Galton’s sweet peas

Sir Francis Galton discusses in an 1877 paper the inheritance of various traits in sweet peas. He categorized parent plants according to the typical diameter of the peas produced. By comparing the peas produced by the parent plant with those produced by the offspring plant he wanted to detect patterns in the inheritance from one generation to the next.

```r
> gp<-read.table("galton_peas_data.q",he=T)
> gp
     parent offspring    sd
    1     21       17.26 1.988
    2     20       17.07 1.938
    3     19       16.37 1.896
    4     18       16.40 2.037
    5     17       16.13 1.654
    6     16       16.17 1.594
    7     15       15.98 1.763

> attach(gp)

> plot(parent,offspring)

> fit<-lm(offspring~parent)
> summary(fit)

Call: lm(formula = offspring ~ parent)
Residuals:
     1     2     3     4     5     6     7
   0.1471 0.1671 -0.3229 -0.08286 -0.1429 0.1071 0.1271

Coefficients:
                     Value Std. Error t value Pr(>|t|)
(Intercept)       12.0772    2.2258  5.4102 0.005147 **
parent            -0.1023    0.0632 -1.6203 0.222114

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

Residual standard error: 0.8135 on 4 degrees of freedom
Multiple R-squared: 0.08157,    Adjusted R-squared: -0.2313
F-statistic: 0.2601 on 1 and 4 DF,  p-value: 0.6241
```


> fit2<-lm(offspring ~ parent, weights=1/sd^2)
> summary(fit2)

Call: lm(formula = offspring ~ parent, weights = 1/sd^2)

Residuals:
    1    2    3    4    5    6    7
 0.08187 0.09162 -0.16747 -0.04067 -0.08950 0.06071 0.06328

Coefficients:
                  Value  Std. Error   t value Pr(>|t|)
(Intercept) 12.79644 0.68106 18.7872 0.0000
parent         0.21000 0.03875  5.4385 0.0029

Residual standard error: 0.2043 on 5 degrees of freedom
Multiple R-Squared: 0.8554
F-statistic: 29.58 on 1 and 5 degrees of freedom, the p-value is 0.002852
Bland’s Apple Shoots

Many types of trees produce two types of morphologically different shoots. Some branches bear no fruits (remain vegetative) year after year and contribute to the size of the tree (these are called long shoots), others seldom exceed 1cm in length but bear flowers from which fruit may arise (these are the short shoots).

The data below were collected by Bland on McIntosh apple trees.

He took samples of long and short shoots and sent them to the laboratory for analysis. One theory is that the difference is given by the number of stem units.

bland<-read.table("apple_shoots_data.q",he=T)

attach(bland)

bland

<table>
<thead>
<tr>
<th>Day</th>
<th>n</th>
<th>average</th>
<th>sd</th>
<th>type.shoot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>10.20</td>
<td>0.83</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>10.40</td>
<td>0.54</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>10.60</td>
<td>0.54</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>12.50</td>
<td>0.83</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>12.00</td>
<td>1.41</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>15.00</td>
<td>0.82</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>88</td>
<td>26.04</td>
<td>0.99</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>91</td>
<td>26.60</td>
<td>0.54</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>94</td>
<td>27.12</td>
<td>1.16</td>
<td>0</td>
</tr>
<tr>
<td>74</td>
<td>97</td>
<td>26.83</td>
<td>0.59</td>
<td>0</td>
</tr>
<tr>
<td>75</td>
<td>100</td>
<td>28.70</td>
<td>0.47</td>
<td>0</td>
</tr>
</tbody>
</table>
> plot(Day[type.shoot==1], average[type.shoot==1], pch=2)  # triangles
> points(Day[type.shoot==0], average[type.shoot==0], pch=3)  # crosses

> yshort<-average[type.shoot==0]
> ylong<-average[type.shoot==1]

> dayshort<-Day[type.shoot==0]
> daylong<-Day[type.shoot==1]
> nshort<-n[type.shoot==0]
> nlong<-n[type.shoot==1]

> fit1<-lm(yshort~dayshort)
> summary(fit1)

Call: lm(formula = yshort ~ dayshort)
Residuals:
       Min        1Q      Median        3Q       Max
-1.144000 -0.252300  0.103800  0.343700  0.888200

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.45882   0.15259  61.971 1.21e-15 ***
  dayshort   0.18721   0.00280  67.279 1.21e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5245 on 46 degrees of freedom
Multiple R-Squared: 0.9899
F-statistic: 4526 on 1 and 46 degrees of freedom, the p-value is 0

> fit2<-lm(ylong~daylong)
> summary(fit2)
Call: lm(formula = ylong ~ daylong)

Residuals:
    Min 1Q Median 3Q    Max
-1.788 -0.5744 0.1902 0.4269 0.9465

Coefficients:
            Value Std. Error t value Pr(>|t|)
(Intercept) 9.9042    0.2578  38.4208 0.0000
   daylong  0.2158    0.0041  53.1299 0.0000

Residual standard error: 0.6744 on 26 degrees of freedom
Multiple R-Squared: 0.9909
F-statistic: 2823 on 1 and 26 degrees of freedom, the p-value is 0

> nshort<-n[type.shoot==0]
> nlong<-n[type.shoot==1]

fit12<-lm(yshort~dayshort,weights=nshort)
fit22<-lm(ylong~daylong,weights=nlong)

> sdshort<-sd[type.shoot==0]
> sdlong<-sd[type.shoot==1]

fit13<-lm(yshort[sdshort>0]~dayshort[sdshort>0],
         weights=1/sdshort[sdshort>0]^2)
fit23<-lm(ylong~daylong,weights=1/sdlong^2)

> summary(fit13)

Call: lm(formula = yshort[sdshort > 0] ~ dayshort[sdshort > 0],
weights = 1/sdshort[sdshort > 0]^2)

Residuals:
Geese counting

Aerial survey methods are regularly used to estimate the number of snow geese in their summer range areas west of Hudson Bay. To obtain estimates, a small aircraft fly over the range and, when a flock of geese is spotted an experienced person estimates the number of
geese in the flock. To investigate the reliability of this method of counting an experiment was conducted in which an airplane carrying two observers flew over $n = 45$ flocks and each observer made an independent estimate of the number of birds. Also, a photograph of the flock was taken so that an exact count could be produced.

```r
> geese<-read.table("geese_data.q",he=T)

> geese

          photo obs1 obs2
     1         56   50  40
     2         38   25  30
     3         25   30  40
     4         48   35  45
                        ....................
     41        325  200  400
     42        114   60  120
     43         83   40  40
     44         91   35  60
     45         56   20  40

> attach(geese)

> fit1<-lm(photo~obs1)

> summary(fit1)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept)   26.64957   8.61448   3.094  0.00347 **
obs1           0.88256   0.07764  11.367 1.54e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
Residual standard error: 44.41 on 43 degrees of freedom
Multiple R-Squared: 0.7503, Adjusted R-squared: 0.7445

> fit2<-lm(photo~obs2)

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 16.16428   | 6.82973 | 2.367    | 0.0225 *   |
| obs2      | 0.76907    | 0.04835 | 15.905   | <2e-16 *** |

---

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

Residual standard error: 33.87 on 43 degrees of freedom
Multiple R-Squared: 0.8547, Adjusted R-squared: 0.8513

> postscript("geese_fit1_plot.ps")
> par(mfrow=c(2,2))
> plot(fit1)
> dev.off()

> postscript("geese_fit2_plot.ps")
> par(mfrow=c(2,2))
> plot(fit2)
> dev.off()
> fit12 <- lm(photo ~ obs1, weight = 1/weights.geese)
> fit22 <- lm(photo ~ obs2, weight = 1/weights.geese)

> photo[29]
[1] 342
As a result of this experiment, the practice of using visual counts of flock size to determine population estimates was discontinued in favor of using photographs.