#<파일 불러오기>

setwd<-("C:/Users/82104/Documents")

getwd()

likes<-read.csv('좋아요수\_최고순위\_1위횟수추가\_단위수정.csv', header=TRUE)

str(likes)

#[좋아요수 분석]

library(car)

scatterplotMatrix(likes,pch=19,col="royalblue",cex=1.2,

regLine = list(method=lm,lty=1,lwd=3,col="salmon"),

smooth = list(smoother=loessLine,spread=FALSE,

lty.smooth=1, lwd.smooth=3,col.smooth="forestgreen"),

main="앨범좋아요수,팬수,최고순위,1위횟수,재생수에 따른 곡좋아요수 변화 ")

likes.lm<- lm(좋아요수~앨범좋아요수+팬수+최고순위+X1위횟수+재생수,data=likes)

likes.lm

summary(likes.lm)

lm(formula = 좋아요수 ~ 앨범좋아요수 + 팬수 + 최고순위 + X1위횟수 +

재생수, data = likes)

likealbum<-lm(좋아요수~앨범좋아요수,data=likes)

summary(likealbum)

plot(likes$좋아요수~likes$앨범좋아요수,

col="royalblue",pch=19,

xlab="앨범 좋아요수",ylab="곡 좋아요수",

main=" 앨범좋아요수와 곡좋아요수의 관계")

abline(likealbum,col="salmon",lwd=2)

likefan<-lm(좋아요수~팬수,data=likes)

likefan

summary(likefan)

plot(likes$좋아요수~likes$팬수,

col="royalblue",pch=19,

xlab="팬수",ylab="곡 좋아요수",

main="팬수와 곡좋아요수의 관계")

abline(likefan,col="salmon",lwd=2)

likealbumfan<-lm(앨범좋아요수~팬수,data=likes)

summary(likealbumfan)

plot(likes$앨범좋아요수~likes$팬수,

col="royalblue",pch=19,

xlab="팬 수",ylab="앨범 좋아요수",

main=" 팬수와 앨범좋아요수의 관계")

abline(likealbumfan,col="salmon",lwd=2)

levels(alllike$장르)

tapply(alllike$팬수,alllike$장르,mean)

tapply(alllike$좋아요수,alllike$장르,mean)

liketop<-lm(좋아요수~최고순위,data=likes)

summary(liketop)

plot(likes$좋아요수~likes$최고순위,

col="royalblue",pch=19,

xlab="최고순위",ylab="곡 좋아요수",

main=" 최고순위와 곡좋아요수의 관계")

abline(likealbum,col="salmon",lwd=2)

like1st<-lm(좋아요수~X1위횟수,data=likes)

summary(like1st)

plot(likes$좋아요수~likes$X1위횟수,

col="royalblue",pch=19,

xlab="1위횟수",ylab="곡 좋아요수",

main=" 1위횟수와 곡좋아요수의 관계")

abline(like1st,col="salmon",lwd=2)

likestream<-lm(좋아요수~재생수,data=likes)

summary(likestream)

plot(likes$좋아요수~likes$재생수,

col="royalblue",pch=19,

xlab="재생수",ylab="곡 좋아요수",

main=" 재생수와 곡좋아요수의 관계")

abline(likestream,col="salmon",lwd=2)

like.lm<-lm(좋아요수~앨범좋아요수+재생수+X1위횟수,data=likes)

plot(like.lm)

flike.regsubsets<-regsubsets(x=좋아요수~앨범좋아요수+재생수+X1위횟수,data=likes,nbest=2)

library(RColorBrewer)

plot(flike.regsubsets,scale="adjr2",col=brewer.pal(9,"Pastel1"),

main="all subsets regression")

like.lm<-lm(좋아요수~재생수+X1위횟수,data=alllike)

summary(like.lm)

scatterplotMatrix(flike,

pch=19,col="royalblue",cex=1.2,

regLine = list(method=lm,lty=1,lwd=3,col="salmon"),

smooth = list(smoother=loessLine,spread=FALSE,

lty.smooth=1, lwd.smooth=3,col.smooth="forestgreen"),

main="1위횟수,재생수에 따른 곡좋아요수 변화 ")

[앨범좋아요수 분석]

#변수별 평균 앨범좋아요수

tapply(likes$앨범좋아요수, likes$장르, mean) #랩>댄스>록>R&B>기타>발라드>팝

tapply(likes$앨범좋아요수, likes$연도, mean) #2022>2020>2021>2018

tapply(likes$앨범좋아요수, likes$성별, mean) #남성>혼성>여성, 눈에 띄는 차이x, 혼성은 두 팀밖에 없기 때문에 단순 평균 앨범좋아요수로 비교 불가능

#일원분산분석

result<-aov(앨범좋아요수~성별, data=likes)

result

summary(result) #p-value=0.926, 유의x, 성별 간 앨범좋아요수의 차이는 존재x

result<-aov(앨범좋아요수~장르, data=likes)

result

summary(result) #p-value=8.16e-10, 유의o, 장르 간 앨범좋아요수의 차이는 존재

reesult<-aov(앨범좋아요수~연도, data=likes)

result

summary(result) #p-value=0.0142, 유의o, 연도 간 앨범좋아요수의 차이는 존재

#상관분석

cor(likes$앨범좋아요수, likes$좋아요수) #0.3414813

with(likes, cor.test(앨범좋아요수, 좋아요수)) #p값 0.05보다 작음. 즉, 유의한걸로 나옴!!! 좋아요수와 앨범좋아요수 간에는 상관관계 존재

cor(likes$앨범좋아요수, likes$팬수) #0.7803542

with(likes, cor.test(앨범좋아요수, 팬수)) # p-value < 2.2e-16. 즉, 유의한걸로 나옴!!! 좋아요수와 앨범좋아요수 간에는 상관관계 존재

cor(likes$앨범좋아요수, likes$X1위횟수) #-0.05071845

with(likes, cor.test(앨범좋아요수, X1위횟수)) #p-value= 0.6163, 상관관계 존재x

cor(likes$앨범좋아요수, likes$재생수)

with(likes, cor.test(앨범좋아요수,재생수)) #p값=1.689e-06, 유의, 상관관계 존재

#회귀진단

library(car)

likes.lm<-lm(앨범좋아요수~좋아요수, data=likes)

par(mfrow=c(2,2))

plot(likes.lm)

shapiro.test(resid(likes.lm)) #p값=5.469e-08<0.05, 정규성 조건 충족x

ncvTest(likes.lm)#p값=0.46>0.05, 등분산 조건 충족

durbinWatsonTest(likes.lm)#p값=0.658>0.05, '자기상관이 0이다'라는 귀무가설 기각x, 독립성 가정 충족

likes.lm<-lm(앨범좋아요수~팬수, data=likes)

par(mfrow=c(2,2))

plot(likes.lm)

shapiro.test(resid(likes.lm)) #p값=0.002<0.05, 정규성 조건 충족x

ncvTest(likes.lm)#p값=7.949e-11<0.05, 등분산 조건 충족x

durbinWatsonTest(likes.lm)#p값=0.972>0.05, '자기상관이 0이다'라는 귀무가설 기각x, 독립성 가정 충족

likes.lm<-lm(앨범좋아요수~X1위횟수, data=likes)

par(mfrow=c(2,2))

plot(likes.lm)

shapiro.test(resid(likes.lm)) #p값=2.988e-09<0.05, 정규성 조건 충족x

ncvTest(likes.lm)#p값=0.050517>0.05, 등분산 조건 충족

durbinWatsonTest(likes.lm)#p값=0.186>0.05, '자기상관이 0이다'라는 귀무가설 기각x, 독립성 가정 충족

likes.lm<-lm(앨범좋아요수~재생수, data=likes)

par(mfrow=c(2,2))

plot(likes.lm)

shapiro.test(resid(likes.lm)) #p값=8.147e-07<0.05, 정규성 조건 충족x

ncvTest(likes.lm)#p값= 0.97569>0.05, 등분산 조건 충족

durbinWatsonTest(likes.lm)#p값=0.718>0.05, '자기상관이 0이다'라는 귀무가설 기각x, 독립성 가정 충족

[재생수 분석]

#일원분산분석

likes.aov<-aov(재생수~장르, data=likes)

summary(likes.aov)

likes.aov2<-aov(재생수~성별, data=likes)

summary(likes.aov2)

likes.aov3<-aov(재생수~연도, data=likes)

summary(likes.aov3)

#이원분산분석

likes.aov4<-aov(재생수~장르\*성별, data=likes)

summary(likes.aov4)

likes.aov5<-aov(재생수~장르\*연도, data=likes)

summary(likes.aov5)

likes.aov6<-aov(재생수~성별\*연도, data=likes)

summary(likes.aov6)

t4<-tapply(likes$재생수, likes$성별, mean)

t4

#장르별 재생수(평균값, 더미변수를 활용한 회귀 분석)

result32<-aov(재생수~장르, data=likes)

summary(result32)

song.lm<-lm(재생수~장르, data)

t1<-tapply(likes$재생수, likes$장르, mean)

t1

#팬수 재생수 그래프, 일원분석

result30<-aov(재생수~팬수, data=likes)

summary(result30)

#상관계수, 서브셋 지정

cor.test(~재생수+팬수, data=likes, subset=(성별==”M”))

cor.test(~재생수+팬수, data=likes, subset=(성별==”F”))