

# Homework 1

[ECON 4753] — *University of Arkansas*

## Review Questions

### *Question 1*

- (a) What does  $\sum_{i=1}^5 (i - 3)$  equal?
- (b) Calculate the sample mean,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , where the sample observations are  $x = (2, 7, 10, 6, 8)$

### *Question 2*

This question is based on our review of statistics. Say you observe a sample of workers from a firm with sample size  $n = 100$ . You observe their wages  $w_i$  and want to estimate the average wage at the firm. You estimate the following statistics in your sample:  $\bar{w} = 17.53$  and  $\text{var}(w) = 4.2$ .

- (a) Given this information what is the (approximate) sample distribution of the sample mean?
- (b) Form a 95% confidence interval for your sample mean. Interpret this in words.
- (c) Another student claims the average worker earns \$17. Using your confidence interval, would you reject this null with a 5% significance level?

## R Coding

This assignment will explore a sample of homes in Boston suburbs. It comes from the paper [Hedonic housing prices and the demand for clean air](#) which tries to estimate how much people are willing to pay to live in homes with cleaner air.

To use this dataset, use the function `read.csv` with the url [https://raw.githubusercontent.com/kylebutts/UARK\\_4753/main/Homework/HW1/data/housing\\_df.csv](https://raw.githubusercontent.com/kylebutts/UARK_4753/main/Homework/HW1/data/housing_df.csv). Remember that you need to include the code that loads the dataset into your R Markdown file.

This data set has the following variables:

Variable	Info
MEDV	Median value of owner-occupied homes in \$1000's
CRIM	Per capita crime rate by town
ZN	Proportion of residential land zoned for lots over 25,000 sq. ft.
INDUS	Proportion of non-retail business acres per town
CHAS	Charles River indicator (=1 if census tract touches river)
NOX	Nitric oxides concentration (parts per 10 million)
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	weighted distances to five Boston employment centres
RAD	index of accessibility to radial highways
TAX	Full-value property-tax rate per \$10,000
PTRATIO	Pupil-teacher ratio by town
B	Formula involving % Black
LSTAT	% lower status of the population

### Question 1

Look in the Environment panel of R studio, how many variables and how many observations are in this data set?

### Question 2

Each observation is a town in the suburbs or Boston. First, we want to get a sense of the distribution of the *median* value of owner-occupied homes of towns in the Boston suburb at this time.

Use the function `mean()` and `sd()` to find the average median value of owner-occupied homes in \$1000's. Report the R code and number to two digits.

### *Question 3*

To practice with R coding. Use the functions `sqrt()`, `sum()`, `^2`, `length()`, and `mean()`, but not `var()` or `sd()`, calculate the sample standard deviation median value of owner-occupied homes in \$1000's. Report the R code and number to two digits.

### *Question 4*

Use the `hist()` function, create a histogram of NOX pollution. Give this graph a nice title.

### *Question 5*

Now with a sense of the distribution of the two variables of interest. Let's make a scatter plot of MEDV on the x-axis and NOX on the y-axis. You can use the `plot(x = _____, y = _____)` function for this. Include the plot in the output.

For this question, we are going to practice making high-quality reports. When presenting our work to stakeholders, we want it to look good.

- The stakeholders want readable axis titles. Use the function arguments `xlab = ""`, and `ylab = ""` to improve the axis labels.
- Our figure needs a title. Use the chunk option `fig.cap = ""` to describe the relationship between neighborhood NOX levels and home prices.
- The stakeholders do not care about the code used to generate the figure. Let's hide it for this question using the chunk option `echo = FALSE`