

M3A1 Fast Technologies Part 3 – Probability Distributions

Connie Smith has asked you to do some preliminary analysis of the data in the *Fast Technologies* data found in the M3A1 Fast Data file.

Download and open the workbook M3A1 Fast Data file. Immediately save the workbook with the following name format using your own name: **LastName_FirstName_M3A1**

1. The worksheet, **Disc Response**, contains the recorded response times in microseconds of 400 2TB SSDs selected at random from the production line.
 - a. Find and report the minimum value, max value, the range, mean, and standard deviation.

How to do it:

- [Finding the arithmetic mean](#)
- [Finding the range of a set of values](#)
- [Find the standard deviation](#)

- b. Construct a frequency distribution and **histogram** (use 14 bins and round your bin width **up** to the nearest tenth of a microsecond) and use them to draw a conclusion. Remember neither a bar or column chart are true histograms which do not have spaces between the columns/bars. Include the cumulative frequency curve on a second axis.

How to do it #1b: Cumulative Frequency Chart using Excel Data Analysis Histogram Tool
https://youtu.be/i4dop_HMKDc

- c. Do the data on this worksheet appear to be normally distributed? Why or why not?
 - d. Could any of the response times be considered outliers? Justify your answer with appropriate calculations.

How to do it: Finding Outliers https://youtu.be/V9g_FFDJgHM

- e. Using the values for the mean and standard deviation of the data, and assuming the population from which the data are drawn is normally distributed:
 - i. Using the normal distribution, find the probability that the response time will exceed 5.20. What is the probability the response time will be less than 4.80?
 - ii. Using the empirical data, what is the actual percentage of response times that are greater than 5.20 or are less than 4.80?
 - iii. How do the normal probability calculations compare to the actual occurrences? What do you conclude?

How to do it #1e: Finding Empirical and Normal Probabilities <https://youtu.be/mrUKDE6-Zrg>

2. Using the data in the tab **2 TB SSD Test** for Model x:

- a. Each sample SSD is tested (observed) 100 times and either passes or fails each observation.
 - i. What is the fraction (%) of fails for each of the 30 samples?
 - ii. What distribution might be appropriate to model the failure of an individual SSD?
How to do it: Hint: what type of experiment has only two possible outcomes: success or failure.

- b. Is a normal distribution an appropriate assumption for the sampling distribution of the mean all 30 samples (n=3000)? Why or why not? Could you use the normal distribution for an n of 100? (Hint: test to see if you can use the normal approximation)

How to do it #2b: Can you use the normal approximation? https://youtu.be/g_iliMLWXak

- c. Using all the data, what fraction of SSDs fails the test? Using this probability, what is the probability of having x failures in the next **100** SSDs tested, for $x = 0$ to $x = 20$?
- d. Find the expected value of the distribution (for $x = 0$ to $x = 20$ only).

How to do it #2c & 2d: Binomial Distribution https://youtu.be/6Y3l6POC_Sg

3. Fast conducted a survey of SSD brand preference of 300 potential customers and the results are shown in the **Gender-Brand** tab.
 - a. Construct a cross-tabulation (also known as a contingency table) using Pivot tables of the data summarizing the counts for each brand for each gender.
 - b. Complete the contingency table to include all marginal probabilities.
 - c. Find the conditional probabilities of each brand given gender and each gender given brand.
 - d. Complete the following table of probabilities. Put the probabilities expressed as a decimal with **three** significant digits in the blank column on the right.

Probabilities	
P(Female)	
P(Male)	
P(Fast)	
P(Carbon)	
P(Sonic)	
P(Female and Carbon)	
P(Female and Fast)	
P(Female and Sonic)	
P(Male and Carbon)	
P(Male and Fast)	
P(Male and Sonic)	
P(Female or Carbon)	
P(Female or Fast)	
P(Female or Sonic)	
P(Male or Carbon)	
P(Male or Fast)	
P(Male or Sonic)	

- e. Are the “events” gender and brand independent? Why or why not? Show calculations to support your conclusion.

How to do it #3a-e: Conditional Probabilities <https://youtu.be/-W0CF-TyJFE>

4. Worksheet **3 TB SSD** contains large sample test data on the new drive. The time to failure was recorded in hours.
 - a. Determine the type of distribution of the test data by creating a histogram with 30 bins and a bin width of 150 hours. Include the Cumulative Relative Frequencies on the graph using a second y-axis.

How to do it:

- Cumulative Frequency Chart using Excel Data Analysis Histogram Tool
https://youtu.be/i4dop_HMKDc
- Help picking the distribution type: http://www.drdawnwright.com/?page_id=17578

b. Using the empirical data, answer the following questions:

- i. What is the mean time to failure of the drives tested?
- ii. What is the probability a drive will fail before 3000 hours?
- iii. What is the probability a drive will last more than 4200 hours?
- iv. What is the probability a drive will last between 4200 and 4500 hours?

How to do it: Finding Empirical and Normal Probabilities <https://youtu.be/mrUKDE6-Zrg>