Ch.3-Ex.6: (a) I apply the model (3.4) in Example 3.3 to the hot dogs data. That is,

\[ B_1, \ldots, B_{20} \sim \text{i.i.d.} N(\mu, \sigma) \]
\[ M_1, \ldots, M_{17} \sim \text{i.i.d.} N(\mu + \delta_M, \sigma) \]
\[ P_1, \ldots, P_{17} \sim \text{i.i.d.} N(\mu + \delta_P, \sigma) \]

Then, the null hypotheses is \( H_0: \delta_P = 0 \). And the alternative hypotheses is \( H_1: \delta_P \neq 0 \).

The model can also be written as

\[ Y_i = \beta_0 + \beta_m X_{B,i} + \beta_p X_{P,i} + \epsilon + i \]

where \( X_{M,i} \) and \( X_{P,i} \) are indicators of Meat hot dogs and Poultry hot dogs respectively. Then, the null hypotheses is \( H_0: \beta_p = 0 \). And the alternative hypotheses is \( H_1: \beta_p \neq 0 \).

(b) I will use \( \bar{Y}_P - \bar{Y}_B \) or \( \hat{\beta}_p \) to test this hypothesis.

(c) If \( H_0 \) is true, the statistic I use should be around 0, since the true mean difference is 0.

(d) First download data from http://lib.stat.cmu.edu/DASL/Datafiles/Hotdogs.html, and import the data into R.

R codes:

```r
> hotdogs<-read.table('hotdogs.txt',header=T)
> hotdogs.fit<-lm(Calories~Type,data=hotdogs)
> summary(hotdogs.fit)
```

Call:
\[ \text{lm(formula = Calories ~ Type, data = hotdogs)} \]
Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-51.706</td>
<td>-18.492</td>
<td>-5.278</td>
<td>22.500</td>
<td>36.294</td>
</tr>
</tbody>
</table>

Coefficients:

|                    | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------|----------|------------|---------|---------|
| (Intercept)        | 156.850  | 5.246      | 29.901  | < 2e-16 *** |
| TypeMeat           | 1.856    | 7.739      | 0.240   | 0.811   |
| TypePoultry        | -38.085  | 7.739      | -4.921  | 9.4e-06 *** |

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 23.46 on 51 degrees of freedom
Multiple R-squared: 0.3866, Adjusted R-squared: 0.3626
F-statistic: 16.07 on 2 and 51 DF, p-value: 3.862e-06

We can see that the $\beta_p = -38.085$, and its standard deviation is 7.739. $-38.085/7.739 = 4.92$
Thus, it is 4.92 SD’s off.

(e) Generally, if it is beyond 2 SD’s off, we should consider it is not around 0, which means we should reject the null hypotheses. Hence, we can draw the conclusion that the mean calorie content of Poultry hot dogs is not equal to the mean calorie content of Beef hot dogs.

(f) For Meat hot dogs, the null hypotheses changes to $H_0: \beta_m = 0$.
From the summary of linear model fit, we can see that $\beta_m = 1.856$ and its standard deviation is 7.739. $1.856/7.739 = 0.24$
Since it is only 0.24 SD’s off 0, we can still accept the null hypotheses. Hence, it is plausible that there is no difference between the mean calorie content of Meat hot dogs and Beef hot dogs.

Ch.3-Ex.8 (a) We are required to analyze sodium of three type hotdogs.
R codes:
```r
> stripchart(Sodium~Type,data=hotdogs,pch=1,xlab='sodium',main='hotdogs')
> par(mfrow=c(3,1))
> plot(density(hotdogs$S[hotdogs$T=='Beef'],bw=20),xlim=c(200,600),
+ xlab='sodium',main='Beef')
> plot(density(hotdogs$S[hotdogs$T=='Meat'],bw=20),xlim=c(200,600),
+ xlab='sodium',main='Meat')
> plot(density(hotdogs$S[hotdogs$T=='Poultry'],bw=20),xlim=c(200,600),
+ xlab='sodium',main='Poultry')
```
hotdogs

sodium

Beef

Meat

Poultry

200 300 400 500 600
(b) Similar to mode (3.4), we can write model for analyzing sodium three type hotdogs.

\[ B_1, \ldots, B_{20} \sim \text{i.i.d.} N(\mu, \sigma) \]
\[ M_1, \ldots, M_{17} \sim \text{i.i.d.} N(\mu + \delta_M, \sigma) \]
\[ P_1, \ldots, P_{17} \sim \text{i.i.d.} N(\mu + \delta_P, \sigma) \]

Here, \( B_i, M_i, \) and \( P_i \) denote sodium of hotdogs.
\( \mu \) is the mean sodium of beef hot dogs.
\( \delta_M \) is the mean difference between beef and meat hotdogs.
\( \delta_P \) is the mean difference between beef and poultry hotdogs.
\( \sigma \) is the standard deviation among every type.

(c) Or the previous model can be rewritten as:

\[ Y_i = \beta_0 + \beta_m X_{B,i} + \beta_p X_{P,i} + \epsilon_i \]
$Y_i$: sodium of the $i$'th hotdog  
$\beta_0$: the mean sodium of beef hot dogs.  
$\beta_m$: the mean difference between beef and meat hotdogs.  
$\beta_p$: the mean difference between beef and poultry.  
Here, $X_{B,i}$ and $X_{P,i}$ are indicator variables.  

$$X_{B,i} = \begin{cases} 
1 & \text{if the } i^{\text{th}} \text{ hotdog is Meat,} \\
0 & \text{otherwise}
\end{cases}$$

$$X_{P,i} = \begin{cases} 
1 & \text{if the } i^{\text{th}} \text{ hotdog is Poultry,} \\
0 & \text{otherwise}
\end{cases}$$

(d) Similar to model (3.7)  

$$Y = XB + E$$

$$Y = (Y_1, \ldots, Y_{54})'$$  
$$B = (\mu, \delta_M, \delta_P)'$$  
$$E = (\epsilon_1, \ldots, \epsilon_{57})'$$

(e) R codes:

```r
> hotdogs.fit2<-lm(Sodium~Type,data=hotdogs)
> summary(hotdogs.fit2)
```

Call: lm(formula = Sodium ~ Type, data = hotdogs)

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-274.53</td>
<td>-75.00</td>
<td>-14.34</td>
<td>77.07</td>
<td>243.85</td>
</tr>
</tbody>
</table>

Coefficients:

|                     | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------|----------|------------|---------|---------|
| (Intercept)         | 401.15   | 21.13      | 18.988  | <2e-16  *** |
| TypeMeat            | 17.38    | 31.17      | 0.558   | 0.5795  |
| TypePoultry         | 57.85    | 31.17      | 1.856   | 0.0692  .|

---

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 94.48 on 51 degrees of freedom  
Multiple R-squared: 0.06517,  Adjusted R-squared: 0.02851  
F-statistic: 1.778 on 2 and 51 DF,  p-value: 0.1793

(f) From the R output, we can find the estimate value and standard deviation of the parameters.
<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>401.5</td>
<td>21.13</td>
</tr>
<tr>
<td>$\beta_m$</td>
<td>17.38</td>
<td>31.17</td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>57.85</td>
<td>31.17</td>
</tr>
</tbody>
</table>

(g) R codes:

```r
> m<-c(401.15, 17.38, 57.85)
> s<-c(21.13, 31.17, 31.17)
> par(mfrow=c(2,2))
> x<-seq(m[1]-3*s[1], m[1]+3*s[1], length=40)
> plot(x, dnorm(x, m[1], s[1]), type='l', xlab=expression(mu), ylab='likelihood')
> x<-seq(m[2]-3*s[2], m[2]+3*s[2], length=40)
> plot(x, dnorm(x, m[2], s[2]), type='l', xlab=expression(delta[M]),
+ ylab='likelihood')
> x<-seq(m[3]-3*s[3], m[3]+3*s[3], length=40)
> plot(x, dnorm(x, m[3], s[3]), type='l', xlab=expression(delta[P]),
+ ylab='likelihood')
```

![Graphs showing likelihood distributions for $\mu$, $\delta_M$, and $\delta_P$.]
(h) Interpretation is that $\mu$ is likely somewhere around 401.5, plus or minus about 42 or so; $\delta_M$ is somewhere around 17.38, plus or minus about 62 or so; $\delta_P$ is somewhere around 57.85, plus or minus about 62 or so. In particular, there is no strong evidence that Meat hot dogs have, on average, more or fewer sodium than Beef hot dogs; but there is some evidence that Poultry hot dogs have more.

Ch.3-Ex.10 First download data from http://lib.stat.cmu.edu/DASL/Datafiles/IceCream.html, and import the data into R.

The model I want to establish is $\mathbf{IC} = \mathbf{XB} + \mathbf{E}$, where

$\mathbf{IC} = (IC_1, \ldots, IC_{30})'$

$\mathbf{B} = (\beta_0, \beta_1)'$

$\mathbf{E} = (\epsilon_1, \ldots, \epsilon_{30})'$

$\mathbf{X} = \begin{pmatrix} 1 \\ temp_1 \\ 1 \\ temp_2 \\ \vdots \\ 1 \\ temp_{30} \end{pmatrix}$

R codes:

```r
> icecream.fit<-lm(IC~temp, data=icecream)
> summary(icecream.fit)
```

Call:
```
  lm(formula = IC ~ temp, data = icecream)
```

Residuals:
```
     Min  1Q Median  3Q    Max
-0.069411 -0.024478 -0.007371 0.029126 0.120516
```

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 0.2068621 | 0.0247002 | 8.375 | 4.13e-09 *** |
| temp | 0.0031074 | 0.0004779 | 6.502 | 4.79e-07 *** |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.04226 on 28 degrees of freedom
Multiple R-squared: 0.6016, Adjusted R-squared: 0.5874
F-statistic: 42.28 on 1 and 28 DF, p-value: 4.789e-07

So the model I estimate is $IC_i = 0.207 - 0.003temp_i + \epsilon_i$. $\beta_0$ is somewhat around 0.207, plus or minus about 0.05 or so; $\beta_1$ is somewhat around 0.003, plus or minus
about 0.001 or so.
If temperature increases by about 5 F, about $5 \times 0.003 = 0.015$ to increase.

```r
> plot(icecream$temp, icecream$IC, main = 'icecream')
> abline(0.207, 0.003)
```