Binomial Distribution

Instructor, Nero Chan Zhen Yu

Forward School

Sales Probability

- As a sales manager, you analyze the sales records for all salesperson under your supervision.
- William has a success rate of 75% and averages 10 sales calls per day. Peter has a success rate of 45% but averages 16 calls per day.
- What is the probability of each salesperson making 6 sales during any given day?

Forward School

Binomial Experiment

- A binomial experiment has the following characteristics:

1. The process consists of a sequence of n trials
2. Only two exclusive outcomes are possible for each trial. One outcome is called "success" and the other a "failure"
3. The probability of a success denoted by p, does not change from trial to trial. The probability of failure is $1-p$ and is also fixed from trial to trial.
4. The trials are independent; the outcome of previous trials do not influence future trials.

Forward School

Binomial Experiment

- For example, let's consider an experiment involving 5 trials; $n = 5$
- Our interest is in the number of 'successes' in n trials
- We can establish a discrete random variable X to represent the number of successes in our trials

$$X = \#\ number\ of\ successes\ in\ n\ trials$$

$$x = 0,1,2,3,4,5$$

- We could have 0 successes all the way up to 5 successes

#success	#failure
0	5
1	4
2	3
3	2
4	1
5	0

Forward School

Binomial Experiment

Trial	1	2	3	4	5
Outcome	S	F	F	S	F

2 successes, 3 failures

Trial	1	2	3	4	5
Outcome	S	S	F	S	S

4 successes, 1 failure

Forward School

Binomial Experiment

Trial	1	2	3	4	5
Outcome	S	S	F	S	S
Outcome	F	S	S	S	S
Outcome	S	S	S	F	S
Outcome	S	F	S	S	S
Outcome	S	S	S	S	F

These are all valid outcomes with 4 successes and 1 failure

C(5,4) = 5

Forward School

Binomial Experiment (n=5)

n	#successes	#failures	Combination	Total
5	0	5	C (5,0)	1
5	1	4	C (5,1)	5
5	2	3	C (5,2)	10
5	3	2	C (5,3)	10
5	4	1	C (5,4)	5
5	5	0	C (5,5)	1

Forward School

Quality Control Problem

- A (very poor) manufacturer is making a product with a 20% defect rate. If we select 5 randomly chosen items at the end of the assembly line, what is the probability of having 1 defective item in our sample?
- Hint: It is NOT $\frac{1}{5}$ or 0.2 or 20% !
- Note: The success that we are interested in this problem is a defective product

So trials are n=5 and #successes x=1 $C\,(5,1) = 5$

Forward School

Quality Control Problem

Reminder: A success in this problem is a defective product

Trial	1	2	3	4	5
Outcome	F	F	S	F	F
Outcome	S	F	F	F	F
Outcome	F	F	F	S	F
Outcome	F	S	F	F	F
Outcome	F	F	F	F	S

Forward School

Quality Control Problem

Reminder: A success in this problem is a defective product

P = 0.2 (success)
1 - p = 0.8 (failure)

$C(5,1) = 5$

Trial	1	2	3	4	5	Probability
Outcome	0.8	0.8	0.2	0.8	0.8	0.08192
Outcome	0.2	0.8	0.8	0.8	0.8	0.08192
Outcome	0.8	0.8	0.8	0.2	0.8	0.08192
Outcome	0.8	0.2	0.8	0.8	0.8	0.08192
Outcome	0.8	0.8	0.8	0.8	0.2	0.08192

$C(n, x)p^x(1-p)^{n-x}$ $\quad$ $C(5,1) \times 0.2^1 \times (0.8)^4 = 5 \times 0.08192 = 0.4096$

Forward School

General Formula

$$C(n, x)p^x(1 - p)^{n-x}$$

n = number of trials
x = number of successes
p = probability of success in any trial

$n = 5$
$x = 1$
$p = 0.2$

$C(5,1) \; X \; 0.2^1 \; X(1 - 0.2)^{5-1}$

$C(5,1) \; X \; 0.2^1 \; X(0.8)^4$

0.4096 or 41%

```
from scipy.stats import binom
binom.pmf(1,5,0.2)
> 0.40959
```

Probability of 1 success or 1 defective product

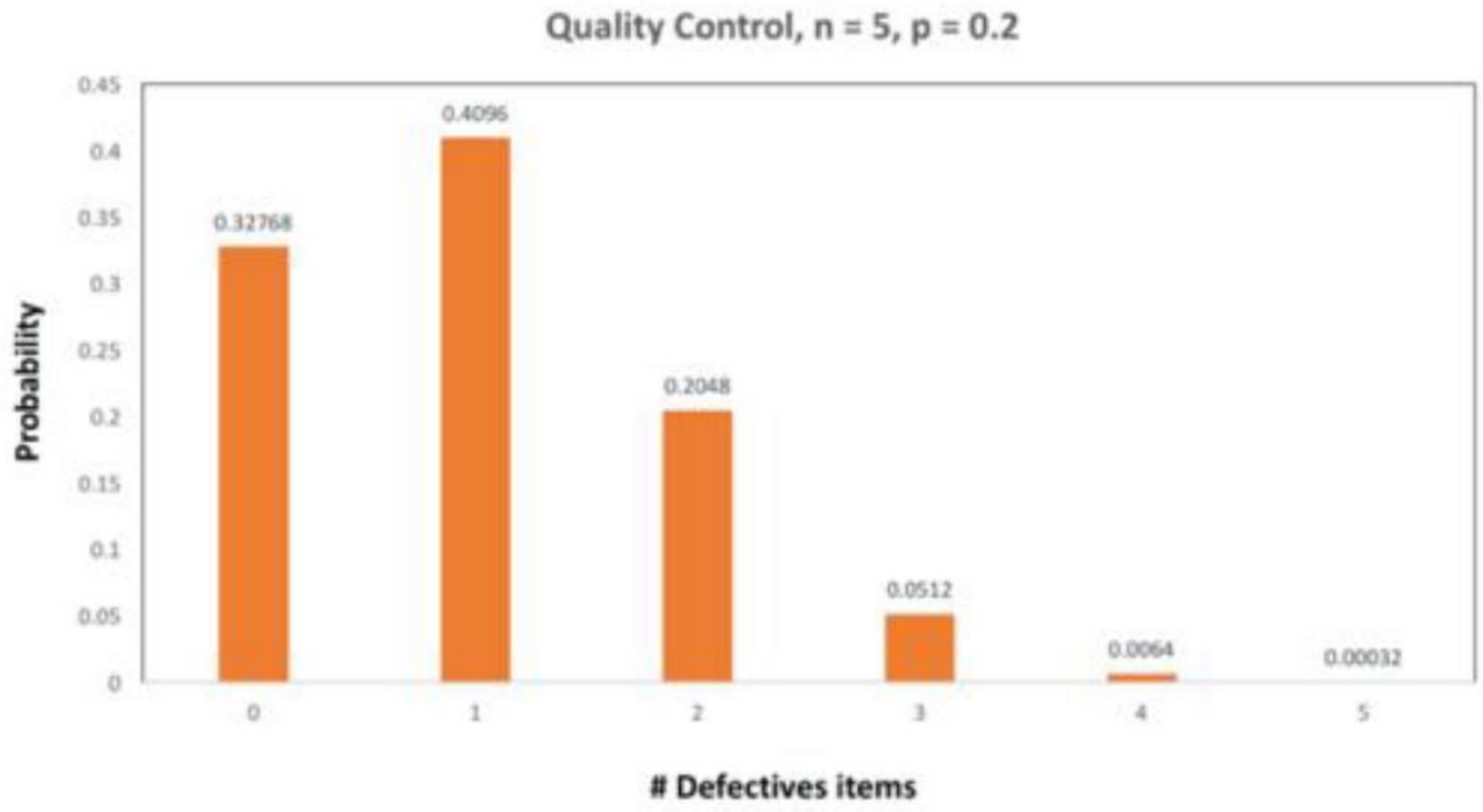
Forward School

Quality Control Problem

n	#successes	#failures	calculation	Probability	#defectives
5	0	5	$C(5,0)\ X\ 0.2^0\ X(0.8)^5$	0.32768	0
5	1	4	$C(5,1)\ X\ 0.2^1\ X(0.8)^4$	0.4096	1
5	2	3	$C(5,2)\ X\ 0.2^2\ X(0.8)^3$	0.2048	2
5	3	2	$C(5,3)\ X\ 0.2^3\ X(0.8)^2$	0.0512	3
5	4	1	$C(5,4)\ X\ 0.2^4\ X(0.8)^1$	0.0064	4
5	5	0	$C(5,5)\ X\ 0.2^5\ X(0.8)^0$	0.00032	5

Forward School

Binomial Distribution



Forward School

Binomial Distribution

- In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n trials is given by the *probability mass function*:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, \ldots, n$, where

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

Forward School

Sales Probability (Revisited)

- As a sales manager, you analyze the sales records for all salesperson under your supervision.
- William has a success rate of 75% and averages 10 sales calls per day. Peter has a success rate of 45% but averages 16 calls per day.
- What is the probability of each salesperson making 6 sales during any given day?

Sales Probability - William

$$C(n,\ x)p^{x}(1\ -\ p)^{n-x}$$

n = number of trials $\quad$ $n = 10$
x = number of successes $\quad$ $x = 6$
p = probability of success in any trial $\quad$ $p = 0.75$

$C(10,5)\ X\ 0.75^{6}\ X\ (\ 1 - 0.75)^{10\text{-}6}$

$C(10,15)\ X\ 0.75^{6}\ X\ (0.25)^{4}$

0.146 or 15%

```
from scipy.stats import binom
binom.pmf(6,10,0.75)
> 0.1459
```

Probability of 6 sales

Forward School

Sales Probability - Peter

$$C(n,\ x)p^x(1\ -\ p)^{n-x}$$

n = number of trials
x = number of successes
p = probability of success in any trial

$n = 16$
$x = 6$
$p = 0.45$

$C(16,6)\ X\ 0.45^6\ X\ (\ 1 - 0.45)^{16-6}$

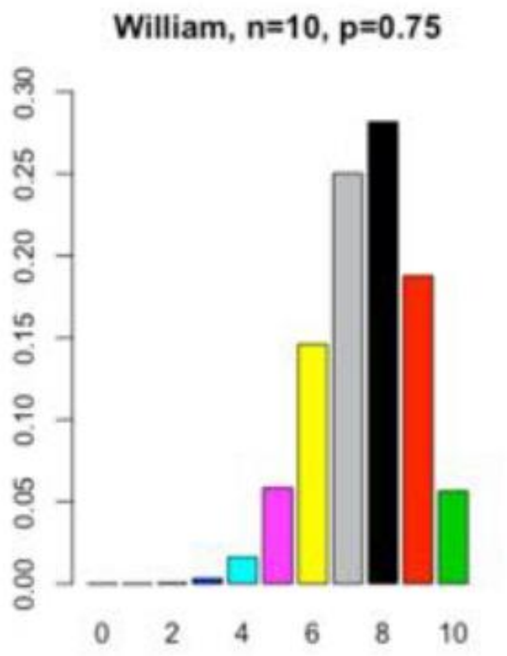
$C(16,6)\ X\ 0.45^6\ X\ (0.55)^6$

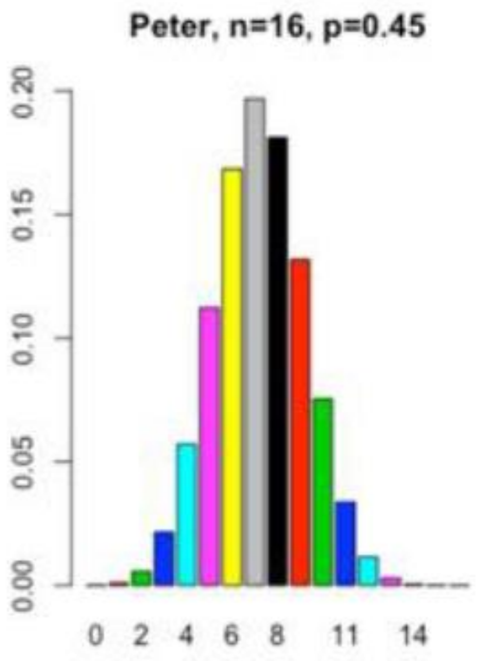
0.168 or 17%

```
from scipy.stats import binom
binom.pmf(6,16,0.45)
> 0.1684
```

Probability of 6 sales

Forward School





William $(x = 6)$ = 0.146

Peter $(x = 6)$ = 0.168

What about probability of making 8 sales in any given day?

William $(x = 8)$ = 0.282

Peter $(x = 8)$ = 0.181

Forward School

Conclusion

- Even though Peter has a lower success rate in terms of making a sale, his greater number of sales calls per day overcomes the lower success rate.
- Therefore, the greater probability of making 6 sales per day belongs to Peter, even though William’s success rate is higher.

Forward School

Binomial Distributions

Mean and Standard Deviation

Example

As a sales manager, you analyse the sales records for all salesperson under your guidance

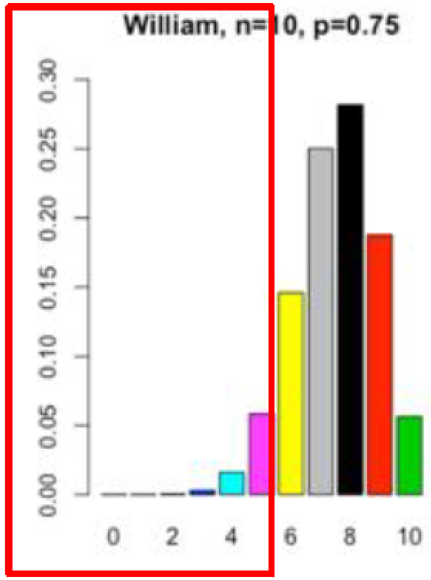
William has a success rate of 75% and averages 10 sales called per day. Peter has a success rate of 45% but averages 16 calls per day.

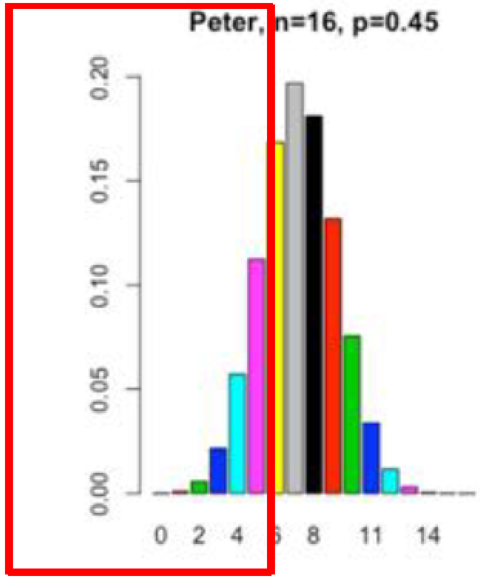
What is the probability that each salesperson makes at least **6 sales on any given day**?

Forward School

Question

What is the mean and standard deviation #daily sales for each employee?





n = number of trals (calls)

p = probability of success in any call

q = (1-p) probability of failure in any call

	n	p	q
William	10	0.75	0.25
Peter	16	0.45	0.55

Forward School

Binomial Mean (Expected) Value

$$\mu = n \cdot p$$

	n	p	q
William	10	0.75	0.25
Peter	16	0.45	0.55

William's mean # daily sales : $\mu = 10 \times 0.75 = 7.5$

Peter's mean # daily sales : $\mu = 16 \times 0.45 = 7.2$

Therefore we expect 7.5 sales and 7.2 sales per day respectively

Forward School

Binomial Standard Deviation

$$\sigma = \sqrt{n.p.q}$$

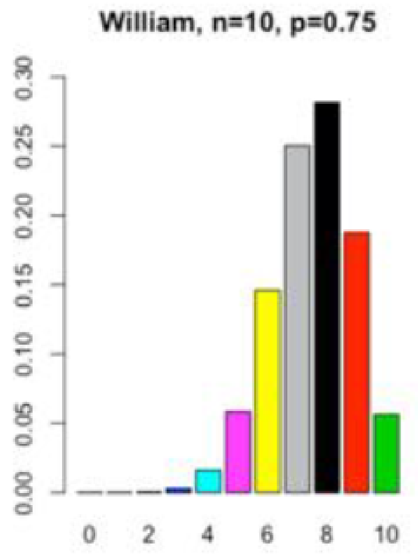
	n	p	q
William	10	0.75	0.25
Peter	16	0.45	0.55

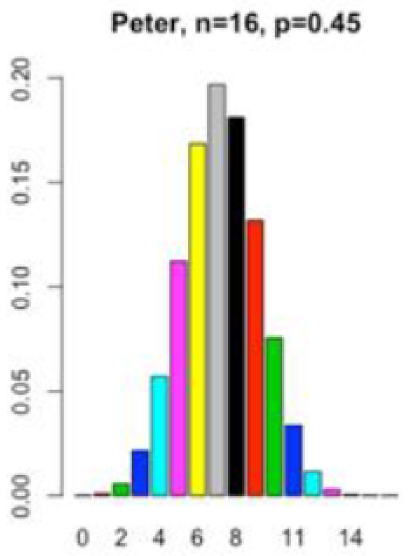
William's standard deviation # daily sales : $\sigma = (10 \text{ X } 0.75 \text{ X } 0.25)^{½} = 1.37$

Peter's standard deviation # daily sales : $\sigma = (16 \text{ X } 0.45 \text{ X } 0.55)^{½} = 1.99$

Forward School

Mean and Standard Deviation for each employee





	μ	σ
William	7.5 daily sales	1.37 daily sales
Peter	7.2 daily sales	1.99 daily sales

Forward School

Example

In June 2013, a poll conducted on where 46% of Americans said the following statement most closely matched their views:

"Human form that we are now has been similar for the past 10,000 years or so"

If you would to replicate this survey in 2014, with sample of 78 similar people, what is expected numbers of respondents that would agree with this statement?

Forward School

Example – mean & standard deviation

$\mu = n \cdot p$

$\mu = 78 \times 0.46$

$= 35.88$

$= 36$

$$\sigma = \sqrt{n.p.q}$$

$\sigma = 78 \times 0.46 \times 0.54$

$\sigma = 4.40$

Forward School

$$\sum(prob) = 1$$

0 → 1 **Cumulative Probability**

$X < 5$

$X \geq 6 \quad = 1 - (X < 6)$

Forward School

Cumulative Probability

P (0 ≤ X ≤ 5)

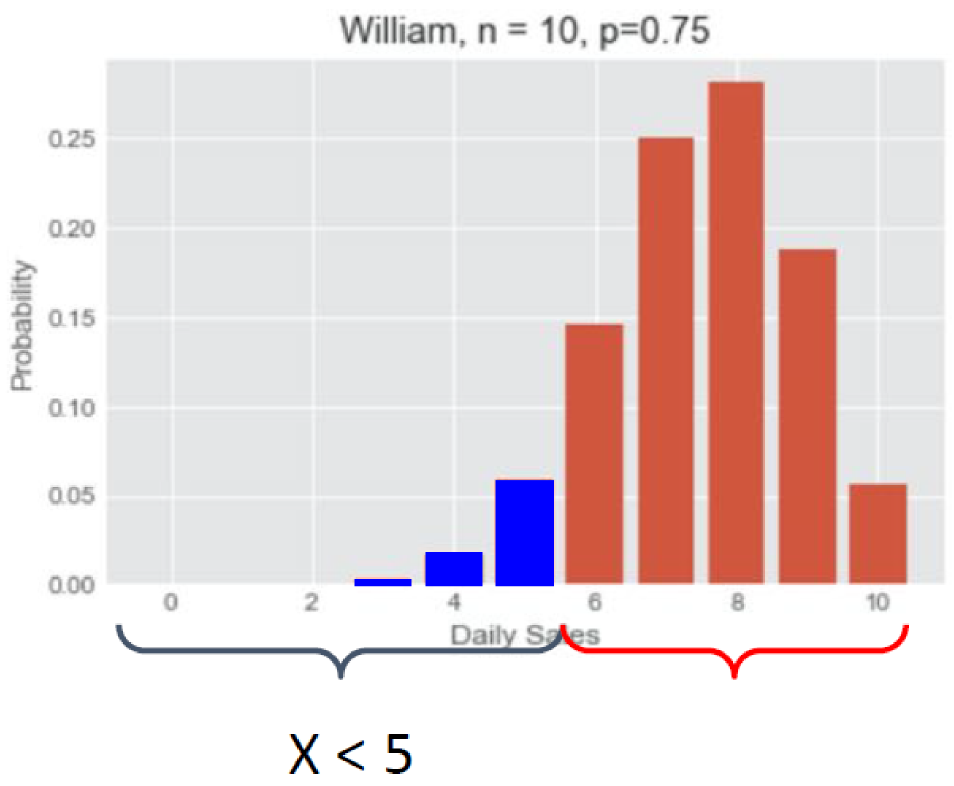
```python
from scipy.stats import binom
binom.cdf(5,10,0.75)
> 0.078
```

P (0 ≤ X ≤ 5) = 0.078

P (X ≥ 6)

```python
from scipy.stats import binom
1 - binom.cdf(5,10,0.75)
> 0.922
```

P (X ≥ 6) = 0.922



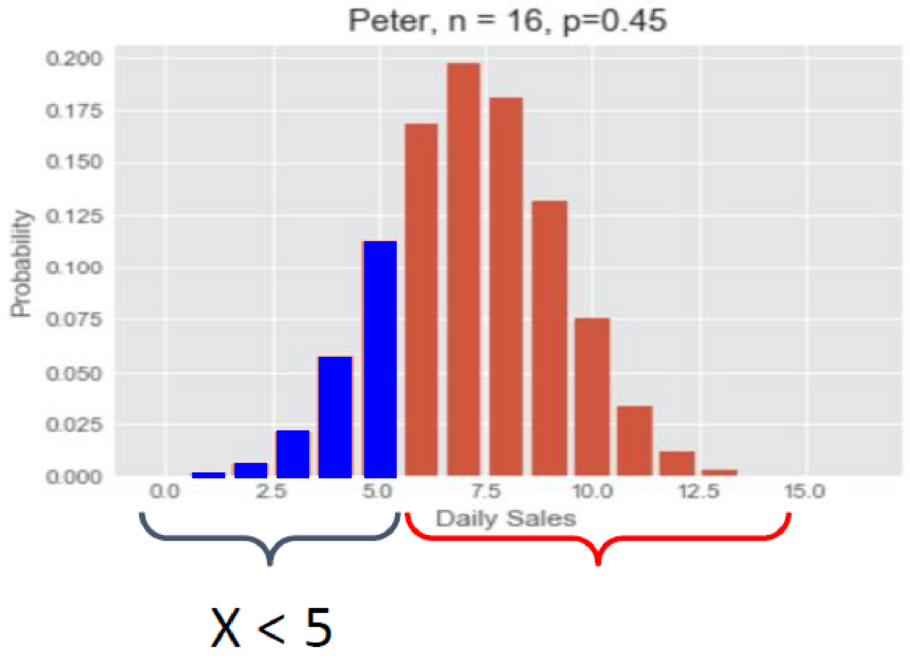
Forward School

Cumulative Probability

$P\ (0 \leq X \leq 5)$

```
from scipy.stats import binom
binom.cdf(5,16,0.45)
> 0.198
```

$P\ (0 \leq X \leq 5) = 0.198$



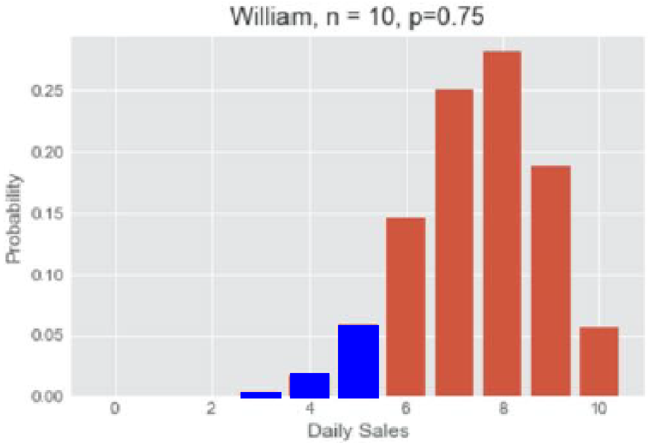
$P\ (X\ \geq 6)$

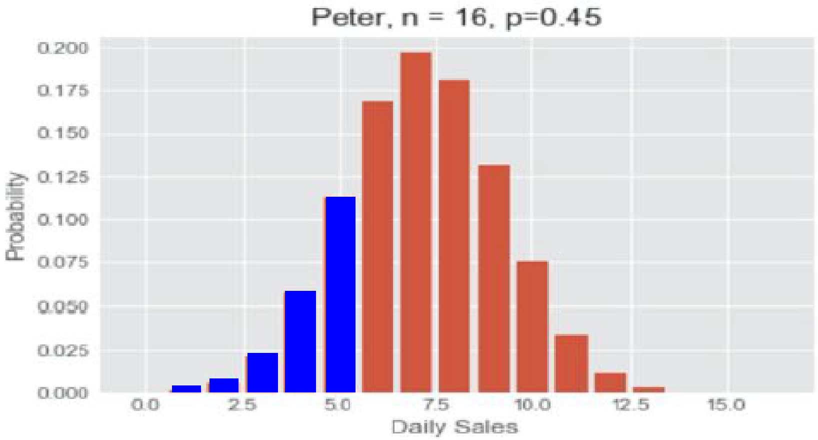
```
from scipy.stats import binom
1 - binom.cdf(5,16,0.45)
> 0.802
```

$P\ (X\ \geq 6) = 0.802$

Forward School

What is the probability of at least 6 daily sales for each person?





$P\ (X\ \geq 6) = 0.922$

$P\ (X\ \geq 6) = 0.802$

Therefore if daily sales quota is 6, William has a better chance of meeting that quota even if it appears that he is doing “less work”

Forward School

Binomial Distribution in scipy

- scipy.stats.binom
- The probability mass function for binom is:

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

```
binorm.pmf(k, n, p, loc=0) Probability Mass Function
binorm.cdf(k, n, p, loc = 0) Cumulative Density Function
binorm.ppf(q, n, p, loc = 0) Percent Point Function
binorm.rvs(n, p, loc=0, size=1, random_state=None) Generate
random normal numbers
```

Forward School

`binorm.pmf(k, n, p, loc=0)`

- `binom.pdf` returns binomial probabilities. There are three required arguments: the value (s) for which to compute the probability (k), the number of trials (n), and the success probability for each trial (k).
- For example, here we find the complete distribution when n=5 and p = 0.1.

```
from scipy.stats import binom
import numpy as np
binom.pmf(np.arange(0,5),5,0.1)
```

Forward School

`binorm.cdf(k, n, p, loc = 0)`

- The function `binom.cdf` is useful for summing consecutive binomial probabilities. With n = 5 and p = 0.1, here are some example calculations.

$\Pr\{X \le 2\} =$ `binom.cdf (2,5,0.1)` = 0.99144

$\Pr\{X \ge 3\} = 1 - \Pr\{X \le 2\} =$ `1-binom.cdf (2,5,0.1)` = 0.00856

$\Pr\{1 \le X \le 3\} = ?$

An alternative is to use binom.pmf in conjunction with sum

$\Pr\{X \le 2\} =$ `sum(binom.pmf(np.arange(0,3),5,0.1))`

$\Pr\{X \ge 3\} =$ `sum(binom.pmf(np.arange(3,6),5,0.1))`

$\Pr\{1 \le X \le 3\} =$ `sum(binom.pmf(np.arange(1,4),5,0.1))` = 0.409

Forward School

binorm.ppf(q, n, p, loc = 0)

- We can also find the quantiles of a binomial distribution. For example, here is the 90th percentile of a binomial distribution with n = 200 and p = 0.3. The function `binom.ppf` finds the quantile.

```
binom.ppf(0.9, 200, 0.3)
[1] 68
```

Forward School

```
binorm.rvs(n, p, loc=0,
size=1, random_state=None)
```

- The last function for the binomial distribution is used to take random samples. Here is a random sample of 20 binomial random variables drawn from the binomial distribution with n = 10 and p = 0.5.

```
binom.rvs(10, 0.5, size=20, random_state=1)
[1] array([5, 6, 0, 4, 3, 3, 4, 4, 5, 5, 5, 6, 4, 7, 2,
6, 5, 5, 3, 4])
```

Forward School

Exercise

- Suppose there are twelve multiple choice questions in an History class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.