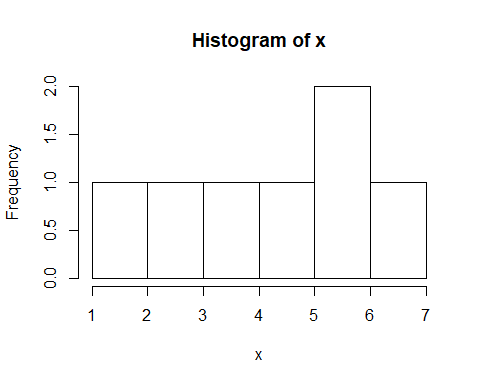
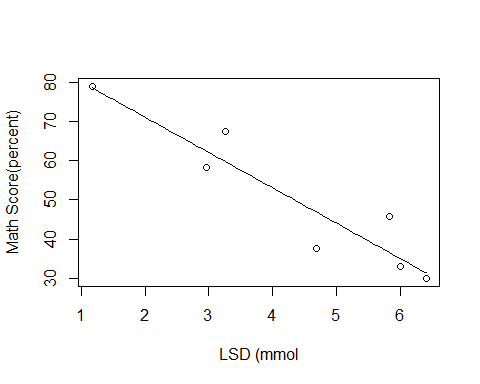
title: “HW 1” author: “Louis Jochems” date: “January 23, 2019” output: word\_document

```

#Question 1  
math<-read.csv("math\_scores.csv")  
x=math$LSD\_concentration  
hist(x)

###b)  
modelq1=lm(math$MATH\_score~math$LSD\_concentration)  
coef(modelq1)

## (Intercept) math$LSD\_concentration   
## 89.123874 -9.009466

confint(modelq1)

## 2.5 % 97.5 %  
## (Intercept) 71.00758 107.240169  
## math$LSD\_concentration -12.87325 -5.145685

# slope and intercept  
a<-89.123874  
b<- -9.009466  
# 95 CI intervals  
#2.5 <- -12.87325  
#97.5 <- -5.415685  
  
###c) Metrics of model fit  
#Rsquared   
y=math$MATH\_score  
y\_hat=a+b\*math$LSD\_concentration  
  
r2<-function(y\_hat,y) {  
 RSS<-sum((((y\_hat))-(y))^2)  
 TSS<-sum(((y)-(mean(y)))^2)  
 return(1-RSS/TSS)}  
  
  
#RMSE   
rmse=function(y\_hat,y) {return(sqrt(mean((y-y\_hat)^2)))}  
  
#return R2  
r2(y\_hat,y)

## [1] 0.877835

0.877835

## [1] 0.877835

#return RMSE  
rmse(y\_hat,y)

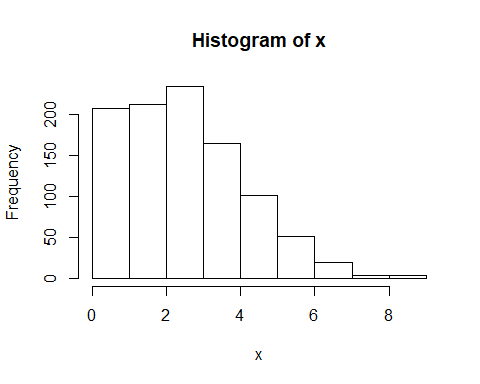
## [1] 6.022355

6.022355

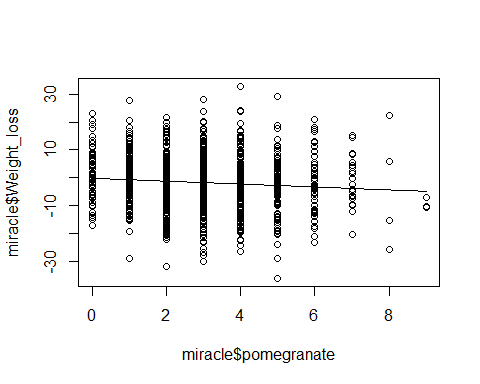
## [1] 6.022355

##A) 85 = 89.123874+ -9.009466\*x  
# -4.123874 = -9.009466\*x  
# x= 0.458   
  
## Based on the parameter estimates of this regression model, there needs to be a max level of 0.458 ug/kg tissue of LSD to ensure a score of >85%.   
  
## B) Based on these data, his model seems show that LSD concentration predicts Math Score fairly well. There is a high R^2 and then sig p value, and the range of the confidence intervals is large and far from zero. However, this indicates a low sample size (n=7).  
   
## C) A normal distribution might be inappropriate to model these data because of a small sample size and these data are continuous, but will only fall between 0 and 100.

#Question 2   
miracle <- read.csv("miracle\_food.csv")  
x=miracle$pomegranate  
hist(x)



##a)  
plot(miracle$Weight\_loss~miracle$pomegranate); curve(-0.1789802+ -0.5251053\*x, add=T)



##b) parameter estimates  
modelq2=lm(miracle$Weight\_loss~miracle$pomegranate)  
coef(modelq2)

## (Intercept) miracle$pomegranate   
## -0.1789802 -0.5251053

confint(modelq2)

## 2.5 % 97.5 %  
## (Intercept) -1.408937 1.0509767  
## miracle$pomegranate -0.886420 -0.1637906

#slope and intercept   
a <- -0.178902  
b <- -0.5251053  
  
  
##c) metrics of fit  
y=miracle$Weight\_loss  
y\_hat=a+b\*miracle$pomegranate  
  
  
r2<-function(y\_hat,y) {  
 RSS<-sum((((y\_hat))-(y))^2)  
 TSS<-sum(((y)-(mean(y)))^2)  
 return(1-RSS/TSS)}  
  
  
#RMSE   
rmse=function(y\_hat,y) {return(sqrt(mean((y-y\_hat)^2)))}  
  
#return R2  
r2(y\_hat,y)

## [1] 0.008083812

0.008083812

## [1] 0.008083812

#return RMSE  
rmse(y\_hat,y)

## [1] 9.961044

9.961044

## [1] 9.961044

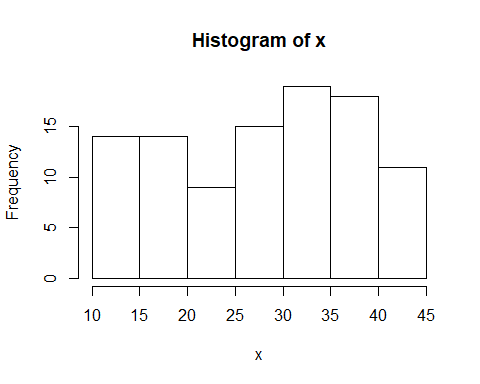
### Answer: Disagree. R squared is low (and plus a poor predictor of fit anyways) indicating that there is little association between the two variables, there is high variance (both explained and unexplained) in the data, residual standard error is high, and CI intervals are only slightly negative.

##A)  
#math  
a<-89.123874  
b<- -9.009466  
y=math$MATH\_score  
y\_hat=a+b\*math$LSD\_concentration  
n\_math=nrow(math)  
mae=function(y\_hat,y) {  
 ABS<-(sum(abs((y)-(y\_hat))))  
 n <- n\_math  
 return(ABS/n)  
 }  
  
#miracle  
a<- -0.1789802  
b<- -0.5251053  
y=miracle$Weight\_loss  
y\_hat=a+b\*miracle$pomegranate  
n\_miracle=nrow(miracle)  
mae=function(y\_hat,y) {  
 ABS<-(sum(abs((y)-(y\_hat))))  
 n <- n\_miracle  
 return(ABS/n)  
}  
  
  
##return mae   
mae(y\_hat, y)

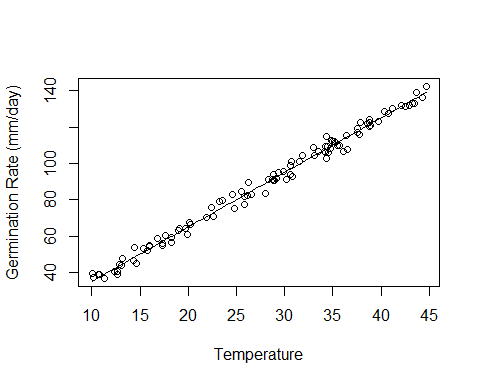
## [1] 7.981461

##B)  
#Metric Math Miracle   
#RMSE 6.022355 9.961044  
#R2 0.877835 0.008083812  
#MAE 4.980145 7.981461  
  
  
##Answer: Mean absolute error for both models shows that it gives about equal weight to all errors compared to RMSE that is higher for both models because of given weight to larger errors, especially with the case of the Miracle model. These metrics (with the exception of the R^2 for the LSD & Math Scores model) indicate that our models are fairly weak with a considerable amount of error.

temperature <- runif(100, min=10,max=45)  
slope <- 3  
intercept <- 5  
sd <- 3  
x=temperature; hist(x)



Germination\_rate <- rnorm(mean=intercept+slope\*temperature, n=100, sd=3)



##B)  
modelq4=lm(Germination\_rate~temperature)  
coef(modelq4)

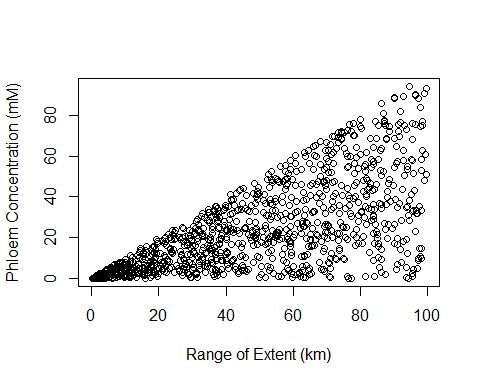
## (Intercept) temperature   
## 5.608691 2.985191

summary(modelq4)

##   
## Call:  
## lm(formula = Germination\_rate ~ temperature)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9262 -1.7046 0.3302 1.7492 6.7490   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.60869 0.85080 6.592 2.2e-09 \*\*\*  
## temperature 2.98519 0.02862 104.310 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.829 on 98 degrees of freedom  
## Multiple R-squared: 0.9911, Adjusted R-squared: 0.991   
## F-statistic: 1.088e+04 on 1 and 98 DF, p-value: < 2.2e-16

## return   
a<- 4.528822   
b<- 3.021530  
  
## C)   
#Answer: The slope is similar to the slope I set and the intercept is about 0.5 off of my parameter estimate, likely due to the considerable Residual Std. Error of about 3.3 (I set a sigma of 3).

##A)  
predictor <- runif(1000,0,100)  
response <- runif(1000,0, predictor)



##B)A particular biologica phenomena that could explain these data is the increasing range extent of a plant species results in the increase in variance of a particular trait, such as sap content of a tree species. Among a population within a given area, the variation of sap content might have considerably less variation as a selective pressure may particular narrow range of sap content phenotypes among those individuals. When you increase the range of the species in question (especially if it's a generalist species) then you may observe more variation in sap content due to more likely variation in microhabitats, or selective pressures, that could result in a high variation of sap content for this species as a whole. There tend to be many spatial and temporal (autocorrelation) examples wherein heteroscedasticity increases as a function of increase of the predictor variable.

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.