# ECON 3640–001

## Problem Set 4

**Marcio Santetti**

Spring 2022

# Problem 1

Using the `AmesHousing` package, do the following:

```
library(tidyverse)
library(AmesHousing)
library(janitor)


ames_data ← ames_raw

ames_data ← ames_data %>%
  clean_names()
```

(a) Select the `lot_area` column, showing data on lot sizes (in square feet). Draw its histogram.

(b) Given that this data set has information on **all** houses in Ames, Iowa between 2006 and 2010, we may assume that it brings the whole *population* data. Calculate the average lot size ($\mu$) for this population.

(c) Now, run a sampling procedure (using a `for` loop), extracting 5,000 samples of size `n = 50` from the population data on lot areas. Compute the sample means $\bar{x}$ of these samples and store them in an array called `samples_50`.

(d) Given that in part (c) you've computed a summary statistic from a population, plot a histogram of the sampling distribution of these $\bar{x}$ values. **Hint**: transform your `samples_50` array in a `tibble` first.

(e) Now, run the same procedure as in part (c), this time increasing the sample size to `n = 1,000`. Plot a histogram reflecting the sampling distribution of $\bar{x}$.

(f) Compare the population mean ($\mu$) with the means of the two sampling distributions for $\bar{x}$ you've computed in parts (c) and (e).

(g) Compare the population variance ($\sigma^2$) with the variances of the two sampling distributions for $\bar{x}$ you've computed in parts (c) and (e).

(h) Run the same procedure as you've done in parts (c)—(g), this time with a different summary statistic: the population and sample **median**.

## Problem 2

Answer the following questions using the `qnorm()` and `rnorm()` functions in `R`:

(a) What is the total area to the **left** of 1 for a random variable $X \sim \mathcal{N}(2, 1)$? In other words, what is *P(X < 1)*?

(b) What is the quantile corresponding to a probability of 85% to its **left** for a random variable $X \sim \mathcal{N}(3, 1)$?

(c) For a Standard Normal distribution, what **quantile** corresponds to a probability of 2.5% to its left?

(d) For a Standard Normal distribution, what **quantile** corresponds to a probability of 5% to its left?

(e) For a Standard Normal distribution, what **quantile** corresponds to a probability of .5% to its left?

## Problem 3

A statistician took a random sample of 50 observations from a population with a population standard deviation ($\sigma$) of 25 and computed the sample mean to be 100.

(a) Estimate the population mean with 90% confidence.

(b) Repeat part (a) using a 95% confidence level.

(c) Repeat part (a) using a 99% confidence level.

(d) Describe the effect on the confidence interval estimate of increasing the confidence level.

## Problem 4

The mean of a random sample of 25 observations from a normal population with a standard deviation ($\sigma$) of 50 is 200.

(a) Estimate the population mean with 95% confidence.

(b) Repeat part (a) changing the population standard deviation to 25.

(c) Repeat part (a) changing the population standard deviation to 10.

(d) Describe what happens to the confidence interval estimate when the standard deviation is decreased.

## Problem 5

A random sample of 25 was drawn from a normal distribution with a standard deviation ($\sigma$) of 5. The sample mean is 80.

(a) Determine the 95% confidence interval estimate of the population mean.

(b) Repeat part (a) with a sample size of 100.

(c) Repeat part (a) with a sample size of 400.

(d) Describe what happens to the confidence interval estimate when the sample size increases.

## Problem 6

The following data represent a random sample of 9 marks (out of 10) on a statistics quiz. Estimate the population mean with 90% confidence.

$$7\ 9\ 7\ 5\ 4\ 8\ 3\ 10\ 9$$

## Problem 7

For the following parts, calculate the value of the *test statistic*, set up the *rejection region*, determine the *p-value* and *interpret* the result.

(a) $H_0$: $\mu = 1,000$; $H_a$: $\mu \neq 1,000$

$\sigma = 200$, $n = 100$, $\bar{x} = 980$, $\alpha = 0.05$

(b) $H_0$: $\mu = 50$; $H_a$: $\mu > 50$

$\sigma = 5$, $n = 9$, $\bar{x} = 51$, $\alpha = 0.03$

## Problem 8

Many Alpine ski centers base their projections of revenues and profits on the assumption that the average Alpine skier skis exactly four times per year. To investigate the validity of this assumption, a random sample of 63 skiers is drawn and each is asked to report the number of times he or she skied the previous year. If we assume that the standard deviation ($\sigma$) is 2, can we infer at the 10% significance level that the assumption is wrong? Assume the sample mean to be 5.3 times per year.

## Problem 9

University bookstores order books that instructors adopt for their courses. The number of copies ordered matches the projected demand. However, at the end of the semester, the bookstore has too many copies on hand and must return them to the publisher. A bookstore has a policy that the proportion of books returned should be kept as small as possible. The average is supposed to be less than 10%. To see whether the policy is working, a random sample of book titles was drawn, and the fraction of the total originally ordered that are returned is recorded and listed here. Can we infer at the 10% significance level that the mean proportion of returns is less than 10%?

4 15 11 7 5 9 4 3 5 8

## Problem 10

Companies that sell groceries over the Internet are called e-grocers. Customers enter their orders, pay by credit card, and receive delivery by truck. A potential e-grocer analyzed the market and determined that the average order would have to exceed $85 if the e-grocer were to be profitable. To determine whether an e-grocery would be profitable in one large city, she offered the service and recorded the size of the order for a random sample of customers. Can we infer from these data that an e-grocery will be profitable in this city?