GRM - WORKSHOP 1: R SESSION

1.1 January flows on the Kootenai river

```
> riverflow <- read.table('Riverflow.txt',head=T)
> plot(riverflow$Flow_at_Libby,riverflow$Flow_at_Newgate)
> title("River flow")
> riverflow.lm <- lm(Flow_at_Newgate ~ Flow_at_Libby, data
= riverflow)
> summary(riverflow.lm)

Call:
lm(formula = Flow_at_Newgate ~ Flow_at_Libby, data = riverf
low)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6796 -1.4361  0.1316  1.0881  3.4132

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     9.5137     1.2239   7.773 8.58e-06 ***
Flow_at_Libby   0.4748     0.0344  13.802 2.73e-08 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.785 on 11 degrees of freedom
Multiple R-squared:  0.9454,  Adjusted R-squared:  0.9404
F-statistic: 190.5 on 1 and 11 DF,  p-value: 2.729e-08

> abline(riverflow.lm)
```
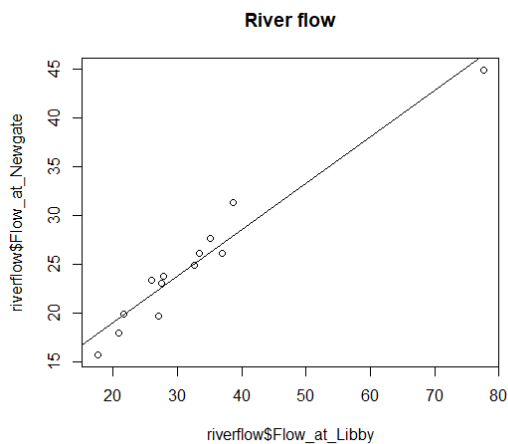

QUIZ 1.1a:
# State an odd looking feature of the data set!
   The point at the top right is an outlier.

State the model being fitted
   E(Flow_at_Newgate) = α + βFlow_at_Libby

```
> # predict missing values at Newgate:
> Flow_at_Libby_with_missing_Newgate <- c(42.0,24.0,38.0,49
.4,24.6,24.2)
> new <- data.frame(Flow_at_Libby= Flow_at_Libby_with_missi
ng_Newgate)
> predict(riverflow.lm,newdata=new,interval = "confidence")
       fit      lwr      upr
1 29.45346 28.14939 30.75753
2 20.90785 19.64082 22.17488
3 27.55444 26.38891 28.71996
4 32.96665 31.28822 34.64509
5 21.19271 19.94824 22.43717
6 21.00280 19.74343 22.26218

> Newgate.ests <- predict(riverflow.lm,newdata=new)
> plot(riverflow$Flow_at_Libby,riverflow$Flow_at_Newgate)
> points(Flow_at_Libby_with_missing_Newgate,Newgate.ests,pc
h=2)
> title("River flow with Newgate estimates")
> abline(riverflow.lm)
```


River flow with Newgate estimates
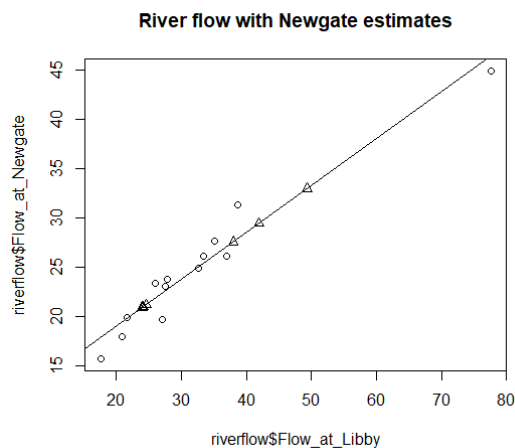
QUIZ 1.1b:
What are the predicted values at Newgate for the missing values in the data set?

From predict we get:

```
1 29.45346
2 20.90785
3 27.55444
4 32.96665
5 21.19271
6 21.00280
```

1.2 Calibration of a flame photometer

```
> flame <- read.table('Flame.txt',head=T)
> plot(flame$Sodium_concentration,flame$Photometer_reading)
> title("Flame photometer")
> flame.lm <- lm(Photometer_reading ~ Sodium_concentration,
data = flame)
> summary(flame.lm)
```

```
Call:
lm(formula = Photometer_reading ~ Sodium_concentration, dat
a = flame)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1718 -0.7786  0.1262  0.5077  1.2556

Coefficients:
                       Estimate Std. Error t value
(Intercept)           -0.900858   0.674560  -1.335
Sodium_concentration   0.415726   0.004791  86.775
                       Pr(>|t|)
(Intercept)              0.224
Sodium_concentration 7.11e-12 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9284 on 7 degrees of freedom
Multiple R-squared:  0.9991,  Adjusted R-squared:  0.9989
F-statistic:  7530 on 1 and 7 DF,  p-value: 7.108e-12

> abline(flame.lm)
```
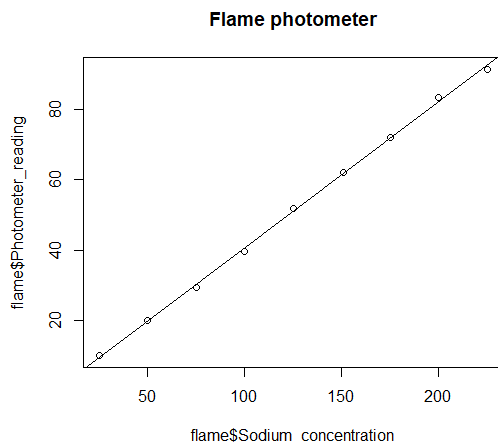

QUIZ 1.2:
State the model being fitted
   E(Photometer_reading) = α + βSodium_concentration

Is the slope parameter significantly different from zero?
   Yes – significance probability is very low (and has three stars)

Is the intercept parameter significantly different from zero?
   No - significance probability is high (0.224) so we can remove the
intercept (no stars!)

1.3 Comparison of two methods for measuring the bitterness of beer

```
> beer <- read.table('Beer.txt',head=T)
> plot(beer$OD_concentration, beer$HPLC_concentration)
> title("Bitterness of beer")
> beer.lm <- lm(HPLC_concentration ~ OD_concentration, data
= beer)
> summary(beer.lm)
```

```
Call:
lm(formula = HPLC_concentration ~ OD_concentration, data =
beer)

Residuals:
    Min      1Q  Median      3Q     Max
-50.021 -11.605  -2.997  12.197  54.456

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -221.07924   12.05987  -18.33   <2e-16
OD_concentration    3.59816    0.03828   93.98   <2e-16

(Intercept)      ***
OD_concentration ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.36 on 28 degrees of freedom
Multiple R-squared:  0.9968,  Adjusted R-squared:  0.9967
F-statistic:  8833 on 1 and 28 DF,  p-value: < 2.2e-16
```
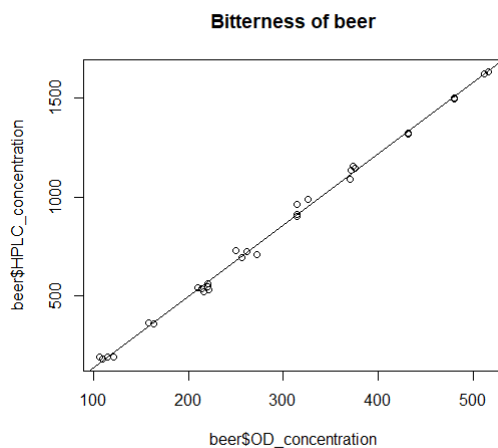
> abline(beer.lm)

Bitterness of beer

QUIZ 1.3:
State the model being fitted
$$E(HPLC\_concentration) = \alpha + \beta OD\_concentration$$

Is the slope parameter significantly different from zero?
Yes – significance probability is very low (and has three stars)

Is the intercept parameter significantly different from zero?
Yes – significance probability is very low (and has three stars)