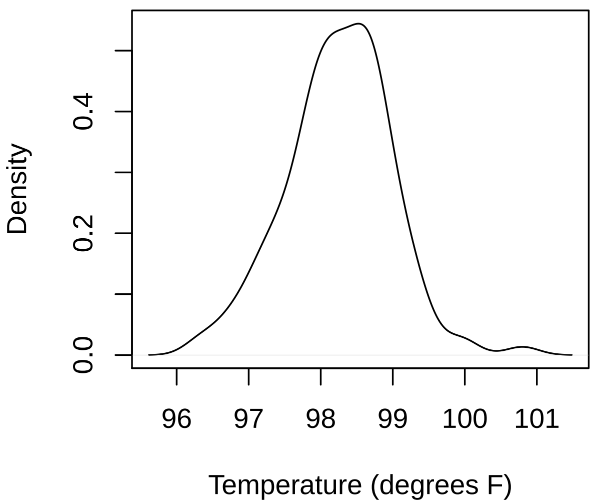
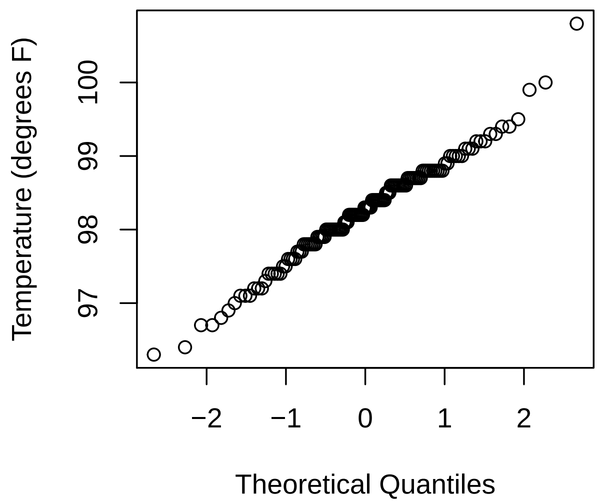
Body Temperature

German physician, Carl Reinhold August Wunderlich (1815-1877), is known for his measurement of body temperature. He established that 98.6 degrees Farenheit is the normal body temperature.
In 1992, Mackowiak, Wasserman, and Levine published an article in JAMA, disputing 98.6 as “normal”.

We examine data from Allen Shoemaker containing 130 measurements of body temperature.

Let’s begin by examining the distribution of body temperature. The density plot of temperature looks roughly normal. And, the normal quantile plot or temperature is close to a straight line. Both of these plots indicate that temperature is normally distributed.





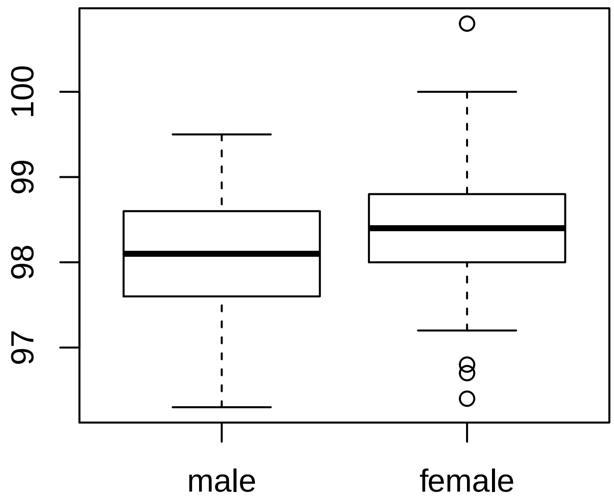
The average temperature is 98.25 and the SD is 0.73. We can carry out a test of the null hypothesis that the true average body temperature is 98.6, out test statistics is -5.43, which follows a t-distribution with 129 degrees of freedom. We would reject the the null hypothesis that 98.6 is the typical body temperature.

The relationship between temperature and gender

Shoemaker also collected information on the gender of the subjects. We ask the question whehter gender can help explain temperature? That is: Is some of the variablility in temperature due to a difference between males and females?

The variable gender is a categorical variable that indicates whether the subject is male or female. In R, these variables can be stored as a factor-type variable. How would we examine the explanatory capacity of gender?

A boxplot is one way to do this. We see that the temperatures for the females appear to be slighlty higher than for males. Is this difference real?



One way to answer this question is to use gender to explain temperature and test whether the addition of this variable explains a significant amount of the variability in temperature. Another, equivalent way is to test whether the difference between the average temperatures for males and females is statisically significant. This second approach can be carried out by a two-sample t-test. However, we also show another way to do this using linear regression.

```
with(body, t.test(temperature[ genderF == "male"],
                  temperature[ genderF == "female"],
                  var.equal = TRUE, paired = FALSE))
```

```
##
##  Two Sample t-test
##
## data:  temperature[genderF == "male"] and temperature[genderF == "female"]
## t = -2.285, df = 128, p-value = 0.02393
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53964 -0.03882
## sample estimates:
## mean of x mean of y
##     98.10     98.39
```

Least squares fit with categorical explanatory variables

This t-statistic can also be derived from the least squares fit of temperature to gender. However, to carry out this fit we must take into account that gender is a categorical variable.

We can do this using, what Fox calls a dummy variable, which is 1 for males and 0 for females (or vice versa). R does this for us automatically, if our variable is a factor type with two levels.

Essentially, a least squares fit of temperature to gender is simply a way to fit group means to the response variable. That is, we explain temperature by the average temperature for females or by the average temperature by males, i.e., by 98.1 or by 98.4.

We call lm as follows:

```
lm.g = lm(temperature ~ genderF, data = body)

summary(lm.g)
```

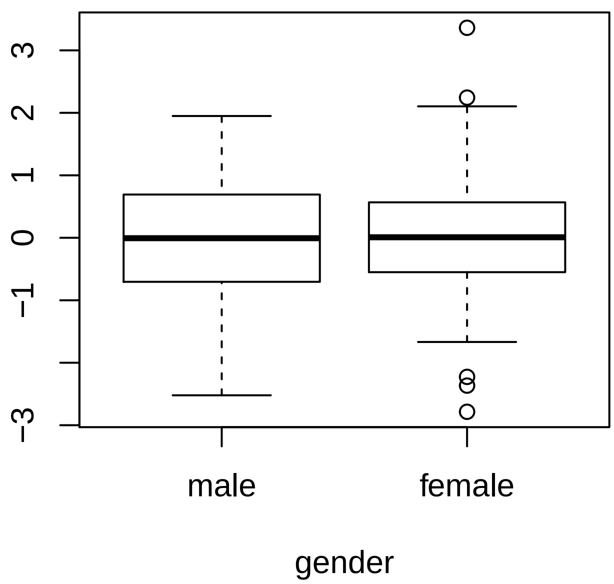
```
##
## Call:
## lm(formula = temperature ~ genderF, data = body)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9938 -0.4715  0.0062  0.4062  2.4062
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    98.1046     0.0895 1096.30   <2e-16 ***
## genderFfemale   0.2892     0.1266    2.29    0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.721 on 128 degrees of freedom
## Multiple R-squared:  0.0392, Adjusted R-squared:  0.0317
## F-statistic: 5.22 on 1 and 128 DF,  p-value: 0.0239
```

How do we read this summary? The intercept is the average body temperature for the males. The average for the females is $98.1046 + 0.2892 = 98.4$. So the difference between the average for males and females is 0.2892.

Notice that the t-statistic for the female coefficient and the p-value are the same (opposite sign for the statistic) as for the earlier t-test.

We can make residual plots. In this case, it makes sense to make boxplots of the residuals for gender.

```
boxplot(rstandard(lm.g) ~ body$genderF, xlab = "gender")
```



Notice the F-statistic in the least squares summary. We explain this next.

Variability in temperature

Let's examine the variability. Without knowledge of gender, the variability in temperature (the Total Sum of Squares) is:

```r
with(body, sum( (temperature - mean(temperature))^2 ))
```

```
## [1] 69.34
```

If we know gender, we can use the average temperature for men and for women to explain temperature, then the variability in temperature reduces to:

```r
with(body[body $ genderF == "male", ],
     sum( (temperature - mean(temperature))^2 ) ) +
  with(body[body $ genderF == "female", ],
     sum( (temperature - mean(temperature))^2 ) )
```

```
## [1] 66.63
```

What is the incremental change in variability between these two approaches? How does this incremental change compare to the variability that remains unexplained by gender?

The incremental change is 2.7. The incremental sum of squares has 1 degree of freedom. Why?

The unexplained variability is 66.63. This sum of squares has 130 - 2 degrees of freedom. Why?

Consider the ratio of these two sums of squares:

(128 / 1) * (incremental SS / unexplained SS) = 5.2

According to our normal theory, this ratio has an F-distribution with 1 and 128 degrees of freedom. The p-value is then

```r
pf(5.2, df1 = 1, df2 = 128, lower.tail = FALSE)
```

```
## [1] 0.02424
```

The p-value is 0.02. Do we reject the null hypothesis that there is no difference between males and females?

Relationship between temperature and heart rate

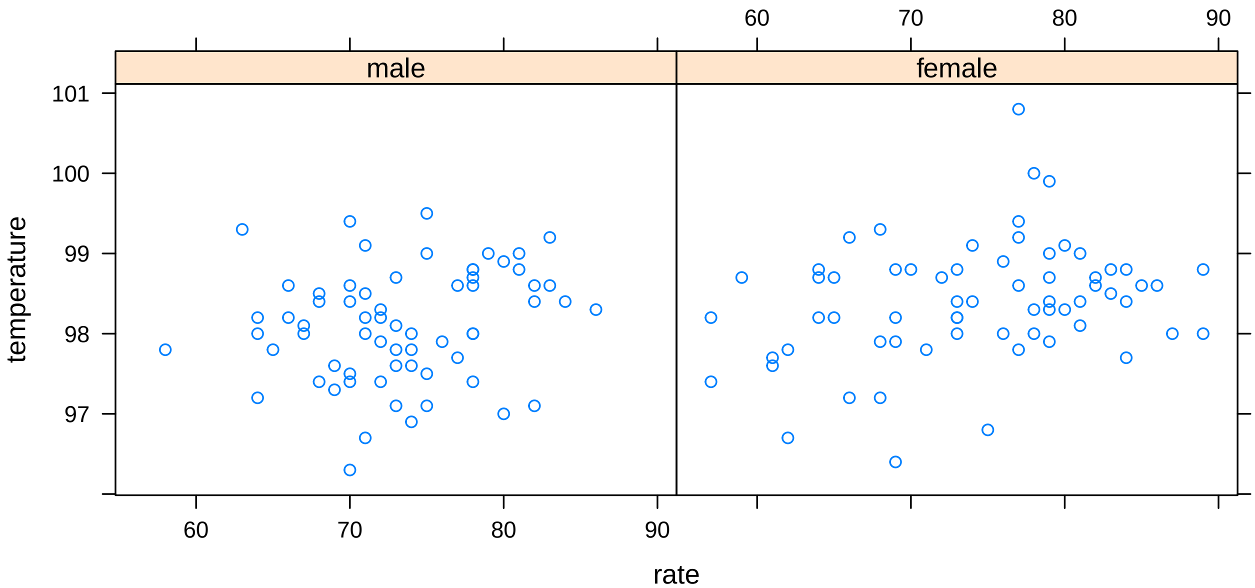
Could a higher heart rate by a mechanism of the body to generate higher body temperatures?

This question leads to other more specific questions such as:

- Is the relationship between heart rate and temperature the same for both genders?
- Is there a linear association between temperature and heart rate for all people, regardless of gender?
- Since we saw that women on average have higher temperatures, we might wonder: Is the relationship between temperature and heart rate explained by two parallel lines, where the female line sits above the male line? That is, do the lines have the same slope, but different intercepts?
- Or, do men and women have two different temperature ~ heart rate lines entirely?

Examine the relationship graphically

```r
library(lattice)
xyplot(temperature ~ rate | genderF, data = body)
```



The correlation between heart rate and temperature is 0.2537. There appears to be a weak linear association between temperature and heart rate for both males and females.

Fitting parallel lines

We can fit a line to temperature and rate, allowing for a different intercept for males and females.

```r
lm.p = lm(temperature ~ genderF + rate, data = body )

summary(lm.p)
```

```
##
## Call:
## lm(formula = temperature ~ genderF + rate, data = body)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8636 -0.4562  0.0184  0.4737  2.3342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  96.25081    0.64872  148.37   <2e-16 ***
## genderFfemale  0.26941    0.12328    2.19   0.0307 *
## rate          0.02527    0.00876    2.88   0.0046 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.702 on 127 degrees of freedom
## Multiple R-squared:  0.0983, Adjusted R-squared:  0.0841
## F-statistic: 6.92 on 2 and 127 DF,  p-value: 0.00141
```

Fitting two lines with different slopes

We want to check whether the relationship between temperature and heart rate is better described by lines with different slopes for males and females. We do this by adding an interaction term to the model.

```
lm.s = lm(temperature ~ genderF + rate + genderF:rate, data = body )

summary(lm.s)
```

```
##
## Call:
## lm(formula = temperature ~ genderF + rate + genderF:rate, data = body)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8582 -0.4463  0.0128  0.4733  2.3312
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        96.39789    1.10303   87.39   <2e-16 ***
## genderFfemale       0.04422    1.36866    0.03     0.97
## rate                0.02326    0.01499    1.55     0.12
## genderFfemale:rate  0.00306    0.01851    0.17     0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.704 on 126 degrees of freedom
## Multiple R-squared:  0.0984, Adjusted R-squared:  0.077
## F-statistic: 4.59 on 3 and 126 DF,  p-value: 0.00441
```

We see that the difference in the two slopes is not significant and the model of parallel lines adequately describes the relationship.