Midterm Exam 1, Isaac Hancock, ST371

Import Section

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import stemgraphic as stg
import statistics
from scipy.stats import binom
from matplotlib_venn import venn3,venn2
```

Assume total population is over 40 years of age C is cancer and D is diagnosed. Table is filled in as the problem is solved

Table	C	C'	Total
D	0.039	0.057	0.096
D'	0.011	0.893	0.904
Total	0.05	0.95	1

$$P(C) = 0.05, P(D|C) = 0.78, P(D|C') = 0.06$$

Problem 1

Need to find $P(D)$

$$P(D|C) = \frac{P(D \cap C)}{P(C)} \rightarrow P(D \cap C) = P(D|C)P(C) = 0.78 * 0.05 = 0.039 \rightarrow P(D \cap C) =$$

$$P(D|C') = \frac{P(D \cap C')}{P(C')} = P(D \cap C') = P(D|C')P(C') = 0.06 * 0.95 = 0.057 \rightarrow P(D \cap C'$$

$$P(D) = P(D \cap C') + P(D \cap C) \rightarrow P(D \cap C') + P(D \cap C) = 0.057 + 0.039 = 0.096$$

Problem 2

We need to find $P(C|D)$ $P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{0.039}{0.096} = 0.406$

Midterm Exam 1, Isaac Hancock, ST371

Problem 3

Blue is the result while green is what is left. From left to right it is very easy to see how the two left venn diagrams when joined in a union $(\cup)$ result in the third venn diagram

In [11]:

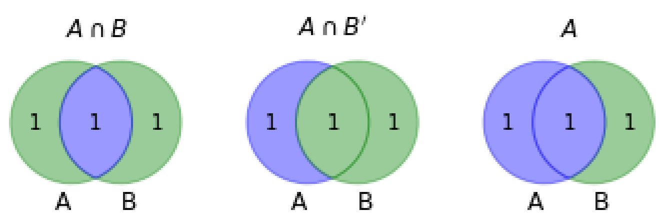
```
fig,axes = plt.subplots(1,3)
set1 = set(['A', 'C'])
set2 = set(['B', 'C'])

v = venn2([set1, set2], ('A', 'B'), ax = axes[0])
v.get_patch_by_id('A').set_color('green')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('blue')
axes[0].title.set_text('$A \cap B$')

v = venn2([set1, set2], ('A', 'B'), ax = axes[1])
v.get_patch_by_id('A').set_color('blue')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('green')
axes[1].title.set_text('$A \cap B\'$')


v = venn2([set1, set2], ('A', 'B'), ax = axes[2])
v.get_patch_by_id('A').set_color('blue')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('blue')
axes[2].title.set_text('$A$')

plt.show()
```

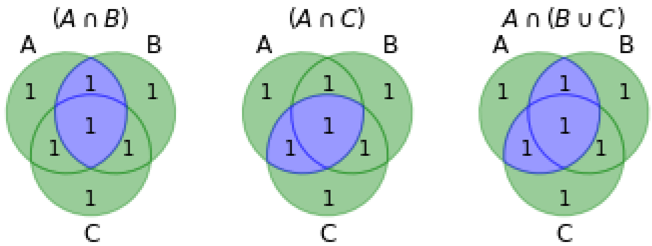


Midterm Exam 1, Isaac Hancock, ST371

Problem 4

We can demonstrate this by using 3 different venn diagrams. One to show what $A \cap (B \cup C)$ and another two to show both $(A \cap B)$ and $(A \cap C)$. Blue can be used to show what the selected region is. From left to right it is very easy to see that an union ($\cup$) of the two regions on the left results in the region displayed in the 3rd venn diagram

In [12]:
```
fig,axes = plt.subplots(1,3)
set1 = set(['A', 'AC','BA','M'])
set2 = set(['B', 'BC','BA','M'])
set3 = set(['C', 'AC','BC','M'])
v = venn3([set1, set2, set3], ('A', 'B','C'), ax = axes[0])
v.get_patch_by_id('A').set_color('green')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('green')
v.get_patch_by_id('111').set_color('blue')
v.get_patch_by_id('101').set_color('green')
v.get_patch_by_id('110').set_color('blue')
v.get_patch_by_id('011').set_color('green')
v.get_patch_by_id('001').set_color('green')
axes[0].title.set_text('$(A \cap B)$')
v = venn3([set1, set2, set3], ('A', 'B','C'), ax = axes[1])
v.get_patch_by_id('A').set_color('green')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('green')
v.get_patch_by_id('111').set_color('blue')
v.get_patch_by_id('101').set_color('blue')
v.get_patch_by_id('110').set_color('green')
v.get_patch_by_id('011').set_color('green')
v.get_patch_by_id('001').set_color('green')
axes[1].title.set_text('$(A \cap C)$')
v = venn3([set1, set2, set3], ('A', 'B','C'), ax = axes[2])
v.get_patch_by_id('A').set_color('green')
v.get_patch_by_id('B').set_color('green')
v.get_patch_by_id('C').set_color('green')
v.get_patch_by_id('111').set_color('blue')
v.get_patch_by_id('101').set_color('blue')
v.get_patch_by_id('110').set_color('blue')
v.get_patch_by_id('011').set_color('green')
v.get_patch_by_id('001').set_color('green')
axes[2].title.set_text('$A \cap (B \cup C)$')
plt.show()
```



$(A \cap B)$
$(A \cap C)$
$A \cap (B \cup C)$
A
B
C
1

Midterm Exam 1, Isaac Hancock, ST371

Problem 5

Can be described 1 - the probability of succedding at all tests.DeMorgans.
$P(1 \cup 2 \cup 3 \cup 4) = P(1' \cap 2' \cap 3' \cap 4')' = 1 - P(1' \cap 2' \cap 3' \cap 4')$

$P(1') = 1 - P(1) = 1 - 0.01 = 0.99$

$P(2') = 1 - P(2) = 1 - 0.03 = 0.97$

$P(3') = 1 - P(3) = 1 - 0.02 = 0.98$

$P(4') = 1 - P(4) = 1 - 0.01 = 0.99$

$P(S) = P(1')P(2')P(3')P(4') = 0.99 * 0.97 * 0.98 * 0.99 = 0.932$

Solve for $1 - P(S) = 1 - 0.932 = 0.068$

Problem 6

Find the probability of succeeding in both and then subtract that from 1. Demorgans
$P(2 \cup 3) = P(2' \cap 3')' = 1 - P(2' \cap 3')$

$P(2' \cap 3') = P(2')P(3') = 0.97 * 0.98 = 0.951 \rightarrow 1 - 0.951 = 0.049$

Problem 7

Discrete random variables. $P(F) = 0.068$ $P(S) = 1 - P(F) = 1 - 0.068 = 0.932$ We assume that if we fail any one of the functional tests than we must reject the chip.Thus we know the number of chips being tested and the probability that any test fails.
$E_{Failed}(X) = N_{Chips} * P_{Failure} = 100 * 0.068 = 6.8 = 6$ Chips. Here we are flooring the value because the 7th chip hasnt quite failed

Midterm Exam 1, Isaac Hancock, ST371

Problem 8

Working with Binomial pmf. 100 choose 10 = 1.7310309e+13 Solution is $1.7310309 * 10^{13} * (0.068)^{10} * (0.932)^{90} = 0.06469$ We can validate this value in python to find the probability of exactly 10 chips failing assuming that the probability of failure is 0.068 or any one test failing.

In [4]:
```
binom.pmf(10,100,0.068)
```

Out[4]:
```
0.06468834954985812
```

Problem 9

Cars	0	1	2	3
P	0.19	0.38	0.29	0.15

This is incorrect because the total probability must be = 1 for a given pmf to be valid.

$$\sum P_{Cars} = 0.19 + 0.38 + 0.29 + 0.15 = 1.01$$

Problem 10

We know that $P(A) + P(A') = 1$. If A is it will rain tomorrow $P(A) + P(A') = 0.52 + 0.4 = 0.92$ This is incorrect because $0.92 \neq 1$

Problem 11

In a valid pmf, $p(x) \geq 0$ for all values of x. This means that the probability of an event occuring must be greater than 0 and the probability 3 or more mistakes is negative. As $-0.25 < 0$ this is not valid.

Midterm Exam 1, Isaac Hancock, ST371

Problem 12

The assumption made here is that hearts and black cards are independent. For this to be true $P(H|B) = P(B)$ however we know that hearts are red and not black thus $P(H|B) = 0$ and $P(B) = 0.5$ and $0 \neq 0.5$ thus proving that this statement is incorrect. In addition, if two events are independent $P(A \cap B) = P(A) \cdot P(B)$ however $P(B \cap H) = 0$ because no heart is black and $P(B) \cdot P(H) = \frac{1}{9}$ as mentioned above. Since $0 \neq \frac{1}{9}$, this also shows that this statement is incorrect

Problem 13

In [5]:
```
data = [123,116,122,110,175,126,125,111,118,117]
mean = statistics.mean(data)
median = statistics.median(data)
print("Mean: %0.1f\nMedian: %0.1f" % (mean,median))
```

```
Mean: 124.3
Median: 120.0
```

Problem 14

If we were to graph this data set on say a stem and leaf plot, we would see that the data is skewed to the right so that their are more higher values than lower values. This will cause the mean to be larger than the median and thus is what is responsible for the substantial difference between the two. The largest reason for this skew is the outlier of 175 that dramatically pulls up the mean.

Problem 15

$P(X = n) = F(n) - F(n-1) \rightarrow P(T = 5) = F(5) - F(4) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$ years

Midterm Exam 1, Isaac Hancock, ST371

Problem 16

Find $P(T > 3)$. We can assume that the max value is 7 because any number greater than or equal to 7 has the same value on the CDF so it doesnt matter. In other words our assumption is that the upper limit for this CDF is 7

$$P(T > 3) = P(3 < T \leq 7) \rightarrow P(a < X \leq b) = F(b) - F(a) \rightarrow P(3 < T \leq 7) = F(7) -$$

years

Problem 17

$$P(1 \leq T < 6) = P(0 < T \leq 5) \rightarrow P(a < X \leq b) = F(b) - F(a) \rightarrow P(0 < T \leq 5) = F$$

years

Problem 18

$$P(T \leq 5 | T \geq 2) \rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} \rightarrow P(T \leq 5 | T \geq 2) = \frac{P(T \leq 5 \cap T \geq 2)}{P(T \geq 2)} \rightarrow \frac{P(2 \leq T \leq 5)}{P(T \geq 2)} =$$

We need to find $P(1 < T \leq 5)$ and $P(1 < T \leq 7)$

$$P(a < X \leq b) = F(b) - F(a) \rightarrow P(1 < T \leq 5) = F(5) - F(1) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

$$P(a < X \leq b) = F(b) - F(a) \rightarrow P(1 < T \leq 7) = F(7) - F(1) = 1 - \frac{1}{4} = \frac{3}{4}$$

$$P(T \leq 5 | T \geq 2) = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

Midterm Exam 1, Isaac Hancock, ST371

x	0	1	2	3	4
f(x)	0.41	0.37	0.16	0.05	0.01

Problem 19

For a given probability distrobution to be a pmf it must satisfy several conditions:

1. $p(x) \geq 0$ for all values of x
2. $P(0 \leq X \leq 20) = 1$

We must evaluate these conditions.

First Condition

We can look at the pmf table above and see that this condition is me

Second Condition

$$P(0 \leq X \leq 20) = \sum_x p(x) = (0.41 + 0.37 + 0.16 + 0.05 + 0.01) = 1$$

Since both conditions are met this is a valid pmf

Problem 20

$$E(x) = \mu_x = \sum_x x \cdot p(x) = 0(0.41) + 1(0.37) + 2(0.16) + 3(0.05) + 4(0.01) = 0.88$$

Midterm Exam 1, Isaac Hancock, ST371

Problem 21

New values are $E_{Cost}(X) = 10 - 0.05X$

Thus the expected price per 10 meters is $10 - 0.05(0.88) \rightarrow E_{Cost} = \$9.956$

This can also be confirmed by calculating new values for the PMF where units are Cost per 10 meters instead of number of imperfections

$P_{Cost}(X) = 10 - 0.05X$

$P_{Cost}(X = 1) = 10 - 0.05(0) = 10$

$P_{Cost}(X = 1) = 10 - 0.05(1) = 9.95$

$P_{Cost}(X = 2) = 10 - 0.05(2) = 9.90$

$P_{Cost}(X = 3) = 10 - 0.05(3) = 9.85$

$P_{Cost}(X = 4) = 10 - 0.05(4) = 9.80$

x	10	9.95	9.9	9.85	9.8
f(x)	0.41	0.37	0.16	0.05	0.01

$$E(x) = \mu_x = \sum_x x \cdot p(x) = 10(0.41) + 9.95(0.37) + 9.9(0.16) + 9.85(0.05) + 9.8(0.01)$$

Problem 22

For this problem we need to use bimodal distrobutions. The probability of a pipework failure due to operator error is $P_{Failure} = 0.3$ and the probability of pipework failure due to something else is $P_{Success} = 0.7$.

$$E(X_F) = \mu_x = \sum_x x \cdot p(x) = 20(0.3) = 6$$

Midterm Exam 1, Isaac Hancock, ST371

Problem 23

We know that $P(18 \leq X)$ is the same as $P(X = 18) + P(X = 19) + P(X = 20)$. We can plot the binomial distrobution and figure out thos values. We can also validate them by hand

$b(20;20;0.3) = 1 * (0.3^20)(0.7)^0 = 34.86784401 * 10^{-12}$

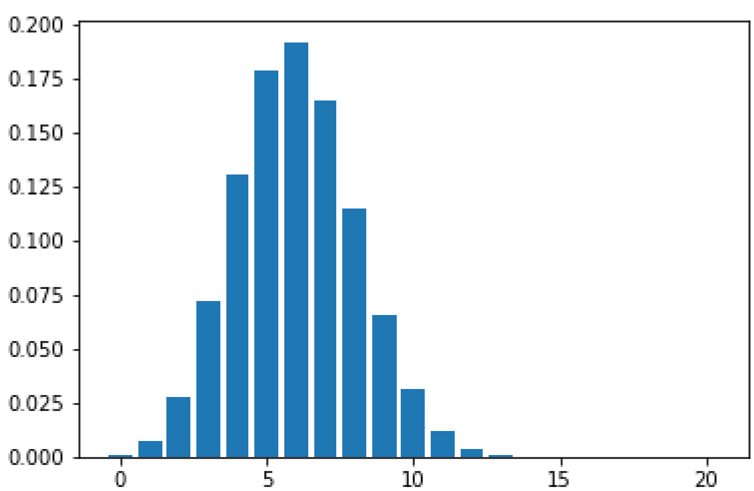
$b(19;20;0.3) = 20 * (0.3^19)(0.7)^1 = 1.627166054 * 10^{-9}$

$b(18;20;0.3) = 190 * (0.3^18)(0.7)^2 = 36.06884753 * 10^{-8}$

Thus the probability is
$34.86784401 * 10^{-12} + 1.627166054 * 10^{-9} + 36.06884753 * 10^{-8} = 3.773088142 * 10^{-8}$

In [6]:
```
dist = binom.pmf(k=range(0,21), n=20,p=0.3)
plt.bar(range(0,21),dist)
vals_greater_than_18 = dist[-3:]
print("Probability of at least 18 failing due to operator error is: ", (sum(va
```

```
Probability of at least 18 failing due to operator error is:  3.77308814237099
95e-08
```



Midterm Exam 1, Isaac Hancock, ST371

Problem 24

We know that $P(4 \geq X)$ is the same as $P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$. We can use the previously plotted binomial distribution to determine this value.

$$b(4;20;0.3) = 4845*(0.3^4)(0.7)^16 = 0.130421$$

$$b(3;20;0.3) = 1140*(0.3^3)(0.7)^17 = 0.071604$$

$$b(2;20;0.3) = 190*(0.3^2)(0.7)^18 = 0.0278459$$

$$b(1;20;0.3) = 20*(0.3^1)(0.7)^19 = 0.0068393$$

$$b(0;20;0.3) = 1*(0.3^0)(0.7)^20 = 0.000797923$$

Our total probability is the sum of these values so $0.130421+0.071604+0.0278459+0.0068393+0.000797923 = 0.237508123$

We can now validate this with python

In [7]:
```
vals_less_than_equal_to_4 = dist[0:5]
print(vals_less_than_equal_to_4)
print("The probability that there are no more than 4 out of 20 failures in this
```

```
[0.00079792 0.00683934 0.02784587 0.07160367 0.13042097]
The probability that there are no more than 4 out of 20 failures in this cas
e due to operator error is 0.237508
```

Midterm Exam 1, Isaac Hancock, ST371

Problem 25

For a certain plant the probability of exactly 5 failing due to operator error if the plants failure rate due to operator error can be calculated below and validated with python. 20 choose 5 is 15,504.

$b(5; 20; 0.3) = 15504 * (0.3^5)(0.7)^15 = 0.178863$

In [8]:

```
p_failure = binom.pmf(5,20,0.3)
print("The probability of exactly 5 parts failing due to operator error is %f"
```

```
The probability of exactly 5 parts failing due to operator error is 0.178863
```

17.89% is a very low percentage which would indicate that this plant does not experience the 30% failure rate for pipelines due to operator error as the figure above mentions. It would seem from this analysis and data that this plant experiences a failure rate that is much lower than 30%.

Midterm Exam 1, Isaac Hancock, ST371

Problem 26

The probability of 1 engine failing is $P(F) = 0.4$ and $P(S) = 0.6$

Two Engine Plane

For a 2 engine play the probability of failure is both engines failing as in all other circumstances at least 1 engine will be running. We do not care about $P(X = 1)$ or $P(X = 2)$ as we just need to find $P(X = 0)$ and subtract that from 1. This is $P_{Failure} = P(F) * P(F) = 0.4 * 0.4 = 0.16$. Thus the probability of success is $P_{Success} = 1 - P_{Failure} = 1 - 0.16 = 0.84$

In [15]:

```
engine_fail = binom.pmf(2,2,0.4)
print("Probability of success for 2 engine plane is %f" %(1-engine_fail))
```

```
Probability of success for 2 engine plane is 0.840000
```

Four Engine Plane

Since this is a larger set, we will need to make use of the binomial distribution. The probability of failure can be defined by the following sets where each position is the status of a plane engine: $\{FFFS\}, \{FFSF\}, \{FSFF\}, \{SFFF\}, \{FFFF\}$. This just proves what we know to be true of a binomial distribution. We can calculate it by hand and then validate with a binomial model.

$$4 * 0.4^3(0.6)^1 + 0.4^4 = 0.1792 \rightarrow 1 - 0.1792 = 0.8208$$

In [9]:

```
p_success_4 = binom.pmf(4,4,0.6)
p_success_3 = binom.pmf(3,4,0.6)
p_success_2 = binom.pmf(2,4,0.6)
print("Probability of success for 4 engine plane is %f" %(p_success_2+p_succes
```

```
Probability of success for 4 engine plane is 0.820800
```

Based on this series of calculations, one would experience a greater chance of success with a 2 engine plane than a 4 engine plane

In [ ]: