The Miller Lumber Company is a large retailer of lumber and paint, as well as of plumbing, electrical, and other household supplies. During a two-week period, in-store surveys were conducted and addresses of customers were obtained. The addresses were then used to identify the metropolitan area census tracts in which the customers reside. At the end of the survey period, the total number of customers who visited the store from each census tract within a 10-mile radius was determined and relevant demographic information for each tract (average income, number of housing units, etc.) was obtained. Other potential predictors of customer counts were constructed from maps, including distance from census tract to nearest competitor and distance to store. Initial screening of the potential predictor variables was conducted which led to the retention of five predictor variables

- X_1: Number of housing units,
- X_2: Average income, in dollars
- X_3: Average housing unit age, in years
- X_4: Distance to nearest competitor, in miles
- X_5: Distance to store, in miles
- Y : Number of customers who visited store from census tract

The data for a portion of the $n = 110$ census tracts are found in file `miller.csv`.

a) Fit a Poisson simple regression model with predictor X_1.

b) Plot the model and the fitted equation.

c) Predict the average number of customers who visit the store from an address with 600 housing units.

d) Find a 95% CI on that average.

e) Fit a multiple Poisson regression model with all predictors in the data set. Compare AIC of the simple and multiple regression models.

```
# miller.r

rm(list=ls())
setwd("C:/Users/USC Guest/Downloads2")
d1 = read.csv("miller.csv",header=T)
head(d1)
names(d1)=c("housing","income","hage","competitor","distance","customers")
d1$income <- d1$income/1000;

# Poisson simple reg model
#===================================================================

m1 = glm(customers ~ housing,d1,family = poisson)
summary(m1)

# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept) 1.9366164  0.0813853  23.796  < 2e-16 ***
#housing     0.0007131  0.0001095   6.514 7.31e-11 ***

#(Dispersion parameter for poisson family taken to be 1)
#    Null deviance: 422.22  on 109  degrees of freedom
#Residual deviance: 379.56  on 108  degrees of freedom
#AIC: 827.6

# based on the Deviance, the fit of this model is not so good.

# plot lambda vs housing
plot(customers~housing,d1)
lines(sort(d1$housing), sort(fitted(m1)))
grid()


# predict customers when housing = 600
newval <- data.frame(housing = 600)
yhat <- predict(m1,newval,type="response")
yhat
# 10.6384

# CI for lambda when housing = 600

alpha <- 0.05
yhat <- predict(m1, newval, se.fit=T, type="link")
lower95 = exp(yhat$fit - qnorm(1-alpha/2)*yhat$se.fit)
lower95 = round(lower95,3)
upper95 = exp(yhat$fit + qnorm(1-alpha/2)*yhat$se.fit)
upper95 = round(upper95,3)
cat("(", lower95,",",upper95,")\n")
# ( 10.023 , 11.291 )
```

```
# Poisson multiple regression
#===================================================================

m2 <- glm(customers ~ .,d1,family=poisson)
summary(m2)
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
#(Intercept)  2.942e+00  2.072e-01  14.198  < 2e-16 ***
#housing      6.058e-04  1.421e-04   4.262 2.02e-05 ***
#income      -1.169e+01  2.112e+00  -5.534 3.13e-08 ***
#hage        -3.727e-03  1.782e-03  -2.091   0.0365 *
#competitor   1.684e-01  2.577e-02   6.534 6.39e-11 ***
#distance    -1.288e-01  1.620e-02  -7.948 1.89e-15 ***

#(Dispersion parameter for poisson family taken to be 1)
#    Null deviance: 422.22  on 109  degrees of freedom
#Residual deviance: 114.99  on 104  degrees of freedom
#AIC: 571.02

# plot fitted vs response

yhat <- predict(m2,d1,type="response")
d2 = data.frame(d1,yhat)
aux = c(0,35)
plot(yhat~customers,d2,xlim=aux,ylim=aux)
abline(0,1,lty=2,col="red")
grid()

# This model shows smaller Residual deviance and AIC value.
# the fit of this model is MUCH better.
```

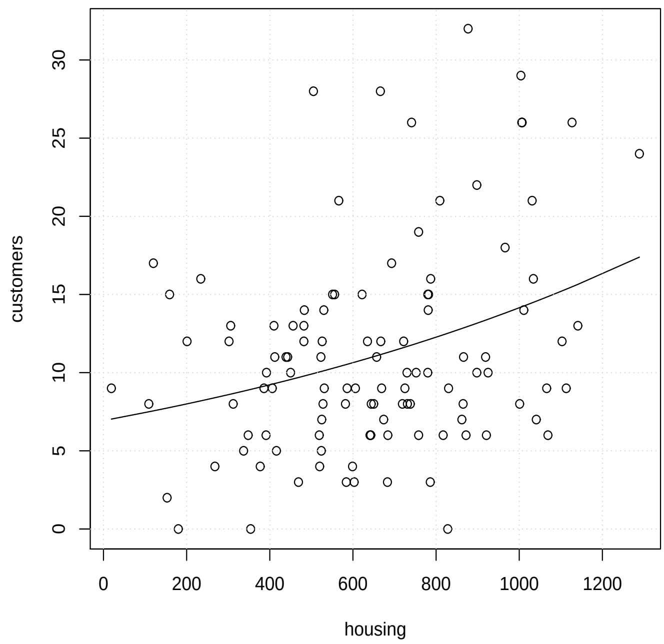

Figure 1: Fitted equation for the Poisson simple regression model

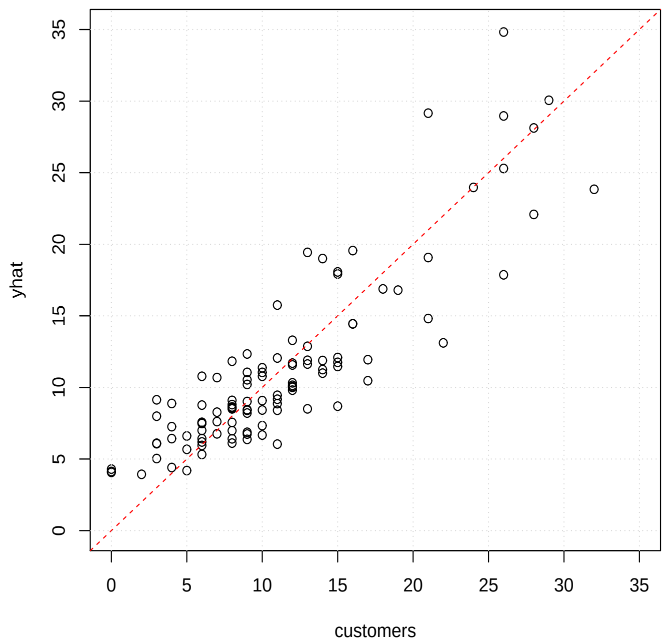

Figure 2: Fitted values vs response - Multiple Poisson regression model