

```
#####
##回帰分析
#####
x <- read.csv("setagaya_mansion.csv")
sman = data.frame(price=x[[1]], walk=x[[3]], area=x[[5]], str=ifelse(x[[6]]=="RC",
1, 0), kenpei=x[[7]], youseki=x[[8]], tikunen=x[[9]], kyuko=x[[10]]) #構造はダミー変
数0-1に置き換える.
plot(sman) #まずはプロット! priceとareaの相関が高そう.

#####
##単回帰
sman.lm <- lm(price ~ area, data=sman) #回帰分析はこの一行でOK
plot(sman$area, sman$price, xlab="area", ylab="price") #とにかく結果をプロット!
abline(sman.lm, lwd=1, col="blue")
summary(sman.lm)

sman.lm$coefficients #推定された回帰係数の取り出し
confint(sman.lm) #回帰係数の信頼区間

#回帰分析の診断
par(mfrow=c(2, 2))
plot(sman.lm)
par(mfrow=c(1, 1))

#予測値の取り出し
sman.predict <- predict(sman.lm) #観測点での予測値
(sman.predict_new <- predict(sman.lm, data.frame(area=c(34, 48, 95)))) #新しい点で
の予測値

#予測値の信頼区間を求める.
area.plot <- seq(min(sman$area)*0.9, max(sman$area)*1.1, by=1)
sman.con <- predict(sman.lm, data.frame(area=area.plot), interval="confidence")
#注意! Y|Xの分位点ではなく、回帰直線 ( $\hat{\beta}X$ ) の信頼区間

xlim <- c(min(area.plot), max(area.plot)) #xの表示範囲
ylim <- c(min(sman$price)*0.9, max(sman$price)*1.1) #yの表示範囲

plot(area.plot, sman.con[, 1], type="l", xlim=xlim, ylim=ylim, ylab="", xlab="")
par(new=T)
plot(area.plot, sman.con[, 2], type="l", xlim=xlim, ylim=ylim, col="red", ylab="",
xlab="")
par(new=T)
plot(area.plot, sman.con[, 3], type="l", xlim=xlim, ylim=ylim, col="red", ylab="",
xlab="")
par(new=T)
plot(sman$area, sman$price, xlim=xlim, ylim=ylim, ylab="price", xlab="area (m^2)",
col="blue")
par(new=F)

#残差の取り出し
sman.residuals <- residuals(sman.lm) # 残差
shapiro.test(sman.residuals) #残差が正規分布に従っているか検定
qqnorm(sman.residuals) #正規分布の分位点と照合. 対角線に近いほど正規分布.

#定数項をはずして解析
sman.lm2 <- lm(price ~ area -1, data=sman) # -1をformulaに加えると切片項を除外
summary(sman.lm2)
AIC(sman.lm2) #AICの計算
c(AIC(sman.lm), AIC(sman.lm2)) # 切片項ない方がAICが良かった.

#####
##重回帰
#築年数を加えてみる (切片項は含める)
sman.lm3 <- lm(price ~ area + tikunen, data=sman)
summary(sman.lm3) # Residualsが小さくなっている.
c(AIC(sman.lm), AIC(sman.lm3)) # AICが改良されている.
shapiro.test(sman.lm3$residuals) #残差の正規性検定

#構造(str)を含めてみる
sman.lm4 <- lm(price ~ area + tikunen + walk + str, data=sman)
summary(sman.lm4) # strは意味なさそう
```

```
sman.lm5 <- lm(price ~ area + tikunen + walk, data=sman)
c(AIC(sman.lm4), AIC(sman.lm5)) #実際, strはないほうが予測力がある.

sman.lmall <- lm(price ~., data=sman) #priceをすべての説明変数へ回帰してみる.
sman.lmAIC <- step(sman.lmall) #変数を一つ削除したり一つ追加したりして自動的にAIC
#最小のモデルを探索する関数. グリーディーに解くので本当の最適性は保証されない.
summary(sman.lmAIC)
#三変数+切片のモデルが最終的に選ばれた. すべての変数が有意.
shapiro.test(sman.lmAIC$residuals) #最後に正規性の検定. 残差の正規性は棄却され
#ない (alpha=0.05).
qqnorm(sman.lmAIC$residual) #正規分布に近い
```