ØAMET2200 · Fall 2019
Worksheet 2

Instructor: Fenella Carpena

August 30, 2019

Exercise 1 Tesla, an automotive company that manufactures electric cars, would like to learn about the preferences of its current and potential future consumers.

Specifically, the company has developed a software update that improves the performance of Model S cars sold in 2017. It would like to learn how owners of the 2017 Model S would prefer to have the software updated: either for free at the nearest Tesla dealer, or paying a service fee to have a Tesla representative come to their home. Suppose that because of set-up costs, Tesla would like to offer customers only one of these options.

(a) What is the population of interest in this context?

(b) Tesla would like to conduct a survey to ask customers what they prefer. Listed below are some of the ideas that were proposed for conducting the survey.

  (i) Send the survey via e-mail to all 2017 Model S owners who live in Oslo. What are some concerns with this approach?

  (ii) Draw a random sample of 5000 people from a list of all 2017 Model S owners. What is this resulting sample called?

  (iii) Survey all customers who purchased their car with blue exterior paint. Is this a random sample? Do you think results will be very different from a simple random sample?

(c) A member of the data analytics team at Tesla argues that a simple random sample may be problematic because a small fraction of customers live far away from a Tesla dealers, and these customers have a lot at stake about whether the service can be provided at their home. Because of this, she proposes that stratified sampling should be used. Is she correct?

Exercise 2 In the year 2020, Tesla is planning to release a new model: the Tesla Roadster. Suppose that Tesla wants to learn whether it should offer a new long-distance battery for this upcoming model.

(a) What is the population of interest here?

(b) Is it reasonable to expect that this population could be used a sampling frame?

(c) Suppose Tesla wished to collect detailed information using an in-person survey. To minimize the survey costs, a member of the analytics team suggests using clustered sampling. What are the potential disadvantages of this approach?

Exercise 3 (Stine and Foster, Chapter 13) The website Hotels.com asks users to rate hotels that they have visited on a scale from 0 (worst) to 5 (best). Ninety-nine guests gave the *Grand Hotel* in Oslo a 4.5 out of 5 rating.

(a) What is the population of interest?

(b) What is the population parameter of interest?

(c) What is the sampling frame?

(d) What is the sample size?

(e) What is the sampling design? Was randomization employed?

(f) What are the potential sources of bias or any problems with the sample?

Exercise 4 (GPS chips) A manufacturer of GPS chips use a Highly Accelerated Life Test (HALT) to check that the chips are manufactured as designed. Engineers take a random sample of 20 chips, and each chip undergoes HALT. Each sampled chip then gets a HALT score (an integer 1 to 16), which represents the number of tests the chip underwent until it failed. Suppose that the engineers who designed the production line say that when manufacturing is functioning properly, on average the chips should have a HALT score of 7 with a SD of 4.

(a) Let $\overline{X}$ be the average HALT score in the sample of 20 chips (i.e., the sample average). If the manufacturing process is working as designed, what is $E(\overline{X})$? What is $Var(\overline{X})$? Assume that HALT scores are independent.

(b) Suppose that we find a kurtosis $K_4 \approx 1$. Do our data meet the sample size condition? What model describes the sampling distribution of $\overline{X}$?

(c) Suppose that the production line is shut down when the sample mean HALT score is less than 6 or more than 8, what is the probability of Type I error?

(d) Suppose the production managers tolerance for Type I error is 5%. What is the control limit that he/she should set?

Exercise 5 Identify the following mistakes as either Type I or Type II errors.

(a) A jury convicts an innocent defendant

(b) A retailer fails to stock fashion items that become popular in the coming season

(c) A diagnostic test fails to detect the presence of a serious virus infection

(d) A company hires an application who is not qualified for the position

Exercise 6 (Adapted from Stine and Foster, Chapter 14) The manager of a warehouse monitors the volume of shipments made by the delivery team (e.g., to ensure that delivery team is not stealing from or damaging the packages). A sample of 25 packages is selected and weighted every day. The average weight of each package should be $\mu = 22$ pounds and $\sigma = 5$ pounds.

Mark each of the following statements below as True, False, or Uncertain, and provide a brief explanation for your answer.

(a) A Type I error occurs if the mean weight μ and the standard deviation σ do not change.

(b) If the average weight of packages falls and the manager does not take action, then the manager has committed a Type II error.

(c) If the manager would like a 1% chance of a Type I error, he/she should set the control limits to 12 pounds and 32 pounds (i.e., $12 < \overline{X} < 32$).

Exercise 7 (Stine and Foster, Chapter 14) A bottler carefully weighs bottled coming off its production line to check that the system is filling the bottles with the correct amount of beverage. By design, the system is set to slightly overfill the bottles to allow for random variation in the packaging. The content weight is designed to be 1,020 g with a SD of 8 g. Assume the weights are normally distributed.

A proposed system shuts down the facility if the contents of a bottle weigh less than 1,000 g. What is the chance of a Type I error with this system?

Exercise 8 (Stine and Foster, Chapter 14) Auto manufacturers buy car components from suppliers who deliver them to the assembly line. A manufacturer can schedule staff time to check a sample of these components for defects; the staff can check either 5 parts per day (everyday) or 25 parts during the weekend (when the assembly line is slowed).

(a) If the manufacturer is interested in finding a flaw that shows up rarely, which approach will have a better chance of finding it?

(b) What if the manufacturer is concerned about a problem that will be evident and persistent?

Exercise 9 (Stata) For this exercise, we will historical data of the S&P 500. The S&P 500 is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the US. The data is built-in Stata, and you can open it by using the command `sysuse sp500`

(a) Take a look at the variables in the dataset. Explain what each variable means and how it is measured (e.g., dollars, percent).

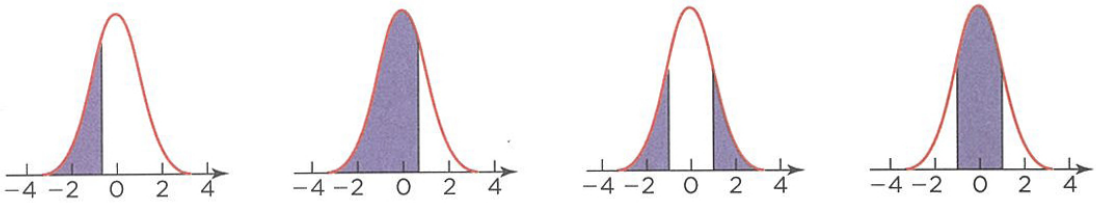
(b) Which variables in the data are string? Which are numeric?

(c) What dates are covered in the data?

(d) Consider the variable `close`. What is its average, minimum, and maximum?

(e) Draw a line graph which shows how the closing price of the S&P 500 has changed over time. *Hint: To make a line graph, the command is* `twoway line`.

(f) Select a simple random sample of 100 observations from the data. What is the average closing price in this subset? Do you find that it is close to the average you found in part in (d)?



z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\|Z\| > z)$	$P(\|Z\| \leq z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	0.32	0.68	0.64	0.36
0.4959	0.31	0.69	0.62	0.38
0.5244	0.30	0.70	0.60	0.40
0.5828	0.28	0.72	0.56	0.44
0.6433	0.26	0.74	0.52	0.48
0.6745	0.25	0.75	0.50	0.50
0.7063	0.24	0.76	0.48	0.52
0.7388	0.23	0.77	0.46	0.54
0.7722	0.22	0.78	0.44	0.56
0.8064	0.21	0.79	0.42	0.58
0.8416	0.20	0.80	0.40	0.60
0.8779	0.19	0.81	0.38	0.62
0.9154	0.18	0.82	0.36	0.64
0.9542	0.17	0.83	0.34	0.66
0.9945	0.16	0.84	0.32	0.68
1.0364	0.15	0.85	0.30	0.70
1.0803	0.14	0.86	0.28	0.72
1.1264	0.13	0.87	0.26	0.74
1.1750	0.12	0.88	0.24	0.76
1.2265	0.11	0.89	0.22	0.78
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999