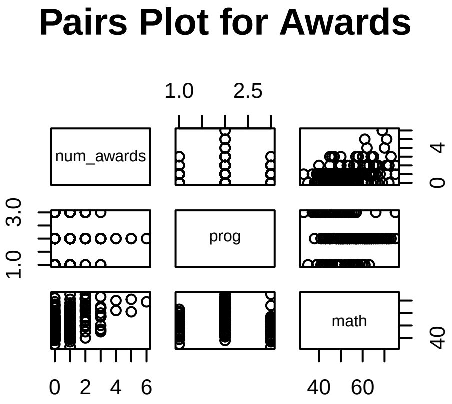
# Assignment 2 - Exercise 3

## Exercise 3. School awards

### a)

```r
#Read in the data and extract the relevant column
awards=read.table("data/awards.txt",header=TRUE)
pairs(awards, main="Pairs Plot for Awards")
```



```r
#Simple Poisson regression without math
awardsglm = glm(num_awards~prog, family=poisson, data=awards)
summary(awardsglm)
```

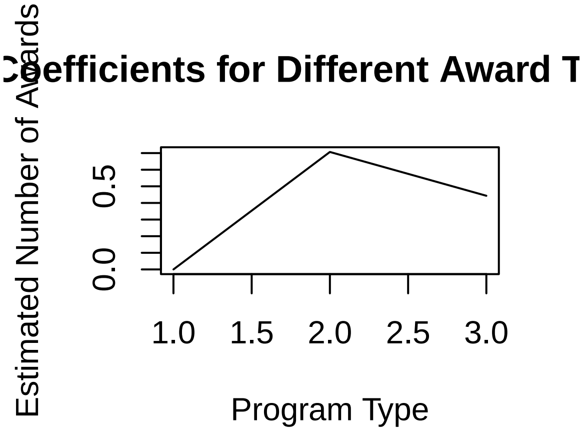
```
##
## Call:
## glm(formula = num_awards ~ prog, family = poisson, data = awards)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.3485     0.2311  -1.508    0.131
## prog          0.1543     0.1047   1.474    0.141
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 228.83  on 199  degrees of freedom
## Residual deviance: 226.65  on 198  degrees of freedom
## AIC: 520.97
##
## Number of Fisher Scoring iterations: 5
```

```r
#Poisson regression with program as a factor
awards$prog=factor(awards$prog)
awardsglm2 = glm(num_awards~prog, family=poisson, data=awards)
summary(awardsglm2)
```

```
##
## Call:
```

```
## glm(formula = num_awards ~ prog, family = poisson, data = awards)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.5486     0.1961  -2.797  0.00515 **
## prog2         0.7068     0.2158   3.275  0.00106 **
## prog3         0.4432     0.2463   1.799  0.07199 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 228.83  on 199  degrees of freedom
## Residual deviance: 216.10  on 197  degrees of freedom
## AIC: 512.42
##
## Number of Fisher Scoring iterations: 5
```

```
plot(c(0,coef(awardsglm2)[2:3]),type="l",xlab="Program Type",ylab="Estimated Number of Awards",main="Co
```



```
program1 = data.frame(prog="1")
program2 = data.frame(prog="2")
program3 = data.frame(prog="3")
num_awards_total = sum(awards$num_awards)
estimated_num_awards_p1 = predict(awardsglm2,program1,type="response")*num_awards_total
estimated_num_awards_p2 = predict(awardsglm2,program2,type="response")*num_awards_total
estimated_num_awards_p3 = predict(awardsglm2,program3,type="response")*num_awards_total
```

The expected number of awards for programs 1, 2, and 3 are 112.0888891, 227.2571429, and 174.6 respectively.

```
program1_data = awards$num_awards[awards$prog=="1"]
program2_data = awards$num_awards[awards$prog=="2"]
program3_data = awards$num_awards[awards$prog=="3"]
ks.test(program1_data, program2_data, program3_data)
```

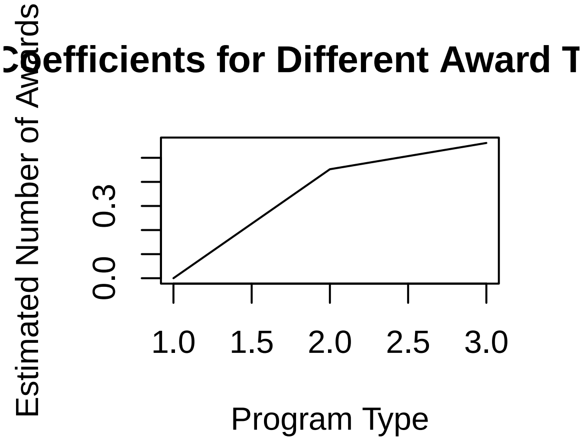
```
## Warning in ks.test.default(program1_data, program2_data, program3_data):
## Parameter(s) ignored
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  program1_data and program2_data
## D = 0.24127, p-value = 0.007523
## alternative hypothesis: two-sided
```

```
#awards$math=factor(awards$math)
awardsglm3 = glm(num_awards~prog+math, family=poisson, data=awards)
summary(awardsglm3)
```

```
##
## Call:
## glm(formula = num_awards ~ prog + math, family = poisson, data = awards)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -2.372577   0.475525  -4.989 6.06e-07 ***
## prog2        0.452621   0.224746   2.014   0.0440 *  
## prog3        0.561720   0.247482   2.270   0.0232 *  
## math         0.035779   0.008344   4.288 1.80e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 228.83  on 199  degrees of freedom
## Residual deviance: 198.05  on 196  degrees of freedom
## AIC: 496.36
##
## Number of Fisher Scoring iterations: 5
```

```
plot(c(0,coef(awardsglm3)[2:3]),type="l",xlab="Program Type",ylab="Estimated Number of Awards",main="Co
```



```
program1 = data.frame(prog="1", math=56)
program2 = data.frame(prog="2", math=56)
program3 = data.frame(prog="3", math=56)
num_awards_total = sum(awards$num_awards)
estimated_num_awards_p1 = predict(awardsglm3,program1,type="response")*num_awards_total
estimated_num_awards_p2 = predict(awardsglm3,program2,type="response")*num_awards_total
estimated_num_awards_p3 = predict(awardsglm3,program3,type="response")*num_awards_total
```

The expected number of awards for programs 1, 2, and 3 are 134.1454372, 210.9340399, and 235.2489717 respectively.