Figur 1: Data from 1980-2010 of average marriage age for men and women in Denmark and the corresponding divorce rate.

1 Marriage Age and Divorce Rate

This exercise deals with a univariate regression analysis of data regarding the average marriage age of both men and women and the divorce percentage in Denmark from 1980-2011. The data comes from www.statistikbanken.dk. It is a public site with data published by Danmarks Statistik, which is *the central authority on Danish statistics*[1].

In Figure 1 the data for the years 1980-2010 are seen. There is one observation for each year. We can see that within the time period the average age of marriage for both men and women is increased, whereas the divorce percentage within the first five years of marriage has decreased. In this exercise we will investigate the (obvious) correlation between the average age of men and women and the divorce percentage.

1.1 Tasks on univariate data set

Task 1   Write up the two models that will describe the relationship between marriage age for both men and women and divorce rate. This should be done in terms of the General Linear Model.

Task 2   In Table 1 the resulting models of the following SAS code are seen. `avgmen`, `avgwomen` and `dp` are the average marriage age of men, average marriage age of women and the divorce percentage respectively.

```
Title 'Model Men';
proc reg data = divorce;
```

---

[1]www.dst.dk

	Model Men	Model Women
θ_1	30.11	27.63
θ_2	-0.60	-0.56
SS_{res}	75.76	75.76
$\hat{\sigma}^2$	0.64	0.65

Tabel 1: Parameters of the two models.

```
model dp = avgmen/r influence;
run;

Title 'Model Women';
proc reg data = divorce;
model dp = avgwomen/r influence;
run;
```

Do these models correspond to the ones you found in the previous task?

Task 3 In 2011 the average marriage age for men was 34.8 and 32.4 for women. The divorce percentage was 8.56. The error of the residual of the marriage age observation are 0.757 and 0.771 respectively. What is the expected value of the divorce rate for 2011? Calculate the confidence and prediction intervals for this new observation for both models.

Task 4 Figure 2 shows the data points, the fitted model and the confidence and prediction intervals of the data set.

What observations would you expect have a high influence in the model estimation? This could be trends within the data or outliers? From the plot, what observations would you expect have high influence on the estimate of the model parameter? (Note: The prediction intervals may be a bit difficult to see on the plot - ask the TA if you are in doubt.)

Task 5 The DFBETAS values of the SAS run is given underneath. Does your expectation from the last task prove correct? (Hint: What interval should the value of DFBETAS lie within?)

```
Obs Intercept avgmen
1    -0.9033    0.8722
2    -0.2075    0.1995
3    -0.0740    0.0709
4    -0.0641    0.0611
5    -0.0824    0.0781
```

6	0.0545	-0.0515
7	0.1110	-0.1042
8	0.0957	-0.0892
9	0.2330	-0.2157
10	0.3600	-0.3301
11	0.0562	-0.0501
12	-0.0380	0.0336
13	0.0167	-0.0138
14	0.0137	-0.0092
15	0.0030	0.0069
16	0.0090	-0.0178
17	0.0301	-0.0429
18	0.0586	-0.0745
19	0.0736	-0.0866
20	0.0367	-0.0425
21	-0.0321	0.0367
22	-0.0800	0.0897
23	-0.1029	0.1145
24	-0.2638	0.2856
25	-0.3005	0.3220
26	-0.1199	0.1281
27	0.1947	-0.2081
28	0.4234	-0.4491
29	0.1863	-0.1953
30	0.0446	-0.0470
31	0.1617	-0.1700

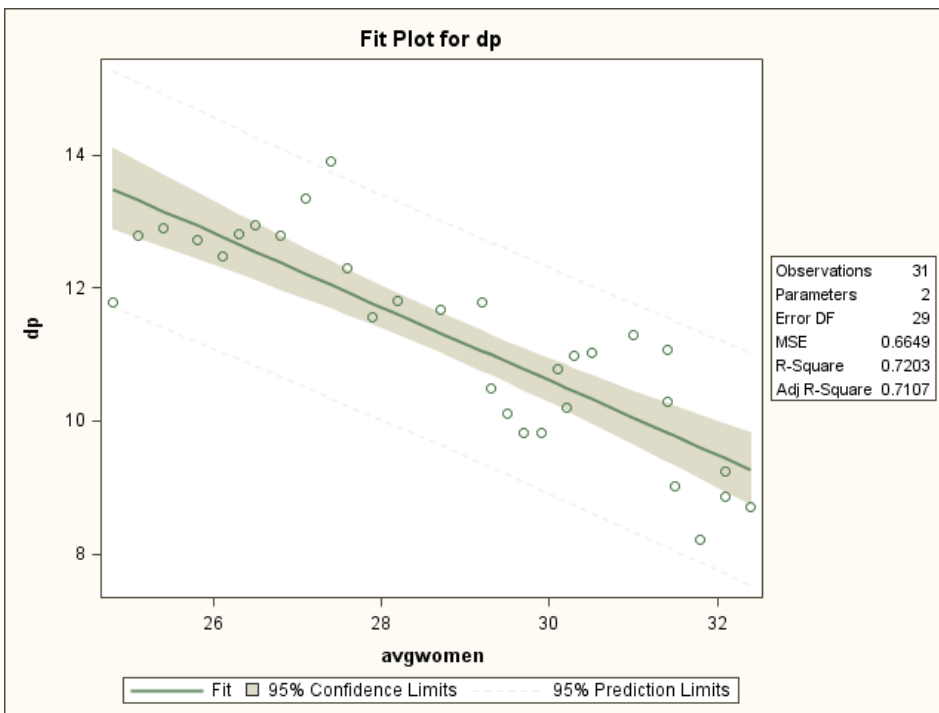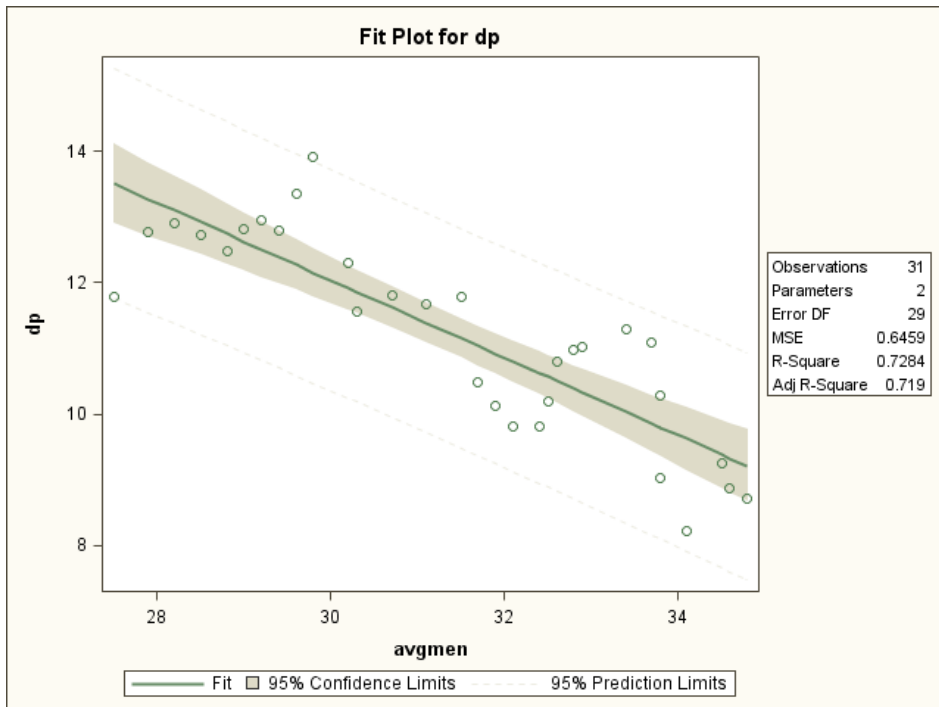**Task 6** Based on the previous tasks; Describe in words the relationship between marriage age and divorce rates.

Figur 2: Fit plots from SAS output.