Assignment 1

Kshitij Kavimandan, Pablo Alves, Pooja Mangal (Group 15)

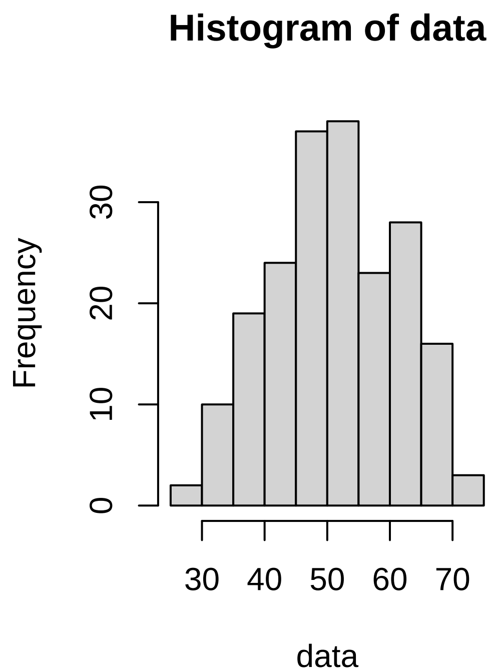
20 February 2024

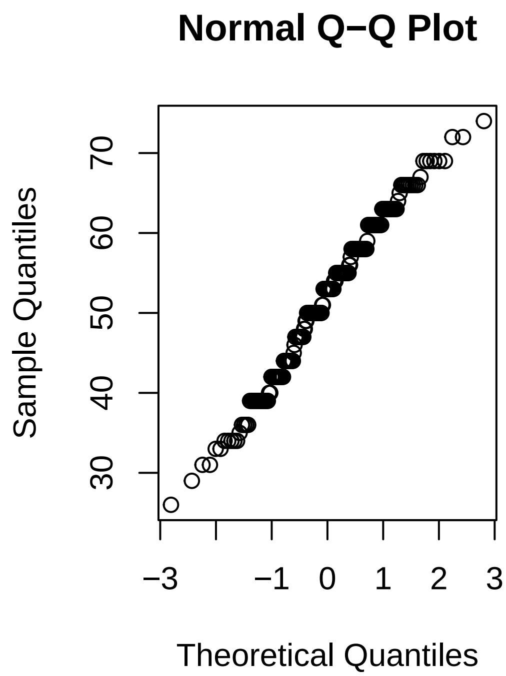
Exercise 1. Ice cream

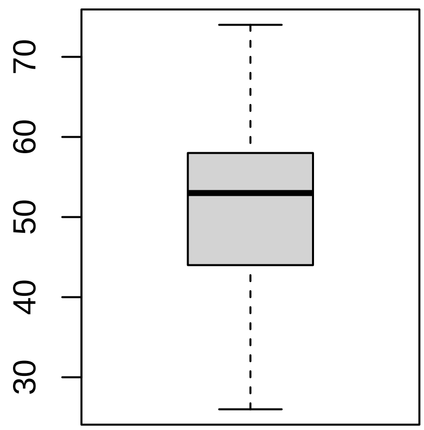
a) Relevant plots and normality assessment:

```r
# Load the dataset
ice_cream_dataset <- read.csv("Ice_cream.csv")
data = ice_cream_dataset$video
library("drcarlate")
```

```r
# Do basic plots
par(mfrow=c(1,2)); hist(data); qqnorm(data); boxplot(data)
```







The histogram of the videogame scores looks normally distributed because it is roughly symetrical around 50, with values getting smaller towards its extremes. Despite this, the histogram bar for values between 60 and 65 is higher than what the corresponding analytical normal distribution should have.

Secondly, the Q-Q Plot also shows that our sample is densely populated in the center and its shape is very close to a line, which is also consistent with normality.

Finally, the boxplot is also consistent with normality, despite its median being not symetrical with its surrounding quantiles. As shown with the histogram, this can be explained with the additional values which are more frequently present in that particular region.

In short, our inspection of these basic plots suggest that this sample is normally distributed.

Because we are assuming by definition that our sample is normally distributed, we can apply the 68-95-99.7% rule. Because our sample has a mean of 51.85 and a standard deviation of 9.9, we can then see that the desired 97% confidence interval for the mean should lie inside the 6 sigma interval, which has a length of 59.9. With this rough estimation, our 97% confidence value for our mean is 51.85 +- 29.7, which is akin to the [22.15,81.55] interval.

To get the 97% CI, we need to evaluate the margin of error with the formula

margin = T(0.97)*(std(x)/sqrt(N))

where T is the t distribution with N-1 degrees of freedom, N our sample size (200)…

```
n <- length(data)
margin <- qt(0.97, df=n-1)*sd(data)/sqrt(n)
min_mean = mean(data) - margin
max_mean = mean(data) + margin
length = (max_mean - min_mean)/2; length
```

```
## [1] 1.324289
```

```
paste("97 CI: ", mean(data), " +-", length)
```

```
## [1] "97 CI:  51.85  +- 1.32428938516609"
```

Thus, the 97% CI for the mean is [50.52571,53.1749], with a length of 2.648579

In order to compute the bounded 97% CI for mu, we need to use the following formula:

```
m_calculator<-function(p, data) sqrt(sd(data)/length(data)*(4*norminv(p+(1-p)/2)))
margin = m_calculator(p=0.97,data=data)
min_mean = mean(data) - margin
max_mean = mean(data) + margin
length = max_mean - min_mean; length
```

```
## [1] 1.311055
```

```
paste("97 CI: ", mean(data)," +-", length)
```

```
## [1] "97 CI:  51.85  +- 1.31105538574369"
```

In order to compute the neccesary number of samples needed for a particular maximun length of the CI we can use the following formula

```
n_calculator <- function(p,length,data) sd(data)*(2*norminv(p+(1-p)/2)/length)^2
samples_needed = n_calculator(p=0.97,length=3,data=data)
paste("For a margin of: ", 3 ," we need", samples_needed)
```

```
## [1] "For a margin of:  3  we need 20.7227503006415"
```

To determine the 97% bootstrap CI we can apply the following code

```
p = 0.97
B = 100

Tstar = numeric(B)
```

```r
for(i in 1:B){Tstar[i]=mean(sample(data,replace=TRUE))}

Tstar985 = quantile(Tstar,0.985)
Tstar15 = quantile(Tstar, 0.015)

c(2*mean(Tstar) - Tstar985,2*mean(Tstar)-Tstar15)
```

```
##    98.5%     1.5%
## 50.11395 53.22650
```

```r
d = ((2*mean(Tstar)-Tstar15) - (2*mean(Tstar) - Tstar985))
paste("Length:",d)
```

```
## [1] "Length: 3.11255"
```

```r
paste("Margin on each side:",d/2)
```

```
## [1] "Margin on each side: 1.556275"
```

Our Bootstrap margin is bigger than our previous margins, which is consistent with the fact that the Bootstrap method does not assume normality, and thus is expected to have a bigger uncertainty on its estimation.

Exercise 2. Hemoglobin in trout

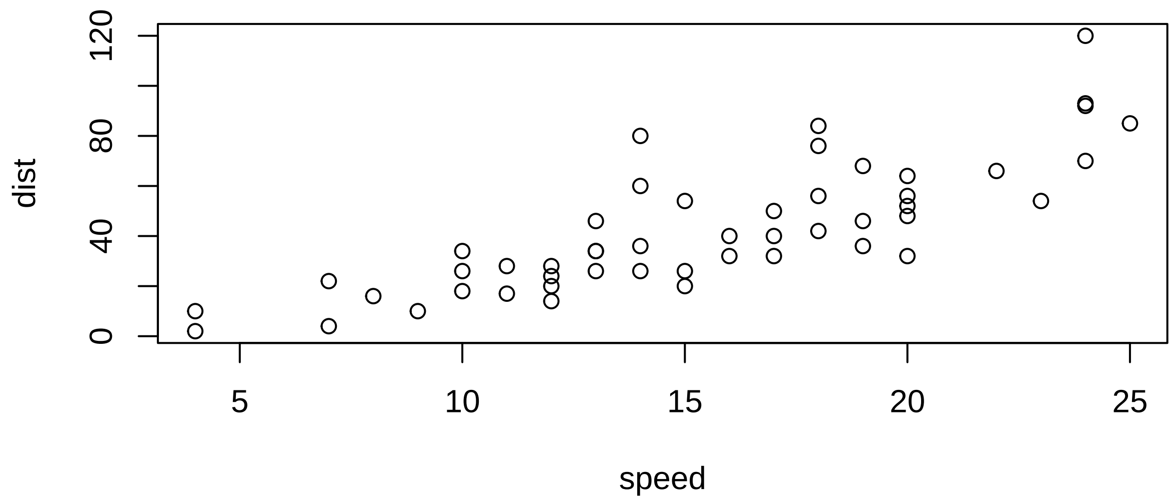
a) R-code for randomization process:

Exercise 3. Sour cream

a) Analyzing the data in a three-way experiment without interactions:

Figures

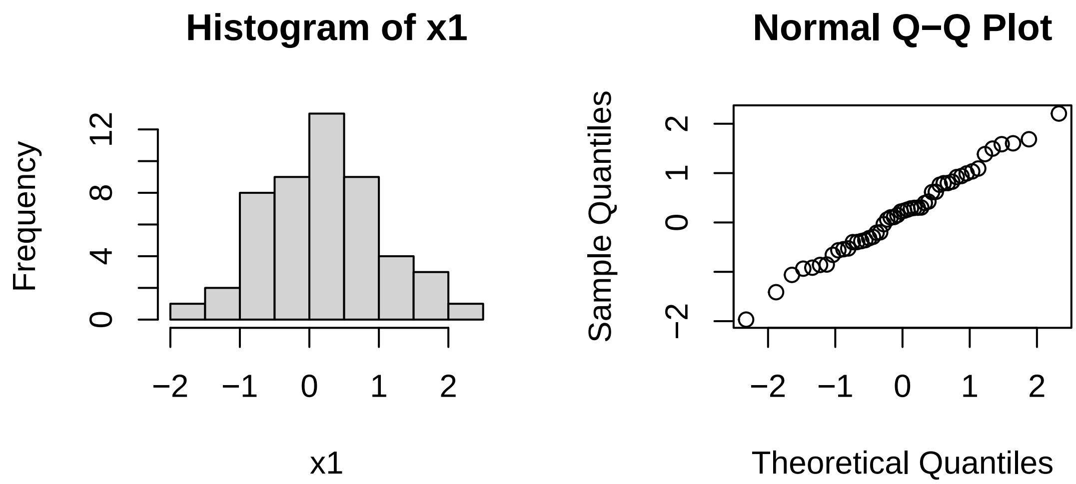
You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot. Use knitr options to style the output of a chunk. Place options in brackets above the chunk. Other options with the defaults are: the `eval=FALSE` option just displays the R code (and does not run it); `warning=TRUE` whether to display warnings; `tidy=TRUE` wraps long code so it does not run off the page.

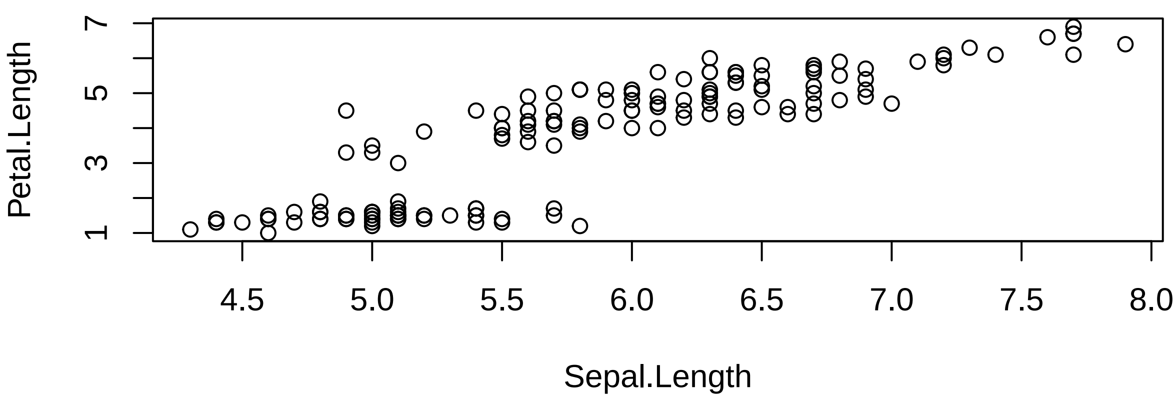
You can control the size and placement of figures. For example, you can put two figures (or more) next to each other. Use `par(mfrow=c(n,m))` to create `n` by `m` plots in one picture in R. You can adjust the proportions of figures by using the `fig.width` and `fig.height` chunk options. These are specified in inches, and will be automatically scaled down to fit within the handout margin. Chunk option `fig.align` takes values `left`, `right`, or `center` (to align figures in the output document).

```
par(mfrow=c(1,2)); x1=rnorm(50); hist(x1); qqnorm(x1)
```



You can arrange for figures to span across the entire page by using the `fig.fullwidth` chunk option.

```
plot(iris$Sepal.Length,iris$Petal.Length,xlab="Sepal.Length",ylab="Petal.Length")
```



More about chunk options can be found at https://yihui.name/knitr/options/.

Equations

To produce mathematical symbols, you can also include LaTeX expessions/equations in your report: inline $\frac{d}{dx}\left(\int_0^x f(u)\,du\right) = f(x)$ and in the display mode:

$$\frac{d}{dx}\left(\int_0^x f(u)\,du\right) = f(x).$$

To be able to use this functionality, LaTeX has to be installed.

Footnotes

Here is the use of a footnote[1].

Images

Want an image? This will do it. To depict an image (say, `my_image.png` which should be in your current working directory), use this command

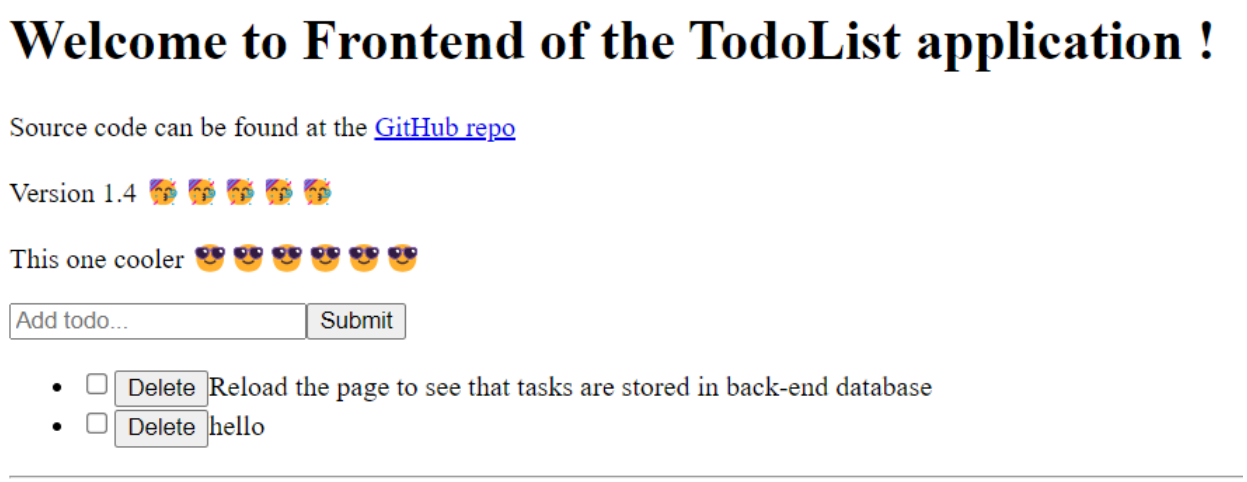


Figure 1: caption for my image

Tables

Want a table? This will create one (note that the separators *do not* have to be aligned).

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

You can also make table by using knit's `kable` function:

Table 2: A knit kable.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1

[1]This is a footnote.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2

Block quote

> This will create a block quote, if you want one.

Verbatim

```
This text is displayed verbatim/preformatted.
```

Links

Links: http://example.com, in-text link to Google.

This is a hyperlink.

This

is where the hyperlink jumps to.

Itimization, italicized and embolded text

- Single asterisks italicize text *like this*.
- Double asterisks embolden text **like this**.

One more way to italicize and embold: *italic* and **bold**.