Question 1c) Use `[ ]` to select `Name` and `Year` **in that order** from the `baby_names` table.

Then repeat the same selection using the `.loc` notation instead.

```
In [13]: name_and_year1= baby_names[['Name','Year']]
         name_and_year1.head()
```

```
Out[13]:        Name  Year
         0      Mary  1910
         1     Annie  1910
         2      Anna  1910
         3  Margaret  1910
         4     Helen  1910
```

```
In [14]: name_and_year2 = baby_names.loc[:,['Name', 'Year']]

         name_and_year2.head()
```

```
Out[14]:        Name  Year
         0      Mary  1910
         1     Annie  1910
         2      Anna  1910
         3  Margaret  1910
         4     Helen  1910
```

Question 2a) A coin is flipped 10 times. How many possible outcomes have exactly 2 heads? Use LaTeX (not code) in the cell below to show all of your steps and fully justify your answer.

Note: In this class, you must always put your answer in the cell that immediately follows the question. DO NOT create any cells between this one and the one that says *Write your answer here, replacing this text.*

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

$n = 10$

$k = 2$

$\implies \binom{10}{2} = \frac{10!}{2!(10-2)!} = \frac{10!}{2!*8!} = \frac{9*10}{2} = \frac{90}{2}$

$\implies \boxed{\binom{10}{2} = 45}$

Question 2b) What is the probability that if I roll two 6-sided dice they add up to **at most** 9? Use LaTeX (not code) in the cell directly below to show all of your steps and fully justify your answer.

There are 6 possible ways to sum greater than 9:

10: (4,6), (5,5), (6,4)

11: (5,6), (6,5),

12: (6,6)

This implies that there is 1/6 chance to roll a sum greater than 9.

Therefore, to roll a sum at most 9 $\implies 1 - \frac{1}{6} = \boxed{\dfrac{5}{6}}$

Question 2c) Suppose you show up to a quiz completely unprepared. The quiz has 10 questions, each with 5 multiple choice options. You decide to guess each answer in a completely random way. What is the probability that you get exactly 3 questions correct? Use LaTeX (not code) in the cell directly below to show all of your steps and fully justify your answer.

$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$k = 3$

$n = 10$

$p = \frac{1}{5}$

$\implies P(3) = \binom{10}{3} (\frac{1}{5})^3 (1 - \frac{1}{5})^{10-3}$

$\implies (\frac{10!}{3!(10-3)!})(\frac{1}{5})^3(\frac{4}{5})^7$

$\implies (\frac{8*9*10}{6})(\frac{1}{125})(.2097)$

$\implies \boxed{P(3) = .2013}$

Question 3a) We commonly use sigma notation to compactly write the definition of the arithmetic mean (commonly known as the average):

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The ith *deviation from average* is the difference $x_i - \bar{x}$. Prove that the sum of all these deviations is 0 that is, prove that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ (write your full solution in the box directly below showing all steps and using LaTeX).

$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i) - \sum_{i=1}^{n}(\bar{x})$

$\implies \sum_{i=1}^{n}(x_i) - \bar{x}\sum_{i=1}^{n}(1)$

$\implies \sum_{i=1}^{n}(x_i) = x_1 + x_2 + ... + x_n$

$\implies \bar{x} = \frac{(x_1+x_2+...+x_n)}{n}$

$\implies \sum_{i=1}^{n}(1) = n$

$\implies (x_1 + x_2 + ... + x_n) - \frac{(x_1+x_2+...+x_n)}{n}(n)$

$\implies \boxed{(x_1 + x_2 + ... + x_n) - (x_1 + x_2 + ... + x_n) = 0}$

Question 3b) Let $x_1, x_2, \ldots, x_n$ be a list of numbers. You can think of each index i as the label of a household, and the entry x_i as the annual income of Household i.

Consider the function

$$f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$$

In this scenario, suppose that our data points $x_1, x_2, \ldots, x_n$ are fixed and that c is the only variable.

Using calculus, determine the value of c that minimizes $f(c)$. You must use calculus to justify that this is indeed a minimum, and not a maximum.

$f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_ic + c^2)$

$\implies \frac{1}{n}\sum_{i=1}^{n}(x_i^2) - \frac{1}{n}\sum_{i=1}^{n}(2x_ic) + \frac{1}{n}\sum_{i=1}^{n}(c^2)$

$\implies \frac{1}{n}\sum_{i=1}^{n}(x_i^2) - \frac{2c}{n}\sum_{i=1}^{n}(x_i) + \frac{c^2}{n}\sum_{i=1}^{n}(1)$

$f'(c) = \frac{2c}{n}(n) - \frac{2}{n}\sum_{i=1}^{n}(x_i)$

$\implies 2c - \frac{2}{n}\sum_{i=1}^{n}(x_i) = 0$

$\implies 2c = \frac{2}{n}\sum_{i=1}^{n}(x_i)$

$\implies c = \frac{1}{n}\sum_{i=1}^{n}(x_i)$

$\implies c = \bar{x}$

$f''(c) = (\frac{d}{dc})(2c - \frac{2}{n}\sum_{i=1}^{n}(x_i)) = 2$

Since $f''(c) = 2$, a positive number, $\bar{x}$ is the minimum according to the second derivative test.

Question 4b) I have a coin that lands heads with an unknown probability p.

Suppose I toss it 10 times and get the sequence TTTHTHHTTH.

If you toss this coin 10 times, the chance that you get the sequence above is a function of p. That function is called the *likelihood* of the sequence TTTHTHHTTH, so we will call it $L(p)$.

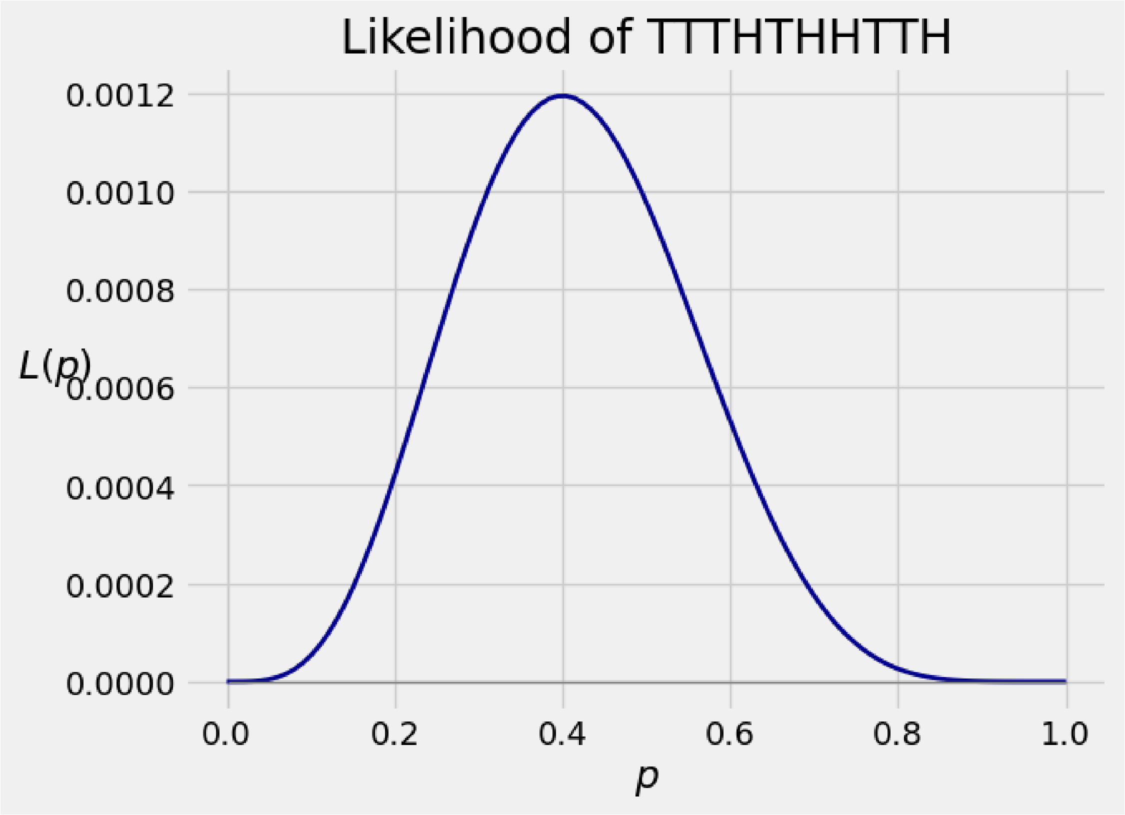
What is $L(p)$ for the sequence TTTHTHHTTH?

Write your answer using LaTeX below (i.e. your answer should be of the form: $L(p)$=some function of p)

$L(p) = p^4(1-p)^6$

Question 4c) Below is a section of code that will help you plot the function $L(p)$ that you defined above. Replace the ellipses with your function of p

```
In [27]: p = np.linspace(0, 1, 100)
         #This creates an array of 100 values equally spaced between 0 and 1

         likelihood = (p**4)*(1-p)**6

         plt.plot(p, likelihood, lw=2, color='darkblue')
         #This plots the likelihood function

         plt.plot([0, 1], [0, 0], lw=1, color='grey')
         #This plots a horizontal axis

         plt.xlabel('$p$')
         #This labels the x axis
         plt.ylabel('$L(p)$', rotation=0)
         #This labels the y-axis

         plt.title('Likelihood of TTTHTHHTTH');
         #This titles the plot
```



Question 4d) The value $\hat{p}$ at which the likelihood function attains its maximum is called the *maximum likelihood estimate* (MLE) of p. Among all values of p, it is the one that makes the observed data most likely.

Using your plot above, what is the value of $\hat{p}$?

Provide a simple interpretation of that value in terms of the data TTTHTHHTTH.

The value of $\hat{p}$ is 0.4. This means that it is very likely that the probability for the coin to land heads is 40%.

Question 4e) Let's prove what you observed graphically above. That is, let's use calculus to find $\hat{p}$.

But wait before you start trying to find the value p where $L'(p) = 0$ (trust us, the algebra is not pretty…)

TIP:
The value $\hat{p}$ at which the function L attains its maximum is the same as the value at which the function $\log(L)$ attains its maximum. To clarify, $\log(L)$ is the composition of log and L: $\log(L)$ at p is $\log(L(p))$. Even though it doesn't make a difference for this problem, log is now and forevermore the log to the base e, not to the base 10.

This tip is hugely important in data science because many probabilities are products and the log function turns products into sums. It's much simpler to work with a sum than with a product.

Armed with that tip use calculus to find $\hat{p}$. You don't have to check that the value you've found produces a max and not a min – we'll spare you that step.

$L(p) = p^4(1-p)^6$

$\implies \log(L) = \log(p^4(1-p)^6)$

$\implies \log(p^4) + \log((1-p)^6) = 4\log(p) + 6\log(1-p)$

$\frac{d}{dp}[4\log(p) + 6\log(1-p)] = \frac{4}{p} - \frac{6}{1-p}$

$\frac{4}{p} - \frac{6}{1-p} = 0$

$\implies 4(1-p) - 6p = 0$

$\implies 4 - 4p - 6p = 0$

$\implies 4 = 10p$

$\boxed{0.4 = p}$