

Home assignment 1

Load libraries

```
library(tidyverse)
library(ggpubr)
```

1.1 “Hand” Calculations

```
obs <- sort(c(3.9, 5.3, 6.1, 4.9, 9.1, 2.8, 3.5, 3.2, 2.6, 5.9))
n = 10
```

```
mean_obs <- (3.9+5.3+6.1+4.9+9.1+2.8+3.5+3.2+2.6+5.9)/n
```

```
var_calc <- tibble(obs = obs) %>%
  mutate(mean = mean_obs) %>%
  mutate(xi_x2 = (obs - mean)^2) %>%
  mutate(sum_xi_x2 = sum(xi_x2)) %>%
  mutate(var = sum_xi_x2/(n-1))
```

```
var_calc
```

```
## # A tibble: 10 x 5
##   obs mean xi_x2 sum_xi_x2 var
##   <dbl> <dbl> <dbl>     <dbl> <dbl>
## 1  2.6  4.73  4.54       35.5  3.94
## 2  2.8  4.73  3.72       35.5  3.94
## 3  3.2  4.73  2.34       35.5  3.94
## 4  3.5  4.73  1.51       35.5  3.94
## 5  3.9  4.73  0.689      35.5  3.94
## 6  4.9  4.73  0.0289     35.5  3.94
## 7  5.3  4.73  0.325      35.5  3.94
## 8  5.9  4.73  1.37       35.5  3.94
## 9  6.1  4.73  1.88       35.5  3.94
## 10 9.1  4.73 19.1       35.5  3.94
```

```
median_position <- 50*(n+1)/100 # 5.5
```

```
median_interpolation <- (obs[5] + 0.5)*(obs[6] - obs[5])
```

```
print(paste0('a) mean: ', mean_obs, ', variance: ', round(var_calc$var[1],2)))
```

```
## [1] "a) mean: 4.73, variance: 3.94"
```

```
print(paste0('b) median: ', median_interpolation))
```

```
## [1] "b) median: 4.4"
```

1.2 Computer Exercise

a) Import dataset

```
cordblood <- read_delim("Data/cordblood.txt", locale = locale(encoding = "latin1"))
head(cordblood)
```

```
## # A tibble: 6 x 7
##   Hospital   Age Sex   Measles Parotitis Rubella Chickenpox
##   <chr>     <dbl> <chr>   <dbl>     <dbl>   <dbl>      <dbl>
## 1 Mölndal    27 male     247       231    119.       1751
## 2 Mölndal    37 female  3291       231    187.       882
## 3 Mölndal    30 female   422       289     67.3       518
## 4 Mölndal    37 female 12946      8346     79.6      2911
## 5 Mölndal    29 female  1164      1235     75.5      1802
## 6 Borås      23 female  1875      2212     49.3      2070
```

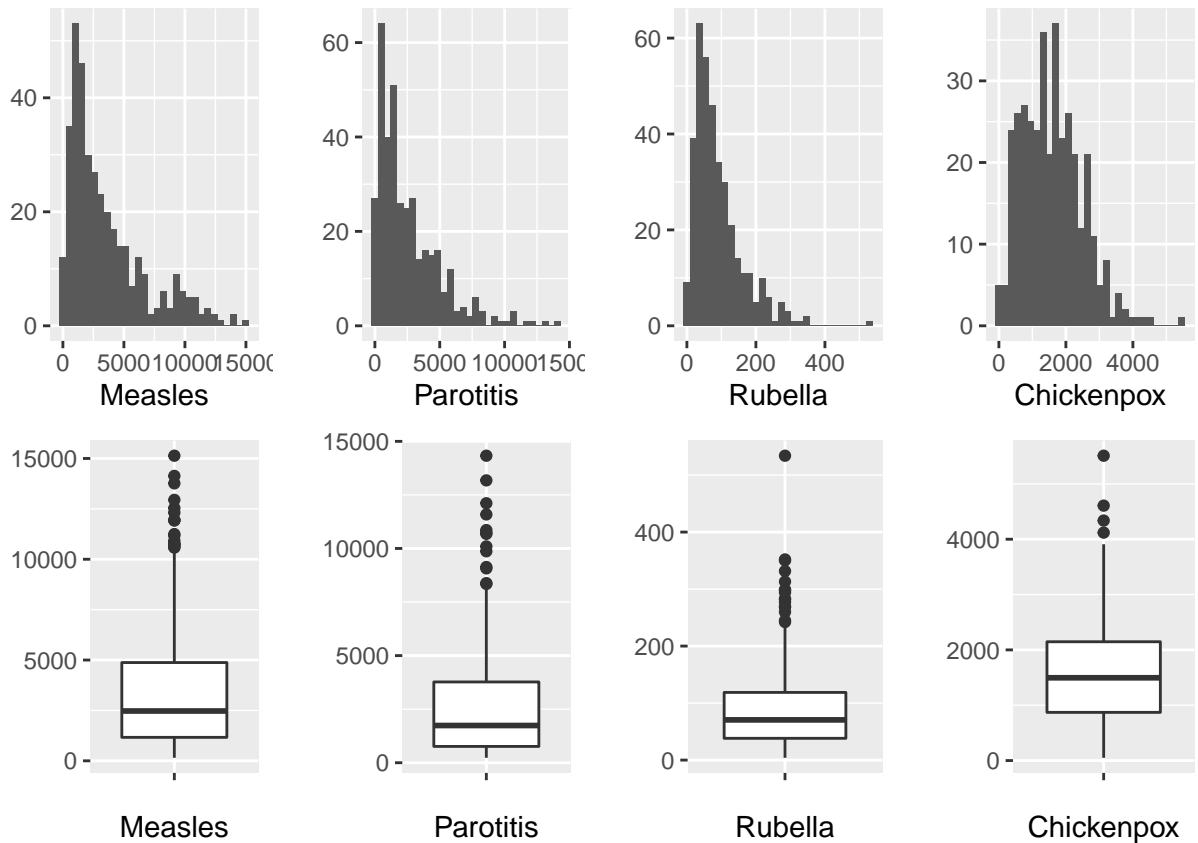
b) Graphs and computations

1 Histograms and boxplots

```
create_hist <- function(data, name) {
  return(
    ggplot(cordblood, aes(data)) +
    geom_histogram() +
    labs(x = name, y = '')
  )
}

create_bp <- function(data, name) {
  return(
    ggplot(cordblood, aes(x = '', data)) +
    geom_boxplot() +
    labs(x = name, y = '')
  )
}

var_names <- c('Measles', 'Parotitis', 'Rubella', 'Chickenpox')
ggarrange(create_hist(cordblood$Measles, var_names[1]),
  create_hist(cordblood$Parotitis, var_names[2]),
  create_hist(cordblood$Rubella, var_names[3]),
  create_hist(cordblood$Chickenpox, var_names[4]),
  create_bp(cordblood$Measles, var_names[1]),
  create_bp(cordblood$Parotitis, var_names[2]),
  create_bp(cordblood$Rubella, var_names[3]),
  create_bp(cordblood$Chickenpox, var_names[4]),
  ncol = 4, nrow = 2)
```



2 Mean, median, variance, interquartile

Comment:

```
calc_values <- function(data, name) {
  df <- tibble(name = name,
               mean = mean(data),
               median = median(data),
               variance = var(data),
               iqr = IQR(data)
             )
  return(df)
}

res_measles <- calc_values(cordblood$Measles, "Measles")
res_parotitis <- calc_values(cordblood$Parotitis, "Parotitis")
res_rubella <- calc_values(cordblood$Rubella, "Rubella")
res_chickenpox <- calc_values(cordblood$Chickenpox, "Chickenpox")

bind_rows(res_measles, res_parotitis, res_rubella, res_chickenpox)
```

```
## # A tibble: 4 x 5
##   name      mean median variance    iqr
```

##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	Measles	3545.	2471	9789271.	3711
## 2	Parotitis	2578.	1737	5999383.	3009
## 3	Rubella	90.1	70.5	5196.	80.5
## 4	Chickenpox	1587.	1496	796967.	1276