# Homework #10

## Problem #1 (just a continuation of last HW's problem)

We will continue with the *Boston* housing data set, from the $MASS$ library.

(c) Based on this data set, provide an estimate, $\hat{\mu}_{med}$, for the median value of *medv* in the population.

(d) We now would like to estimate the standard error of $\hat{\mu}_{med}$. Unfortunately, there is no simple formula for computing the standard error of the median. Instead, estimate the standard error of the median using the bootstrap.

(e) Proceed to use the debiasing approach from class in order to calculate the 95% bootstrap confidence interval for population median, $\mu_{med}$. Interpret the interval.

(f) Based on this data set, provide an estimate for the 0.9-quantile of *medv* in Boston suburbs. Call this quantity $\hat{\mu}_{0.9}$.

(g) Use the bootstrap to estimate the distribution of $\hat{\mu}_{0.9}$. Plot its histogram. Afterwards, proceed to use the debiasing approach from class in order to calculate the 95% bootstrap confidence interval of population 0.9-quantile, $\mu_{0.9}$. Interpret the interval.

## Problem #2

Write your own one-sample $t$-test function for the " $\neq$ " alternative ($H_0 : \mu = \mu_0 \ vs \ H_a : \mu \neq \mu_0$), name it $my.t.test()$, for example. It should execute steps of the testing procedure I outlined in the class. Run your function and $R$'s $t.test()$ function on the same input, compare the outputs. **You can not utilize function** $t.test()$ **inside your own** $my.t.test()$ **function definition**.

Make sure your function takes in just those three parameters as input:

- data vector of quantitative values,
- hypothesized mean value,
- confidence level,

and returns a list with following elements:

- t (**TS value**, has to equal the "$t =$ " from the $t.test()$ call output for same data)
- df (**degrees of freedom**, has to equal the "$df =$ " from the $t.test()$ call output)
- p.value (has to equal the "$p\text{-}value =$ " from the $t.test()$ call output)
- CI (has to be a 2-element vector: first element = first value under "95 percent CI" of $t.test()$ output, second element = second value)
- sample.estimate (just the sample mean, has to equal to the value under 'mean of x' in $t.test()$ output)

Make sure to provide one example for classical confidence level (0.95), and one - for non-typical level (e.g. 0.99, or 0.90). Some code to guide you:

```r
my.t.test <- function(vec, mu, ci.lvl=0.95){
  # Here the floor is yours.
}
my.t.test(c(1:10),mu=5) # Exemplary call and output for your function.
$t
[1] 0.522233
$df
[1] 9
$p.value
[1] 0.6141173
$CI
[1] 3.334149 7.665851
$sample.estimates
[1] 5.5

t.test(c(1:10), mu=5) # Comparative t.test() call.

my.t.test(c(1:10), mu=5, ci.lvl=0.99) # Example for non-typical CI level.

t.test(c(1:10), mu=5, conf.level=0.99) # Compare output with similar t.test() call.
```

## Problem #3

This question deals with data on working hours per week.

a. Conduct a $t$-test on finding out whether the US population average of weekly working hours differs from the standard of $40hrs$. Make sure to formulate the hypotheses (in parameter notation), report both the $p$-value and (Cohen's) effect size. Interpret the result. Does the confidence interval agree with hypothesis test? Explain.
b. Conduct a $t$-test on finding out whether the **male** population average of weekly working hours differs from the standard of $40hrs$. Make sure to formulate the hypotheses (in parameter notation), report both the $p$-value and (Cohen's) effect size. Interpret the result. Does the confidence interval agree with hypothesis test? Does confidence interval also allow to judge the practical significance? Explain.
c. For parts (a) and (b) check if:

    - There are any strong outliers.
    - Given that, should we be concerned about the legitimacy of t-test results? Why/why not?

## Problem #4 (Make sure to use $R$ when appropriate for carrying out the calculations, showing your work)

**9.29**

**9.47**

**9.49**

**9.52**

**9.54 (In addition, provide the Cohen's d value)**