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 [40]:
```
name_and_year1= baby_names.iloc[:, [3,2]]
name_and_year1.head()
```

Out[40]:

	Name	Year
0	Mary	1910
1	Annie	1910
2	Anna	1910
3	Margaret	1910
4	Helen	1910

In [41]:
```
name_and_year2 =  baby_names.loc[:, ['Name','Year']]

name_and_year2.head()
```

Out[41]:

	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.*

$\binom{10}{2}$ Because the coin is flipped 10 times, and we want the outcome that there is exactly 2 heads.

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.

Two six sided die have 36 outcomes. Out of these we can add up to at most 9. Thus, we can subtract the possibilies that add up to 10,11,12. There are 6 possibilies $2 \cdot (6+6), (6+5), (6+4), (5+5), (5+6)$, thus shows 36-6 = 30, where the probability to roll 2 6 sided dice that they add up to at most 9 is 30/36

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.

We can use the binomial distribution property. \$P_x = \$ $\binom{10}{3} (\frac{1}{5})^3 * (\frac{4}{5})^7$

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

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

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).

First we see that the summation notation algebra, where we can split the equation above to:
$\sum_{i=1}^{n}(x_i) - \sum_{i=1}^{n} \bar{x} = 0$.
We can then move the equation around like:
$\sum_{i=1}^{n}(x_i) = \sum_{i=1}^{n} \bar{x}$
We can show that $\sum_{i=1}^{n} \bar{x}$ is equivalent to $\bar{x} + \bar{x} ... n * \bar{x}$. Thus we can show this as $n \cdot \bar{x}$.
Therefore:
$\sum_{i=1}^{n}(x_i) = n \cdot \bar{x}$
$\frac{1}{n}\sum_{i=1}^{n}(x_i) = \bar{x}$
We see that $\bar{x} = \frac{1}{n}\left(x_1 + x_2 + ... + x_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i$ as above and matches the equation

Question 3b) Let $x_1, x_2, \dots, x_n$ 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 $x_1, x_2, \dots, x_n$ 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.

In order to find the minimum we must take the derivitave of the equation respect to c. Then we should set it equal to 0 to determine the value of c. Furthermore, we must take the second derivitave of the equation in order to find if it is concave up/down to determine if it is actually the minimum or maximum.

$f(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$
$f(c)' = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{dc}(x_i - c)^2$
$f(c)' = \frac{1}{n}\sum_{i=1}^{n}2 \cdot (x_i - c)(-1)$
$= \frac{1}{n}\sum_{i=1}^{n}(-2)(x_i - c)$
We can set this equal to 0 and solve for c.
$\frac{1}{n}\sum_{i=1}^{n}(-2)(x_i - c) = 0$
$= \frac{-2}{n}\sum_{i=1}^{n}(-x_i + c) = 0$
$\frac{-2}{n}\sum_{i=1}^{n}(c) + \frac{-2}{n}\sum_{i=1}^{n}(-x_i) = 0$ $\sum_{i=1}^{n}(c) + \sum_{i=1}^{n}(-x_i) = 0$
$\sum_{i=1}^{n}(c) = \sum_{i=1}^{n}(x_i)$
We can show that $\sum_{i=1}^{n}(c)$ is the same as \$n c, \$where $c + c ... c * n$.
$n \cdot c = \sum_{i=1}^{n}(x_i)$
$c = n\sum_{i=1}^{n}(x_i)$ Thus the above equation shows what c equals.
Then now we must show that it is concave up to show that c is the minimum by taking the second derivative.
$f(c)' = \frac{1}{n}\sum_{i=1}^{n}(-2)(x_i - c)$
$f(c)'' = \frac{d}{dc}(\frac{1}{n}\sum_{i=1}^{n}(-2)(x_i - c))$
$\frac{1}{n}\sum_{i=1}^{n}\frac{d}{dc} - 2x_i + 2c$

$\frac{1}{n}\sum_{i=1}^{n}2$
Thus, we show that this equation is concave up with positive 2. Thus we can show that c is the minimum.

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 $L(p)$.

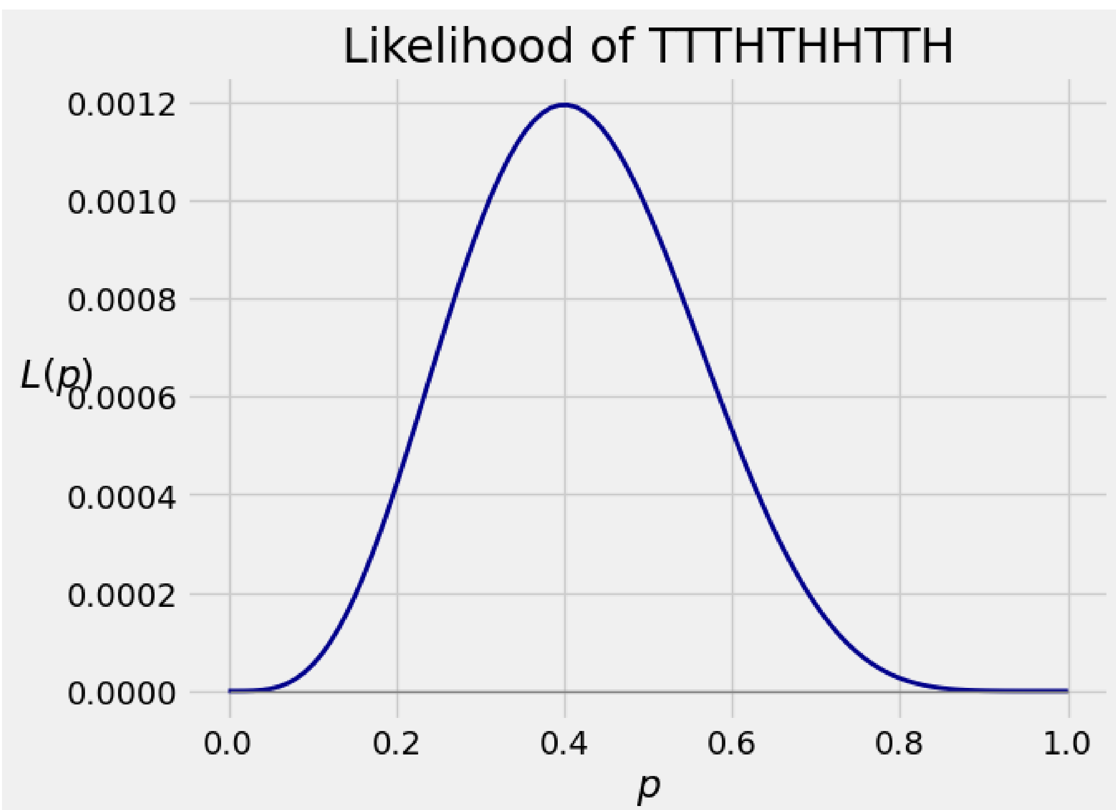
What is $L(p)$ for the sequence TTTHTHHTTH?

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

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

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 [ ]: 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 $\widehat{p}$ 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 $\widehat{p}$?

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

From the data above, we can see that the max is 0.4, or 40% of getting a heads when rolling a two-sided coin, with the sequence of TTTHTHHTTH.

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.

$L(p) = p^4 \cdot (1-p)^6$
Take the log of each side.
$\ln(L(p)) = \ln(p^4 \cdot (1-p)^6)$
$= 4\ln(p) + 6\ln(1-p)$
$L(p)' = \frac{d}{dp} 4\ln(p) + 6\ln(1-p)$
$\frac{4}{p} + \frac{-6}{(1-p)} = 0$
$\frac{-6}{(1-p)} = \frac{4}{p}$
$4 - 4p = 6p$
$4 = 10p$
$\frac{4}{10} = p$
We can see that the answer here matches the maximum on the graph we made above.