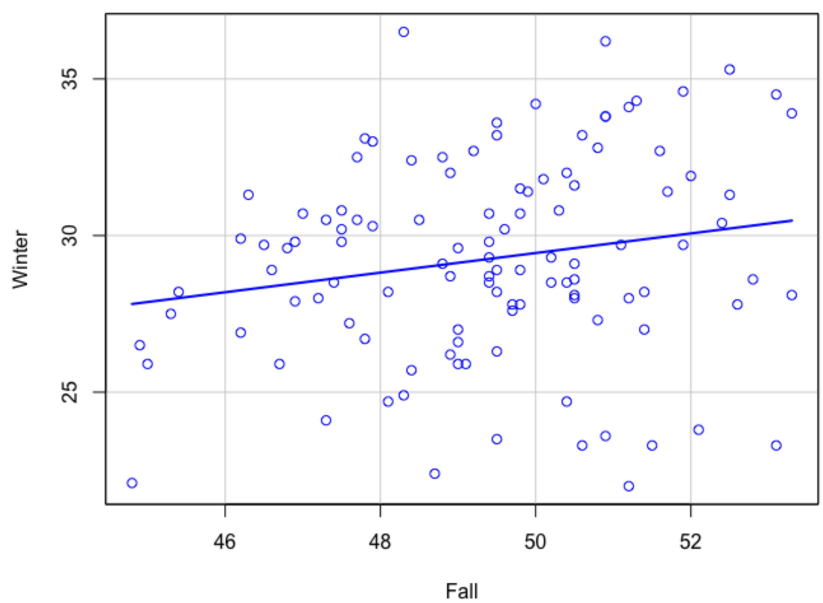
Exercise 1: Do Problem 2.6. Skip 2.6.4
The data file give the mean temperatures in the fall of each year, defined as September 1 to November 30 and the mean temperature in the following winter, defined as December 1 to the end of February in the following calender year. . . These data cover the time period from 1900 to 2010. the question of interest is: Does the average fall temperature predict the average winter temperature.

1. Draw a scatter plot of the response versus the predictor, and describe any pattern you might see.

   Solution:



Code:

```
> df <- ftcollinstemp

> scatterplot(df$fall , df$winter ,
+             regLine = TRUE, boxplots = FALSE,
+             smooth = FALSE, xlab = 'Fall', ylab = 'Winter')
```

It is difficult to infer from a simple scatter plot, the relationship between both variables. There does seem to be a grouping in the center with a slightly positive relationship.

2. Use statistical software to fit the regression of the response on the predictor. Add the fitted line to you graph(done in previous figure). Test the slope to be 0 against a two-sided alternative, and summarize your results.

Solution:
The following, describes a SLR for the response on the predictor.
Code:

```
> Fall_Winter_LM = lm(df$winter ~ df$fall)
> summary(Fall_Winter_LM)

Call:
lm(formula = df$winter ~ df$fall)

Residuals:
    Min      1Q  Median      3Q     Max
-7.8186 -1.7837 -0.0873  2.1300  7.5896

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.7843     7.5549   1.825   0.0708 .
df$fall       0.3132     0.1528   2.049   0.0428 *


Residual standard error: 3.179 on 109 degrees of freedom
Multiple R-squared:  0.0371,    Adjusted R-squared:  0.02826
F-statistic:   4.2 on 1 and 109 DF,  p-value: 0.04284


> 2*pt(2.04973, df=length(df$fall)-2, lower.tail = FALSE)
[1] 0.04279107
```

Stating our hypothesis, If there is a significant linear relationship between the fall temperature and the winter temperature, the slope will not equal zero. Defining our experiment,

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

Computing the test statistic,

$$t = \frac{\hat{\beta}_1}{S.E(\hat{\beta}_1)} = \frac{0.3132}{0.1528} = 2.04973$$

Evaluating the p-value for a t-distribution on 109 df we get $p = 0.04279$. On the $\alpha = .05$ significance level we get that there is a statistically significant relationship between the fall and winter temperatures.

3. Compute or obtain from your computer output, the value of the variability in winter explained by fall and explain what this means.

Solution:
The variability in winter explained by the linear regression on the fall data is the sum of squares for regression(SSreg) and the R^2 value gives us that as a percentage of the total variability,

$$R^2 = \frac{SSreg}{SSreg + RSS} = \frac{SSreg}{SYY} = \frac{42.446}{42.446 + 1101.69} = 0.0371021$$

What this means is our model explains 3.7% of the variability in response variable. We can compute these values using anova(), summary(lm()) or by-hand with r. **Code:**

```
> anova(Fall_Winter_LM)
Analysis of Variance Table

Response: df$winter
            Df  Sum Sq Mean Sq F value  Pr(>F)
df$fall      1   42.45  42.446  4.1995 0.04284 *
Residuals  109 1101.69  10.107
```

Exercise 2: For the data set in problem 2.6, do the following.

1. Calculate a 95% confidence interval for the model intercept.

Solution:
We compute the 95% confidence interval for the model intercept using the following formula,

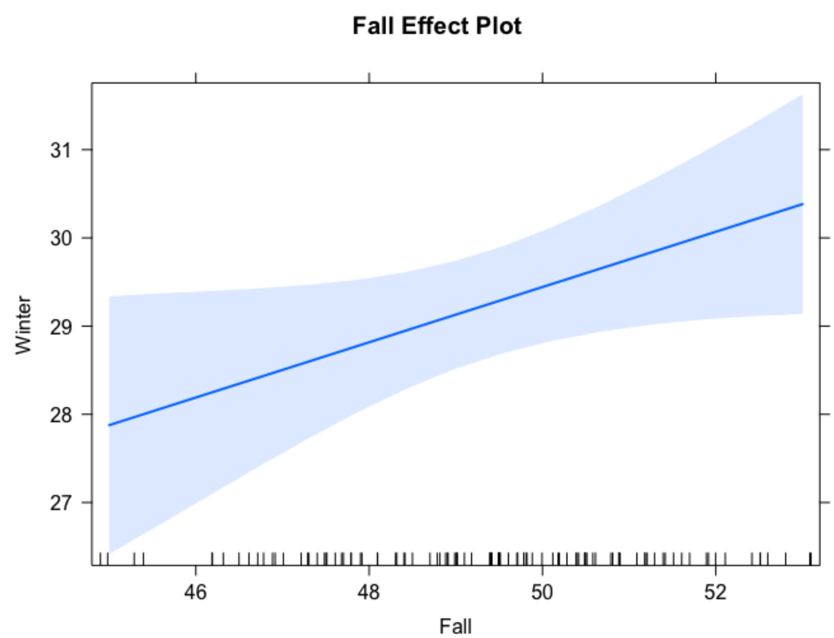
$$\hat{\beta}_0 \pm t_{.05}\hat{\sigma}\left(\frac{1}{n}\frac{\bar{x}^2}{SXX}\right)$$

We can also compute it using r, with the following,
Code:

```
> confint(Fall_Winter_LM, level = .95)
                    2.5 %      97.5 %
(Intercept)  -1.18920049  28.757891
df$fall       0.01028623   0.616052
```

2. Produce an effect plot for the effect of fall on the mean of winter.

Solution:



Code:

```
> plot(Effect("fall",Fall_Winter_LM), main = 'Fall Effect Plot',
                  xlab = 'Fall', ylab = 'Winter')
```

3. What are the values for the smallest and largest residuals.

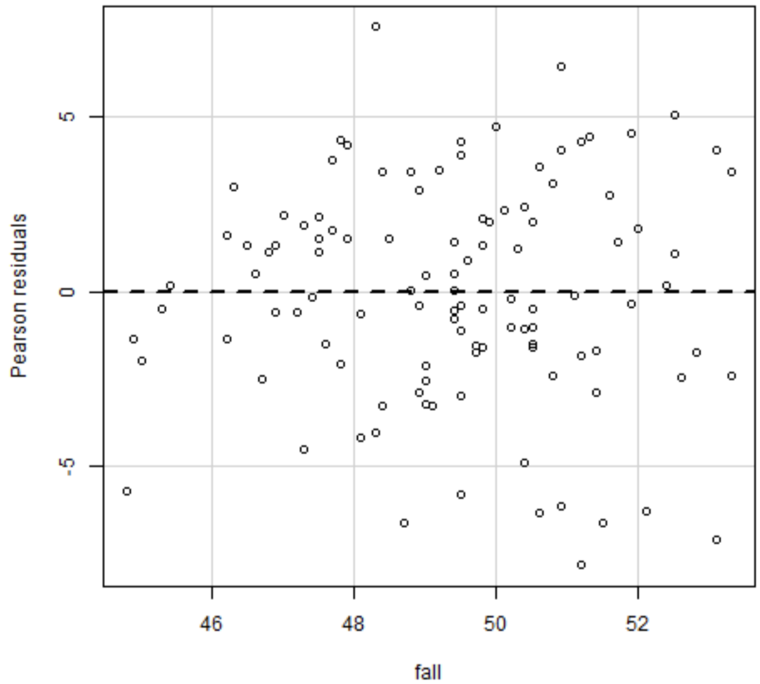
Solution:
Code:

```
> Predicted_values = predict(Fall_Winter_LM, data.frame(fall))

> max(abs(winter - Predicted_values))
[1] 7.818605

> min(abs(winter - Predicted_values))
[1] 0.03300102
```

4. Interpret the residual plot shown below in terms of the appropriateness of the linear model.



Solution:
The residual plot shows no obvious clustering or patterns which would suggest that our model is an appropriate fit.

Exercise 2.13: Heights of mothers and daughters.

1. Compute the regression of *dheight* on *mheight*, and report the estimates, their standard errors, the value of the coefficient of determination, and the estimate of variance. Write a sentence or two that summarizes the results of these computations.

Solution:
Using r, we can fit the regression we get the following,
Code:

```
> df <- Heights
> mheights = df$mheight
> dheights = df$dheight
> LM_Heights = lm(dheights ~ mheights)
> summary(LM_Heights)

Residuals:
    Min      1Q  Median      3Q     Max
-7.397  -1.529   0.036   1.492   9.053

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.91744    1.62247   18.44   <2e-16 ***
mheights     0.54175    0.02596   20.87   <2e-16 ***

Residual standard error: 2.266 on 1373 degrees of freedom
Multiple R-squared:  0.2408,    Adjusted R-squared:  0.2402
F-statistic: 435.5 on 1 and 1373 DF,  p-value: < 2.2e-16
```

Interpreting the linear model, we can say that for every one inch a mother's height increases we can expect her daughter's to increase by approximately .5 inches. The standard errors give us a good idea of were the true regression parameters may be. An R^2 value of .2406 suggests that our model doesn't do an excellent job of explaining all the variance in the data.

2. Obtain a 99% confidence interval for β_1 from the data.

Solution:
Code:

```
> Confint(LM_Heights, level = .99)
               Estimate      0.5 %     99.5 %
mheights       0.541747  0.4747836  0.6087104
```

3. Obtain a prediction and 99% prediction interval for a daughter whose mother is 64 inches tall.

Solution:
Code:

```
> predict(LM_Heights, data.frame(mheights = 64),
          interval = 'prediction', level = .99)

       fit      lwr      upr
1 64.58925 58.74045 70.43805
```

Exercise 4: For the data set in problem 2.13, do the following.

1. Calculate the value of the correlation coefficient r

   Solution:
   Recall that,
   $$r = \frac{SXY}{\sqrt{SXX * SYY}} = \sqrt{R^2}$$
   We can compute each term and use the formula or we can just square root the value given in the summary of the linear model.
   Code:

   ```
   > cor(mheights, dheights)
   [1] 0.4907
   > sqrt(0.2408)
   [1] 0.4907
   ```

2. If the intercept-only model is fit to the data, what is it's estimated intercept $\hat{\beta}_0$?

Solution:
The intercept only model is computed by finding the mean of the response variables. Here is an interesting way of computing it using the lm() command,
Code:

```
> summary(lm(dheights ~ 1))

Call:
lm(formula = dheights ~ 1)

Residuals:
    Min      1Q  Median      3Q     Max
-8.6511 -1.7511 -0.1511  1.8489  9.3489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 63.75105    0.07012   909.2   <2e-16 ***
Residual standard error: 2.6 on 1374 degrees of freedom

> mean(dheights)
[1] 63.75105
```

3. Produce the Anova table for this model. Be sure to include a ”Total” row in your table.

Solution:
Code:

```
> anova(LM_Heights)
Analysis of Variance Table

Response: dheights
            Df Sum Sq Mean Sq F value    Pr(>F)
mheights     1 2236.7 2236.66  435.47 < 2.2e-16 ***
Residuals 1373 7052.0    5.14
Total     1374 9288.7
```

4. If the predictor *mheight* is standardized before fitting the model, the resulting fitted model becomes $E(Y) = 63.75 + .54x$. Interpret the slope and intercept in this model.

Solution:
Since the slope is unchanged from the previous fitting I will assume that the method of standardizing was just centering. When you center your predictor variable, you just subtract the mean from every data. This makes so that the intercept becomes the predicted value for the mean predictor. So for average height women, their daughters are expected to be around 63.75. The slope continues to have the same interpretation as before.