Statistics Project

If there are errors displaying this, [here is a pdf version]().

WARNING: You cannot use the Google sheets that I provide here except as a learning aid. You must submit your own work. If you do use Excel, you MUST submit the Excel sheet (not a PDF/Word/Text document) and all computations should be done with Excel formulas in the sheet, not just numbers pasted. The only data entered should be from the data you create for Part 1. You can use my sheets to help you with this and, of course, ask questions!

You must complete Parts 1 and 2, you can view the grading rubric [here](). Look at this rubric, it is different than what LoudCloud shows.

Problem

Consider one way of modeling a problem. Say you have a device and in any given year, there is a 1 in 6 chance that the device will fail. One question is, “On average, how long will it be before such a device fails?” This sort of problem is what will be modeled and analyzed here.

The problem above can be modeled as follows: Roll a normal fair six-sided die. Rolling a 1 will count as “fail”. Anything else will count as “not-fail”. The random variable we will use will be

X = # of rolls required until a 1 is rolled = “time until failure”

This is a random variable, so in each experiment (for each device), this will produce a value. For example, if we roll ⚅⚁⚃⚃⚀, then X = 5 (we count the final roll).

Part 1 [Excel] (Empirical Analysis)

Simulate 20 repetitions of obtaining a value for X, that is, roll until a 1 is rolled, record the result, and repeat this 20 times. For example, if you roll ⚂⚃⚄⚀, then record 4 since 4 rolls were required. You can complete this using a die or simulating using Excel, Python, or some other option, e.g., [Random.org.]()

- Record your rolls and counts.
- Compute the mean and standard deviation of your values for X.
- Given the modeling problem we started with, interpret the mean and standard deviation you found in terms of how many years are expected before the device fails.

Excel might be used for this part as it makes the calculations, recording of data, etc. very simple. However, it is not required. Here is a [Google sheets example]() for the version of this problem addressed in Topic 7 DQ 1 and 2, this is different from the problem at hand, so don't directly use this for your own work here.

You can check to see if your results are reasonable here.

Part II [Excel] (Theoretical Analysis - Approximations)

Here is a Google sheets version you can look at for help with the Excel part. (Again this sheet is for Topic 7 DQ 1 and 2 so you may learn from it, but don't just copy from it.) **If you utilize Excel or Google Sheets, you must submit the Excel or share the Google Sheet, I will need to see the formulas that you are using.**

Compute approximations to the expected value, variance, and standard deviation of the random variable X. See the additional notes for discussion on these computations, in brief:

- $P(X = i)$ = the probability of rolling something other than a ⚀ $(i - 1)$-times, then rolling a ⚀ on the i^{th} throw. (**You must explain the formula!**)

  Make a table for the first 20 values:

i	1	2	…	20	Totals
$P(X = i)$	$P(X = 1)$	$P(X = 2)$	...	$P(X = 20)$	$\sum_{i=1}^{20} P(X = i)$
$i \cdot P(X = i)$	$1 \cdot P(X = 1)$	$2 \cdot P(X = 2)$	...	$20 \cdot P(X = 20)$	$\sum_{i=1}^{20} i \cdot P(X = i)$
$i^2 \cdot P(X = i)$	$1^2 \cdot P(X = 1)$	$2^2 \cdot P(X = 2)$	...	$20^2 \cdot P(X = 20)$	$\sum_{i=1}^{20} i^2 \cdot P(X = i)$

  Use your table to compute

  - $P(X \leq 20) = \sum_{i=1}^{20} P(X = i)$, the probability that you get the first 1 in at most 20 rolls.

- An approximation of $E[X]$ using the first 20 values for X, namely $E[X] \approx \sum_{i=1}^{20} i \cdot P(X = i)$.
- An approximation to the variance of X using the first 20 terms. For this, use $E\left[X^2\right] \approx \sum_{i=1}^{20} i^2 \cdot P(X = i)$ and then use this and the above to get a rough approximation of $var[X] = E\left[X^2\right] - (E[X])^2$.
- Use the preceding to approximate the standard deviation $\sigma[X] = \sqrt{var[X]}$ using the first 20 terms.

Again, using Excel for this will simplify what needs to be done.

- Repeat the third item of Part 1: Given the modeling problem we started with, interpret $E[X]$ and $\sigma[X]$ in terms of how many years are expected before the device fails.

Optional Part III [Written] (Theoretical Analysis – Exact Computations)

Compute the expected value, variance, and standard deviation of the random variable X. See the additional notes for discussion on these computations, in brief:

- (Expected Value) $E[X] = \sum_{i=1}^{\infty} i \cdot P(X = i)$
- (Variance) $var[X] = E\left[(X - E[X])^2\right] = E\left[X^2\right] - (E[X])^2$
- (Standard Deviation) $\sigma[X] = \sqrt{var[X]}$

Hint: For computing these without directly manipulating the infinite summations that appear in the definitions of $E[X]$ and $var[X]$, let A be the event that a 1 is rolled on the first throw and A' be the complementary event, namely, that a 1 is not rolled on the first throw. It is clear that $E\left[A'\right] = E[1 + X] = 1 + E[X]$ since what is rolled after the first roll is just like starting over. It is also clear that $E[A] = 1$. The Law of Total Expectation (see notes) gives:

$$E[X] = E[A] \cdot P(A) + E\left[A'\right] \cdot P\left(A'\right) = P(A) + (1 + E[X]) \cdot (1 - P(A))$$

This makes it quite simple to find $E[X]$.

A similar “trick” can be used to find $E[X^2]$, here you will use $E\left[A^{'}\right] = E\left[(1 + X)^2\right] = E\left[1 + 2X + X^2\right] = 1 + 2E[X] + E[X^2]$. Again, the Law of Total Expectation gives:

$$E\left[X^2\right] = E[A]\cdot P(A) + E\left[A^{'}\right]\cdot P(A^{'})$$

and from this, it is simple to compute $E\left[X^2\right]$.