---

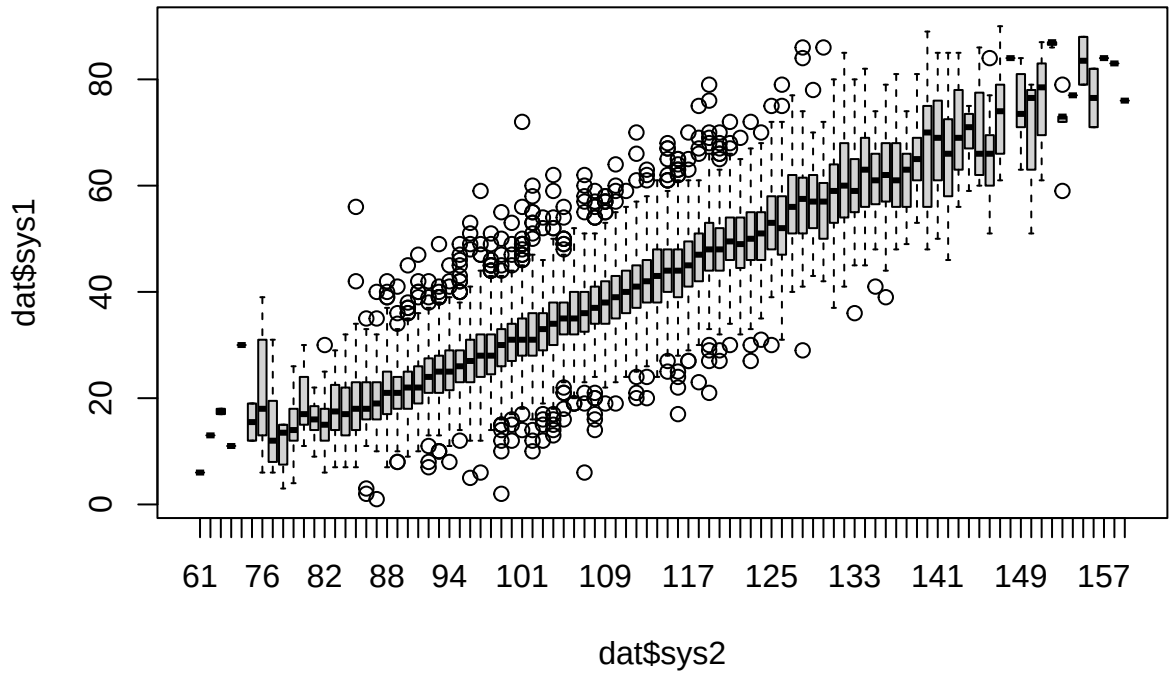
title: “Homework 6” author: “Tauqeer Kasam Rumaney” date: “2022-11-28” output: html_document

Exercise 1: Statistical tests

a) In the KiGGS dataset, perform a 2-sample t-test for paired samples (i.e. dependent samples) for the two variables ‘sys1’ and ‘sys2’. Interpret the results. # We can’t do dependent test since arguments don’t have the same length but if we do indepndent test we get a vaalue, for which null hypothesis holds true.

b) In the KiGGS dataset, select one metric and one binary variable (or create one) and perform a 2-sample t-test. Then do a Mann-Whitney U-test with the same variables, explore the function to do this, and compare the results to the t-test. #Even though the p-values are different for both tests, its enough to deduce that the null hypothesis is wrong and thus discarded.

```
dat_link <- url("https://www.dropbox.com/s/pd0z829pv2otzqt/KiGGS03_06.RData?dl=1")
load(dat_link)
dat <- KiGGS03_06
boxplot(dat$sys1~dat$sys2)
```




```
#t.test(dat$sys1~dat$sys2, mu=0, paired=TRUE, var.equal=FALSE, alternative="two.sided", conf.level=0.95
#Error in t.test.formula(dat$sys1 ~ dat$sys2, mu = 0, paired = FALSE, var.equal = FALSE,  : grouping fa

abc<-na.omit(dat$sys1)
def<-na.omit(dat$sys2)

#shapiro.test(as.numeric(abc))
#Error in shapiro.test(as.numeric(abc)) : sample size must be between 3 and 5000

#t.test(as.numeric(abc), as.numeric(def), alternative = "two.sided", mu = 0  ,paired = TRUE, var.equal
#Error in complete.cases(x, y) : not all arguments have the same length

t.test(as.numeric(abc),as.numeric(def),alternative = "two.sided", mu = 0, paired = FALSE, var.equal = F

##
```

```
##  Welch Two Sample t-test
##
## data:  as.numeric(abc) and as.numeric(def)
## t = -8.0336, df = 29251, p-value = 9.818e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.4092075 -0.8564348
## sample estimates:
## mean of x mean of y
##  36.20795  37.34077
```

```
#This is for independent test
```

```
summary(dat$PPoint)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   43.00   84.00   84.19  126.00  167.00
```

```
summary(dat$OW)
```

```
##   Ost  West
##  5899 11741
```

```
result <- wilcox.test(dat$PPoint~dat$OW,exact = FALSE)
result
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dat$PPoint by dat$OW
## W = 34935901, p-value = 0.3378
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(dat$PPoint~dat$OW)
```

```
##
##  Welch Two Sample t-test
##
## data:  dat$PPoint by dat$OW
## t = 1.0338, df = 12982, p-value = 0.3012
## alternative hypothesis: true difference in means between group Ost and group West is not equal to 0
## 95 percent confidence interval:
##  -0.6859978  2.2172422
## sample estimates:
##  mean in group Ost mean in group West
##           84.69876           83.93314
```

Exercise 2: Study planning (optional)

Compute the sample size using R or G*Power, for a study that investigates the question if biking to the HPI is associated with concentration in class or not.

Think about how you can formulate this study question in terms of variables, and which statistical test (that we have covered in class 6) would be appropriate for this.

a) Which test could you use?

b) Which sample size is necessary, to find a true effect with 80% power at $\alpha = 0.05$ using this statistical test?