This is the eight homework assignment. Students should tick in TUWEL problems they have solved and upload their detailed solutions by 20:00 on Monday December 4, 2023.

1. **One-sample t-test (without R)**
   In the context of a one-sample t-test let $\bar{x} = 2.05$, $s^2 = 4$, $n = 16$ and $H_0 : \mu = 1$. The values of the distribution function F of the $t(n)$-distribution are

t	-2.300	-2.200	-2.100	-2.000	-1.900	-1.800
$F(t)$ (for $n = 15$)	0.018	0.022	0.027	0.032	0.038	0.046
$F(t)$ (for $n = 16$)	0.018	0.021	0.026	0.031	0.038	0.045

   Using only the upper information, what is your conclusion for the following cases:

   (a) at a right-sided test at level 5%
   (b) at a right-sided test at level 1%?
   (c) at a left-sided test at level 5%?
   (d) at a two-sided test at level 5%?

   Briefly explain your choice.

2. **One-sample t-test (with R)**
   Two programmers are developing a computer game in which a BMX rider has to pass a parcours. The goal is to cover as much distance as possible. They are interested in the mean distance covered by a layman, who plays the game for the first time. For that they let some trialists play the game and for each player note the distance covered. The distances (unit meter) are found in the file `dist.Rdata`.

   Test the null hypothesis, that the mean distance covered by a layman is 550 meters using a two-sided one-sample t-test. Set the significance level to $\alpha = 5\%$.

   Proceed as follows:

   (a) Create a histogram to visualize the data. Are the data approximately bell-shaped? Mark the null hypothesis, the mean and the standard error of the mean (i.e., mean $\mp$ one standard error of the mean). Add labels and legends.
   (b) Calculate the t-statistic (without using `t.test()`)
   (c) Calculate the p-value (without `t.test()`). Do you reject the null hypothesis?
   (d) Interpret your result.
   (e) Now perform the test using `t.test()` and compare your results.

3. **Effect of sample size**
   Consider the situation of the previous exercise except that three times the number of laymen played the game (for simplicity, literally repeat each sample 3 times). Call this new data set `dist3x`.

   (a) Perform a t-test. What is your conclusion?

   (b) Represent the data from `dist.Rdata` as well as `dist3x` each in a histogram, arranged below each other (`par(mfrow=c(2,1))`). Mark the mean, the 1se interval around the mean, as well as the value 550 meters.

   (c) Discuss your graphic regarding the outcomes of the tests.

4. **One-sample confidence intervals**

   (a) Let $I = (\bar{X} + q_{\alpha/2} \cdot \text{SEM}, \bar{X} + q_{(1-\alpha/2)} \cdot \text{SEM})$ be Student's $(1-\alpha) \cdot 100\%$-confidence interval for the expectation, while q_α is the α-quantile of the $t(n)$-distribution (for $\alpha \in (0,1)$). Comment on the following statements:

      i. At level $\alpha = 10\%$ the confidence interval I is smaller than at level $\alpha = 5\%$
      ii. The sample size is $n-1$
      iii. If the sample size is doubled, then the width of the confidence interval is halved
      iv. The probability to find the expectation left of I is at the most α
      v. If I overlaps the expectation, then the null hypothesis is not significant.

   (b) Messages are regularly sent from a sender to a receiver. For randomly chosen messages the transfer times were measured and stored in the file `waitingtimes.Rdata`.

      i. Represent the data in a histogram. Is the distribution approximately bell-shaped?
      ii. Construct an approximate 99%-confidence interval for the expectation and plot it into the graphic.
      iii. An engineer claims that the mean transfer time is 1.5 seconds. Are the measurements compatible with this statement?

5. **Confidence interval**
   In the June 1986 issue of Consumer Reports, some data on the calorie content of beef hot dogs is given. Here are the numbers of calories in 20 different hot dog brands:

   $$186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132.$$

   Assume that the numbers are from a normal distribution with mean μ and variance σ^2, both unknown. Use `R` to obtain a 90% confidence interval for the mean number of calories μ.