Question:

The data below show the sugar content (as a percentage of weight) of several national brands of children's and adults' cereals.

Children's cereals: 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, 46.5

Adults' cereals: 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, 3.5, 8.5, 10, 1, 4.4, 1.3, 8.1, 4.7, 18.4

(a) Does it seem reasonable to assume that each sample comes from a normal distribution? Draw Q-Q plots to answer this question.
(b) Can the variances of the two distributions be assumed to be equal? Justify your answer.
(c) Compute an appropriate 90% confidence interval for difference in mean sugar contents of the two cereal types. What assumptions did your make, if any, to construct the interval?
(d) What do you conclude on the basis of your answer in (c)? Can we say that children's cereals have more sugar on average than adult cereals? If yes, by how much? Justify your answers.

Answer:

Question a:

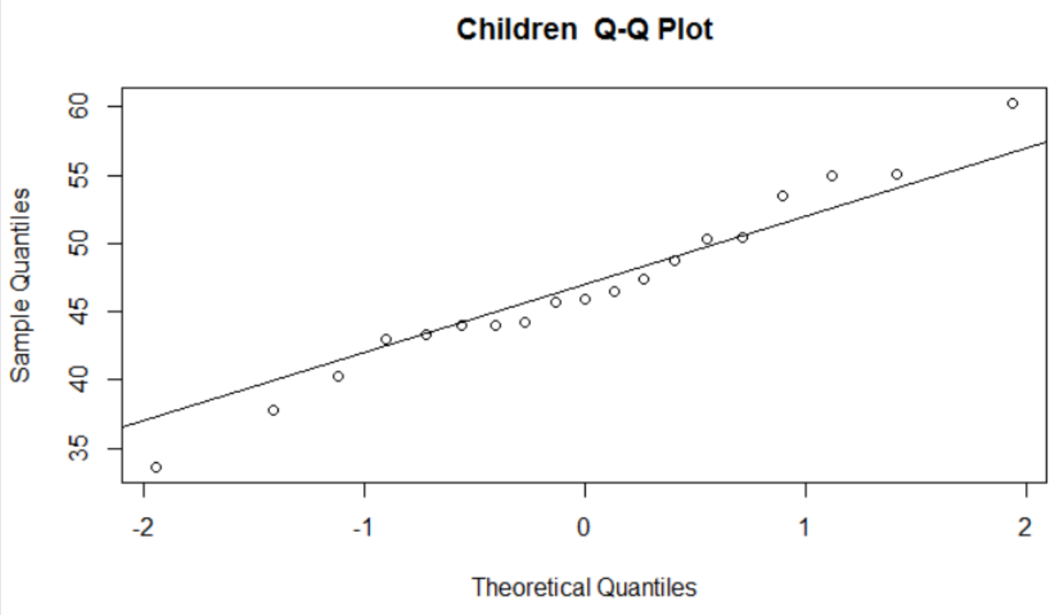
First we input data in to R.

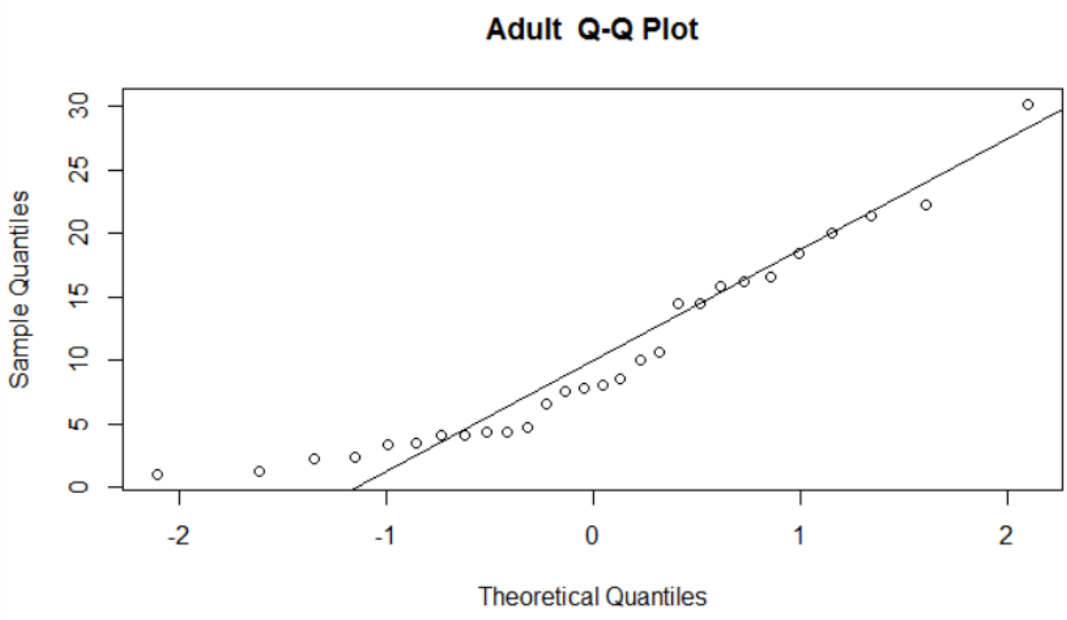
```
children<-
c(40.3,55,45.7,43.3,50.3,45.9,53.5,43,44.2,44,47.4,44,33.6,55.1,48.8,50.4,37.8,60.3,46.5)
adult<-
c(20,30.2,2.2,7.5,4.4,22.2,16.6,14.5,21.4,3.3,6.6,7.8,10.6,16.2,14.5,4.1,15.8,4.1,2.4,3.5,8.5,
10,1,4.4,1.3,8.1,4.7,18.4)
```

And then we call the function:

```
qqnorm(children, main = "Children    Q-Q Plot")
qqline(children)
qqnorm(adult, main = "Adult    Q-Q Plot")
qqline(adult)
```

And we get 2 graphs:





From those graphs we can find that each sample may come from normal distribution. Because for each graph, its plots are close to the line. However, Children sample is more likely to be the normal distribution because there are only few plots far from the line but for Adult sample, more plots are far from the line so Adult sample may not come from normal distribution. Since we need do some calculate in below questions, so when I do question b, c and d, I treat adult distribution as normal distribution.

<u>Question (b):</u>

I will calculate CI for ratio of two normal variances. Based on the knowledge I have learnt, I will use F-distribution to solve the question. Since the question doesn't show the value of alpha, I take the value of alpha by 0.05.

First I will define the degrees of freedom.

```
n<-length(children)
m<-length(adult)
```

And then I will define S^2_x/S^2_y

```
F<-var(children)/var(adult)
```

Third, I will calculate two special points of the distribution graph by use function “qf”. In this step I should use (n_x-1,n_y-1)degrees of freedom. In F1, p=alpha/2 = 0.025. In F2, p=1-alpha/2 =0.975

```
F1<-qf(0.025,n-1,m-1)
F2<-qf(0.975,n-1,m-1)
```

Finally I will get the domain of CI by doing the calculating below:

```
F*1/F1
F*1/F2
```

The answer is:

```
> F*1/F1
[1] 1.756476
> F*1/F2
[1] 0.3102977
```

Because 1 is in [0.3102977, 1.756476], so that we can assume they are equal.

<u>Question(c):</u>

I will calculate CI for u_x-u_y with two independent samples. So I use the knowledge of T-distribution.

Because from question b we assume two variances are equal, so I choose the function:

$$\bar{X} - \bar{Y} \pm t_{\alpha/2}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

where s_p is the *pooled standard deviation*, a root of the pooled variance in (9.11)

and $t_{\alpha/2}$ is a critical value from T-distribution with $(n + m - 2)$ degrees of freedom

$$s_p^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{n+m-2} = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}.$$

Based on those two functions above, I made the function in R:

```
sp=sqrt(((n-1)*var(children)+(m-1)*var(adult))/(n+m-2))
mean(children)-mean(adult)+c(1,-1)*qt((0.95),45)*sp*sqrt(1/n+1/m)
```

The answer is:

```
[1] 40.21467 33.06766
```

Question (d):

From question(c) the answer is [33.06766, 40.21467], which means the probability of mean(children)-mean(adult) between those two values. Because this domain is bigger than 0, so that I can guess that children's cereals have more sugar on average than adult cereals.