Statistical Literacy

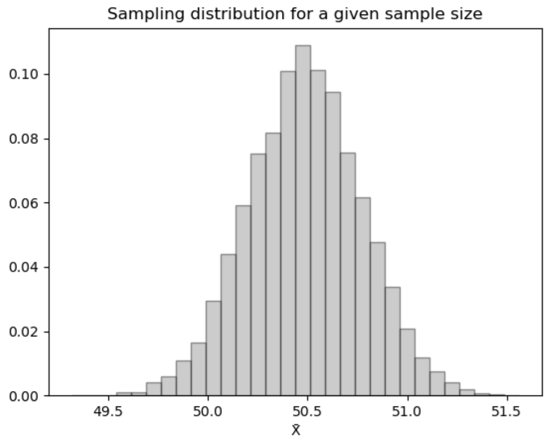
Problem Set 6

Rémi Viné
Due November 6th, 2023

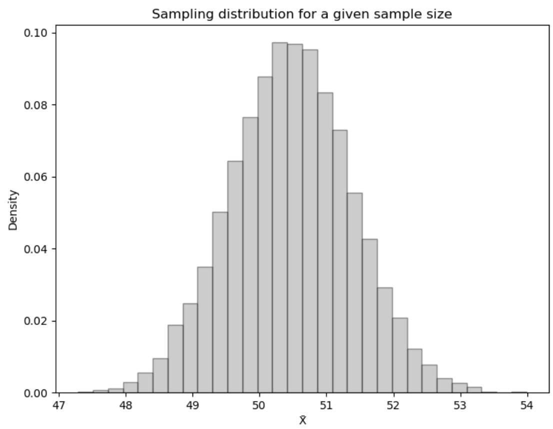
1. **Find population and sampling distributions**

   (a) Among the two sampling distributions depicted below, which sampling distribution has the largest sample size? Justify.

(a)

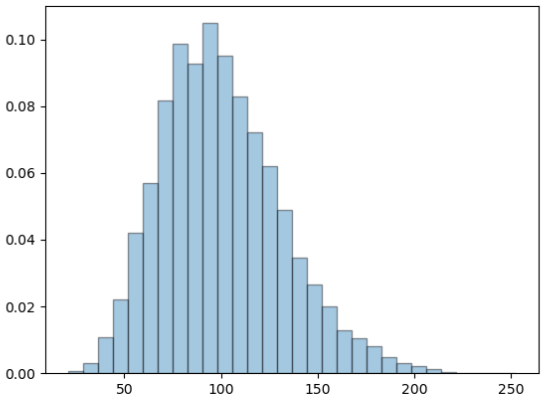


(b)

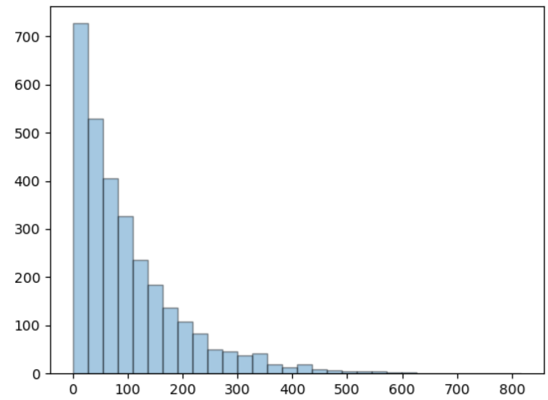


(b) Using the four graphs below, find (i) the population distribution, (ii) the sampling distribution using $n = 100$, (iii) the sampling distribution using $n = 10$, (iv) the sampling distribution using $n = 3$. Note that the sampling distributions here are built from the population distribution, which is one of the four graphs. Justify.

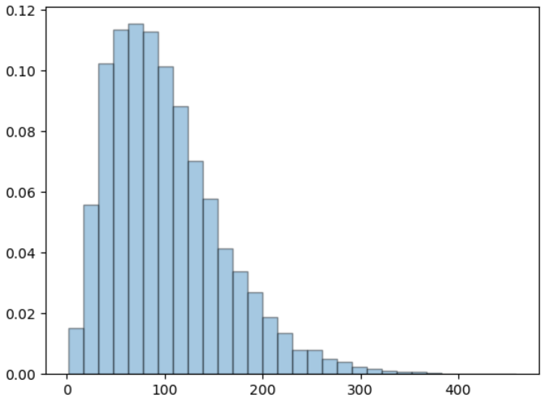
(a)



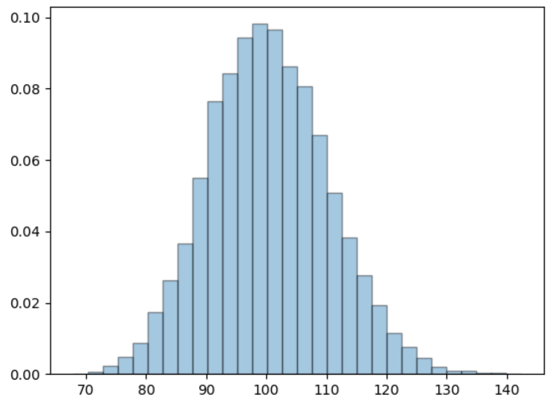
(b)



(c)

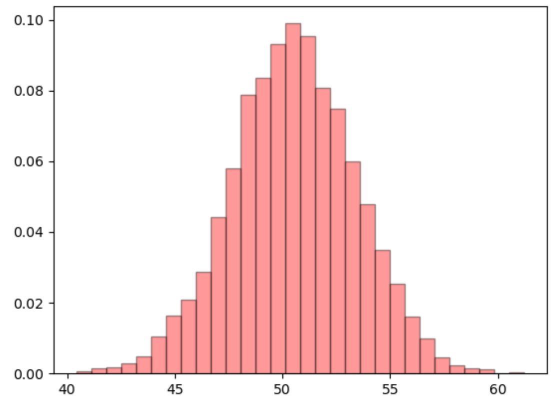


(d)

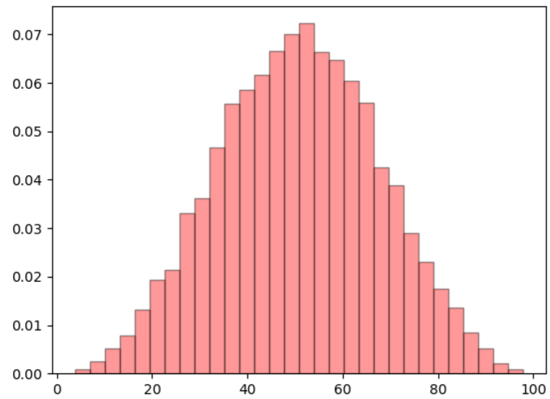


(c) Same exercise for these 4 plots:

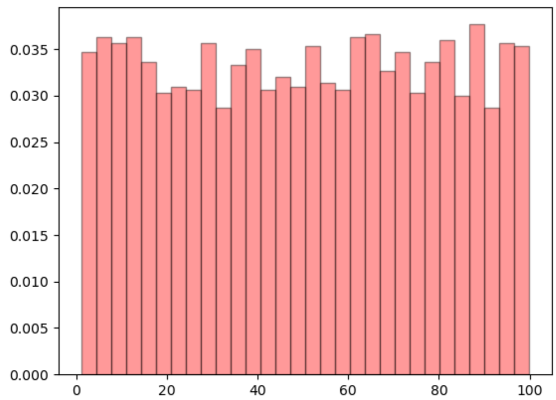
(a)



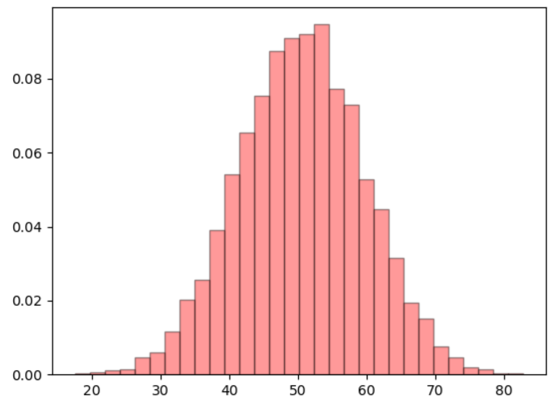
(b)



(c)



(d)



2. **Understanding the foundations.**

   (a) Assume the population is composed of 4 observations, where *i can be* $1, 2, 3, 4,$ *and attached values are* x_1, x_2, x_3, x_4, how many different samples (one observation can appear only once within a sample, and it does not matter whether, for example, x_1 was picked first or picked last in the sample) of size $n = 3$ can be constructed?

   (b) With the same population, how many sample(s) with $n = 4$ can be built?

   (c) With the same population, how many sample(s) with $n = 1$ can be built?

   (d) With the same population, how many sample(s) with $n = 2$ can be built?

   (e) *Difficult.*[1] How many samples of size $n = 30$ can be constructed in a population of $N = 2000$?

   (f) You read the following statement: "By virtue of the Central Limit Theorem, all population distributions are Normal as long as the sample used is large enough ($n > \ or \geq 30$)". Do you agree with it? Comment.

   (g) You read the following statement: "By virtue of the Central Limit Theorem, all sample averages are identical ($\bar{x_1} = \bar{x_2} = ... = \bar{x_m}$) as long as the sample used is large enough ($n > \ or \geq 30$)". Do you agree with it? Comment.

   (h) You read the following statement: "Constructing a sampling distribution, the expected value of the sampling distribution is identical to the average of the population distribution *regardless of the sample size*". Do you agree with it? Comment.

   (i) You read the following statement: "Constructing a sampling distribution, the standard deviation of the sampling distribution (also called the standard error) is identical to the standard deviation of the population distribution *regardless of the sample size*". Do you agree with it? Comment.

   (j) You read the following statement: "Constructing a sampling distribution, the standard error decreases when the sample size increases". Do you agree with it? Comment.

   (k) You read the following statement: "Constructing a sampling distribution from a Normally distributed population distribution, then the sampling distribution is Normal too, even when $n < 30$". Do you agree with it? Comment.

[1] Here, you would need to use the *Binomial coefficient*. As any questions labelled "Difficult", no need to solve it, especially if you have never heard about the binomial coefficient.

3. **Population and Sampling distributions.**

   Assume damages from a storm is modelled by an insurance company as following a Normal distribution with an average of 10000 Swiss Francs and a standard deviation of 2000 Swiss Francs. Assume the insurance company wants to better understand some aspects of the damages and conducts a study using a sample of 100 households.

   (a) Picking one household *in the population*, what is the probability to select a household where damages exceeded 13700 Swiss Francs?

   (b) Picking one household *in the population*, what is the probability to select a household where damages exceeded 10500 Swiss Francs?

   (c) The insurance company wants to know what is the damage value beyond which there are the 20% of the most affected households.

   (d) **Now, let's focus on the sample:** what is the probability to gather a sample ($n = 100$) whose average damage is beyond 13700 Swiss Francs?

   (e) What is the probability to gather a sample ($n = 100$) whose average damage is beyond 10500 Swiss Francs?

   (f) How do these two last probabilities change when the sample size increases? *(for example from $n_1 = 100$ to $n_2 = 200$)*

   (g) It is important for the insurance company not to underestimate the damages and therefore it wants to know whether the sample average is particularly low or not. The insurance company wants to knows beyond what threshold the sample would belong to the bottom 10% of all samples of 100 households. What is the damage value under which there are 10% of all samples with 100 households?