Analysis of Variance for

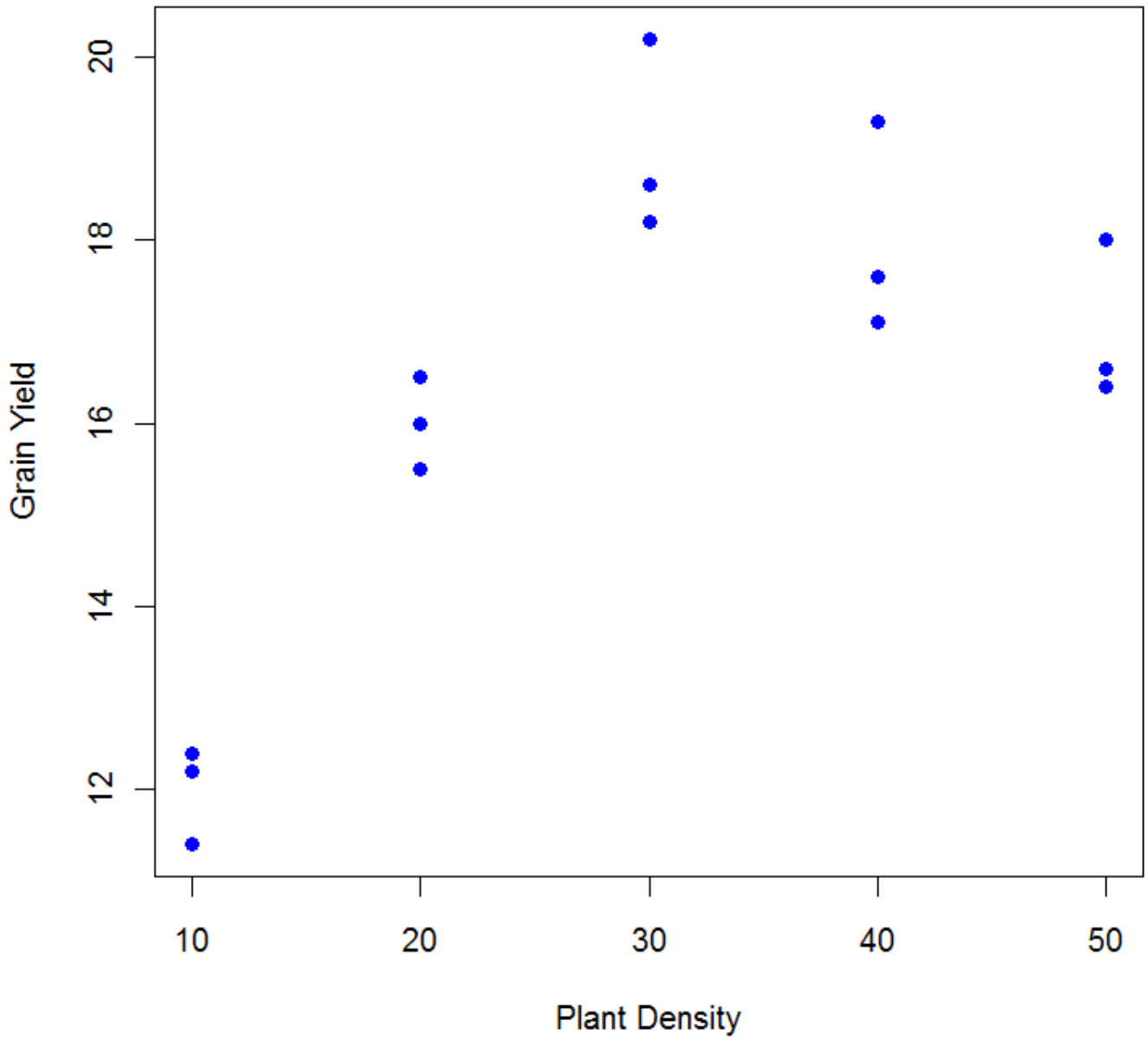
Orthogonal Polynomial Contrasts

```
#An example from "Design of Experiments: Statistical
#Principles of Research Design and Analysis"
#2nd Edition by Robert O. Kuehl

d=read.delim("http://www.public.iastate.edu/~dnett/S511/PlantDensity.txt")

d
   PlantDensity GrainYield
1            10       12.2
2            10       11.4
3            10       12.4
4            20       16.0
5            20       15.5
6            20       16.5
7            30       18.6
8            30       20.2
9            30       18.2
10           40       17.6
11           40       19.3
12           40       17.1
13           50       18.0
14           50       16.4
15           50       16.6
```

```
head(d)
   PlantDensity GrainYield
1            10       12.2
2            10       11.4
3            10       12.4
4            20       16.0
5            20       15.5
6            20       16.5
```

```
tail(d)
    PlantDensity GrainYield
10            40       17.6
11            40       19.3
12            40       17.1
13            50       18.0
14            50       16.4
15            50       16.6
```

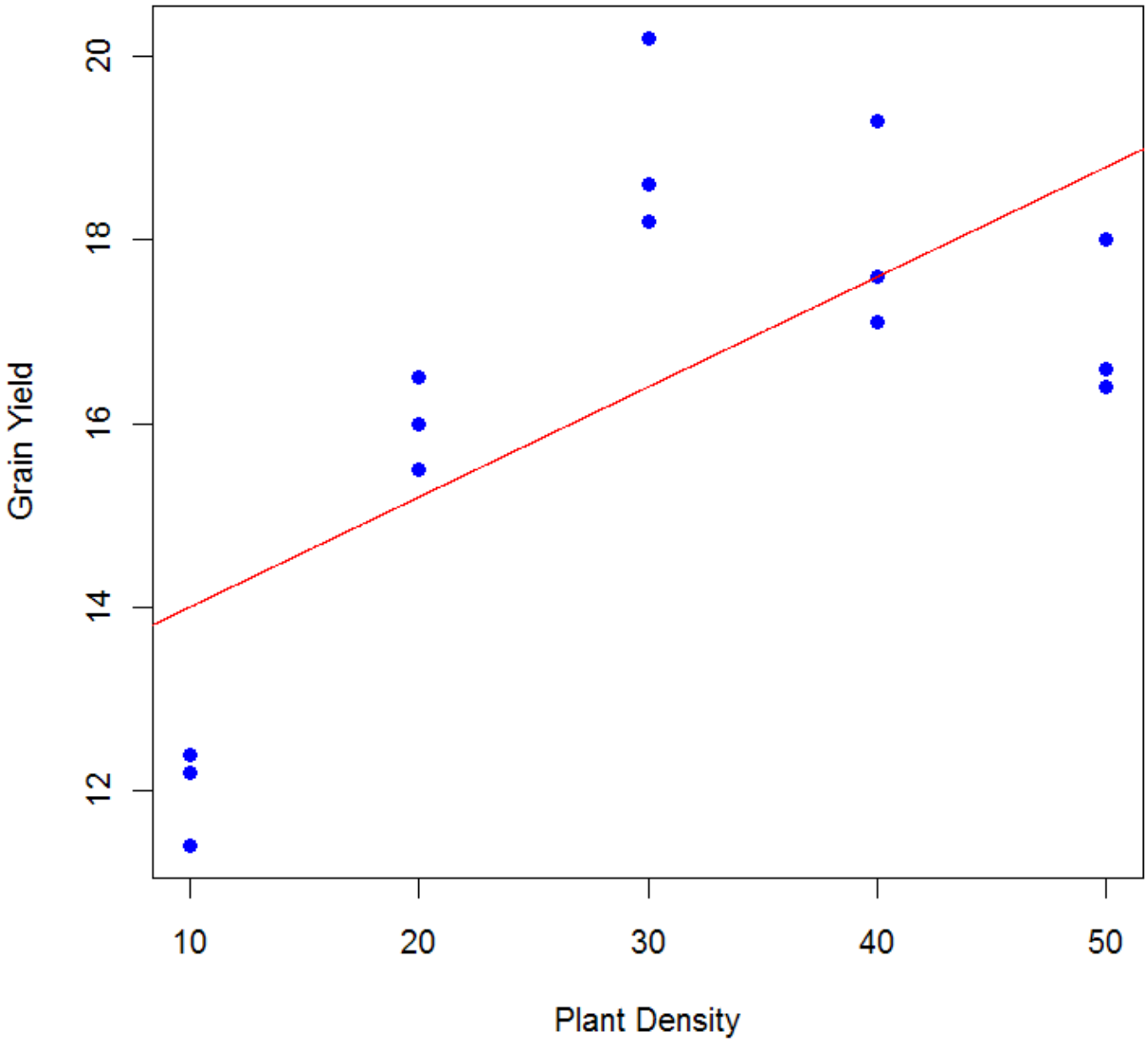
```
names(d)=c("x","y")
head(d)
   x    y
1 10 12.2
2 10 11.4
3 10 12.4
4 20 16.0
5 20 15.5
6 20 16.5

plot(d[,1],d[,2],col=4,pch=16,xlab="Plant Density",
                              ylab="Grain Yield")
```


Grain Yield
12
14
16
18
20
10
20
30
40
50
Plant Density

```
o=lm(y~x,data=d)

model.matrix(o)
   (Intercept)  x
1            1 10
2            1 10
3            1 10
4            1 20
5            1 20
6            1 20
7            1 30
8            1 30
9            1 30
10           1 40
11           1 40
12           1 40
13           1 50
14           1 50
15           1 50
```

```
summary(o)

Call:
lm(formula = y ~ x, data = d)

Residuals:
       Min          1Q      Median          3Q         Max
-2.600e+00 -1.700e+00  1.382e-15  1.500e+00  3.800e+00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.80000    1.20966   10.58  9.3e-08 ***
x            0.12000    0.03647    3.29  0.00586 **

Residual standard error: 1.998 on 13 degrees of freedom
Multiple R-squared: 0.4544,     Adjusted R-squared: 0.4124
F-statistic: 10.82 on 1 and 13 DF,  p-value: 0.005858
```

```
anova(o)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x          1  43.20  43.200  10.825 0.005858 **
Residuals 13  51.88   3.991

#Let's add the best fitting simple linear regression
#line to our plot.

u=seq(0,60,by=.01) #overkill here but used later.
lines(u,coef(o)[1]+coef(o)[2]*u,col=2)
```


Grain Yield
12
14
16
18
20
10
20
30
40
50
Plant Density

```
#The fit doesn't look very good.
#Let's formally test for lack of fit.

o2=lm(y~x+as.factor(x),data=d)

#Could have used 'update' to obtain the same fit.
#o2=update(o,.~.+as.factor(x))
```

```
model.matrix(o2)
   (Intercept)  x as.factor(x)20 as.factor(x)30 as.factor(x)40 as.factor(x)50
1            1 10               0               0               0               0
2            1 10               0               0               0               0
3            1 10               0               0               0               0
4            1 20               1               0               0               0
5            1 20               1               0               0               0
6            1 20               1               0               0               0
7            1 30               0               1               0               0
8            1 30               0               1               0               0
9            1 30               0               1               0               0
10           1 40               0               0               1               0
11           1 40               0               0               1               0
12           1 40               0               0               1               0
13           1 50               0               0               0               1
14           1 50               0               0               0               1
15           1 50               0               0               0               1
```

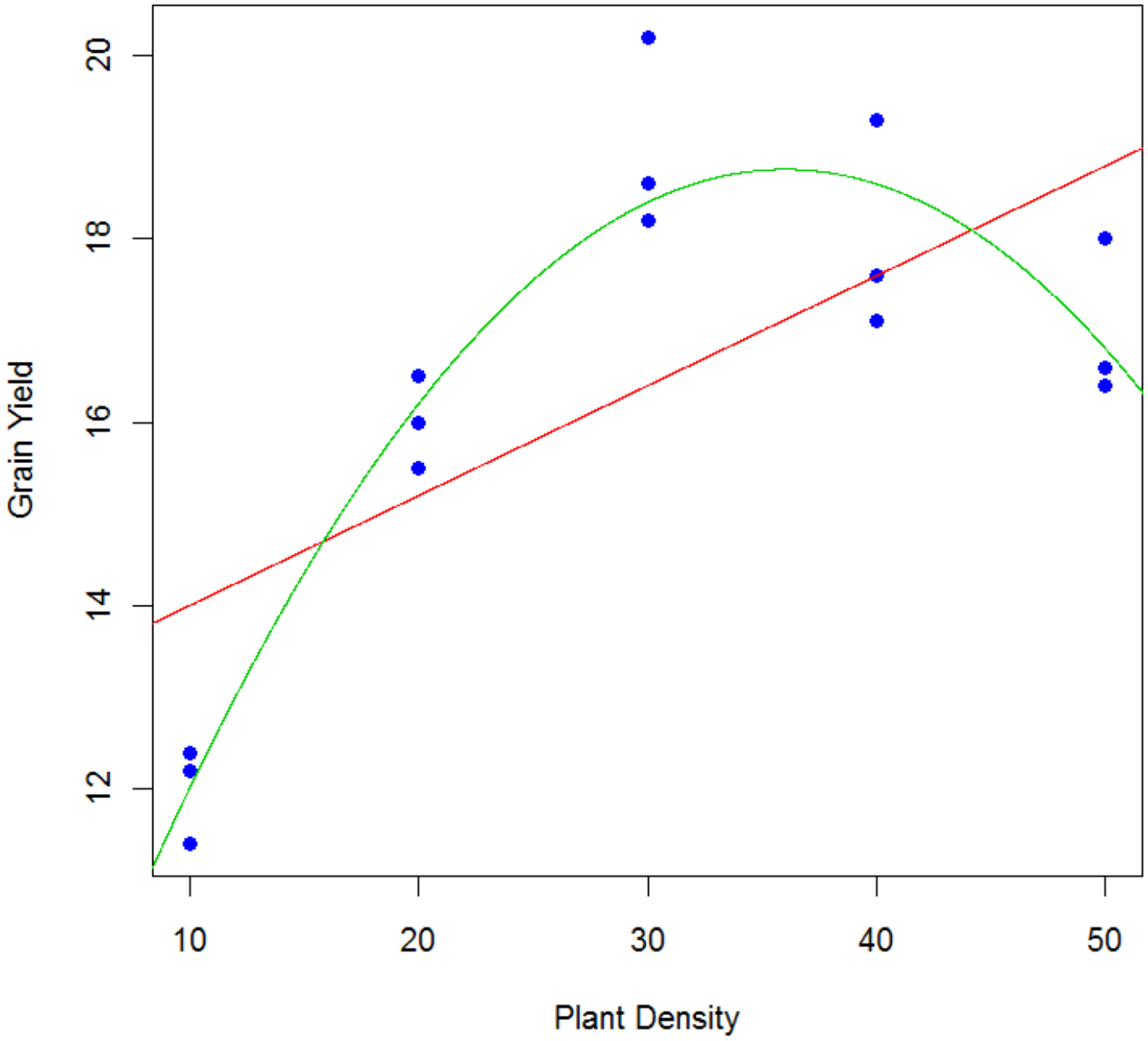
```
summary(o2)

Call:
lm(formula = y ~ x + as.factor(x), data = d)

Residuals:
   Min     1Q Median     3Q    Max
 -0.90  -0.55  -0.40   0.45   1.30

Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.75000    0.63653  16.889 1.11e-08 ***
x               0.12500    0.01765   7.081 3.37e-05 ***
as.factor(x)20  2.75000    0.63653   4.320  0.00151 **
as.factor(x)30  4.50000    0.61156   7.358 2.43e-05 ***
as.factor(x)40  2.25000    0.63653   3.535  0.00540 **
as.factor(x)50       NA         NA      NA       NA

Residual standard error: 0.8649 on 10 degrees of freedom
Multiple R-squared: 0.9213,     Adjusted R-squared: 0.8899
F-statistic: 29.28 on 4 and 10 DF,  p-value: 1.690e-05
```

```
anova(o2)
Analysis of Variance Table

Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)
x             1  43.20  43.200  57.754 1.841e-05 ***
as.factor(x)  3  44.40  14.800  19.786 0.0001582 ***
Residuals    10   7.48   0.748
```

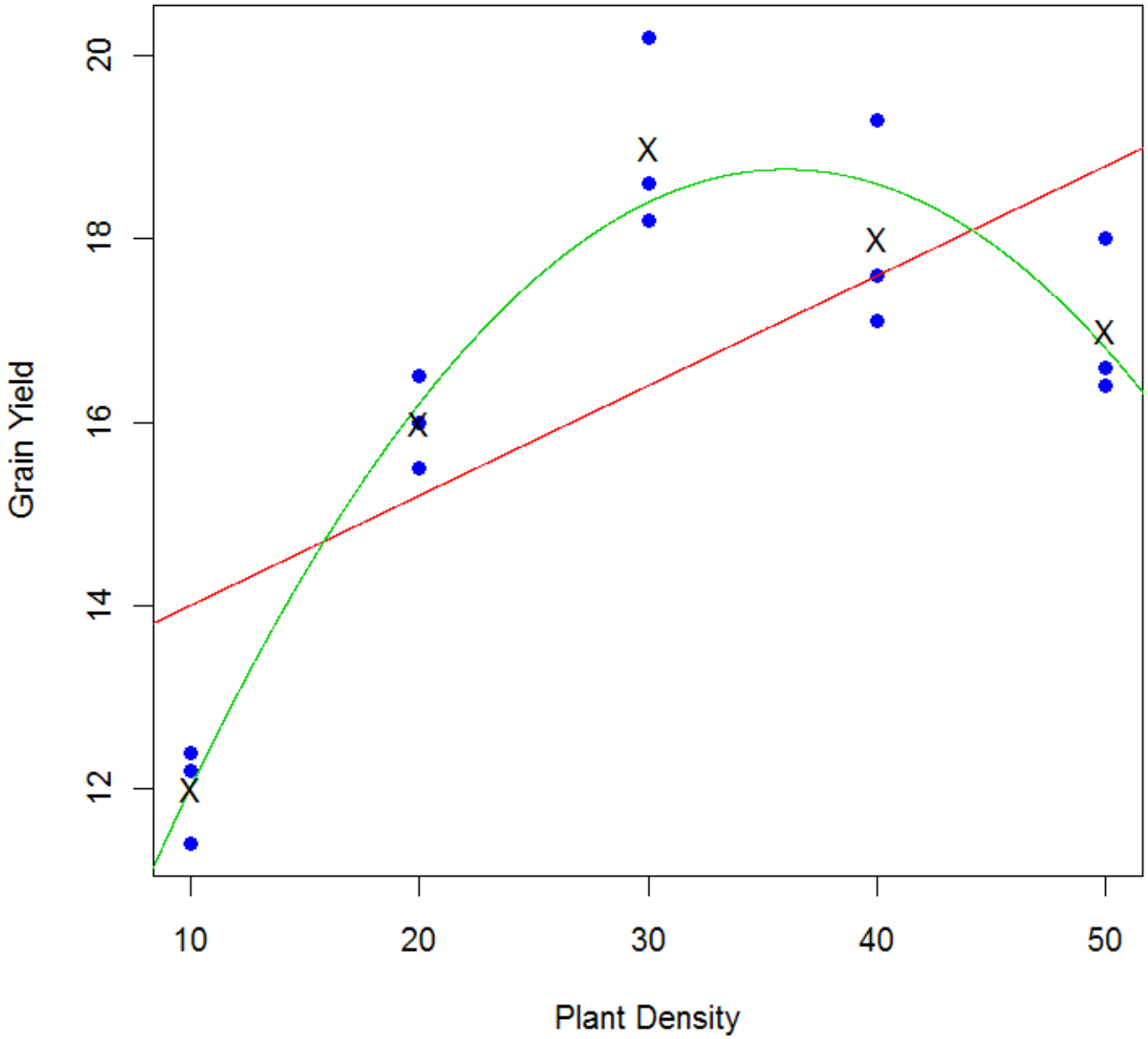
```
#It looks like a linear fit is inadequate.
#Let's try a quadratic fit.

o3=lm(y~x+I(x^2)+as.factor(x),data=d)

model.matrix(o3)
   (Intercept)  x I(x^2) as.factor(x)20 as.factor(x)30 as.factor(x)40  50
1            1 10    100               0               0               0   0
2            1 10    100               0               0               0   0
3            1 10    100               0               0               0   0
4            1 20    400               1               0               0   0
5            1 20    400               1               0               0   0
6            1 20    400               1               0               0   0
7            1 30    900               0               1               0   0
8            1 30    900               0               1               0   0
9            1 30    900               0               1               0   0
10           1 40   1600               0               0               1   0
11           1 40   1600               0               0               1   0
12           1 40   1600               0               0               1   0
13           1 50   2500               0               0               0   1
14           1 50   2500               0               0               0   1
15           1 50   2500               0               0               0   1
```

```
summary(o3)

Call:
lm(formula = y ~ x + I(x^2) + as.factor(x), data = d)

Residuals:
   Min     1Q Median     3Q    Max
 -0.90  -0.55  -0.40   0.45   1.30

Coefficients: (2 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.000000   1.278483   5.475 0.000271 ***
x               0.575000   0.123579   4.653 0.000904 ***
I(x^2)         -0.007500   0.002122  -3.535 0.005403 **
as.factor(x)20  0.500000   0.789515   0.633 0.540750
as.factor(x)30  1.500000   0.872841   1.719 0.116449
as.factor(x)40        NA         NA      NA       NA
as.factor(x)50        NA         NA      NA       NA

Residual standard error: 0.8649 on 10 degrees of freedom
Multiple R-squared: 0.9213,     Adjusted R-squared: 0.8899
F-statistic: 29.28 on 4 and 10 DF,  p-value: 1.690e-05
```

```
anova(o3)
Analysis of Variance Table

Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)
x             1  43.20  43.200 57.7540 1.841e-05 ***
I(x^2)        1  42.00  42.000 56.1497 2.079e-05 ***
as.factor(x)  2   2.40   1.200  1.6043    0.2487
Residuals    10   7.48   0.748

#It looks like a quadratic fit is adequate.
#Let's estimate the coefficients for the best
#quadratic fit.

b=coef(lm(y~x+I(x^2),data=d))

#Let's add the best fitting quadratic curve
#to our plot.

lines(u,b[1]+b[2]*u+b[3]*u^2,col=3)
```


Grain Yield
12
14
16
18
20
10
20
30
40
50
Plant Density

```
#Let's add the treatment group means to our plot.

trt.means=tapply(d$y,d$x,mean)

points(unique(d$x),trt.means,pch="X")
```

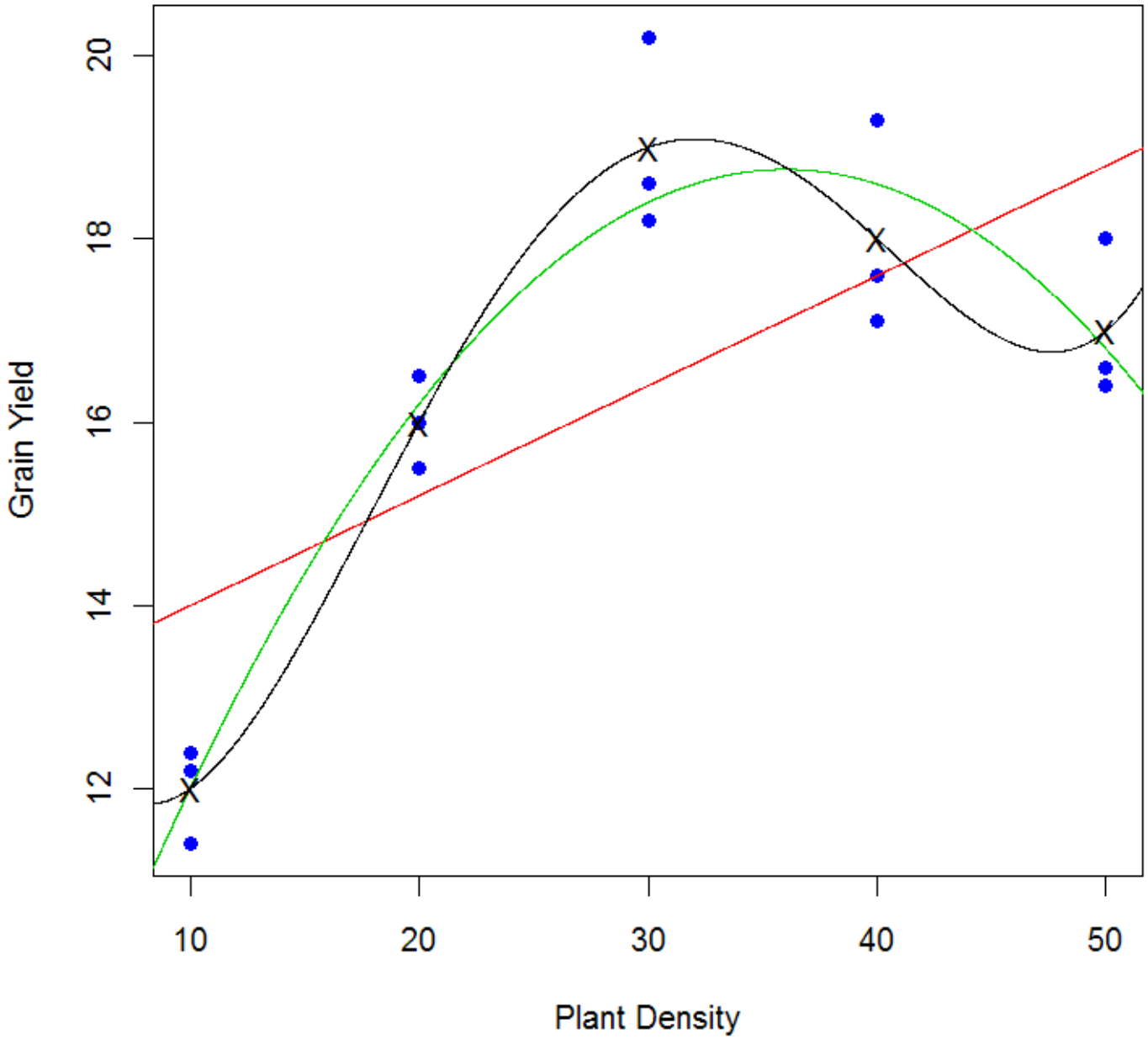

Grain Yield
12
14
16
18
20
10
20
30
40
50
Plant Density

```
#Note that we can break down the 4 degrees of
#freedom for treatment into orthogonal polynomial
#contrasts corresponding to linear, quadratic,
#cubic, and quartic terms.
```

```
o4=lm(y~x+I(x^2)+I(x^3)+I(x^4),data=d)
anova(o4)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1  43.20  43.200 57.7540 1.841e-05 ***
I(x^2)     1  42.00  42.000 56.1497 2.079e-05 ***
I(x^3)     1   0.30   0.300  0.4011    0.5407
I(x^4)     1   2.10   2.100  2.8075    0.1248
Residuals 10   7.48   0.748
```

```
anova(lm(y~as.factor(x),data=d))
Analysis of Variance Table

Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(x)  4  87.60  21.900  29.278 1.690e-05 ***
Residuals    10   7.48   0.748
```

```
#The quartic fit will pass through the treatment
#means.

b=coef(o4)
lines(u,b[1]+b[2]*u+b[3]*u^2+b[4]*u^3+b[5]*u^4,col=1)
```


Grain Yield
12
14
16
18
20
10
20
30
40
50
Plant Density