|  |
| --- |
| title: “Estadística Aplicada II- Taller # 1” |
| author: “David Rojas y Felipe Palomino” |
| date: “31/03/2023” |
| output: |
| html\_document: default |

) Describa las variables e individuos del subgrupo asignado en la base Yarn. Realice el gráfico de dispersión entre las variables y=densidad y la variable x= que presentan la mayor correlación.\ Para el cálculo de estas estadística realizamos una función que agrupaba diferentes descriptivas en una sola tabla, añadiendo a su vez, histogramas, junto con las correlaciones lineales de pearson entre las covariables.

#Importamos la base de datos  
##############################  
options(scipen=999) # Realizamos este cambio en options para trabajar sin notación científica  
#  
library(easypackages) #Librería para la carga de base de datos  
lib\_req<-c('psych','car','lmtest','MASS','xtable','latex2exp',  
 'orcutt','nlme',  
 'mixtools',"alr4","depth","readr",  
 "ddalpha","robustbase","rrcov","zoom",  
 'ggfortify','readxl')# Listado de librerias requeridas por el script  
easypackages::packages(lib\_req)   
data <- read\_excel("C:/Users/sebas/OneDrive/Escritorio/Octavo Semestre/OctavoSemestre/Estadística Aplicada II/Base de datos/data.xlsx")  
#Función creada para las estadística descriptivas  
resumen<- function(x){  
 X<- matrix(0,7,1)  
 resumen<- round(c(mean(x),median(x),min(x),max(x)  
 ,sd(x),quantile(x,0.25),  
 quantile(x,0.75)),4)  
 for( i in 1:7){  
 X[i,]<- resumen[i]  
 }  
 rownames(X)<-c('Media','Mediana','Min','Max','Sd','1st Qu.','3rd Qu')  
 colnames(X)<- (paste(colnames(x)))  
 return(X)  
}  
X<- data  
################ Selección de las variables dadas por el docente  
X<- cbind(X[,1:30],X[,colnames(X)=='density'])  
# EstadC-stica descriptivas  
summary<- matrix(0,31,7)  
for(i in 1:31){  
 summary[i,]<- t(resumen(X[,i]))  
}  
rownames(summary)<-colnames(X)  
colnames(summary)<-c('Media','Mediana','Min','Max','Sd','1st Qu.','3rd Qu')  
print(summary)

## Media Mediana Min Max Sd 1st Qu. 3rd Qu  
## NIR1 3.0915 3.0970 2.8733 3.1948 0.0579 3.0749 3.1238  
## NIR2 3.0674 3.0874 2.6803 3.1762 0.0996 3.0709 3.1151  
## NIR3 3.0056 3.0526 2.4531 3.1329 0.1541 2.9842 3.1014  
## NIR4 2.9009 2.9707 2.2380 3.0996 0.2051 2.8474 3.0260  
## NIR5 2.7564 2.8309 2.0406 3.0249 0.2411 2.6713 2.8987  
## NIR6 2.5992 2.6676 1.8602 2.9345 0.2649 2.4897 2.7686  
## NIR7 2.4557 2.4997 1.7071 2.8884 0.2840 2.2896 2.6405  
## NIR8 2.3365 2.3747 1.5946 2.8716 0.3019 2.1641 2.5185  
## NIR9 2.2482 2.2618 1.5207 2.8850 0.3244 2.0728 2.4643  
## NIR10 2.1909 2.1788 1.4652 2.9260 0.3578 1.9656 2.4368  
## NIR11 2.1574 2.1368 1.4177 2.9908 0.4021 1.8764 2.4382  
## NIR12 2.1426 2.1382 1.3805 3.0716 0.4522 1.8550 2.4615  
## NIR13 2.1445 2.1408 1.3522 3.1508 0.5000 1.8555 2.5006  
## NIR14 2.1597 2.1597 1.3305 3.2129 0.5397 1.8664 2.5519  
## NIR15 2.1778 2.1854 1.3147 3.2603 0.5751 1.8816 2.6067  
## NIR16 2.1811 2.1945 1.3045 3.2946 0.6116 1.8902 2.6454  
## NIR17 2.1670 2.1808 1.2206 3.3168 0.6437 1.8436 2.6576  
## NIR18 2.1424 2.1525 1.1410 3.3254 0.6584 1.7923 2.6396  
## NIR19 2.1042 2.1044 1.1161 3.3066 0.6499 1.7518 2.5851  
## NIR20 2.0450 2.0324 1.1329 3.2575 0.6194 1.7158 2.4873  
## NIR21 1.9613 1.9386 1.1604 3.1723 0.5725 1.6660 2.3504  
## NIR22 1.8536 1.8243 1.1638 3.0267 0.5166 1.5873 2.1876  
## NIR23 1.7424 1.7112 1.1527 2.8334 0.4590 1.5025 2.0294  
## NIR24 1.6474 1.6189 1.1228 2.6243 0.4012 1.4293 1.8937  
## NIR25 1.5621 1.5368 1.1170 2.4256 0.3488 1.3780 1.7713  
## NIR26 1.4846 1.4568 1.0681 2.2585 0.3105 1.3052 1.6629  
## NIR27 1.4115 1.3705 1.0197 2.0801 0.2709 1.2327 1.5629  
## NIR28 1.3242 1.2939 1.0084 1.8021 0.2009 1.1805 1.4481  
## NIR29 1.2470 1.2267 1.0321 1.5330 0.1246 1.1558 1.3263  
## NIR30 1.2089 1.2107 1.0828 1.3570 0.0681 1.1618 1.2605  
## density 33.7507 31.2200 0.0000 100.0000 26.9621 20.2625 50.5000

#Selección de variables con mayor correlación lineal de pearson positiva o negativa  
Y<-cor(X)  
Y[31,]

## NIR1 NIR2 NIR3 NIR4 NIR5 NIR6   
## 0.18785681 0.42415816 0.53165763 0.54980711 0.50034480 0.40385529   
## NIR7 NIR8 NIR9 NIR10 NIR11 NIR12   
## 0.27277376 0.11693283 -0.04258148 -0.17899885 -0.28525901 -0.36425811   
## NIR13 NIR14 NIR15 NIR16 NIR17 NIR18   
## -0.41420050 -0.43972514 -0.45811600 -0.48698764 -0.52031069 -0.54262433   
## NIR19 NIR20 NIR21 NIR22 NIR23 NIR24   
## -0.54768143 -0.53662425 -0.51913210 -0.50698335 -0.49586830 -0.48293275   
## NIR25 NIR26 NIR27 NIR28 NIR29 NIR30   
## -0.49871555 -0.55762713 -0.62007626 -0.65693381 -0.66071886 -0.61289059   
## density   
## 1.00000000

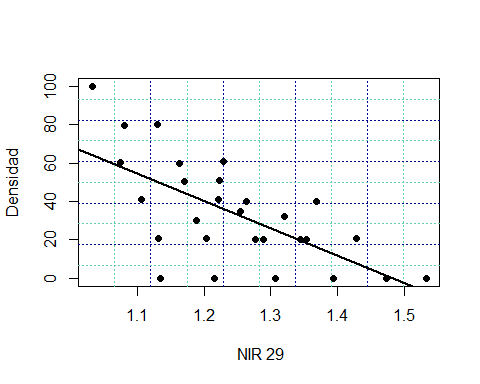
Y<-Y[,-31]  
cor<-c(max(Y[31,]),min(Y[31,]))  
which(Y[31,]==cor[1])

## NIR4   
## 4

which(Y[31,]==cor[2])

## NIR29   
## 29

windows(height=10, width=10)   
#Creación del panel de fondo  
par(mfrow=c(1,1))  
plot(seq(min(X[,29]),max(X[,29]),length.out=30),seq(min(X[,31]),max(X[,31]),length.out=30),type='n',xlab='',ylab='')  
grid(10,10,col=c('aquamarine3','blue4'))  
par(new=T)  
plot(X[,31]~X[,29],ylab='Densidad',xlab=' NIR 29',pch=19,axes=F)  
model<- lm(density~NIR29,data=X)  
abline(model,lwd=2)

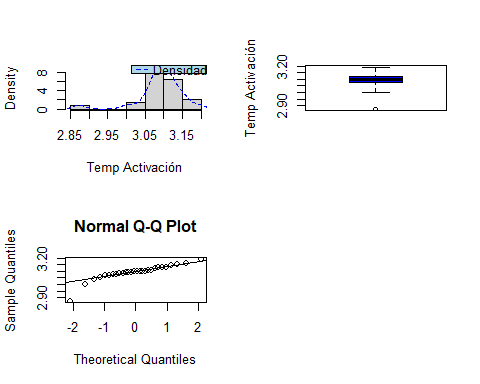


summary(model)

##   
## Call:  
## lm(formula = density ~ NIR29, data = X)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -49.856 -11.991 2.011 13.049 35.533   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 212.02 39.91 5.312 0.0000148 \*\*\*  
## NIR29 -142.96 31.85 -4.488 0.00013 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.62 on 26 degrees of freedom  
## Multiple R-squared: 0.4365, Adjusted R-squared: 0.4149   
## F-statistic: 20.14 on 1 and 26 DF, p-value: 0.0001298

<< echo=TRUE,warning=FALSE, message=FALSE>>=

windows(height=8, width=10) # Nueva ventana gráfica, con dimensiones preestablecidas  
par(mfrow=c(2,2)) # particón dela ventana gráfica  
  
hist(X[,1], freq=F,main="",xlab="Temp Activación")  
lines(density(X[,1]), col="blue",lty=2)  
legend("topright","Densidad estimada",lty=2, col="blue",bg='lightblue')  
boxplot(X[,1],col="Blue", ylab="Temp Activación")  
qqnorm(X[,1]); qqline(X[,1])

 @

La forma asimétrica del histograma y del boxplot, ademas de la distancia de los puntos a la linea recta en el *qqplot*, sugieren que los datos no provienen de una distribución *Normal*. Este resultados son corroborados por el *test de Shapiro-Wilk*.

# Un Test formal de Normalidad  
shapiro.test(X[,1])

##   
## Shapiro-Wilk normality test  
##   
## data: X[, 1]  
## W = 0.84011, p-value = 0.0005942

##### Ejercicio 4.

Al seleccionar un concreto azufrado para la construcción de carreteras en regiones que experimentan congelamiento intenso, es importante que el concreto tenga baja conductividad térmica, a fin de reducir el da?o posterior debido a temperaturas cambiantes. Se ha pensado que la adición de agregados graduados a las mezcla de concreto puede disminuir su conductividad media. Para ello se ha diseñado un experimento que consiste en generar especímenes de prueba bajo las dos condiciones, (con agregados, sin agregados). Los resultados son lo siguientes:

***Con agregados***

0.166, 0.296, 0.249, 0.366, 0.415, 0.388, 0.273, 0.324, 0.413, 0.280, 0.247, 0.130, 0.279, 0.364, 0.402, 0.202, 0.255, 0.315, 0.375, 0.316

***Sin agregados***

0.426, 0.461, 0.476, 0.467, 0.436, 0.509, 0.428, 0.508, 0.457, 0.513, 0.448, 0.523, 0.431, 0.439, 0.395, 0.443, 0.361, 0.480, 0.467, 0.436, 0.471, 0.487, 0.498, 0.538, 0.500

De acuerdo con los datos, incorporar agregados disminuye la conductividad térmica del material?

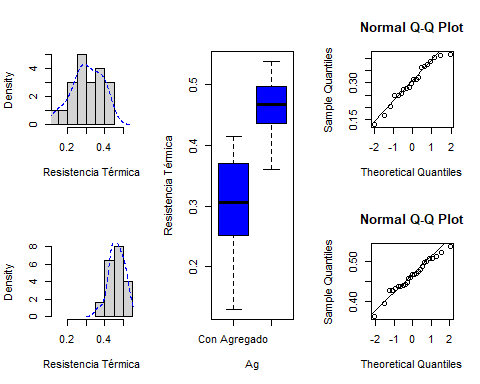
#Ingresando los datos en el formato variable - indicador  
rt = c(0.166, 0.296, 0.249, 0.366, 0.415, 0.388, 0.273, 0.324, 0.413, 0.280, 0.247, 0.130,   
 0.279, 0.364, 0.402, 0.202, 0.255, 0.315, 0.375, 0.316,  
 0.426, 0.461, 0.476, 0.467, 0.436, 0.509, 0.428, 0.508, 0.457, 0.513, 0.448, 0.523,   
 0.431, 0.439, 0.395, 0.443, 0.361, 0.480, 0.467, 0.436, 0.471, 0.487, 0.498, 0.538, 0.500)  
Ag = factor(c(rep("Con Agregado",20), rep("Sin Agregado",25)))

Inicialmente realizaremos un análisis exploratorio, que incluye el calculo de algunos indicadores descriptivos y la representación gráfica de los datos.

# Algunas medidas resumen.  
promedios = tapply(rt,Ag,mean)  
desviacion = tapply(rt,Ag,sd)  
C\_var = desviacion/promedios  
resumen =round(cbind(promedios, desviacion, C\_var),2)  
colnames(resumen)= c("Promedio","Desviación","Coef.Var")  
resumen

## Promedio Desviación Coef.Var  
## Con Agregado 0.30 0.08 0.27  
## Sin Agregado 0.46 0.04 0.09

#Cambiando los datos al formato de columnas  
rt\_ca=rt[Ag =="Con Agregado"] #resistencia térmica con agregado  
rt\_sa=rt[Ag =="Sin Agregado"] #resistencia térmica sin agregado  
  
#Preparando la ventana gráfica  
M = matrix(c(1,2,3,3,4,5),ncol=3,byrow = F) # matriz con la ubicación de 5 gráficos   
windows(height=20, width=15) # Nueva ventana gráfica   
layout(M) # partición dela ventana gráfica según M  
  
# Histogramas de frecuencia con densidad estimada - igualado el rango en el eje X.  
  
hist(rt\_ca, freq=F,main="",xlab="Resistencia Térmica",xlim=range(rt))  
lines(density(rt\_ca), col="blue",lty=2)  
hist(rt\_sa, freq=F,main="",xlab="Resistencia Térmica",xlim=range(rt))  
lines(density(rt\_sa), col="blue",lty=2)  
boxplot(rt~Ag,col="Blue", ylab="Resistencia Térmica")  
qqnorm(rt\_ca); qqline(rt\_ca)  
qqnorm(rt\_sa); qqline(rt\_sa)



Los resultados muestran que la muestra obtenida con la adición de agregados tiene una menor conductividad eléctrica que aquella muestra que no incorporó estos agregados. La representación gráfica también sugiere que la distribución de la variable, en ambos casos, es simétrica y no se aleja mucho de la distribución normal. Lo anterior se corroboró con el test *Shapiro-Wilk* para ambas muestras.

# Un Test de Normalidad  
shapiro.test(rt\_ca)

##   
## Shapiro-Wilk normality test  
##   
## data: rt\_ca  
## W = 0.95628, p-value = 0.4724

shapiro.test(rt\_sa)

##   
## Shapiro-Wilk normality test  
##   
## data: rt\_sa  
## W = 0.97688, p-value = 0.8171

Verificado el supuesto de normalidad para ambas poblaciones, procedemos a desarrollar el test de hipotésis:

los resultados del test son los siguientes:

# Un Test de Hipotesis Paramétrico  
t.test(rt\_ca,rt\_sa,alternative = "less")

##   
## Welch Two Sample t-test  
##   
## data: rt\_ca and rt\_sa  
## t = -8.0981, df = 26.806, p-value = 0.000000005608  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -0.127262  
## sample estimates:  
## mean of x mean of y   
## 0.30275 0.46392

De acuerdo con el si existe una reducción en la conductividad térmica del concreto cuando se adicionan los agregados en la mezcla