STAT-S631

Assignment 4

John Koo

```
dp <- loadNamespace('dplyr')
import::from(magrittr, `%>%`, `%<>%`)
import::from(foreach, foreach, `%dopar%`)
doMC::registerDoMC(parallel::detectCores())
library(ggplot2)
theme_set(theme_bw())
```

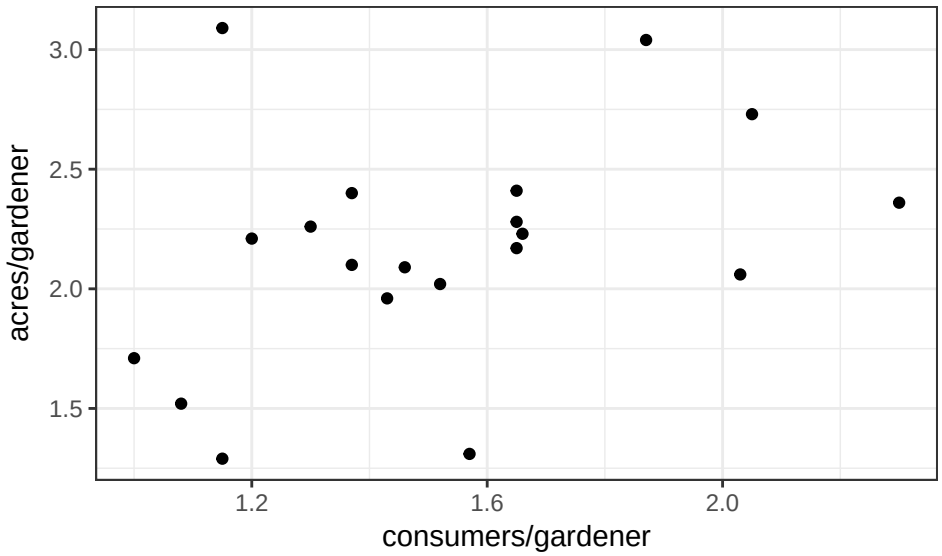
Question 1

Question 2

```
sahlins.df <- read.delim('~/dev/stats-hw/stat-s631/Sahlins.txt', sep = ' ')
```

Part a

```
ggplot(sahlins.df) +
  geom_point(aes(x = consumers, y = acres)) +
  labs(x = 'consumers/gardener', y = 'acres/gardener')
```



From the scatterplot, the data do not appear to be particularly linear, although there appears to be a very slight positive correlation. We can compute this:

```
cor(sahlins.df$consumers, sahlins.df$acres)
```

```
[1] 0.3756561
```

Most of the data appear to be clustered in the center with a ring of points surrounding it. One household has an unusually high value for acres per gardener—it's almost 3 times its value for consumers per gardener (~3 vs ~1).

Part b

```
sahlins.mod <- lm(acres ~ consumers, data = sahlins.df)
summary(sahlins.mod)
```

```
Call:
lm(formula = acres ~ consumers, data = sahlins.df)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8763 -0.1873 -0.0211  0.2135  1.1206

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.3756     0.4684   2.937  0.00881 **
consumers     0.5163     0.3002   1.720  0.10263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4543 on 18 degrees of freedom
Multiple R-squared:  0.1411,    Adjusted R-squared:  0.0934
F-statistic: 2.957 on 1 and 18 DF,  p-value: 0.1026
```

The results indicate that, if we set the conventional value of $\alpha = .05$, we would fail to reject the null hypothesis that $\beta_1 \neq 0$, implying that there is no significant relationship between acres per gardener and consumers per gardener. However, using the same value of α, we reject the null hypothesis that $\beta_0 = 0$, indicating that each household receives some amount regardless of productivity.

The residual standard error $\hat{\sigma}$ is 0.4543

If we remove the 4^{th} data point:

```
sahlins.mod.2 <- lm(acres ~ consumers, data = sahlins.df[-4, ])
summary(sahlins.mod.2)
```

```
Call:
lm(formula = acres ~ consumers, data = sahlins.df[-4, ])

Residuals:
     Min       1Q   Median       3Q      Max
-0.82291 -0.16808  0.03215  0.23505  0.69061

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0000     0.3969   2.519   0.0221 *
consumers     0.7216     0.2514   2.870   0.0106 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3681 on 17 degrees of freedom
Multiple R-squared:  0.3264,    Adjusted R-squared:  0.2868
F-statistic: 8.238 on 1 and 17 DF,  p-value: 0.01061
```

We would reject the null hypothesis for both β_0 and β_1. This would imply that each household receives some base amount but also can work for additional resources.

Part c

Part d