02441 Applied Statistics and Statistical Software

Exercise 2A - Sport

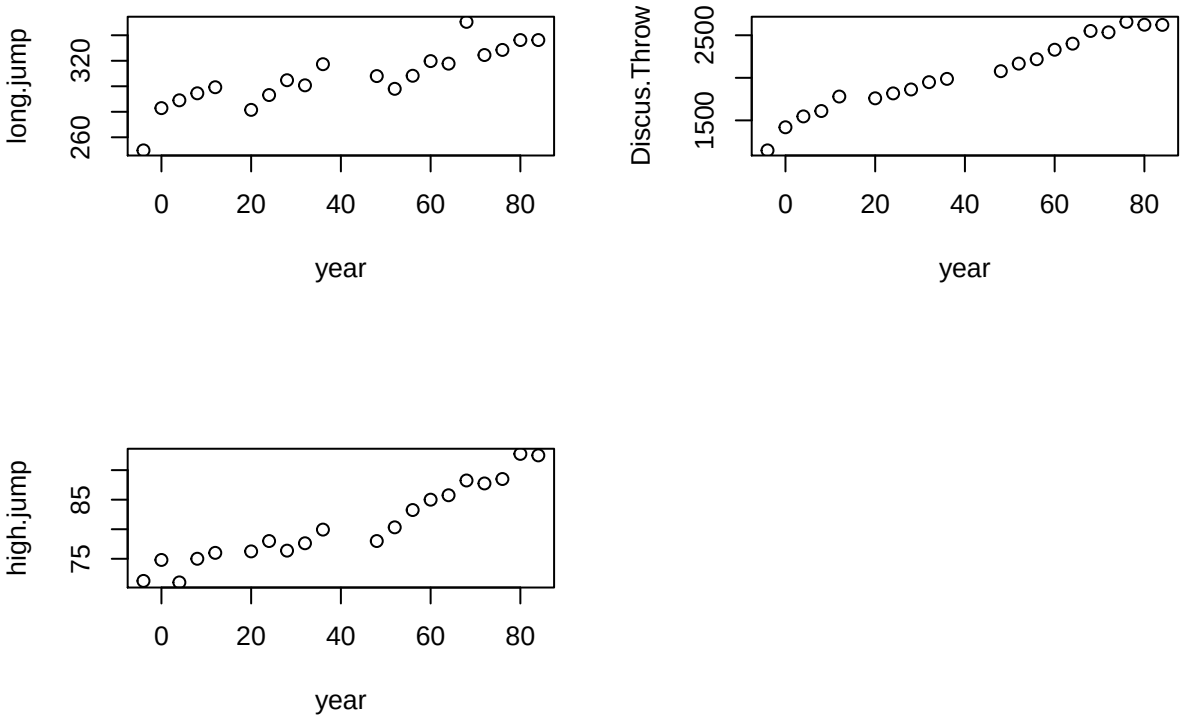
This dataset contains the gold medal performances in the men's long jump, high jump and discus throw for the modern Olympic games from 1900 to 1984. Data are also provided for the 1968 Olympics, but one may wish to omit them from the analyses because that Olympics was quite different from later ones.

Variable name	Description
highjump	winning result in high jum
discusthrow	winning result in discus throw
longjump	winning result in long jump
year	year with reference at 1900

1. Make appropriate plots for the different sport disciplines. Is there a tendency?

Start by loading data

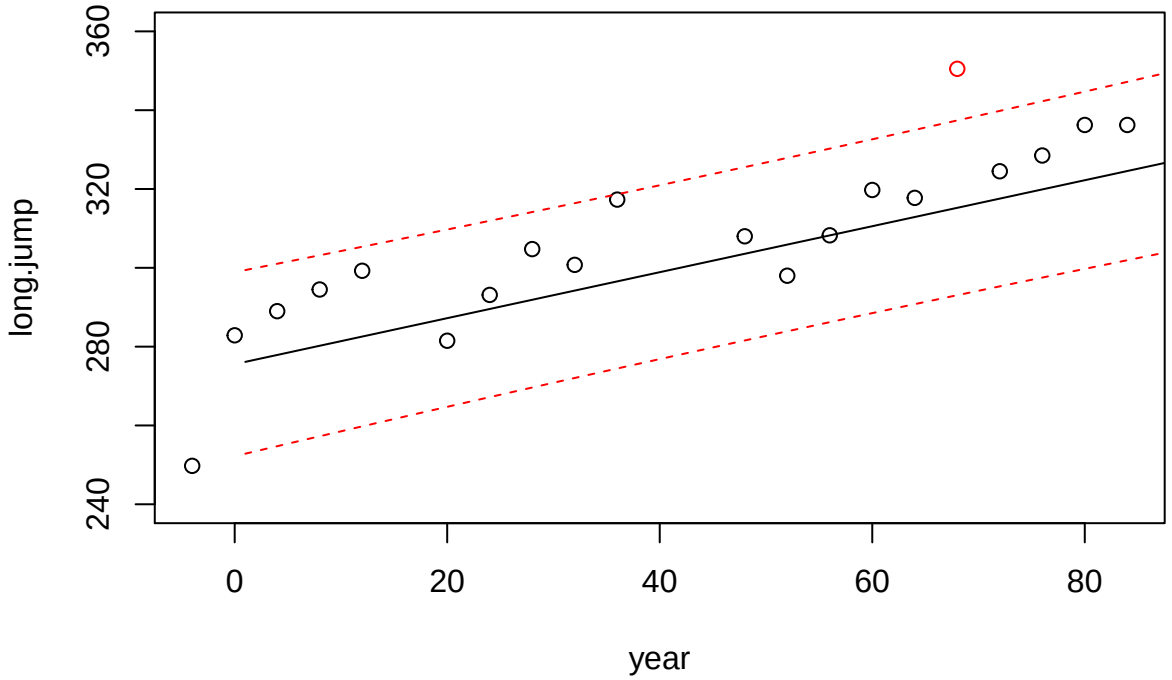
```
sport = read.table('sport.txt', header = TRUE)
par(mfrow=c(2,2))
plot(long.jump~year, data = sport)
plot(Discus.Throw~year, data = sport)
plot(high.jump~year, data = sport)
```



2. It has been suggested that the Mexico City Olympics in 1968 saw unusually good track and field performances, possibly because of the high altitude. To investigate this question we must establish some expected performance level. How would you do this?

Fit linear model without observation for long.jump 1968.

```
lm1 <- lm(long.jump~year, data = sport[sport$year!=68,])
plot(long.jump~year, data = sport[sport$year!=68,], ylim = c(240,360))
x_dat <- data.frame(year = seq(min(sport$year), max(sport$year), length.out = 100))
pred <- predict(lm1, newdata = x_dat, interval = "prediction")
matlines(pred, col = c(1,2,2), lty = c(1,2,2))
points(long.jump~year, data = sport[sport$year==68,], col = 'red')
```



The long jump results for 1968 lie outside the 95% prediction interval. This means that we can conclude that they are significantly different. Remember though that with a 95% prediction interval we expect 5% of all points to lie outside the interval. We therefore have a 5% chance of making a TypeI error.

3. Could you predict the winning result in long jump for the Olympic games in Greece 2004?

```
predict(lm1, newdata = data.frame(year=104), interval = "prediction")
```

```
##        fit      lwr      upr
## 1 347.0246 322.2718 371.7774
```

Don't use your model to extrapolate into far future. Your model might only be valid in the calibrated range. You can also see that the prediction intervals become larger the further you move from $\bar{x}$.