

MATH10282 Introduction to Statistics
Semester 2, 2021/2022
Solutions to the R coursework assignment

Sugar is a major ingredient in many breakfast cereals. The data below give the sugar content as a percentage of weight for 19 brands of cereal for children and 19 brands of cereal for adults.

Children Cereal	Adult Cereal
40.3	20
55	30.2
45.7	2.2
43.3	7.5
50.3	4.4
45.9	22.2
53.5	16.6
43	14.5
44.2	21.4
44	3.3
47.4	6.6
44	7.8
33.6	10.6
55.1	16.2
48.8	14.5
50.4	4.1
37.8	15.8
60.3	4.1
46.6	2.4

(a) The commands

```
x=c(40.3, 55, 45.7, 43.3, 50.3, 45.9, 53.5, 43, 44.2, 44,
47.4, 44, 33.6, 55.1, 48.8, 50.4, 37.8, 60.3, 46.6)
c(min(x),median(x),mean(x),max(x),sd(x))
```

give the following

Minimum = 33.6, Median = 45.9, Mean = 46.8, Maximum = 60.4, SD = 6.418376

[1]

(b) The commands

```
y=c(20, 30.2, 2.2, 7.5, 4.4, 22.2, 16.6, 14.5, 21.4, 3.3,
6.6, 7.8, 10.6, 16.2, 14.5, 4.1, 15.8, 4.1, 2.4)
c(min(y),median(y),mean(y),max(y),sd(y))
```

give the following

Minimum = 2.2, Median = 10.6, Mean = 11.810526, Maximum = 30.2, SD = 8.055977

[1]

(c) The minimum, median, mean and maximum are larger for the sugar content of children cereal. The standard deviation is larger for the sugar content of adults cereal. [1]

(d) The command

```
boxplot(x,y,names=c("Children cereal","Adults cereal"),  
ylab="Sugar content as percentage of weight")
```

gives

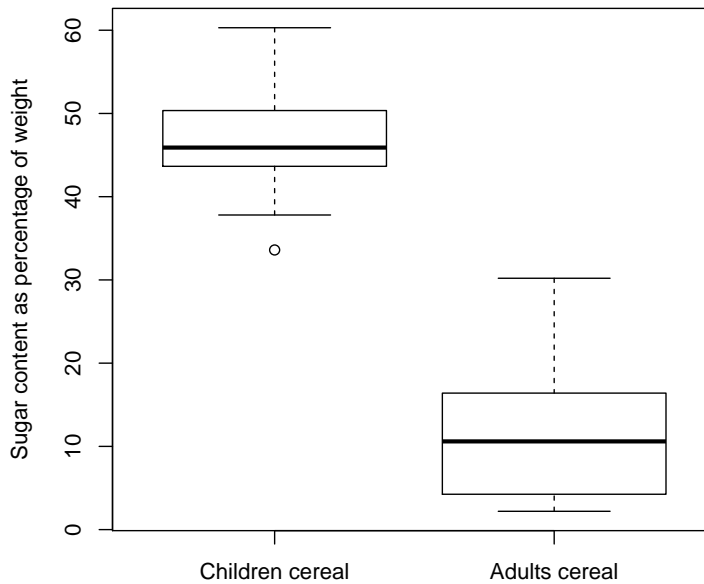


Figure 1: The boxplots of sugar content of children cereal and adults cereal.

The boxplots show that the median, first quartile and third quartile are larger for the sugar content of children cereal. The range and inter quartile range are larger for the sugar content of adults cereal. [1]

(e) The estimates of the two parameters of the normal distribution are

$$\hat{\mu} = \frac{1}{19} \sum_{i=1}^{19} x_i = 46.8$$

and

$$\hat{\sigma} = \sqrt{\frac{1}{19} \sum_{i=1}^{19} (x_i - \hat{\mu})^2} = 6.247189.$$

The command

```
ks.test(x,"pnorm",46.8,6.247189)
```

gives the output

```
One-sample Kolmogorov-Smirnov test

data: x
D = 0.11361, p-value = 0.9669
alternative hypothesis: two-sided
```

Hence, the fit of the normal distribution is adequate. [1]

(f) The estimates of the two parameters of the normal distribution are

$$\hat{\mu} = \frac{1}{19} \sum_{i=1}^{19} y_i = 11.810526$$

and

$$\hat{\sigma} = \sqrt{\frac{1}{19} \sum_{i=1}^{19} (y_i - \hat{\mu})^2} = 7.841113.$$

The command

```
ks.test(y,"pnorm",11.810526,7.841113)
```

gives the output

```
One-sample Kolmogorov-Smirnov test

data: y
D = 0.16917, p-value = 0.6484
alternative hypothesis: two-sided
```

Hence, the fit of the normal distribution is adequate. [1]

(g) The following commands

```
up<-mean(x)-mean(y)+qnorm(0.975)*sqrt(var(x)/19+var(y)/19)
low<-mean(x)-mean(y)-qnorm(0.975)*sqrt(var(x)/19+var(y)/19)
```

give the 95% confidence interval as [30.35802, 39.62093]. Since this interval is entirely positive, the sugar content of children cereal is significantly larger than the sugar content of adults cereal. [2]

(h) The test statistic

$$(\text{mean}(x) - \text{mean}(y)) / \sqrt{\text{var}(x)/19 + \text{var}(y)/19}$$

gives 14.80704. Since this is greater than $z_{0.95} = 1.644854$ there is evidence against the hypothesis that the sugar contents are equal for children cereal and adults cereal. [2]

[Total 10 marks]