Risk analysis lab 2018. 09. 18. (Introduction, part 2.)

1. Colored Dice: Suppose that a game contained a set of six-sided dice with two opposite faces colored red, two opposite faces colored green, and two opposite faces colored yellow. If one of these dice is tossed 20 times, what is the probability that there will be 5 or less green results? What is the probability that there will be 10 or less green results? Write a script that calculates these quantities (Hint). Please use the scipy library.

2. Position and Velocity of a Ball: If a stationary ball is released at a height \( h_0 \) above the surface of the Earth with a vertical velocity \( v_0 \), the position and velocity of the ball as a function of time will be given by the equations

\[
\begin{align*}
    h(t) &= \frac{1}{2} gt^2 + v_0 t + h_0 \\
    v(t) &= gt + v_0
\end{align*}
\]

where \( g \) is the acceleration due to gravity (-9.81 m/s\(^2\)), \( h \) is the height above the surface of the Earth (assuming no air friction), and \( v \) is the vertical component of velocity. Write a function that plots the height and velocity as a function of time for an arbitrary \( h_0 \) and \( v_0 \).

3. The Birthday Problem: The Birthday Problem is stated as follows: if there is a group of \( n \) people in a room, what is the probability that two or more of them have the same birthday? It is possible to determine the answer to this question by simulation. Write a function that calculates the probability that two or more of \( n \) people will have the same birthday, where \( n \) is a calling argument. (Hint: To do this, the function should create an array of size \( n \) and generate \( n \) birthdays in the range 1 to 365 randomly. It should then check to see if any of the \( n \) birthdays are identical. The function should perform this experiment at least 5000 times and calculate the fraction of those times in which two or more people had the same birthday.) Write a test program that calculates and prints out this probability for \( n = 2, 3, ..., 40 \).

4. Linear regression problem: There are many cases in science and engineering where there are noisy sets of data, and we wish to estimate the straight line that “best fits” the data. This problem is called the linear regression problem. Given a noisy set of measurements \((x, y)\) that appear to fall along a straight line, how can we find the equation of the line

\[
y = mx + b
\]

that “best fits” the measurements? If we can determine the regression coefficients \( m \) and \( b \), then we can use this equation to predict the value of \( y \) at any given \( x \) by evaluating the equation for that value of \( x \). A standard method for finding the regression coefficients \( m \) and \( b \) is the method of least squares. This method is named
“least squares” because it produces the line for which the sum of the squares of the differences between the observed $y$ values and the predicted $y$ values is as small as possible. The slope of the least squares line is given by and the intercept of the least squares line is given by

$$m = \frac{(\sum xy) - (\sum x)\bar{y}}{(\sum x^2) - (\sum x)^2}$$

and the intercept of the least squares line is given by

$$b = \bar{y} - m\bar{x}$$

where $\bar{x}$ and $\bar{y}$ is the mean of the $x$ and $y$ values respectively. Write a program that will calculate the least-squares slope $m$ and $y$-axis intercept $b$ for a given set of noisy measured data points $(x, y)$. The data points should be read from file linreg_data.csv using the pandas library (Help).

5. **Correlation Coefficient of Least Squares Fit**: Develop a function that calculates the correlation coefficient of the fit. The input data points $(x, y)$ will be passed to the function in two input arrays, $x$ and $y$. The correlation coefficient is

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[ n(\sum x^2) - (\sum x)^2 \right] \left[ n(\sum y^2) - (\sum y)^2 \right]}}$$

Test your function using linreg_data.csv.