Statistical Methods 2 Homework 1

2023-01-19

Loadings possible useful libraries

```
library(ggplot2)
library(Sleuth3)
library(MASS)
```

Problem 1)

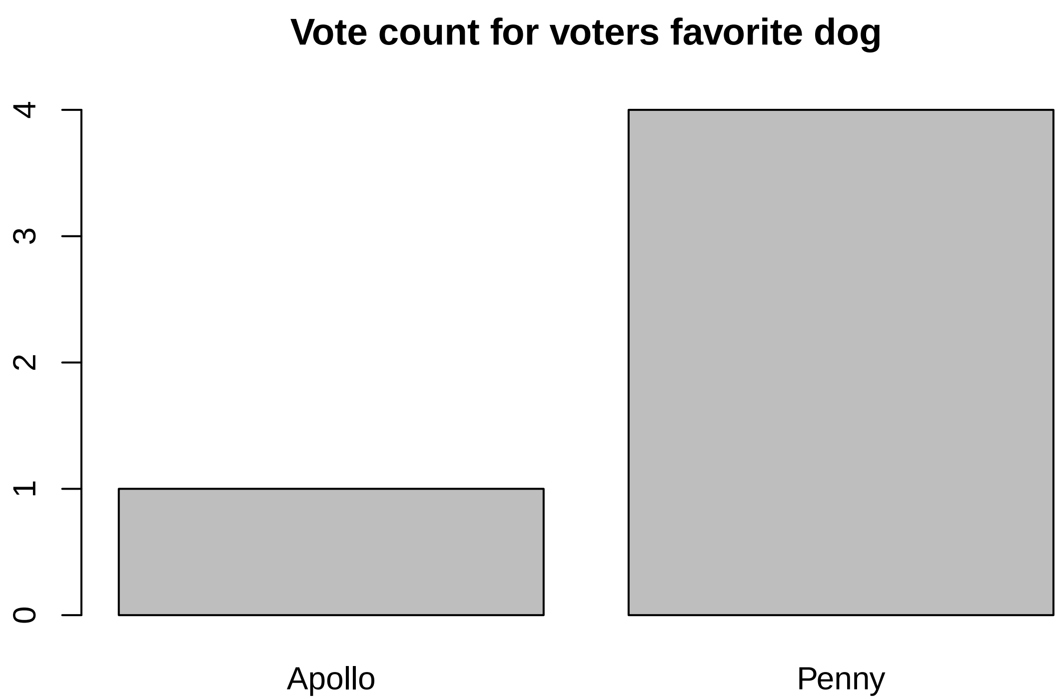
a)

```
#voting for favorite dog
name_of_voter <- c("Luis", "Fob", "Bob", "Rob", "Lob") #unique voters voting
fav_dog_data <- c("Apollo", "Penny", "Penny", "Penny", "Penny") #voters favorite dog
weight_in_kg <- c(1,10,20,30,40) #weight in kg of voter

df <- data.frame(name_of_voter,fav_dog_data,weight_in_kg)

print(df)
```

```
##   name_of_voter fav_dog_data weight_in_kg
## 1          Luis       Apollo            1
## 2           Fob        Penny           10
## 3           Bob        Penny           20
## 4           Rob        Penny           30
## 5           Lob        Penny           40
```

b)

```
plot(x = factor(df$fav_dog_data), main = "Vote count for voters favorite dog")
```



c)

```
mammals <- mammals
fit <- lm(brain ~ body, data=mammals)
summary(fit)
```

```
##
## Call:
## lm(formula = brain ~ body, data = mammals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -810.07  -88.52  -79.64  -13.02 2050.33
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.00440   43.55258    2.09   0.0409 *
## body         0.96650    0.04766   20.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 334.7 on 60 degrees of freedom
## Multiple R-squared:  0.8727, Adjusted R-squared:  0.8705
## F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

d)

```r
B_0 <- fit$coefficients[1]
```

e)

```r
se <- sigma(fit)
```

f)

```r
log_body <- log(mammals$body)
log_brain <- log(mammals$brain)

mammals$log_body <- log_body
mammals$log_brain <- log_brain

mammals
```

```
##                                body   brain    log_body  log_brain
## Arctic fox                    3.385   44.50  1.21935391  3.7954892
## Owl monkey                    0.480   15.50 -0.73396918  2.7408400
## Mountain beaver               1.350    8.10  0.30010459  2.0918641
## Cow                         465.000  423.00  6.14203741  6.0473722
## Grey wolf                    36.330  119.50  3.59264385  4.7833164
## Goat                         27.660  115.00  3.31998733  4.7449321
## Roe deer                     14.830   98.20  2.69665216  4.5870062
## Guinea pig                    1.040    5.50  0.03922071  1.7047481
## Verbet                        4.190   58.00  1.43270073  4.0604430
## Chinchilla                    0.425    6.40 -0.85566611  1.8562980
## Ground squirrel               0.101    4.00 -2.29263476  1.3862944
## Arctic ground squirrel        0.920    5.70 -0.08338161  1.7404662
## African giant pouched rat     1.000    6.60  0.00000000  1.8870696
## Lesser short-tailed shrew     0.005    0.14 -5.29831737 -1.9661129
## Star-nosed mole               0.060    1.00 -2.81341072  0.0000000
## Nine-banded armadillo         3.500   10.80  1.25276297  2.3795461
## Tree hyrax                    2.000   12.30  0.69314718  2.5095993
## N.A. opossum                  1.700    6.30  0.53062825  1.8405496
## Asian elephant             2547.000 4603.00  7.84267147  8.4344635
## Big brown bat                 0.023    0.30 -3.77226106 -1.2039728
## Donkey                      187.100  419.00  5.23164323  6.0378709
## Horse                       521.000  655.00  6.25575004  6.4846352
## European hedgehog             0.785    3.50 -0.24207156  1.2527630
## Patas monkey                 10.000  115.00  2.30258509  4.7449321
## Cat                           3.300   25.60  1.19392247  3.2425924
## Galago                        0.200    5.00 -1.60943791  1.6094379
## Genet                         1.410   17.50  0.34358970  2.8622009
## Giraffe                     529.000  680.00  6.27098843  6.5220928
## Gorilla                     207.000  406.00  5.33271879  6.0063532
```

```
## Grey seal                  85.000  325.00  4.44265126  5.7838252
## Rock hyrax-a                0.750   12.30 -0.28768207  2.5095993
## Human                      62.000 1320.00  4.12713439  7.1853870
## African elephant         6654.000 5712.00  8.80297346  8.6503245
## Water opossum               3.500    3.90  1.25276297  1.3609766
## Rhesus monkey               6.800  179.00  1.91692261  5.1873858
## Kangaroo                   35.000   56.00  3.55534806  4.0253517
## Yellow-bellied marmot       4.050   17.00  1.39871688  2.8332133
## Golden hamster              0.120    1.00 -2.12026354  0.0000000
## Mouse                       0.023    0.40 -3.77226106 -0.9162907
## Little brown bat            0.010    0.25 -4.60517019 -1.3862944
## Slow loris                  1.400   12.50  0.33647224  2.5257286
## Okapi                     250.000  490.00  5.52146092  6.1944054
## Rabbit                      2.500   12.10  0.91629073  2.4932055
## Sheep                      55.500  175.00  4.01638302  5.1647860
## Jaguar                    100.000  157.00  4.60517019  5.0562458
## Chimpanzee                 52.160  440.00  3.95431592  6.0867747
## Baboon                     10.550  179.50  2.35612586  5.1901752
## Desert hedgehog             0.550    2.40 -0.59783700  0.8754687
## Giant armadillo            60.000   81.00  4.09434456  4.3944492
## Rock hyrax-b                3.600   21.00  1.28093385  3.0445224
## Raccoon                     4.288   39.20  1.45582042  3.6686767
## Rat                         0.280    1.90 -1.27296568  0.6418539
## E. American mole            0.075    1.20 -2.59026717  0.1823216
## Mole rat                    0.122    3.00 -2.10373423  1.0986123
## Musk shrew                  0.048    0.33 -3.03655427 -1.1086626
## Pig                       192.000  180.00  5.25749537  5.1929569
## Echidna                     3.000   25.00  1.09861229  3.2188758
## Brazilian tapir           160.000  169.00  5.07517382  5.1298987
## Tenrec                      0.900    2.60 -0.10536052  0.9555114
## Phalanger                   1.620   11.40  0.48242615  2.4336134
## Tree shrew                  0.104    2.50 -2.26336438  0.9162907
## Red fox                     4.235   50.40  1.44338333  3.9199912
```

g)

```r
fit <- lm(log_brain ~ log_body, data=mammals)
summary(fit)
```

```
## 
## Call:
## lm(formula = log_brain ~ log_body, data = mammals)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.71550 -0.49228 -0.06162  0.43597  1.94829 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  2.13479    0.09604   22.23   <2e-16 ***
## log_body     0.75169    0.02846   26.41   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6943 on 60 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9195
## F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16
```

Problem 2)

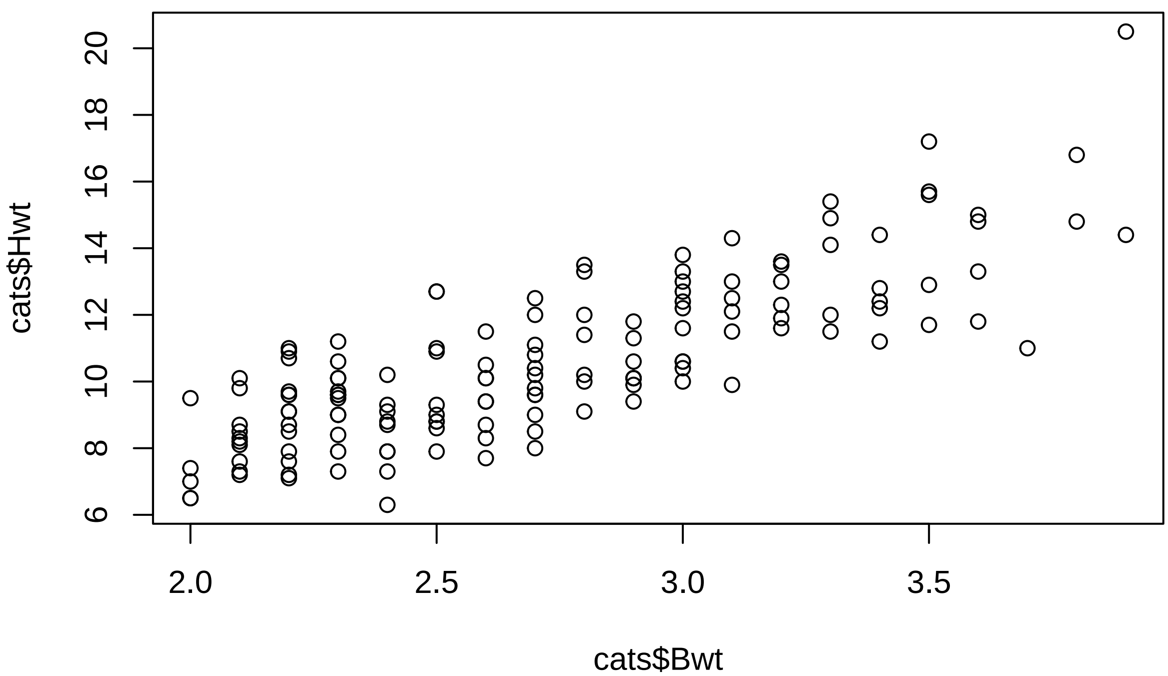
```
cats <- cats
```

a)

The explanatory variable here would be the body weight (Bwt). The response variable would be hearth weight (Hwt). This is because we are trying to investigate if body weight is a good variable to predict heart weight in cats.

b)

```
plot(cats$Bwt,cats$Hwt)
```



There appears to be a somewhat positive linear relationship between cats body weight and heart weight. There appears to be less variance the lower the weight of the cat, and higher variance as the body weight increases. However, this effect doesn't seem too large.

c)

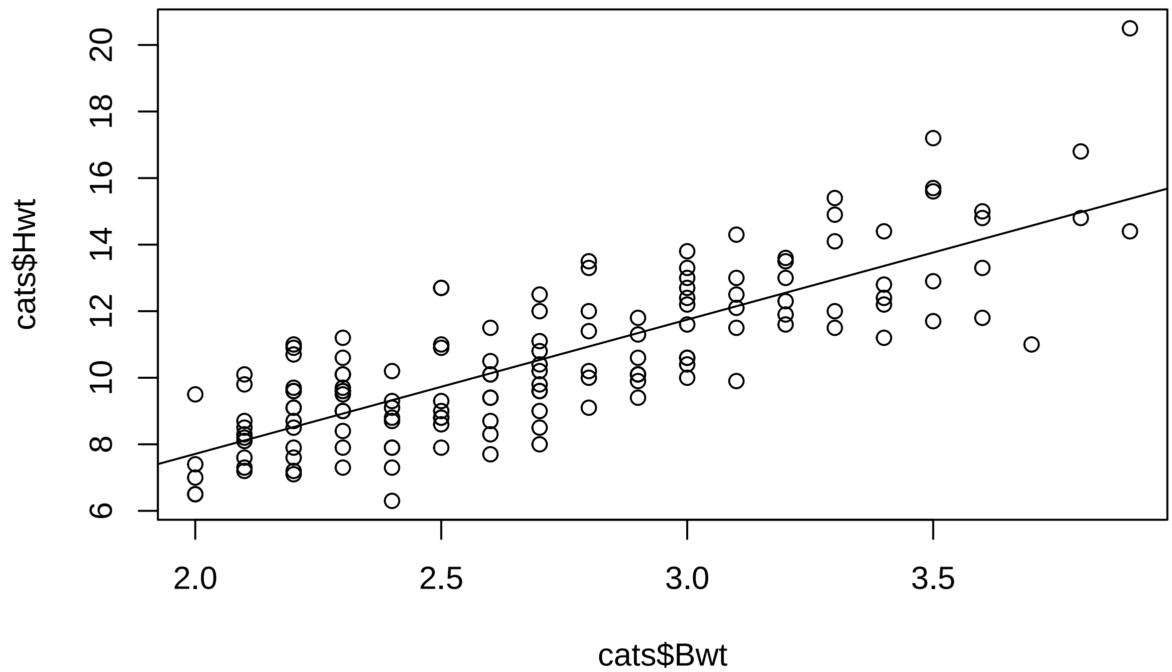
Yes, I think using a linear regression model would be appropriate as it has a fairly positive linear relationship and the variance is fairly consistent (and possibly normal) for the distribution around heart weight. There is also a reasonably large sample size. It also likely reasonably fair to assume independence between observations in this study.

d)

```r
fit <- lm(Hwt ~ Bwt, data = cats)
summary(fit)
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt, data = cats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5694 -0.9634 -0.0921  1.0426  5.1238
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.3567     0.6923  -0.515    0.607
## Bwt           4.0341     0.2503  16.119   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.452 on 142 degrees of freedom
## Multiple R-squared:  0.6466, Adjusted R-squared:  0.6441
## F-statistic: 259.8 on 1 and 142 DF,  p-value: < 2.2e-16
```

```r
plot(cats$Bwt,cats$Hwt)
abline(fit)
```



The intercept parameter was estimated to be $\hat{\beta}_1 = -0.3567$ and $\hat{\beta}_0 = 4.0341$.

e)

In cats, for every 1 unit increase in body weight (Bwt) in kg it is associated with mean heart weight increases (Hwt) by 4.0341 grams. With a base case of 0 kg body weight for a cat, the heart weight would be -.3567 grams, which doesn't make sense to have a negative weight however this is okay as it also doesn't make sense to have a 0 weight.

f)

```
B_1_hypothesis <- 3.75
B_1_estimated <- fit$coefficients[2]
difference <- B_1_estimated - B_1_hypothesis
se <- summary(fit)$coefficients[2, 2]

t_stat <- difference/se
t_stat
```

```
##      Bwt
## 1.135064
```

```r
p_value <- 2*(1-pt(t_stat, fit$df.residual))
p_value
```

```
##      Bwt
## 0.258261
```

$$H_0 : \beta_1 = 4$$

Here, the test statistic = 1.135 and the p-value = .258. We would fail to reject the null hypothesis with an alpha level .05 that B_1 is equal to 4. To determine a linear relationship between the explanatory variable and the response variable, it is not very useful because any non zero number can be considered a non linear relationship.

g)

```r
intercept <- fit$coefficients[1]
predicted <- intercept + B_1_estimated*2.25
predicted
```

```
## (Intercept)
##    8.719979
```

Based on the fitted regression model in part (d), we would predict a mean heart weight of 8.72 grams for a cat with a body weight of 2.25 kg.