMat <- dados %>% select(vars\_num) %>% cor(use = "complete.obs") %>% round(2)  
#corMat = round(cor(dados[,c(3,5,7,9,11,13)]),2)  
#corMat  
vars\_cat = c("se\_doutor\_cont\_1s","se\_mestre\_contabeis\_1s","public\_agrp",  
 "certif\_atual","mercado\_atual","for\_pedag","genero\_1M",  
 "regime\_atual","cat\_adm\_atual","regiao\_geografica","Organ\_acad\_atual",  
 "experiencia\_mercado")  
dados <- dados %>% mutate\_each(funs(factor(.)),vars\_cat)   
#usar interação experiencia\_mercado:tempo\_mercado  
  
### análise bivariada  
extras = c("area\_societaria","area\_custos","area\_gerencial","area\_publica","area\_teoria",  
 "area\_etica","area\_auditoria","area\_pericia","area\_tributaria")  
dados <- dados %>% mutate\_each(funs(factor(.)),extras)  
conc13 = melt(dados[,c(1,3,7)],id="ID",variable.name = "Conceito",value.name = "Nota")  
conc13\_1s = melt(dados[,c(1,2,6)],id="ID",variable.name = "Conceito",value.name = "Aceitacao")  
dconc13 = merge(dados[,c("ID",vars\_num,vars\_cat,extras)],conc13,by="ID",all.y=T)  
dconc13\_1s = merge(dados[,c("ID",vars\_num,vars\_cat,extras)],conc13\_1s,by="ID",all.y=T)  
#  
dconc13$Conceito=mapvalues(dconc13$Conceito, from = c("conceito1\_modelo", "conceito3\_usp"), to = c("Representação", "Ciência Aplicada"))  
dconc13\_1s$Conceito=mapvalues(dconc13\_1s$Conceito, from = c("conc1\_1s", "conc3\_s1"), to = c("Representação-S/N", "Ciência Aplicada-S/N"))  
dconc13$se\_doutor\_cont\_1s = mapvalues(dconc13$se\_doutor\_cont\_1s,from=c("0","1"),to=c("não","sim"))  
dconc13$se\_mestre\_contabeis\_1s = mapvalues(dconc13$se\_mestre\_contabeis\_1s,from=c("0","1"),to=c("não","sim"))  
dconc13$experiencia\_mercado = mapvalues(dconc13$experiencia\_mercado,from=c("0","1"),to=c("não","sim"))  
dconc13$certif\_atual = mapvalues(dconc13$certif\_atual,from=c("0","1"),to=c("não","sim"))  
dconc13$for\_pedag = mapvalues(dconc13$for\_pedag,from=c("0","1"),to=c("não","sim"))  
dconc13$genero\_1M = mapvalues(dconc13$genero\_1M,from=c("0","1"),to=c("F","M"))  
dconc13$regime\_atual2 = mapvalues(dconc13$regime\_atual,from=c("1","2","3"), to=c("Exclus","Parcial","Horista"))  
dconc13$cat\_adm\_atual = mapvalues(dconc13$cat\_adm\_atual,from=c("1","2","3"), to=c("Feder","Estad","Priv"))  
dconc13$regiao\_geografica = mapvalues(dconc13$regiao\_geografica, from=c("1","2","3","4","5"), to=c("SE","S","N","NE","CO"))  
dconc13$Organ\_acad\_atual = mapvalues(dconc13$Organ\_acad\_atual,from=c("1","2","3"), to=c("Univ","Centro","Faculd"))  
  
#  
#se\_doutor\_cont\_1s  
g1<-ggplot(subset(dconc13,!is.na(se\_doutor\_cont\_1s)), aes(y=Nota, x=se\_doutor\_cont\_1s)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Doutor Ciências Contábeis")  
  
g2<-ggplot(subset(dconc13,!is.na(se\_mestre\_contabeis\_1s)), aes(y=Nota, x=se\_mestre\_contabeis\_1s)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Mestre Ciências Contábeis")   
  
#public\_atual --> public\_agrp  
g3<-ggplot(subset(dconc13,!is.na(public\_agrp)), aes(y=Nota, x=public\_agrp)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)+ xlab("Publicações Alto Impacto")   
#  
g4<-ggplot(subset(dconc13,!is.na(experiencia\_mercado)), aes(y=Nota, x=experiencia\_mercado)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8)+xlab("Experiência no Mercado")  
  
#certif\_atual  
g5<-ggplot(subset(dconc13,!is.na(certif\_atual)), aes(y=Nota, x=certif\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Certificação Profissional")  
  
#mercado\_atual  
g6<-ggplot(subset(dconc13,!is.na(mercado\_atual)), aes(y=Nota, x=mercado\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Áreas de Atuação Mercado")  
#for\_pedag  
g7<-ggplot(subset(dconc13,!is.na(for\_pedag)), aes(y=Nota, x=for\_pedag)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Formação Pedagógica")  
#genero\_1M  
g8<-ggplot(subset(dconc13,!is.na(genero\_1M)), aes(y=Nota, x=genero\_1M)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Gênero")  
#  
#regime\_atual  
g9<-ggplot(subset(dconc13,!is.na(regime\_atual2)), aes(y=Nota, x=regime\_atual2)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Regime de Trabalho Docente")  
#cat\_adm\_atual  
g10<-ggplot(subset(dconc13,!is.na(cat\_adm\_atual)), aes(y=Nota, x=cat\_adm\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Categoria Administrativa")  
#regiao\_geografica  
g11<-ggplot(subset(dconc13,!is.na(regiao\_geografica)), aes(y=Nota, x=regiao\_geografica)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Região Geográfica")  
#  
g12<-ggplot(subset(dconc13,!is.na(Organ\_acad\_atual)), aes(y=Nota, x=Organ\_acad\_atual)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) + xlab("Organização Acadêmica")  
#doutor\_tempo\_concl  
#hist(dados$doutor\_tempo\_concl)  
##########################################  
c1<-ggplot(dconc13, aes(y=Nota, x=doutor\_tempo\_concl)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point() +xlab("Tempo Conclusão Doutorado")  
#  
dconc13\_1s$Aceitacao= mapvalues(dconc13\_1s$Aceitacao,from=c(0,1), to=c("Rejeita","Aceita"))  
dconc13\_1s$Aceitacao = factor(dconc13\_1s$Aceitacao,levels = c("Rejeita","Aceita"),ordered = T)  
c2<-ggplot(dconc13\_1s, aes(y=doutor\_tempo\_concl, x=(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação do Conceito") +ylab("Tempo Conclusão Doutorado") + theme(axis.title.y = element\_text(size=10))  
#tempo\_mercado  
#mestr\_tempo\_concl  
#hist(dados$mestr\_tempo\_concl)  
#cor(dados$doutor\_tempo\_concl,dados$mestr\_tempo\_concl,use="complete.obs")  
c3<-ggplot(dconc13, aes(y=Nota, x=mestr\_tempo\_concl)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point() +xlab("Tempo Conclusão Mestrado")  
#cor(dados$conceito1,dados$mestr\_tempo\_concl,use="complete.obs")  
#cor(dados$conceito3,dados$mestr\_tempo\_concl,use="complete.obs")  
#  
c4<-ggplot(dconc13\_1s, aes(y=mestr\_tempo\_concl, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação do Conceito")+ylab("Tempo Conclusão Mestrado")+ theme(axis.title.y = element\_text(size=10))  
#tempo\_mercado  
c5<-ggplot(dconc13, aes(y=Nota, x=tempo\_mercado)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point() +xlab("Tempo de Atuação Mercado")  
#  
c6<-ggplot(dconc13\_1s, aes(y=tempo\_mercado, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação do Conceito")+ylab("Tempo de Atuação Mercado")  
  
#docencia\_tempo  
c7<-ggplot(dconc13, aes(y=Nota, x=docencia\_tempo)) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_point() +xlab("Tempo de Atuação Docência")  
#  
c8<-ggplot(dconc13\_1s, aes(y=docencia\_tempo, x=as.factor(Aceitacao))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab("Aceitação do Conceito")+ylab("Tempo de Atuação Docência")  
#area de atuação - mais de uma resposta, ver se tem alguma de interesse  
#str(dconc13$Conceito)  
labvec <- c("Contabilidade Societária","Contabilidade de Custos",  
 "Contabilidade Gerencial","Contabilidade Pública",  
 "Teoria e Normas Contabilidade","Ética e Legislação Profissional",  
 "Auditoria","Perícia Contábil","Contabilidade Tributária")  
#cbind(extras,labvec)  
for (i in extras) dconc13[,i] = mapvalues(dconc13[,i], from=c("0","1"),to=c("não","sim"))  
#dconc13[,extras[1]]  
  
plot\_extra = function (i) {  
 ggplot(subset(dconc13,!is.na(eval(parse(text = i)))),   
 aes(y=Nota, x=eval(parse(text = i)))) +  
 facet\_wrap(~ Conceito, scales = "fixed") + geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point", shape=8) +xlab(labvec[extras==i])  
 #tab <-ftable(dconc13\_1s$Conceito,dconc13\_1s$Aceitacao,eval(parse(text = paste0("dconc13\_1s$",i))))  
 #pval = chisq.test(table(eval(parse(text = paste0("dados$",i))),dados$conc1\_1s))$p.value  
 #return(list(p,tab,pval))  
 }  
  
myplots <- lapply(extras, plot\_extra)  
#myplots[[4]]  
  
####gráficos   
grid.arrange(g1,g2,g3,g4,ncol=2)



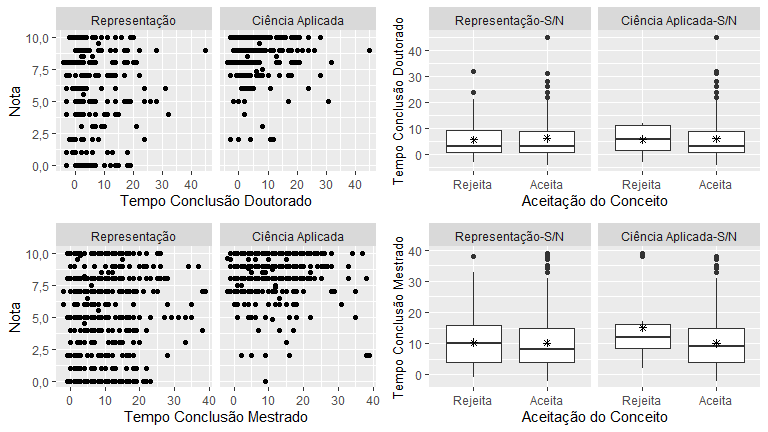
grid.arrange(g5,g6,g7,g8,ncol=2)



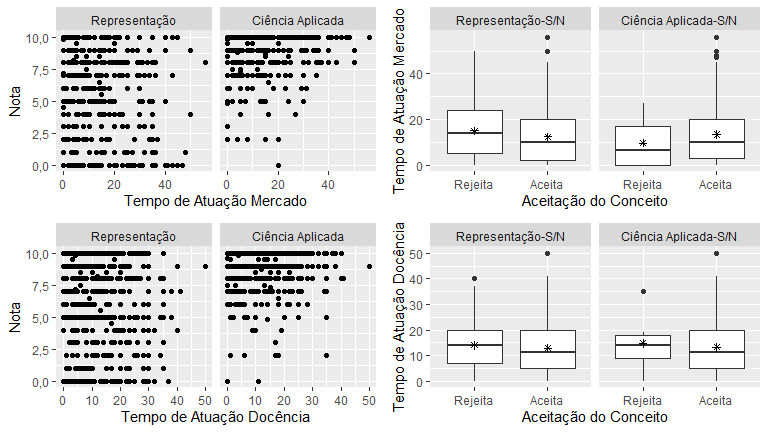
grid.arrange(g9,g10,g11,g12,ncol=2)



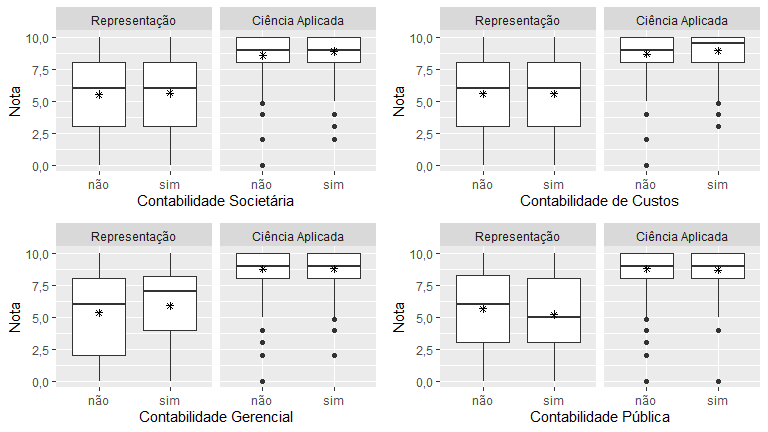
grid.arrange(c1,c2,c3,c4,ncol=2)



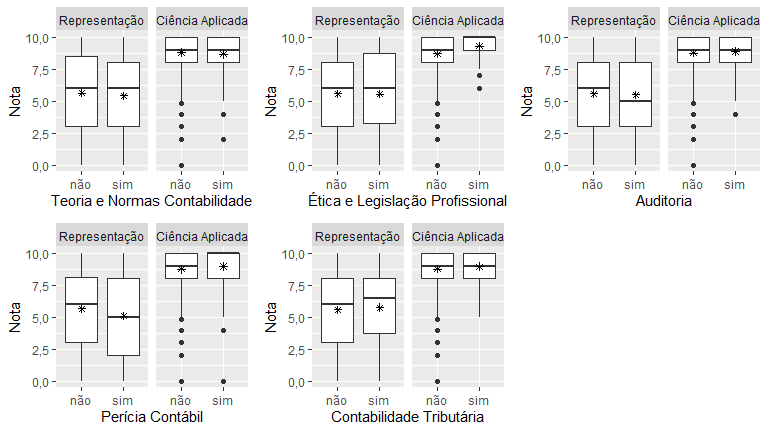
grid.arrange(c5,c6,c7,c8,ncol=2)



grid.arrange(myplots[[1]],myplots[[2]],myplots[[3]],myplots[[4]],ncol=2)



grid.arrange(myplots[[5]],myplots[[6]],myplots[[7]],myplots[[8]],myplots[[9]],ncol=3)



**resultados para o modelo completo Regressão Beta:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Coefficients (mean model with logit link): | | | |  |  |
|  | Estimate | Std. Error | z value | Pr(>|z|) |  |
| (Intercept) | 0.423536 | 0.53393 | 0.793 | 0.427637 |  |
| public\_agrp1-2 | 0.131429 | 0.119473 | 1.1 | 0.271298 |  |
| public\_agrp3-5 | 0.374116 | 0.137062 | 2.73 | 0.006342 | \*\* |
| public\_agrp6 ou mais | 0.398702 | 0.202568 | 1.968 | 0.049041 | \* |
| conceito1 | 0.013582 | 0.017299 | 0.785 | 0.432366 |  |
| conceito3 | -0.00739 | 0.035427 | -0.209 | 0.834679 |  |
| mestr\_tempo\_concl | 0.050202 | 0.012941 | 3.879 | 0.000105 | \*\*\* |
| docencia\_tempo | -0.03754 | 0.01174 | -3.198 | 0.001384 | \*\* |
| Organ\_acad\_atual2 | -0.07805 | 0.151754 | -0.514 | 0.607034 |  |
| Organ\_acad\_atual3 | -0.29106 | 0.143073 | -2.034 | 0.041914 | \* |
| cat\_adm\_atual2 | -0.39089 | 0.149593 | -2.613 | 0.008975 | \*\* |
| cat\_adm\_atual3 | -0.84467 | 0.164174 | -5.145 | 2.68E-07 | \*\*\* |
| nivel\_conh\_EC | -0.0035 | 0.04164 | -0.084 | 0.933057 |  |
| regiao\_geografica2 | -0.03974 | 0.119934 | -0.331 | 0.740414 |  |
| regiao\_geografica3 | -0.91009 | 0.219983 | -4.137 | 3.52E-05 | \*\*\* |
| regiao\_geografica4 | -0.4909 | 0.143643 | -3.417 | 0.000632 | \*\*\* |
| regiao\_geografica5 | -0.3303 | 0.196792 | -1.678 | 0.093266 | . |
| genero\_1M | -0.1327 | 0.115327 | -1.151 | 0.249881 |  |
| parcialIntegral | 0.199844 | 0.131765 | 1.517 | 0.12935 |  |
| se\_doutor\_cont\_1s | 0.324468 | 0.155022 | 2.093 | 0.036346 | \* |
| se\_mestre\_contabeis\_1s | -0.03417 | 0.126765 | -0.27 | 0.787482 |  |
| tempo\_mercado | 0.002673 | 0.005975 | 0.447 | 0.65458 |  |
| for\_pedag | -0.07086 | 0.165038 | -0.429 | 0.667652 |  |
| certif\_atual | 0.166909 | 0.152292 | 1.096 | 0.273087 |  |