This is the sixth homework assignment. Students should tick in TUWEL problems they have solved and upload their detailed solutions by 20:00 on Monday November 20, 2023.

---

1. **Histogram**
   Set `k <- 100` and generate

```
x <- rnorm(sample(k:(2*k),1), runif(1,0,k), rexp(1,1/k))}.
```

   (a) Explain what is realized in `x`.

   (b) Plot a histogram of `x`. Mark its mean in red, its standard deviation in blue and add a legend which explains them both.
   Helpful R-commands: `hist()`, `mean()`, `sd()`, `lines()`, `abline()`, `arrows()`, `legend()`.

2. **Boxplot**
   Two novel randomized algorithms (A and B) are to be compared regarding their running time. Both algorithms were executed n times. The running times (in seconds) are stored in the file `algorithms.Rdata`.

   (a) Set the working directory and load the data using `load()`. Create a boxplot to compare the running times. Color the boxes and add proper notations (axes notations, title etc.). More info via `?boxplot`.

   (b) Comment on the following statements/questions only using the boxplot.
   - The first quartile of the times in A was about?
   - The interquartile range of the times in B is about trice the interquartile range of A.
   - Is the sample size $n = 100$?
   - More than half of the running times in B were faster than 3/4 of the running times in A.
   - At least 50% in A were faster than the 25% slowest in B.

3. **Dataset `trees`**
   `trees` is the `R` dataset package containing diameter, height and volume for 31 black cherry trees. Denote by h the vector that contains the values of the column `Height` and by v the vector that contains the values of the column `Volume` from the dataset `trees`.

   (a) Make the frequency table of the height of black cherry trees.
   The command `table` is useful.

   (b) Compute the sample means and the sample variances for both vectors.

   (c) Compute the covariance and the correlation of these two vectors.
   Use `qqplot` to produce quantile-quantile plot of the two datasets. More info via `?qqplot`.

   (d) Explain the outputs of the commands `fivenum` and `summary`.

4. **Normal distributions and neighborhoods**
Let X be a normal random variable with the expectation 5 and the variance 4. Let $X_1, X_2, \ldots, X_{50}$ be a random sample from this distribution. Consider two statistics, the sample sum S and the sample mean $\bar{X}$, given by

$$S = X_1 + \cdots + X_{50} \qquad \text{and} \qquad \bar{X} = \frac{1}{50}(X_1 + \cdots + X_{50}).$$

   (a) What are the distributions of S and of $\bar{X}$?

   (b) Use R-command `plot` to plot the density of X and the density of $\bar{X}$.

   (c) Which quantiles of the normal distribution $\mathcal{N}(5, 4)$ mark the neighborhoods of the mean that contain 95%, and 99.9% of the probability density? Which values do they take? Add these neighborhoods to the plot. Add the corresponding neighborhoods to the plot of the density of the sample mean.

   (d) Generate a sample of 50 random numbers from $\mathcal{N}(5, 2^2)$ and plot the histogram for this sample.

5. **The CLT - Simulation approach**
We illustrate the Central Limit Theorem for a random sample from exp(1)-distribution by applying the following R-code. Explain the commands used and comment on the result.

```
for(n in c(1,2,3,5,10,25,50,100, 250, 500, 1000, 2000)){
result<-c()
mu=1
sigma=1
for (i in 1:5000){
X<-rexp(n, 1/mu)
result[i] <- (mean(X)-mu)/sigma*sqrt(n)
}
hist<- hist(result, breaks=seq(min(result)-1, max(result)+1,by=0.25), plot=
   FALSE)
ylim<-range(hist$density,dnorm(0))
hist(result, breaks=seq(min(result)-1, max(result)+1,by=0.25), prob=TRUE,
   ylim=ylim, main=paste("n=",n),col="lightblue")
x<-seq(-4,4,by=0.1)
lines(x,dnorm(x),lty=1,lwd=2, col="lightpink")
Sys.sleep(1.5)
}
```