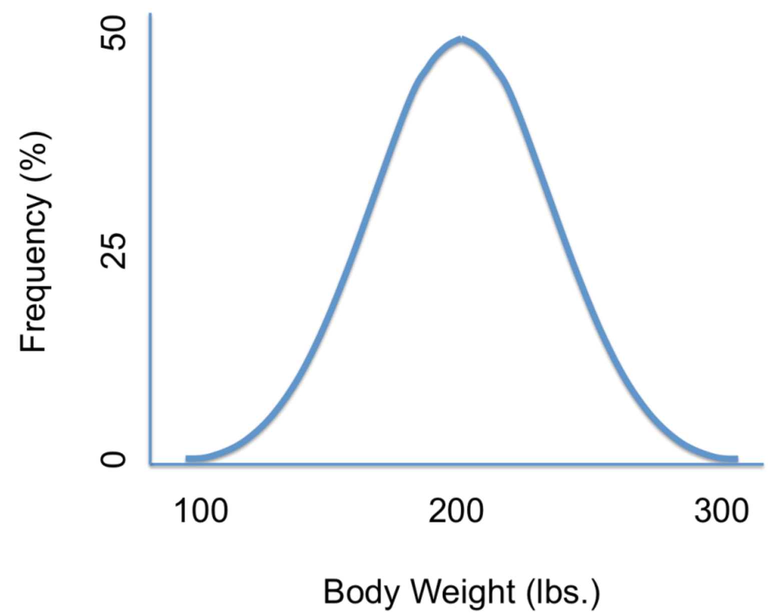
Continuous Probability Distributions



What did we learn form the last class?

Getting a grasp on data

Populations and Samples

- Probabilities
- Discrete distributions

Remember what we did learn

- Discrete probability
  - Random variables and probability distribution
    - How many dogs do you have ?
  - Bivariate distribution
    - How many dogs and cats do you have ?
  - Binomial distribution
    - Flip a coin ten times and count the number of heads
  - Poisson distribution
    - Number of events occurring within a given interval

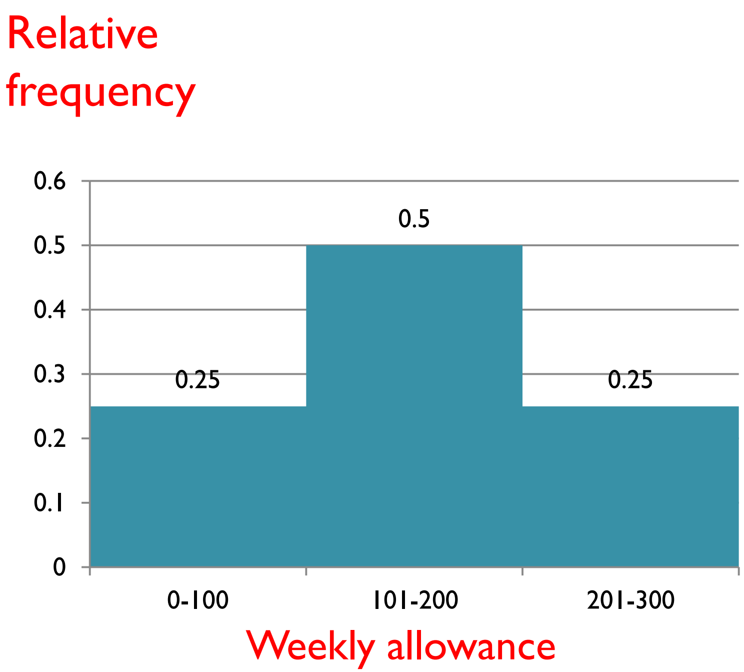
To be complete we must look at continuous probability distributions

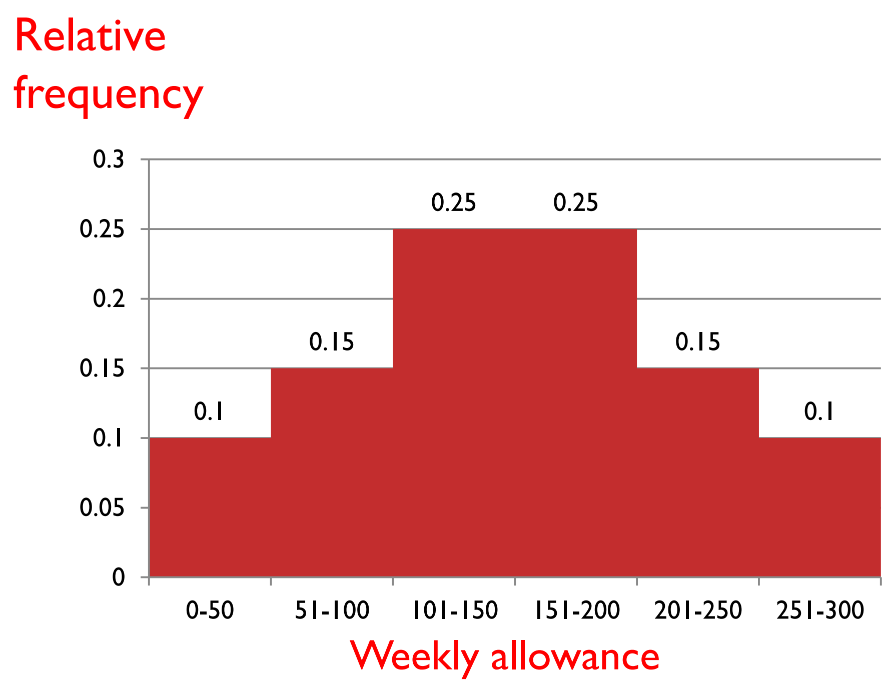
- Uniform distribution
- Normal distribution (z distribution)
- Sampling distribution
  - Mean
  - Proportion

Your weekly allowance

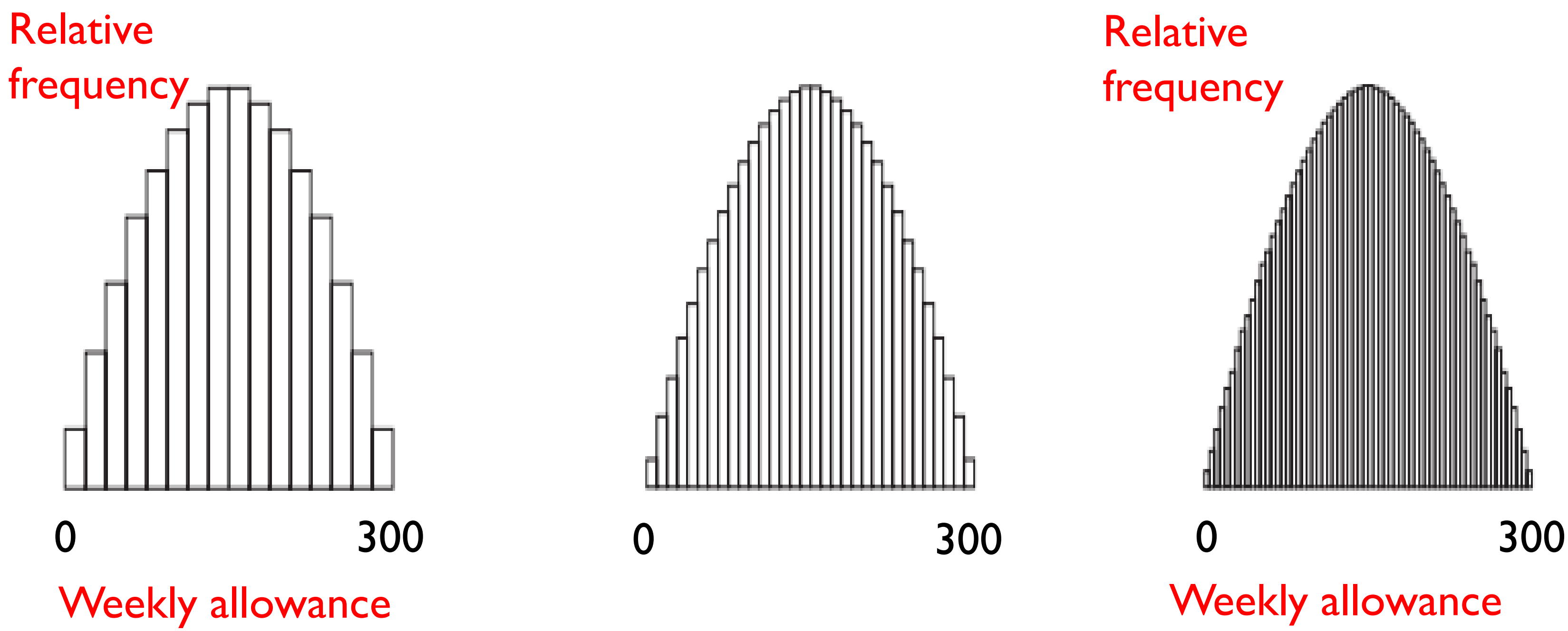
General Results

Detailed Results





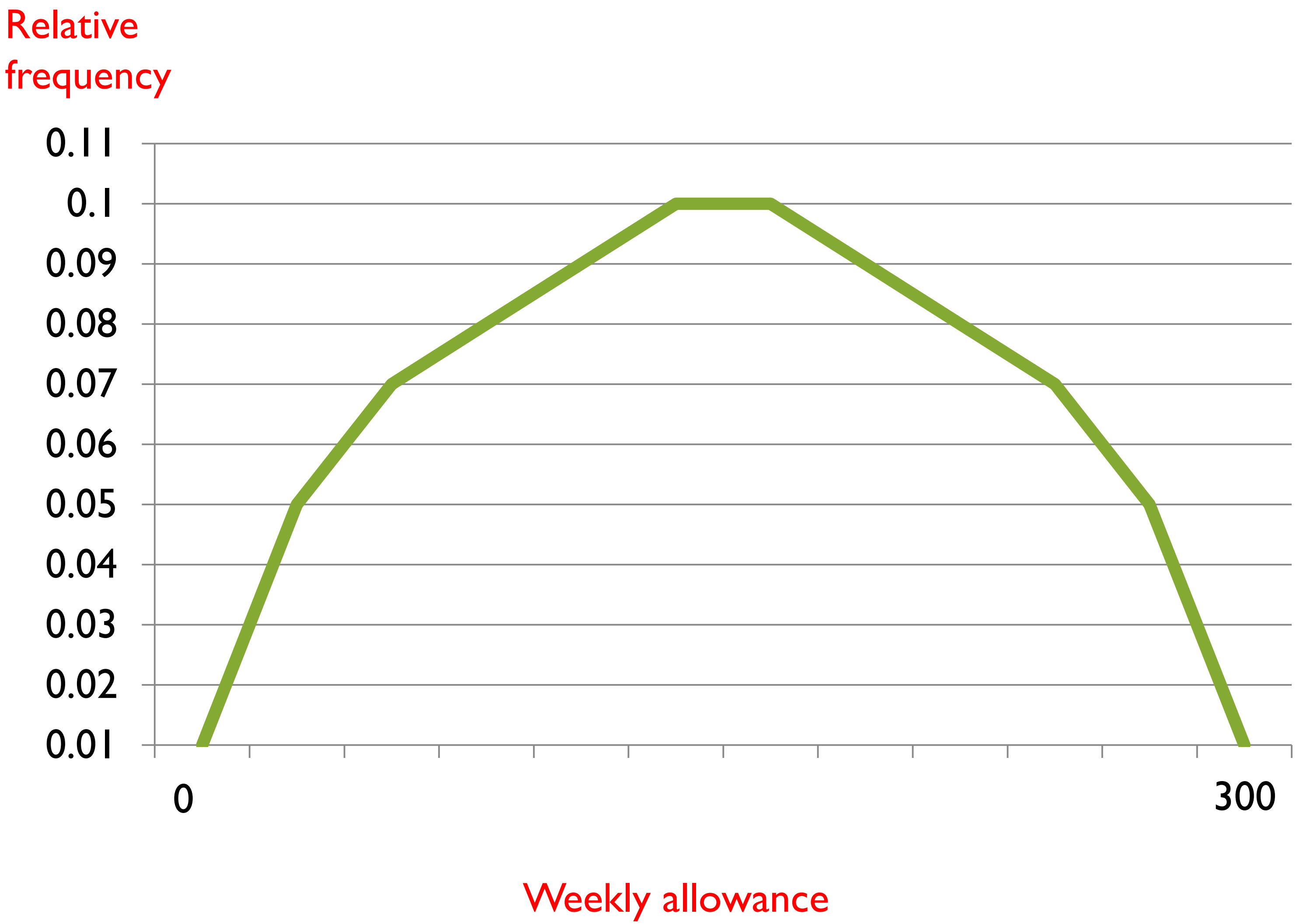
Probability density functions



Continuous

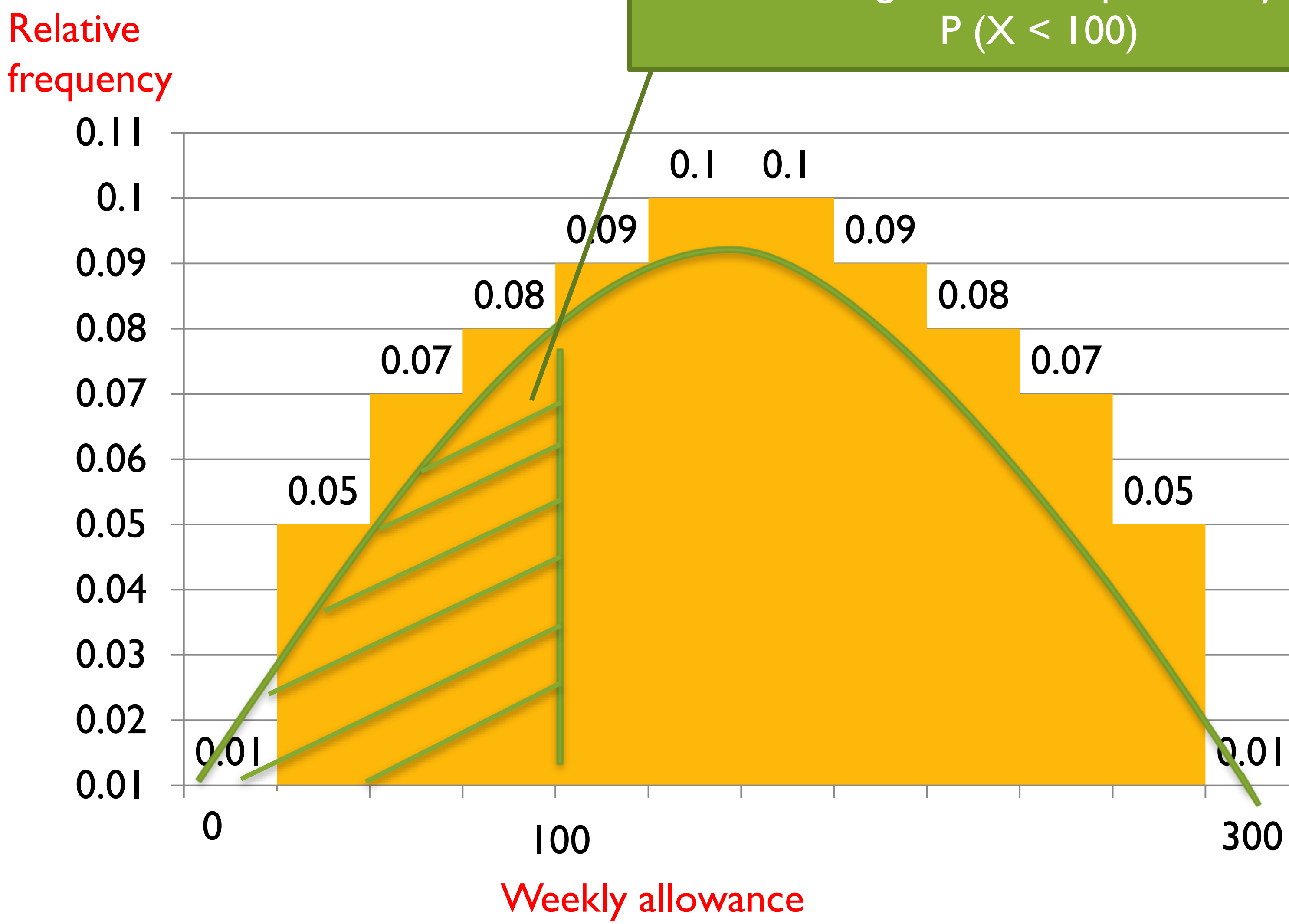
Uncountable

Probability density functions



Probability density functions

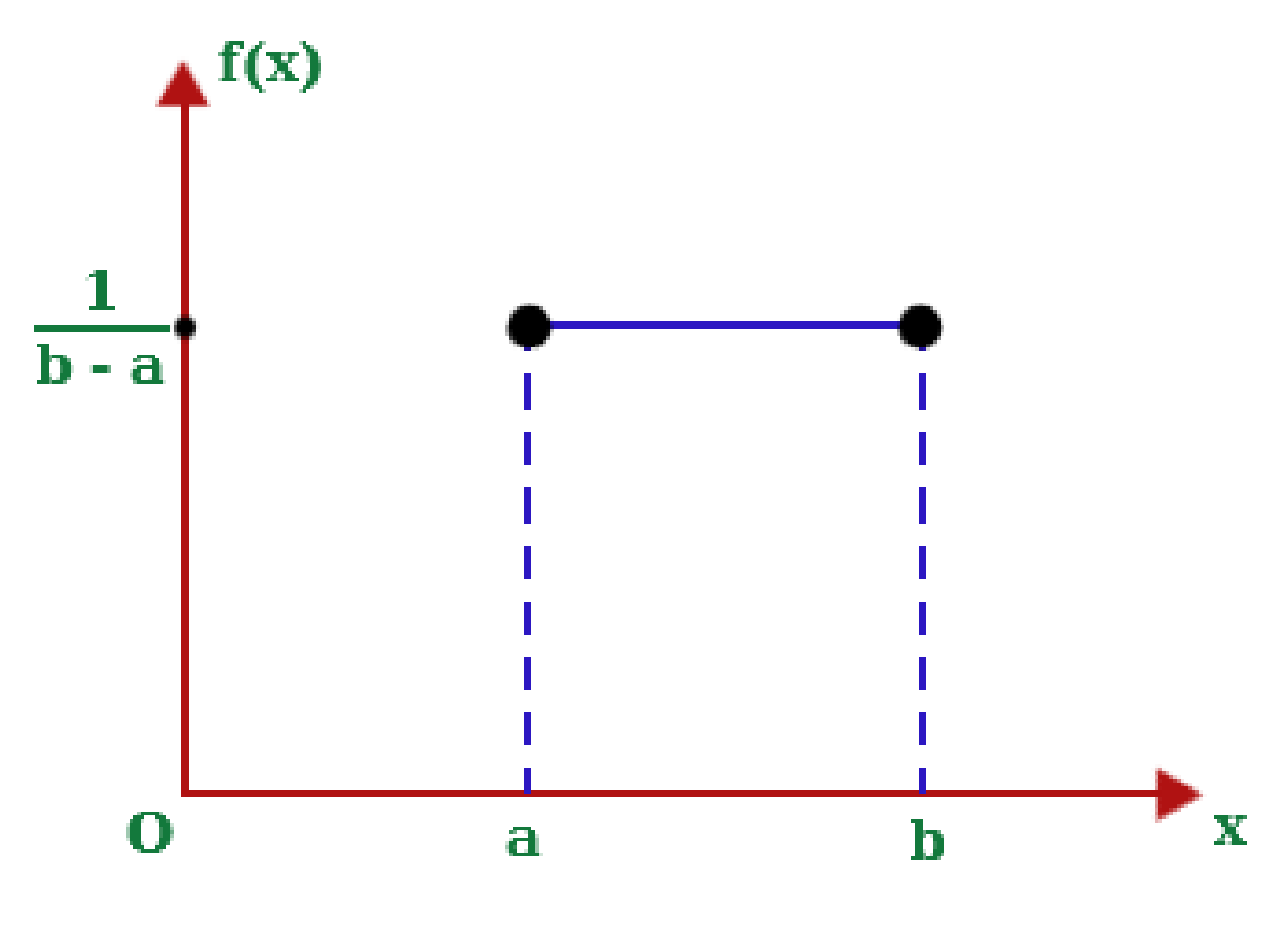
The green area is probability
P (X < 100)



Common continuous probability distribution

- Uniform distribution
- Normal distribution (z distribution)
- Sampling distribution

UNIFORM DISTRIBUTION

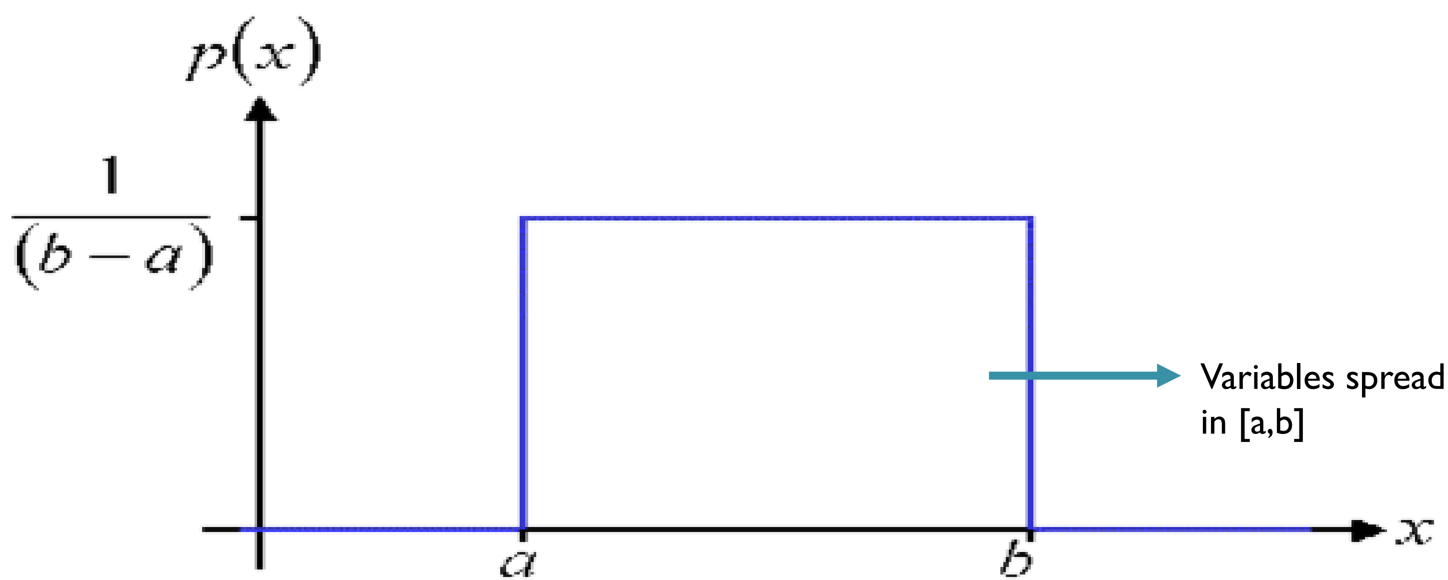


Uniform distribution

- A probability distribution of n random variables spread over an interval [a, b]
- The probability of each one of x variables is 1/(b-a)
- The total area under the rectangular between a and b is 1
- Since it is a special type of continuous probability, the probability of any single value of x is equal to 0

Uniform distribution

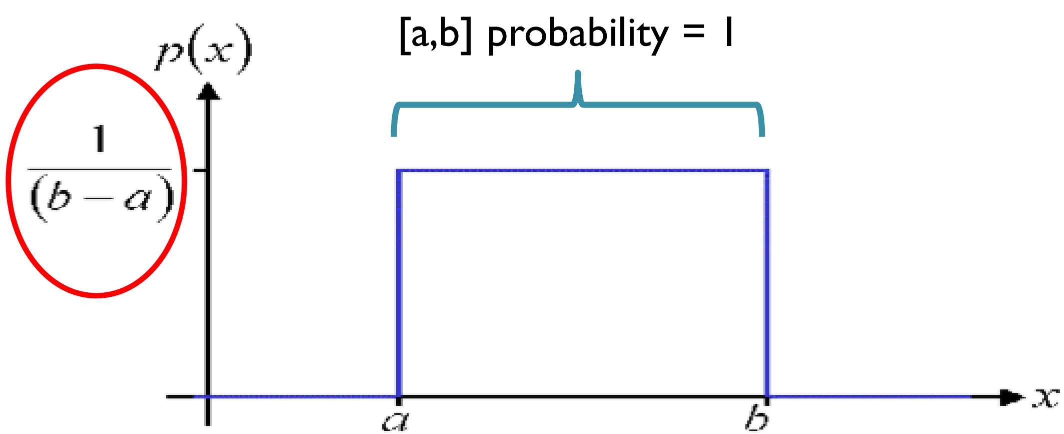
- A probability distribution of n random variables spread over an interval [a, b]



Uniform distribution

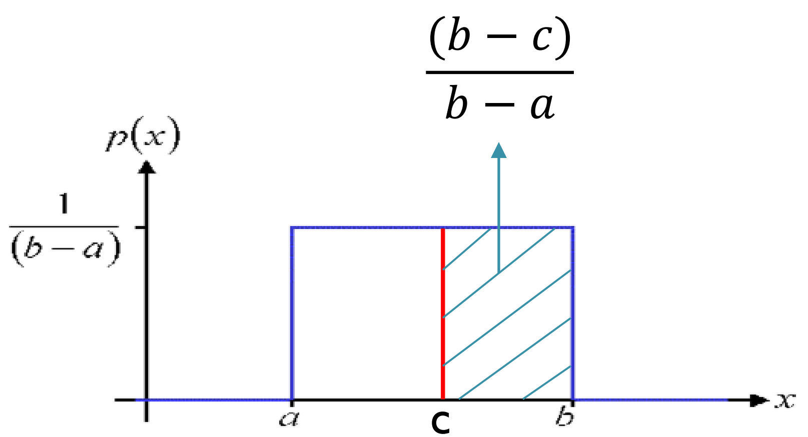
- The total area under the rectangular between a and b is 1
- The probability of each one of x variables is 1/(b-a)

The probability of each one of x variables

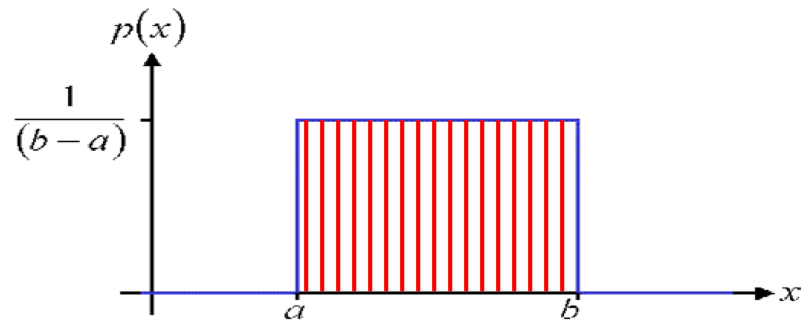


Uniform distribution

- Since it is a special type of continuous probability, the probability of any single value of x is equal to 0



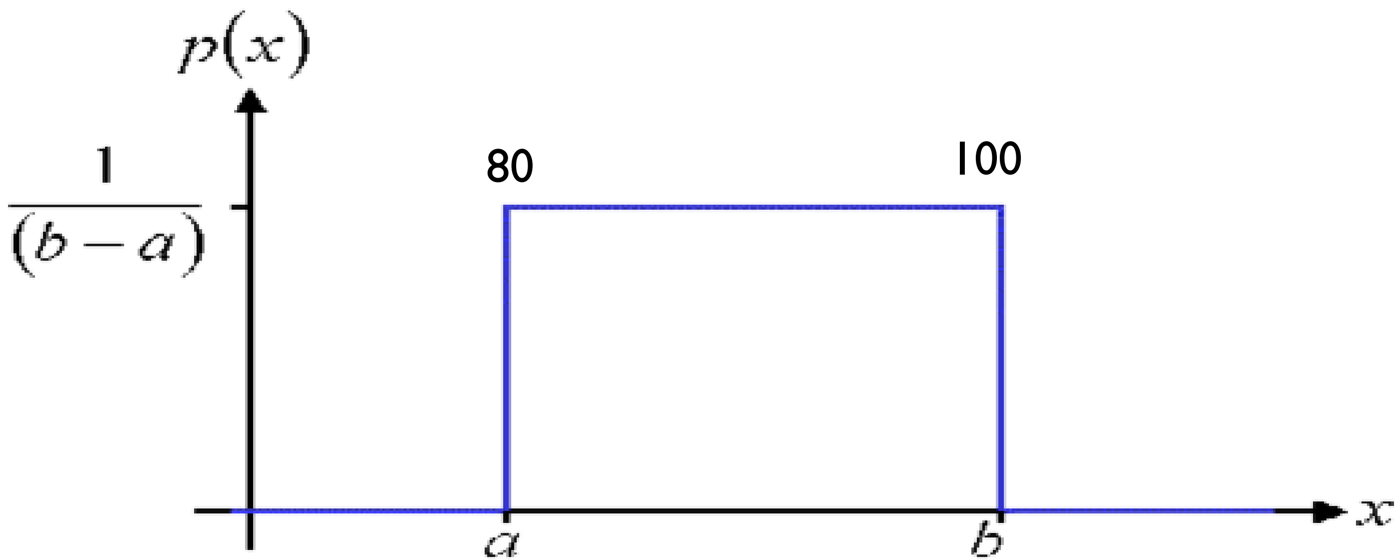
If the data is densely distributed…



probability of any single value of x is equal to 0

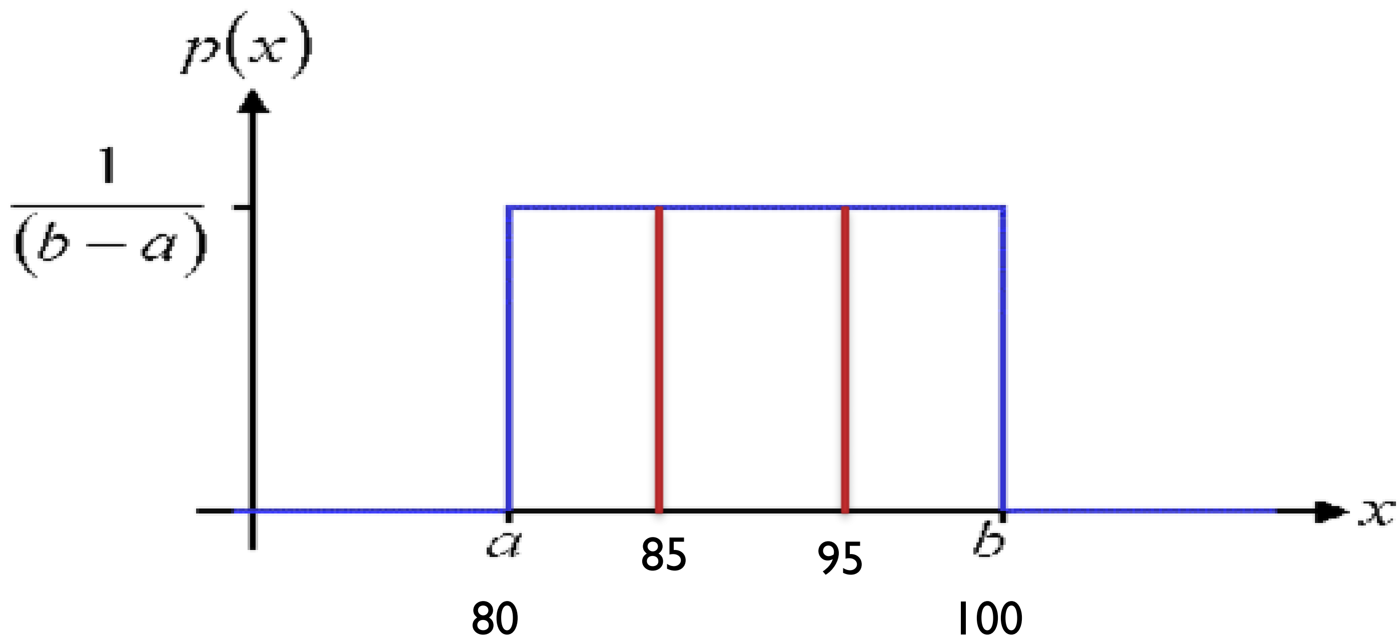
Uniform distribution

Uniform distribution is a probability distribution in which all the outcomes are expected to occur equally.



What is the p (x < 90) ?

Uniform distribution

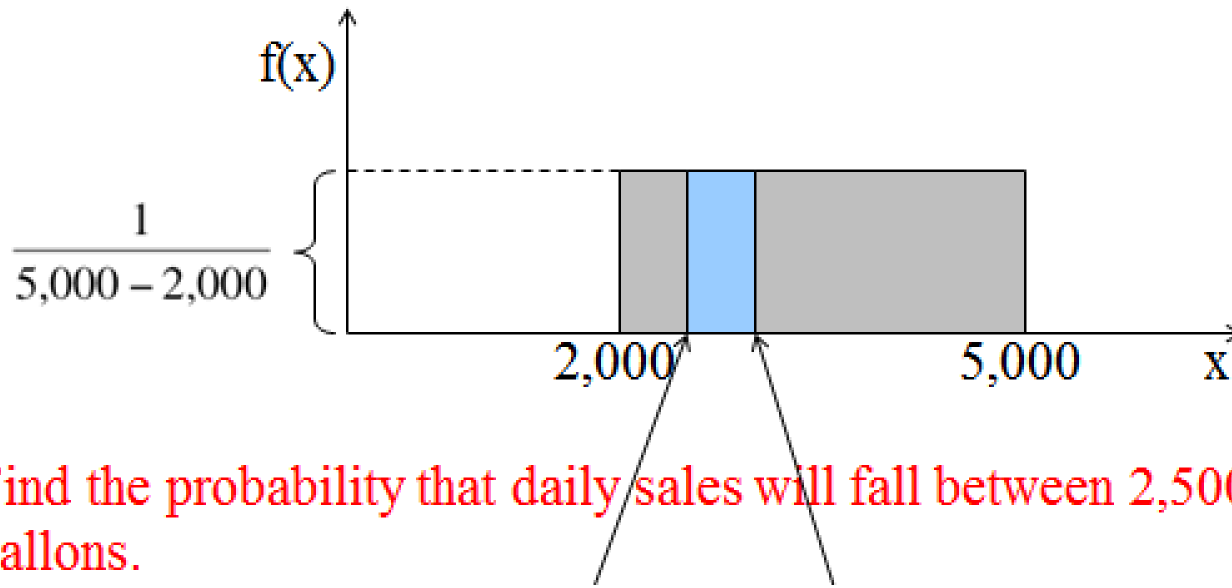


What is the p (85< x < 95) ?

What is the p (x = 87) ?

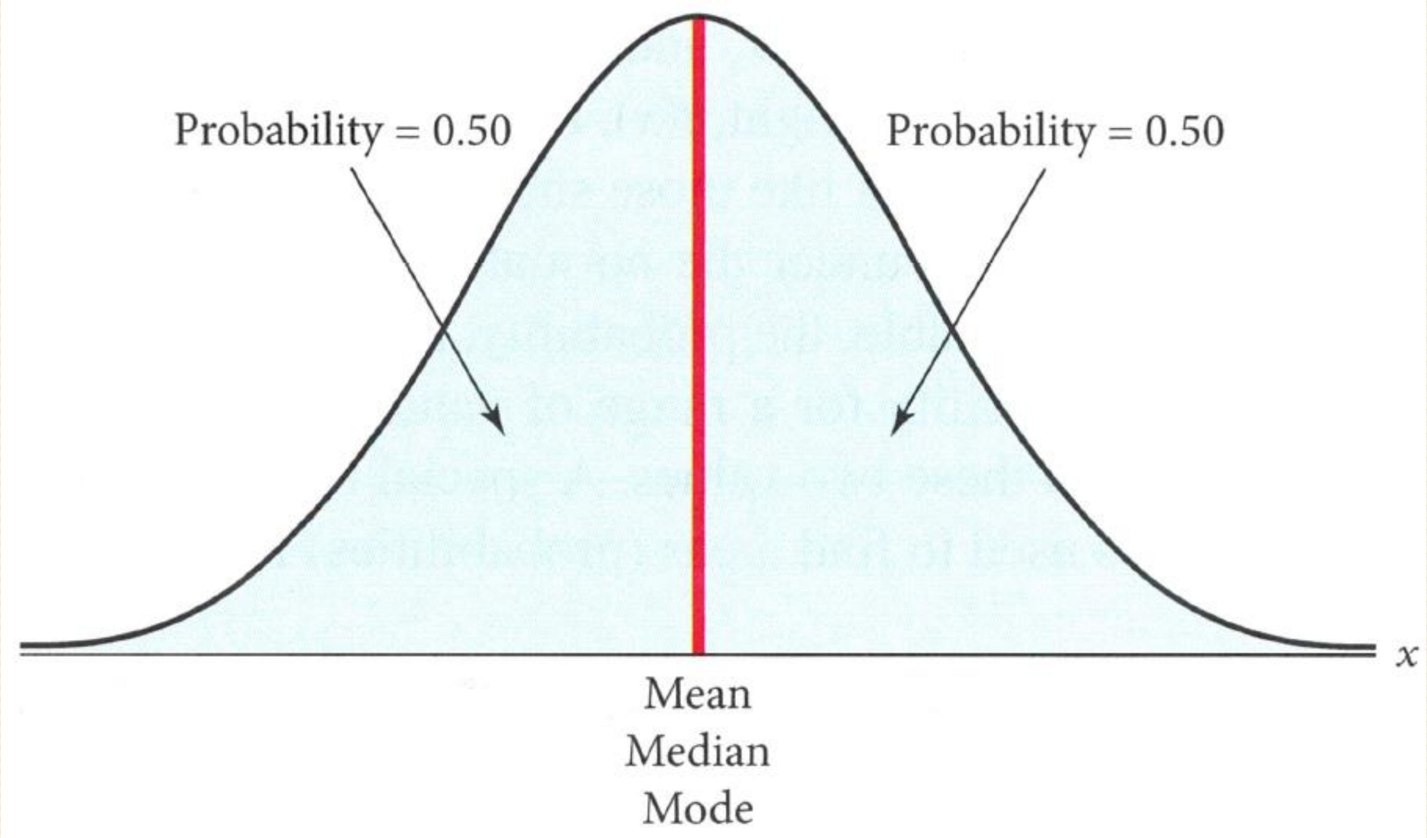
Problem

- The amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.



Find the probability that daily sales will fall between 2,500 and 3,000 gallons.

NORMAL DISTRIBUTION

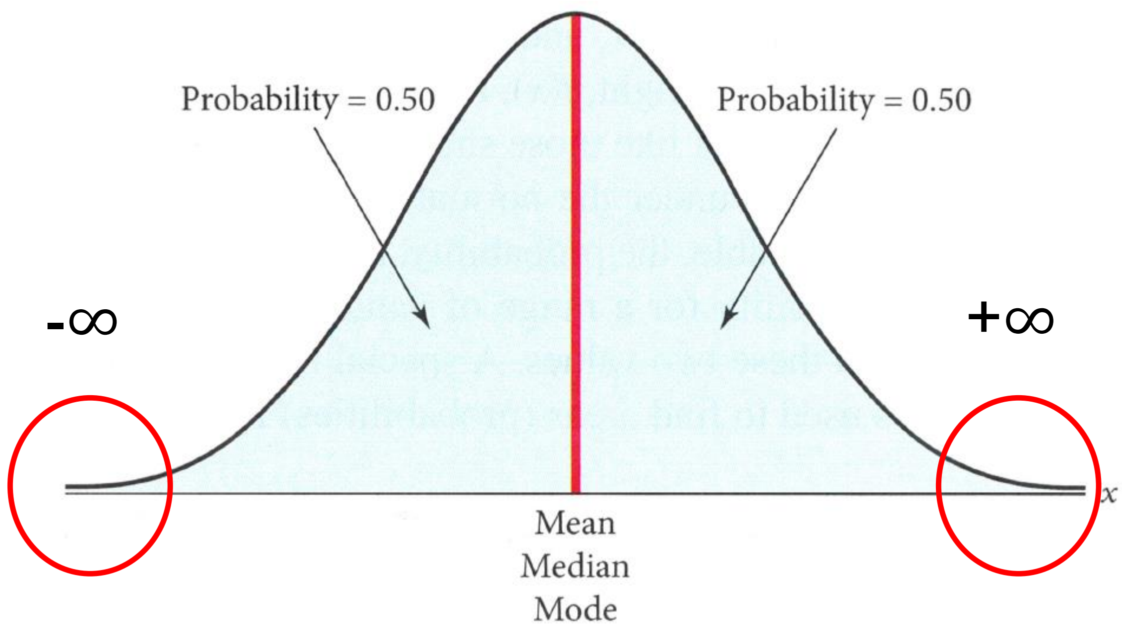


Normal distribution

- Goes to infinity in each direction and often used as an approximation
- It is symmetrical; half the area is to the right of the mean, half to the left. Mean, median, and mode are in the center and equal
- The amount of variation in the random variable determines the height and spread of the normal distribution
- Since it is a special type of continuous probability, the probability of any single value of x is equal to 0

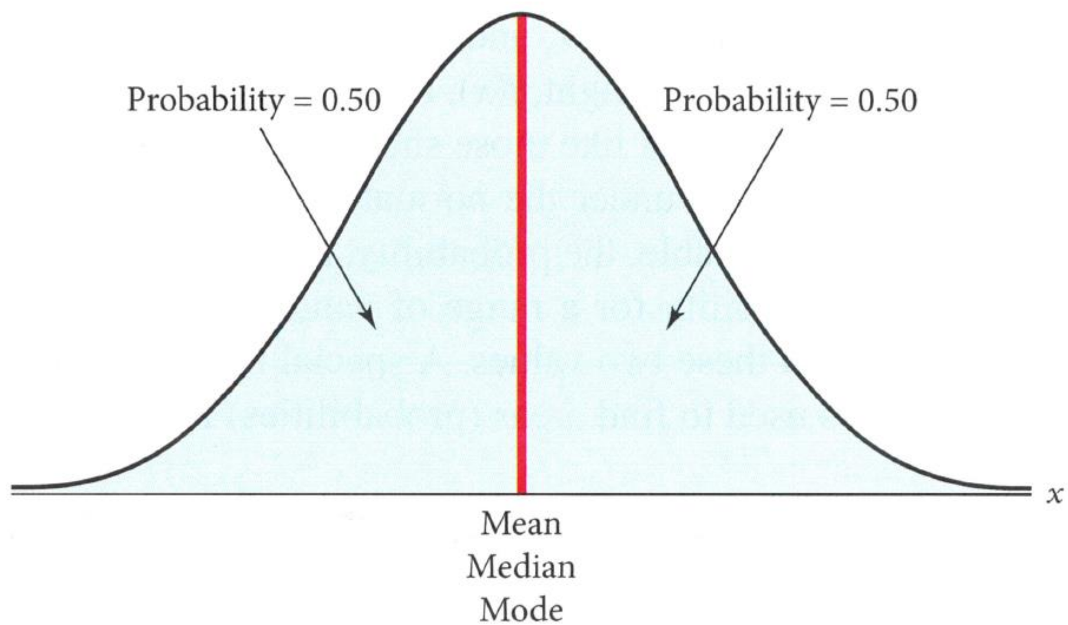
Normal distribution

- Goes to infinity in each direction and often used as an approximation



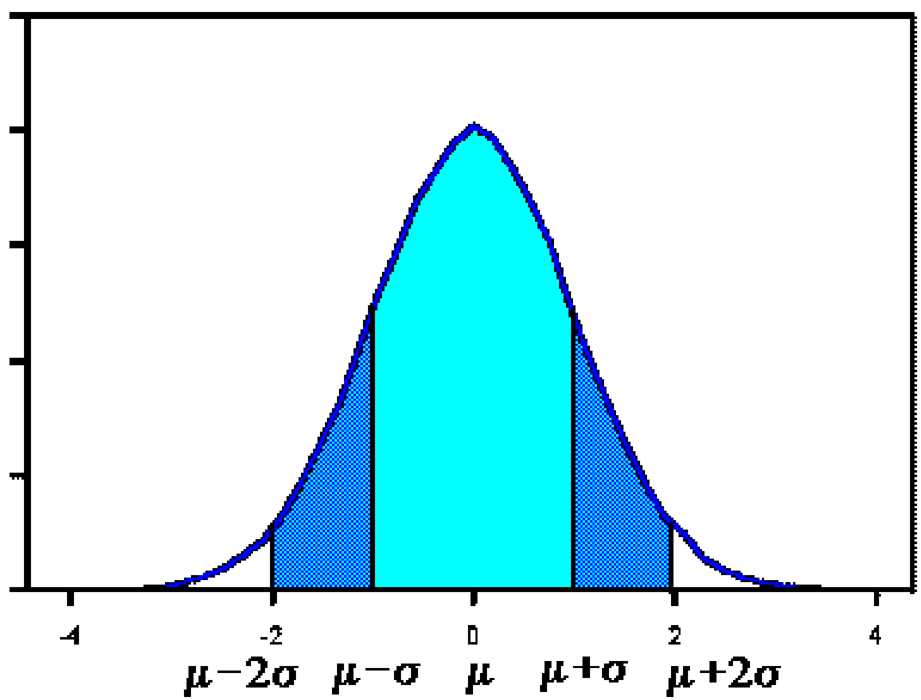
Normal distribution

- It is symmetrical; half the area is to the right of the mean, half to the left. Mean, median, and mode are in the center and equal



Normal distribution

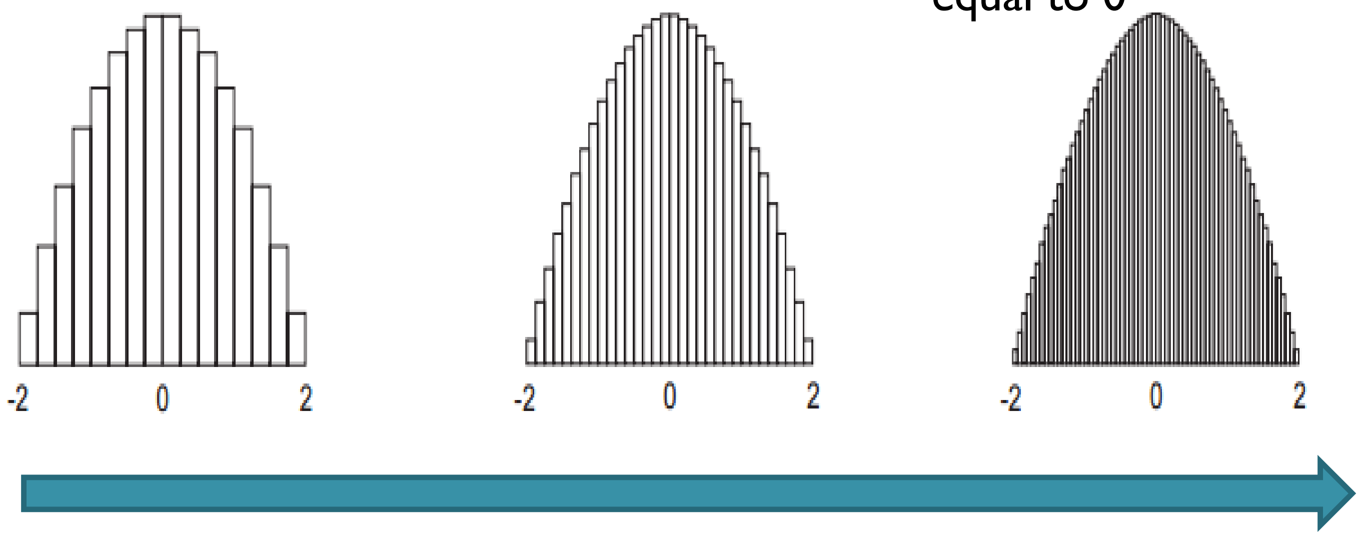
- The amount of variation in the random variable determines the height and spread of the normal distribution



Normal distribution

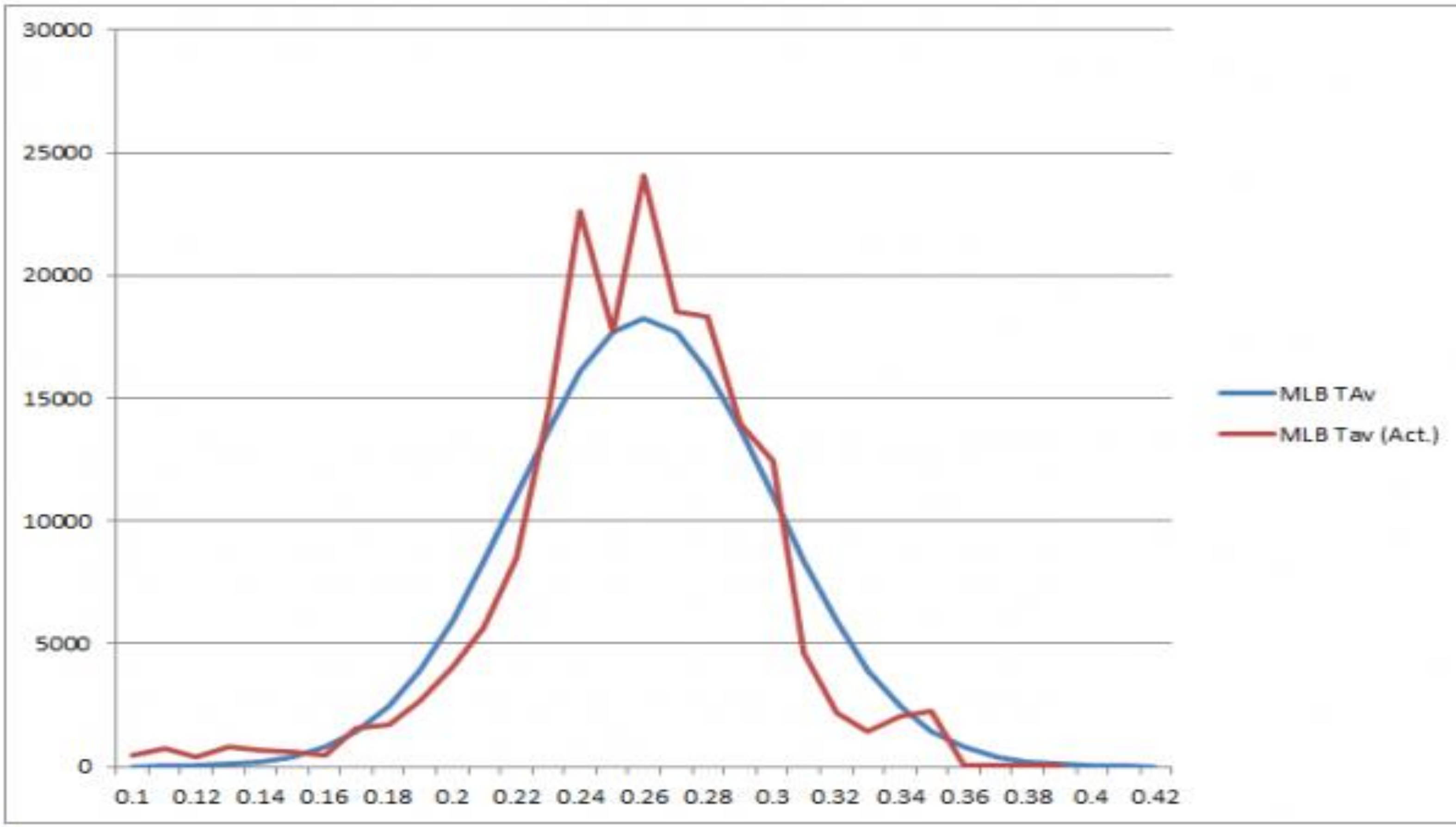
- Since it is a special type of continuous probability, the probability of any single value of x is equal to 0

the probability of any single value of x is equal to 0



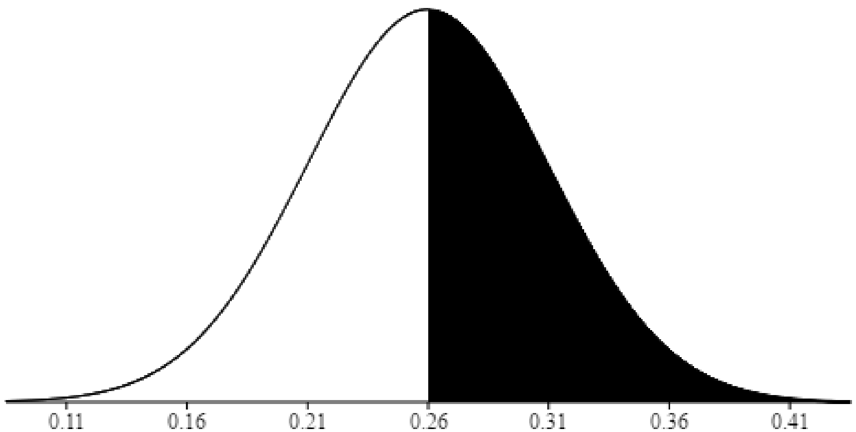
When the interval is divided into a smaller piece forever, we can not find the probability of a particular value

Normal distribution – Major League Baseball (MLB) batting average 1985-2013



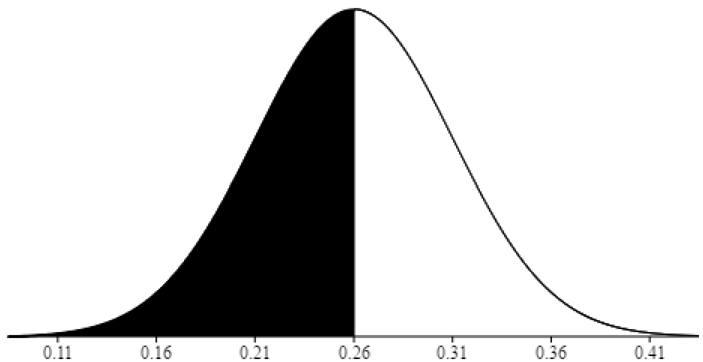
MLB official data: Mean = 0.26 Std = 0.05

Normal distribution – MLB batting average



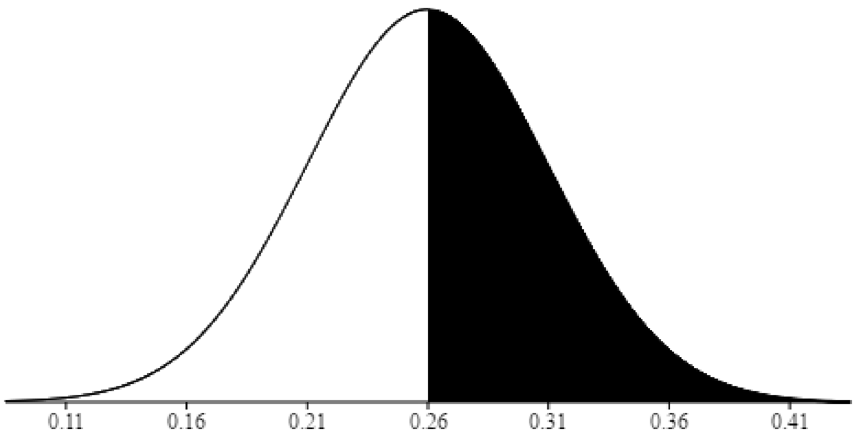
What is the p (0.26≤ x≤∞)

Normal distribution – MLB batting average

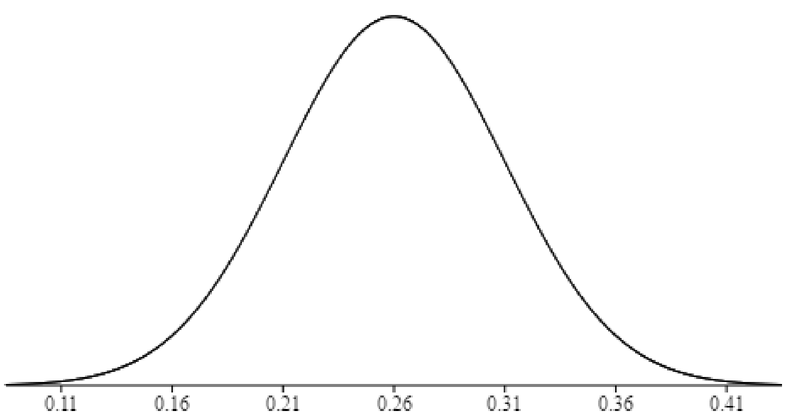


What is the
p ($-\infty \leq x \leq 0.26$)

Normal distribution – MLB batting average

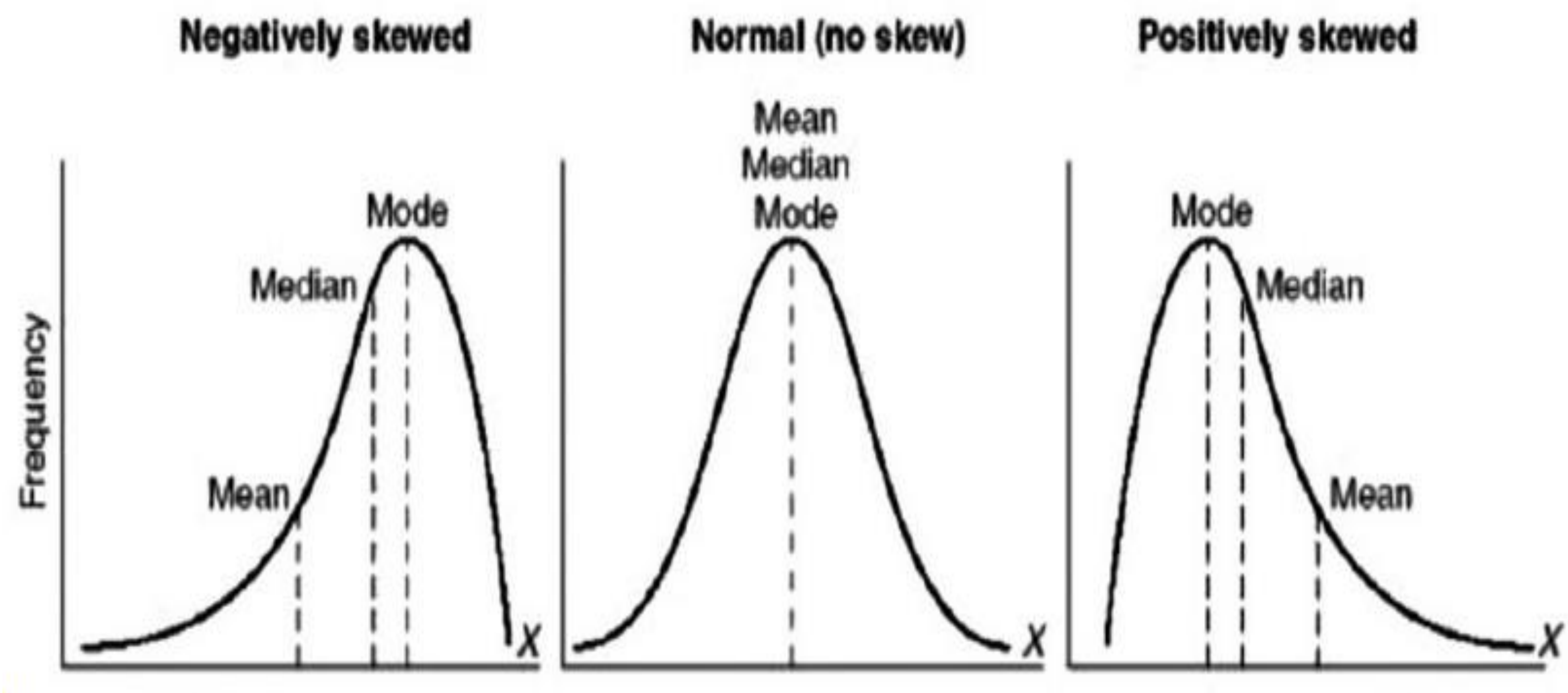


What is the p (∞≥x>0.26)



What is the p (x=0.31)

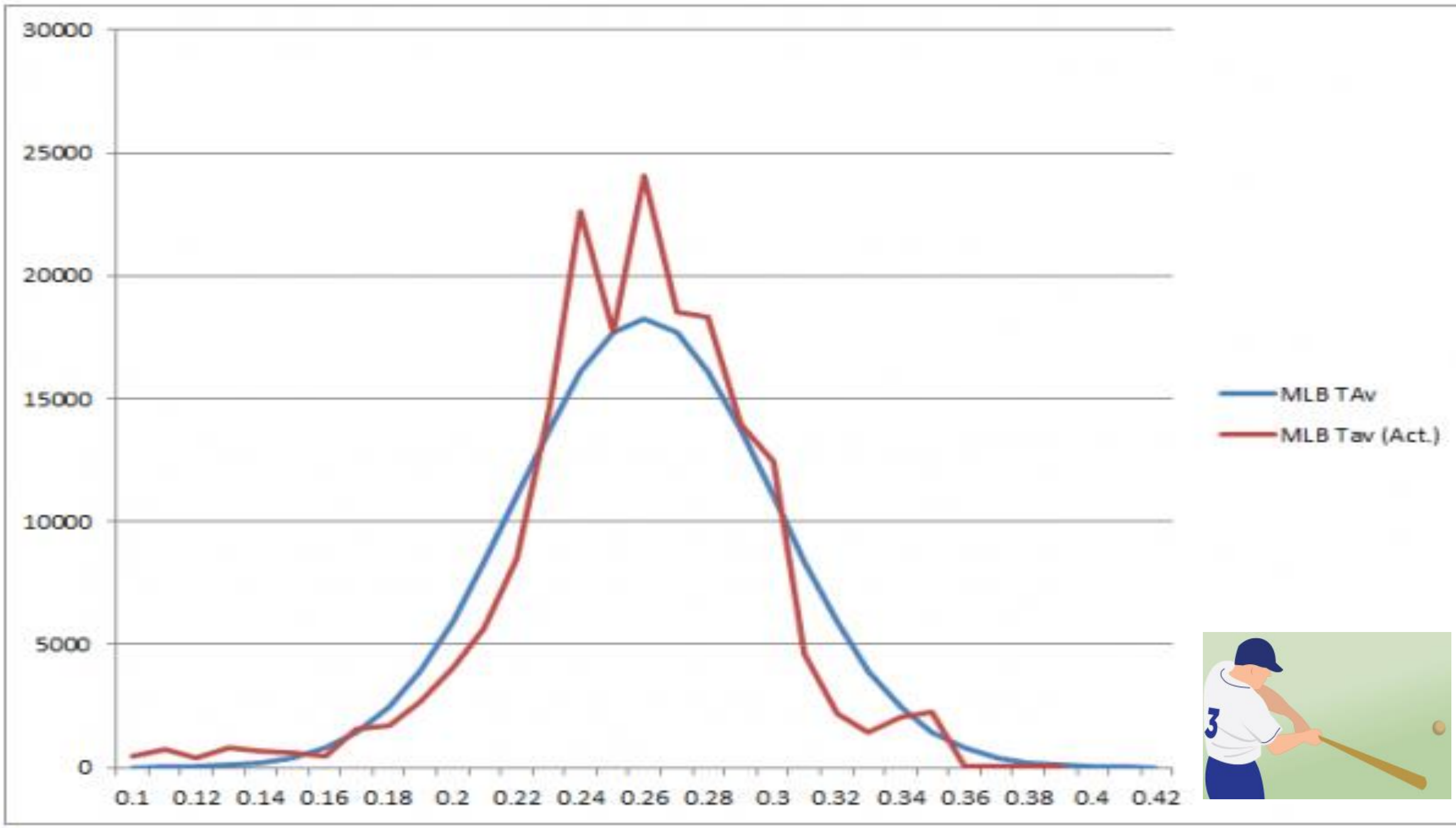
The shape of a normal distribution



There is a long tail in the negative direction on the number line

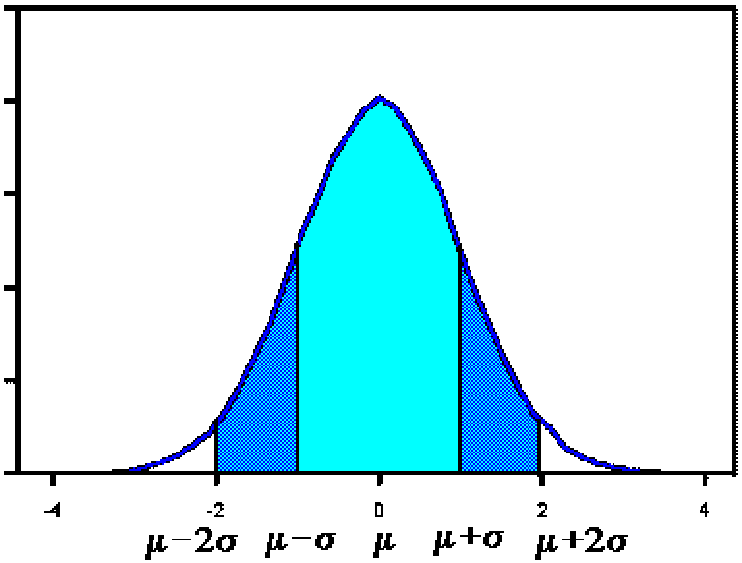
There is a long tail in the positive direction on the number line.

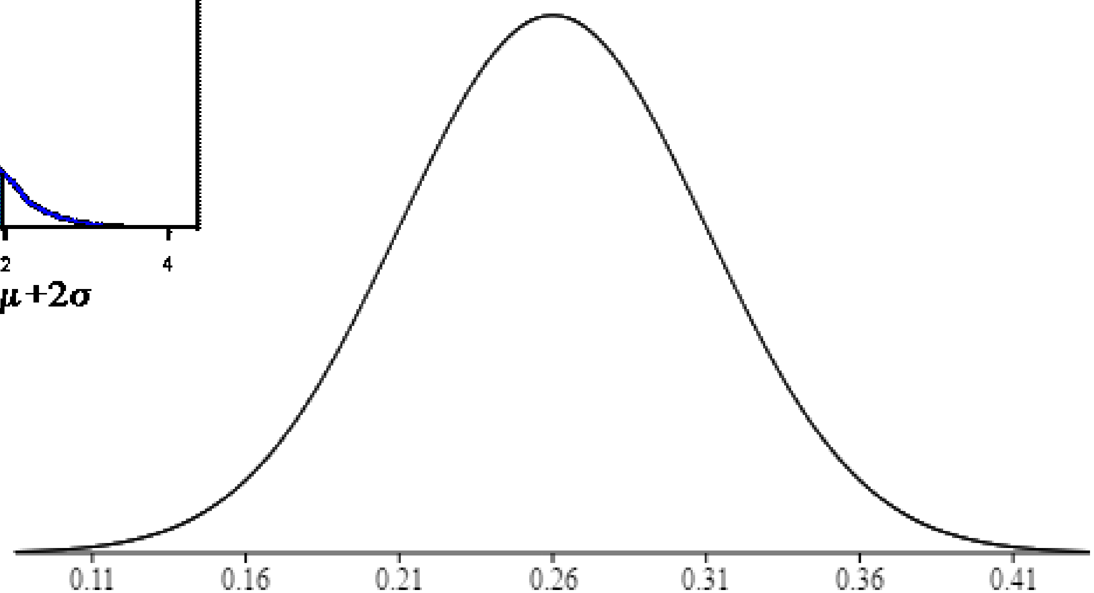
Normal distribution – Major League Baseball (MLB) batting average 1985-2013



MLB official data: Mean = 0.26 Std = 0.05

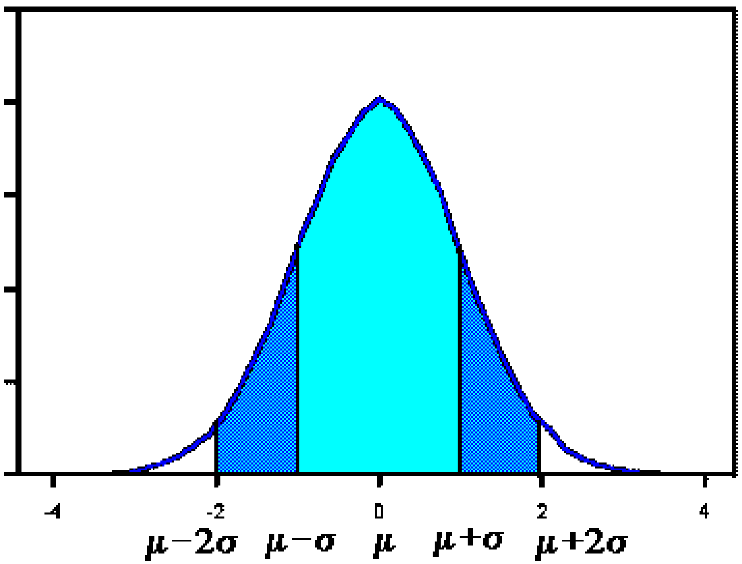
The normal distribution

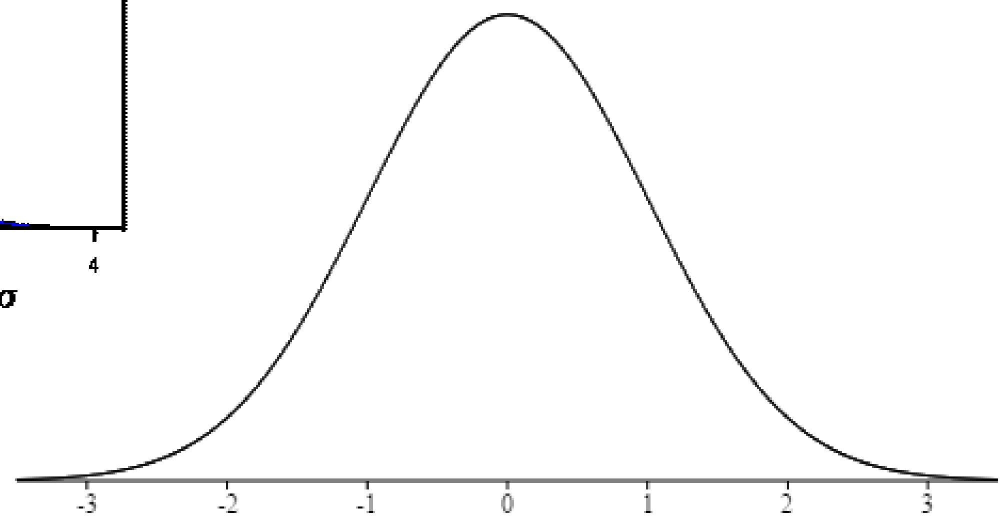




The standard normal distribution

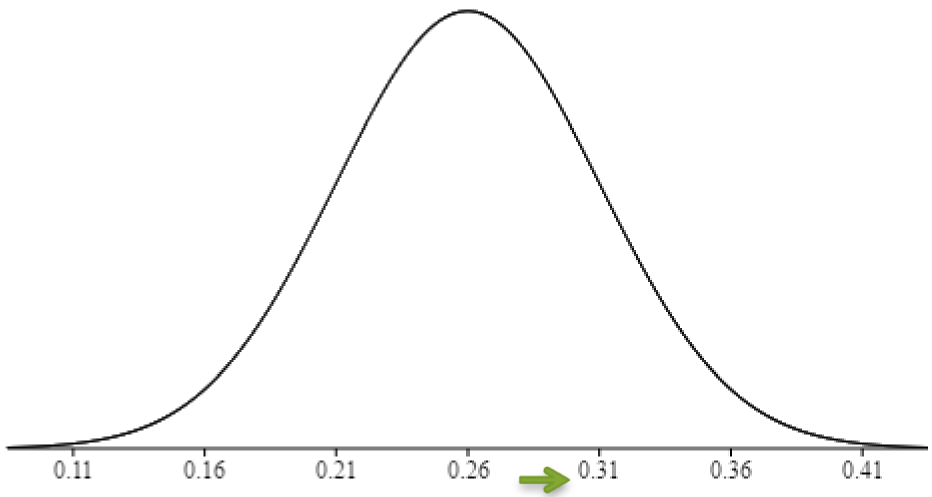
If we have the **standardized situation** of $\mu = 0$ and $\sigma = 1$, then we have:



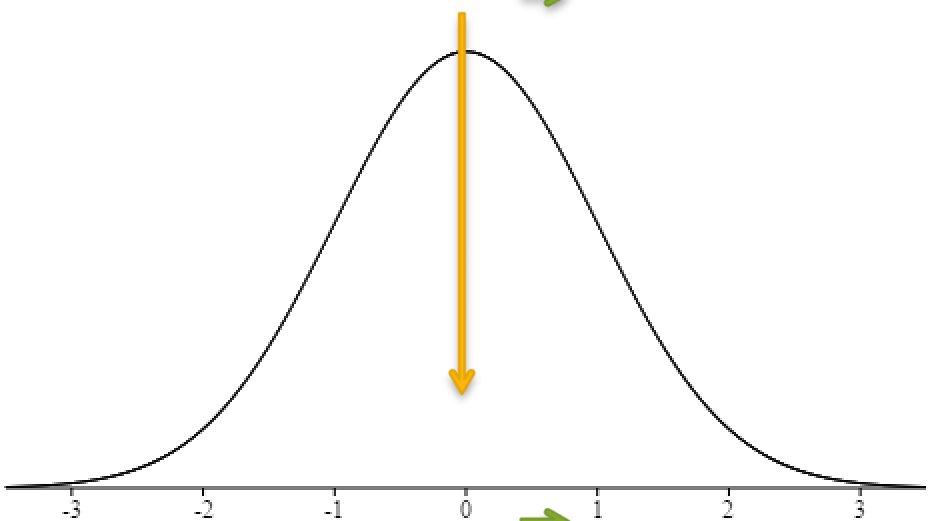


Standard Normal Curve $\mu = 0, \sigma = 1$

Normal distribution- Convert X into Z score

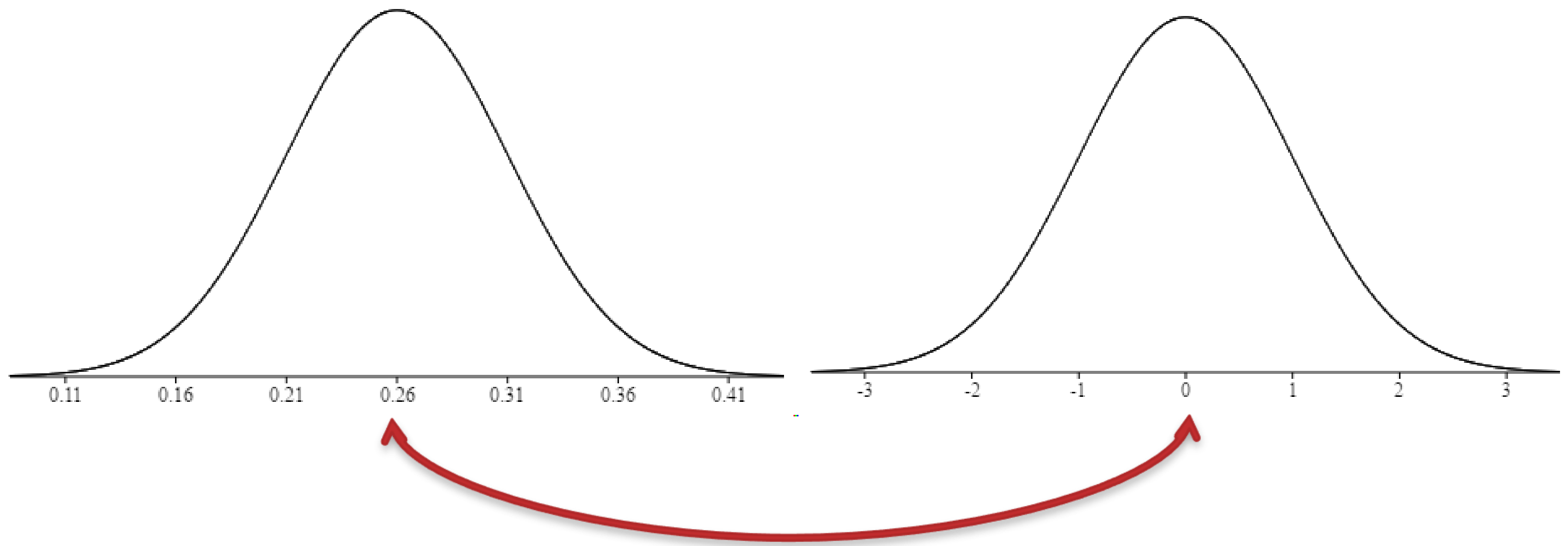


If all the X values in a continuous distribution are converted to Z scores



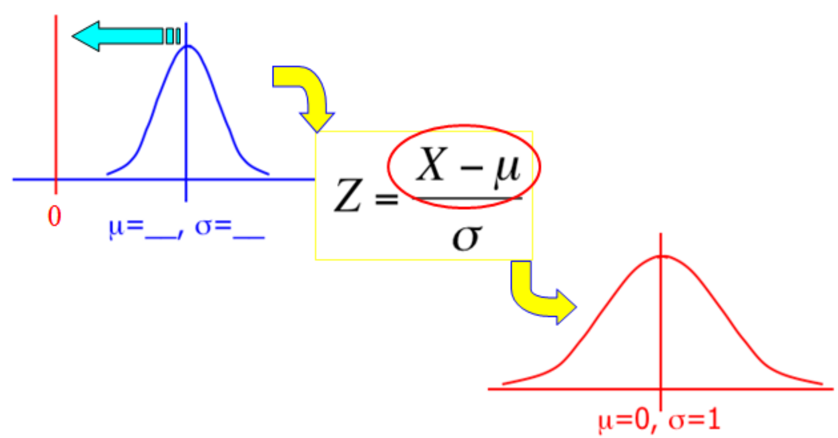
The standardized normal curve will have a mean of 0 and a standard deviation of 1.

Normal distribution- Mean

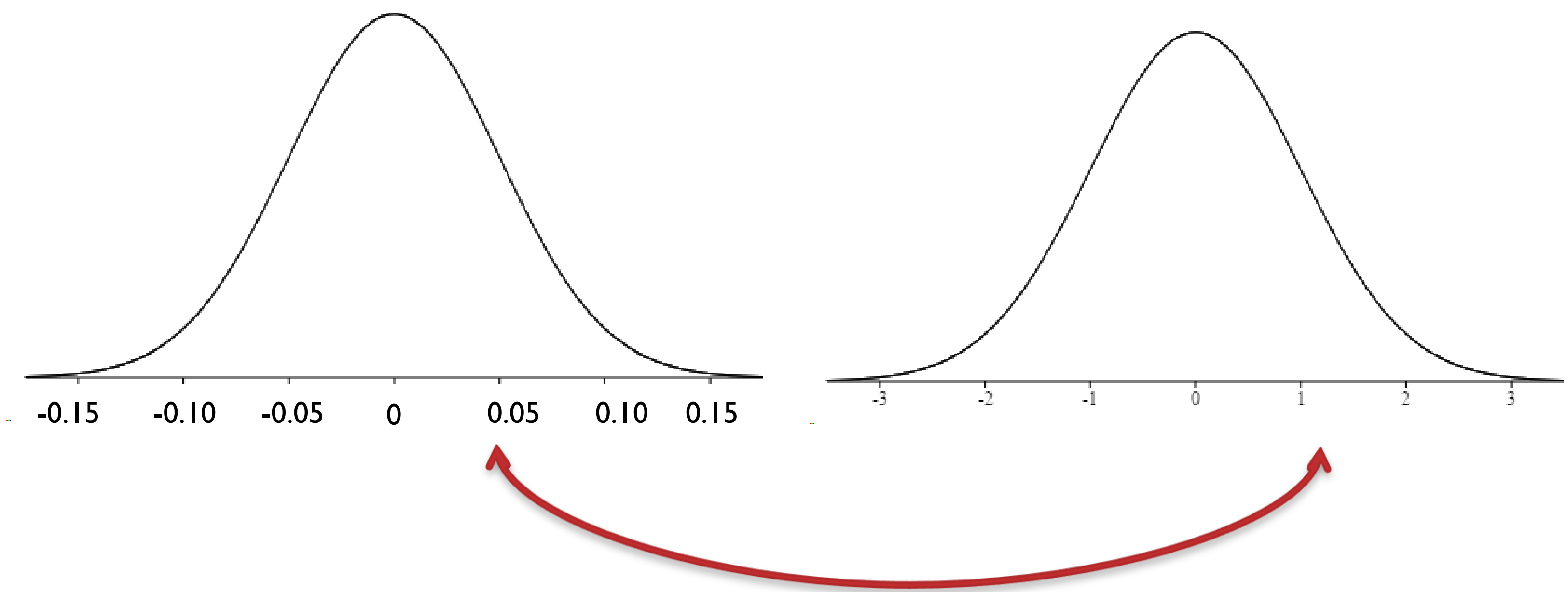


The mean is 0.26

How to convert 0.26 into 0 ?

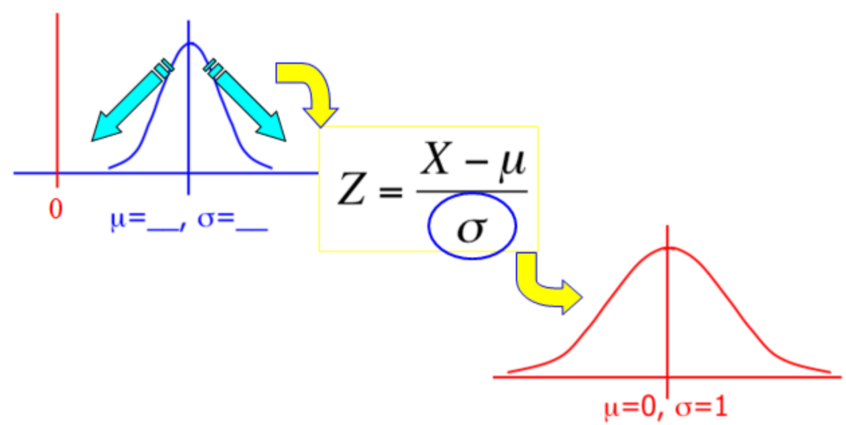


Normal distribution- Variance



The standard deviation is 0.05

How to convert 0.05 into 1 ?



Normal distribution- Z score

$$Z = \frac{X - \mu}{\sigma}$$

- μ is the mean and σ is the standard deviation of the variable X
- This process of transforming any normal distribution to one with a mean of 0 and a standard deviation of 1 is called *standardizing* the distribution

So what do you think happens when?

- You compute a Z score for the case where:
  - x = μ : What is Z score of x= 0.26
  - x = any other value : What is Z score of x = 0.23
  - What is the purpose to calculate Z score ?
    - We use z score to find the probability

Normal distribution- Z table

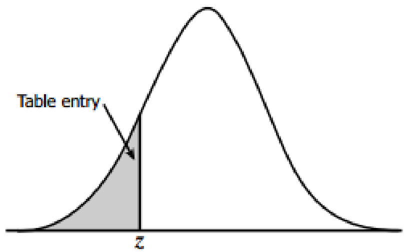


Table entry for z is the area under the standard normal curve to the left of z.

P (Z < -2.95)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019

P (Z < 0.73)

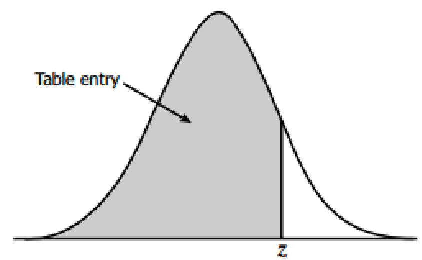


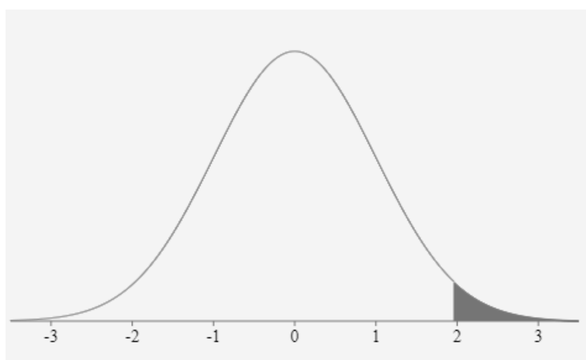
Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852

Normal distribution- Z table

P (z≤1.96)

```
> pnorm(1.96)
[1] 0.9750021
> pnorm(1.96,lower.tail = TRUE)
[1] 0.9750021
```

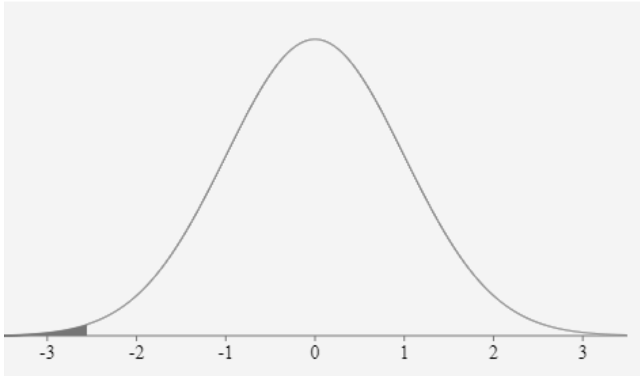


P (z≥1.96)

```
> 1-pnorm(1.96)
[1] 0.0249979
> pnorm(1.96,lower.tail = FALSE)
[1] 0.0249979
```

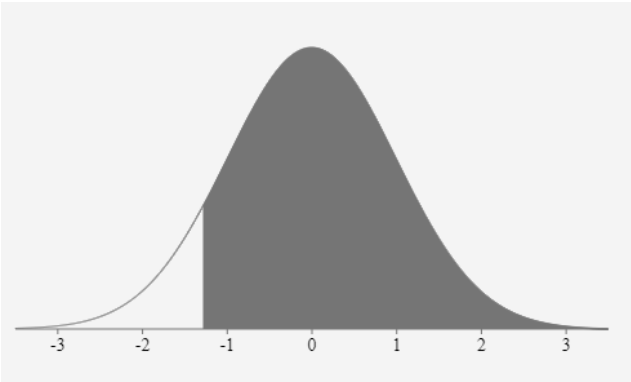
```
> pnorm(1.96)+1-pnorm(1.96)
[1] 1
```

Normal distribution- Z table



P (z≤ -2.56)

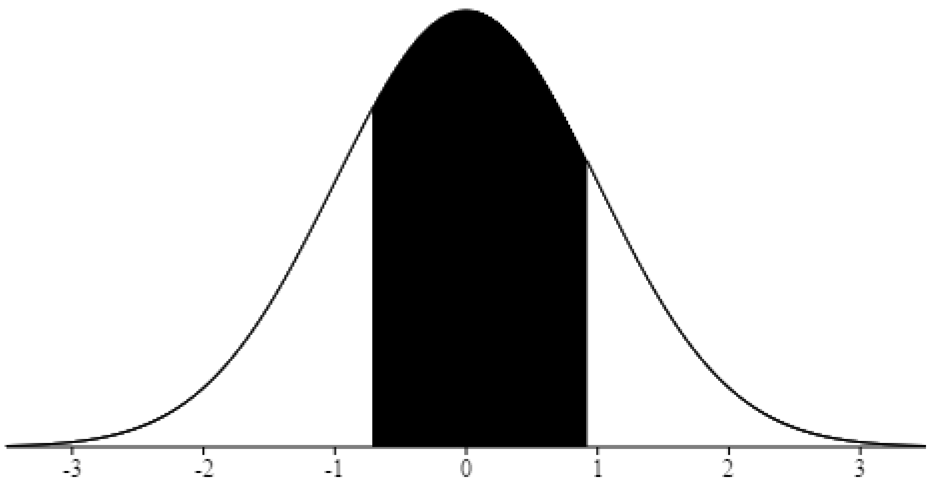
```
> pnorm(-2.56)
[1] 0.005233608
> pnorm(-2.56,lower.tail = TRUE)
[1] 0.005233608
```



P (z≥ -1.28)

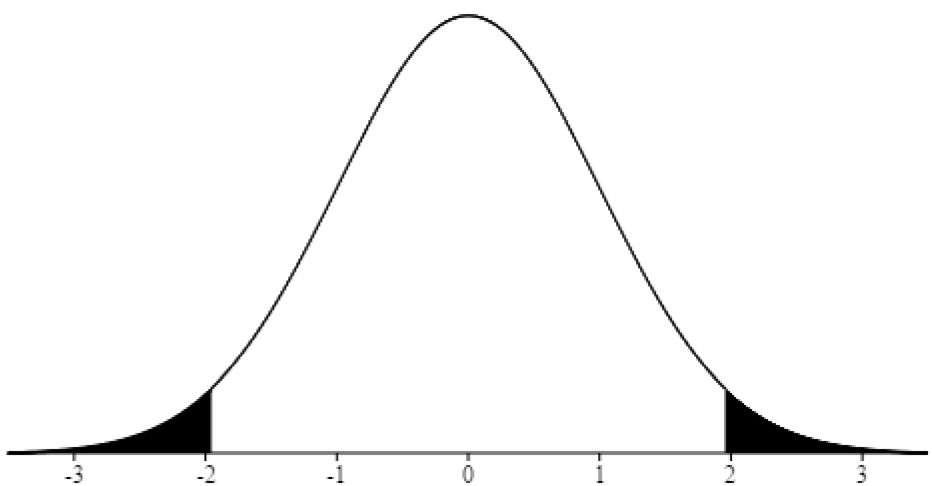
```
> 1-pnorm(-1.28)
[1] 0.8997274
> pnorm(-1.28,lower.tail = FALSE)
[1] 0.8997274
```

Normal distribution- Z table



P (-0.71 ≤ z≤ 0.92)

```
> pnorm(0.92)-pnorm(-0.71)
[1] 0.5823616
```

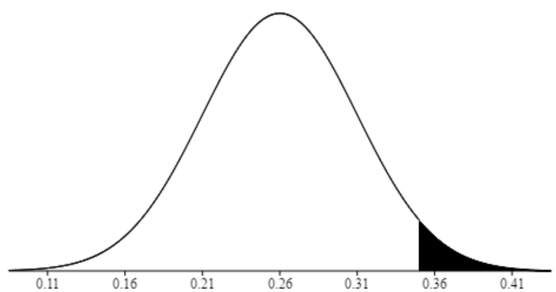


P (z≤ -1.96)

P (z≥ 1.96)

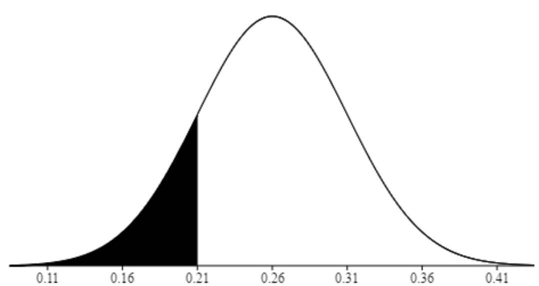
Problem

- Suppose that the batting average in MLB is normally distributed with a mean of 0.26 and a standard deviation of 0.05.
  - What is the probability that a player's batting average more than 0.35




```
> x<-0.35
> mu<-0.26
> sigma<-0.05
> z<-(x-mu)/sigma
> 1-pnorm(z)
[1] 0.03593032
> pnorm(z,lower.tail = FALSE)
[1] 0.03593032
```

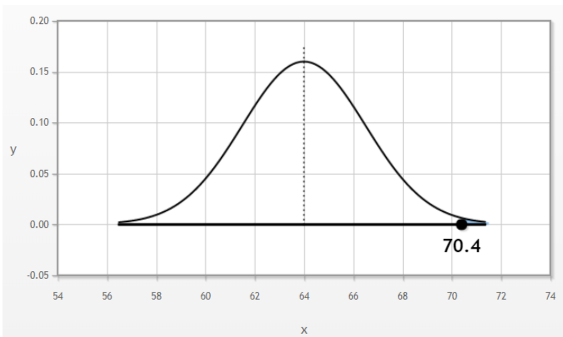
  - What is the probability that a player's batting average less than 0.21




```
> x<-0.21
> mu<-0.26
> sigma<-0.05
> z<-(x-mu)/sigma
> pnorm(z)
[1] 0.1586553
> pnorm(z,lower.tail = TRUE)
[1] 0.1586553
```

Additional exercise- normal distribution

- Assume that the height of women in the US is normally distributed with a mean of 64 inches and a standard deviation of 2.5 inches
  - The probability that a randomly selected woman is taller than 70.4 inches

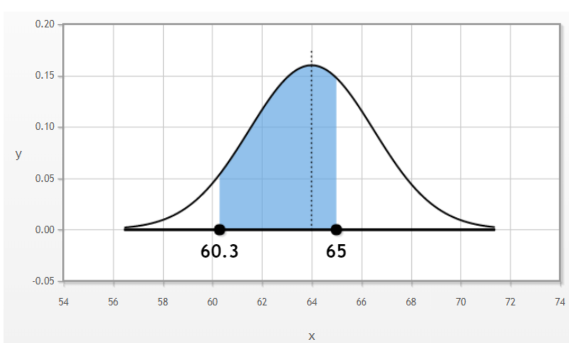


$$P(X > 70.4) = P(X - \mu > 70.4 - 64) = P\left(\frac{X-\mu}{\sigma} > \frac{70.4-64}{2.5}\right)$$

$$P(Z > 2.56) = 0.0052$$

```
> 1-pnorm(2.56)
[1] 0.005233608
> pnorm(2.56,lower.tail = FALSE)
[1] 0.005233608
```

  - The probability that a randomly selected woman is between 60.3 and 65 inches tall.



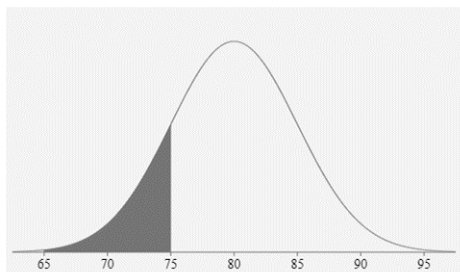
$$P(60.3 < X < 65) = P(60.3 - 64 < X - \mu < 65 - 64) = P\left(\frac{60.3-64}{2.5} < \frac{X-\mu}{\sigma} < \frac{65-64}{2.5}\right)$$

$$P(-1.48 < Z < 0.4) = 0.586$$

```
> pnorm(0.4)-pnorm(-1.48)
[1] 0.5859851
>
```

Additional exercise- normal distribution

- A total of 100 students took the test in a statistical test. Assume the average score is 80, the standard deviation is 5.
  - Number of people with scores between 65 and 75?



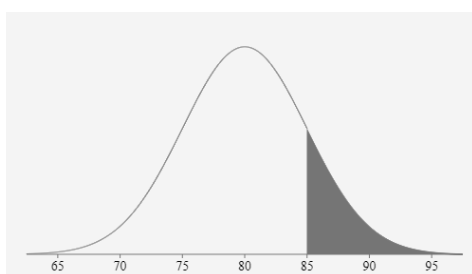
P(65 < X < 75)

$= P(\frac{65-80}{5} < Z < \frac{75-80}{5}) = P(-3 < Z < -1)$

```
> pnorm(-1)-pnorm(-3)
[1] 0.1573054
```

100*0.157 = 15.7≈16

  - Number of people with scores more than 85?



P(X > 85)

$= P(Z > \frac{85-80}{5}) = P(Z > 1)$

```
> 1-pnorm(1)
[1] 0.1586553
> pnorm(1,lower.tail = FALSE)
[1] 0.1586553
```

100*0.159 = 15.9≈16

Normal distribution- Probability to z value

- If we want to use the probability to infer the standard deviation or the mean, we must change the probability back to the Z value.

Z value to probability → pnorm()

Probability to Z value → qnorm()

Normal distribution- Probability to z value

- The probability is 0.95, then what is the z value?

```
> qnorm(0.95)
[1] 1.644854
```

- The probability is 0.01, then what is the z value?

```
> qnorm(0.01)
[1] -2.326348
```

Normal distribution- Probability to z value

- What is the z value if the probability is 0.3?

```
> qnorm(0.3)
[1] -0.5244005
```

- What is the z value if the probability is 0.5?

```
> qnorm(0.5)
[1] 0
```

Additional exercise- normal distribution(qnorm)

- The daily letters received by trading companies are approaching the normal distribution. The probability of an population mean of 80 per day and more than 120 seals is known to be 0.1.

- What the population standard deviation is?

```
qnorm(0.1)
qnorm(0.1, lower.tail = FALSE)
z<-qnorm(0.1, lower.tail = FALSE)
x<- 120
mu<-80
sigma<- (x - mu)/z
sigma
```

```
> qnorm(0.1)
[1] -1.281552
> qnorm(0.1, lower.tail = FALSE)
[1] 1.281552
> z<-qnorm(0.1, lower.tail = FALSE)
> x<- 120
> mu<-80
> sigma<- (x - mu)/z
> sigma
[1] 31.21217
```

Additional exercise- normal distribution(qnorm)

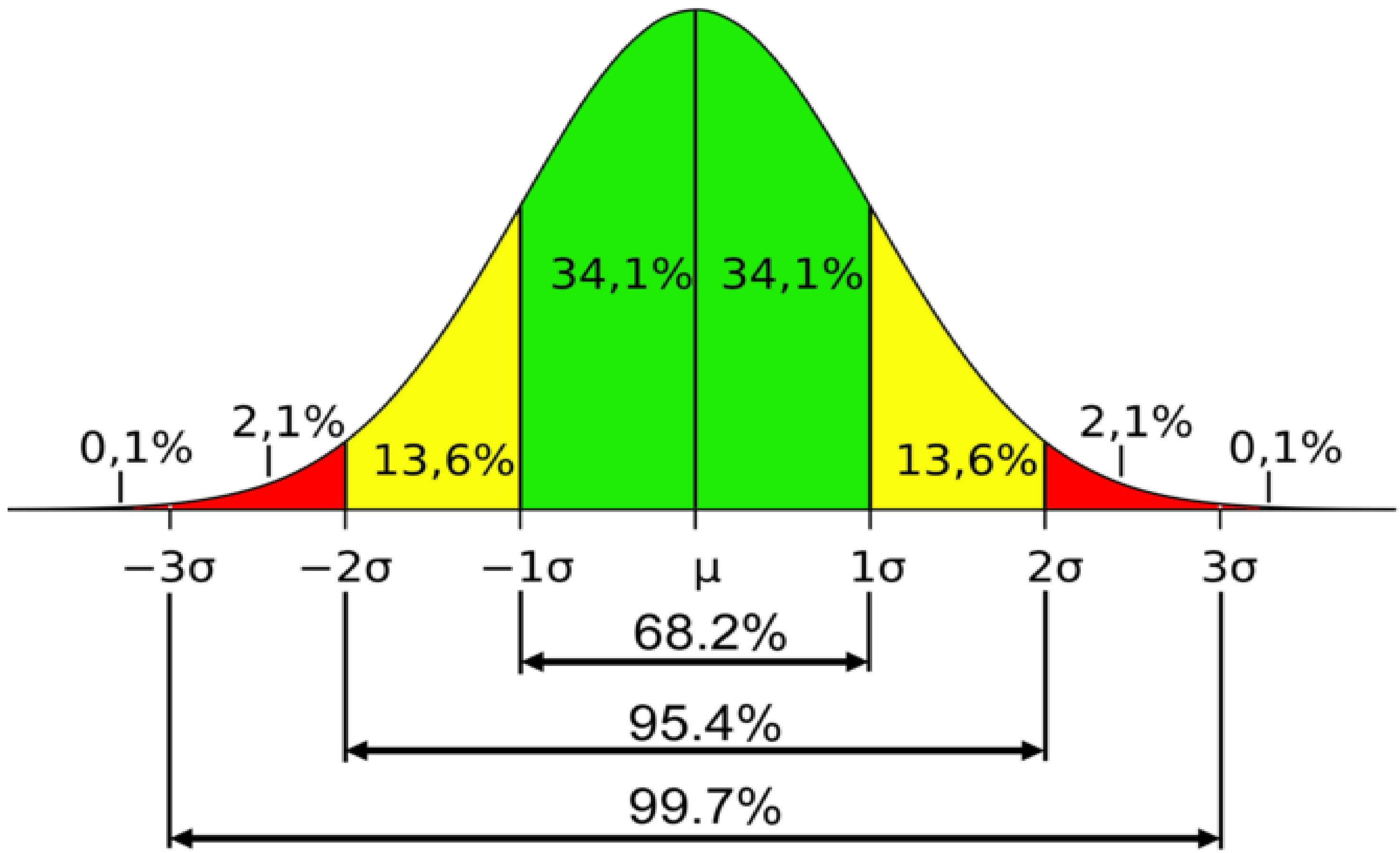
- The standard deviation of the life of electrical appliances is 1 year. If the mean is less than 2 years is known to be 0.0062.

- What is the population mean of electrical appliances life if the z value?

```
qnorm(0.0062)
z<-qnorm(0.0062)
x<- 2
mu<-80
sigma<- 1
mu<-x - (z*sigma)
mu
```

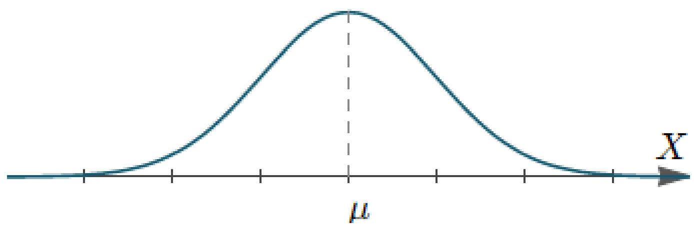
```
> qnorm(0.0062)
[1] -2.500552
> z<-qnorm(0.0062)
> x<- 2
> mu<-80
> sigma<- 1
> mu<-x - (z*sigma)
> mu
[1] 4.500552
```

Normal Distribution and Standard Deviations- Empirical rule



The standard normal distribution

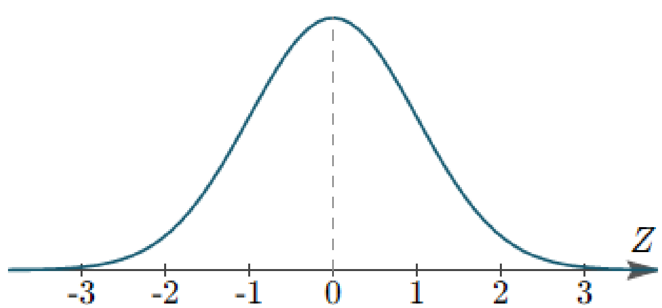
$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\,\sigma^2}$$



$$P(a < X < b) = \int_a^b f(X)\,dx$$

$$Z = \frac{X-\mu}{\sigma}$$

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



Standard Normal Curve $\mu = 0,\, \sigma = 1$

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 0.68269$$

$$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 0.95450$$

$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 0.9973$$

Common continuous probability distribution

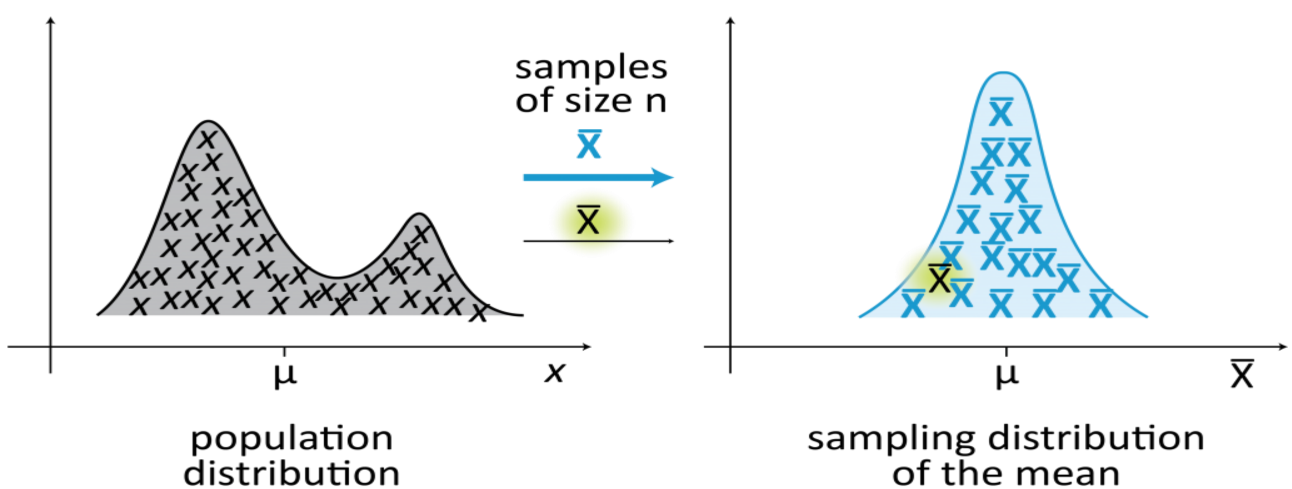
- Uniform distribution ☑
- Normal distribution (z distribution) ☑
- Sampling distribution

SAMPLING DISTRIBUTION

- Sampling distribution of mean
- Sampling distribution of proportion

Sampling distribution – Mean

- Parameters (mean, variance and proportion) are almost always unknown
- Suppose that we draw some possible samples of size *n* from a given population.
  - The probability distribution of this sample statistic is called a sampling distribution.



Calculating Z-Scores with the Sampling Distribution of the Sample Mean

$$Z = \frac{\boxed{X} - \mu}{\sigma}$$

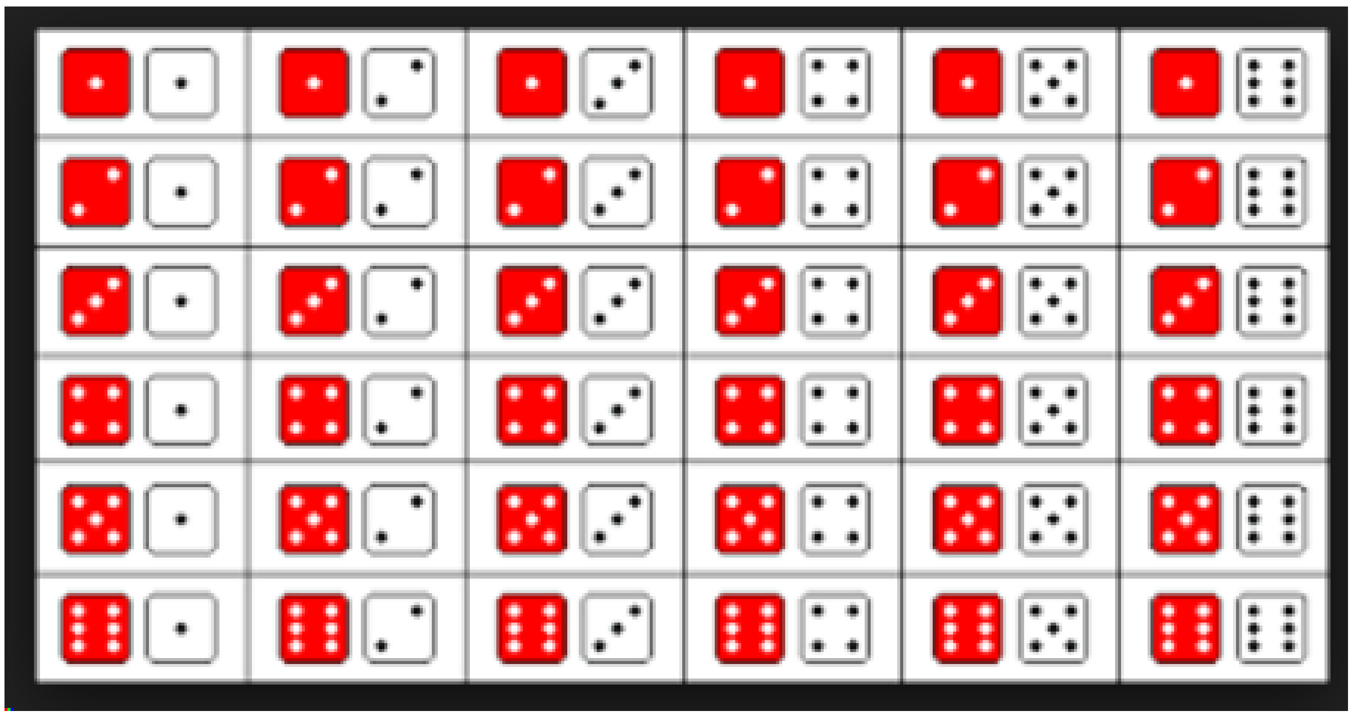
$$Z = \frac{\boxed{\bar{X}} - \mu}{\sigma / \sqrt{n}}$$

You are looking at one random variable x

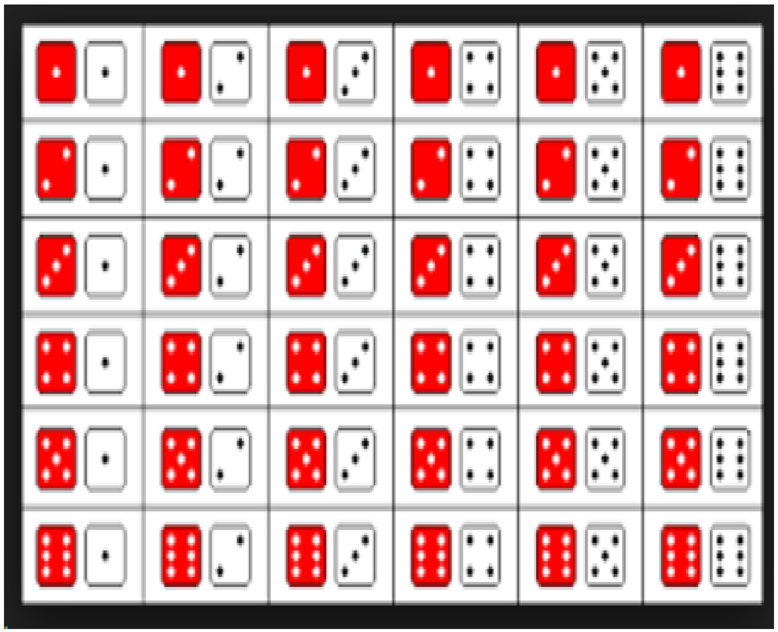
You are looking at sample mean $\bar{x}$

We toss two dices (Sample size =2)

- How many different outcomes we will have ?



Let's compute the sum in each outcome



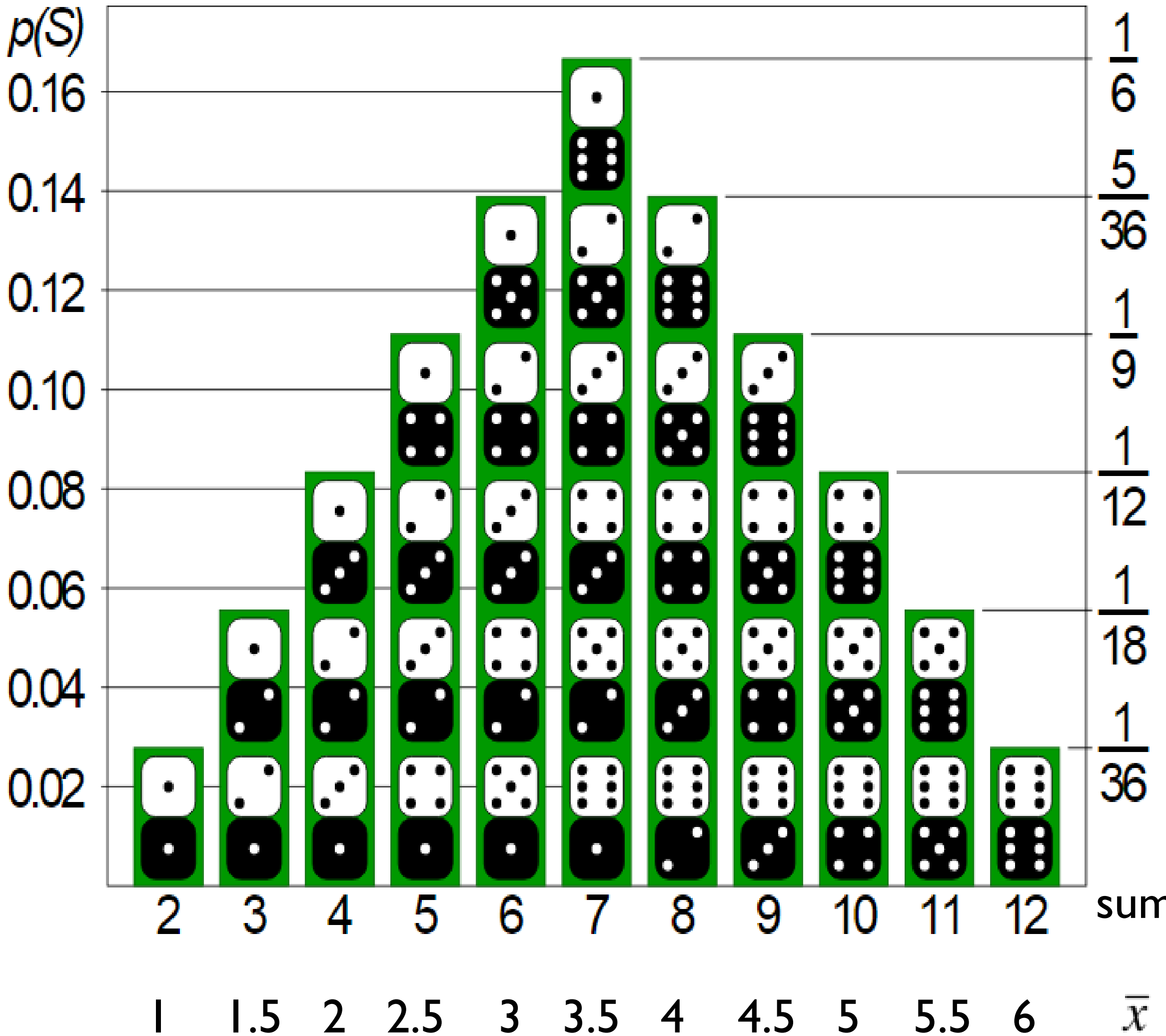
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Note that in these 36 different combinations, some outcomes appear more times than others.

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

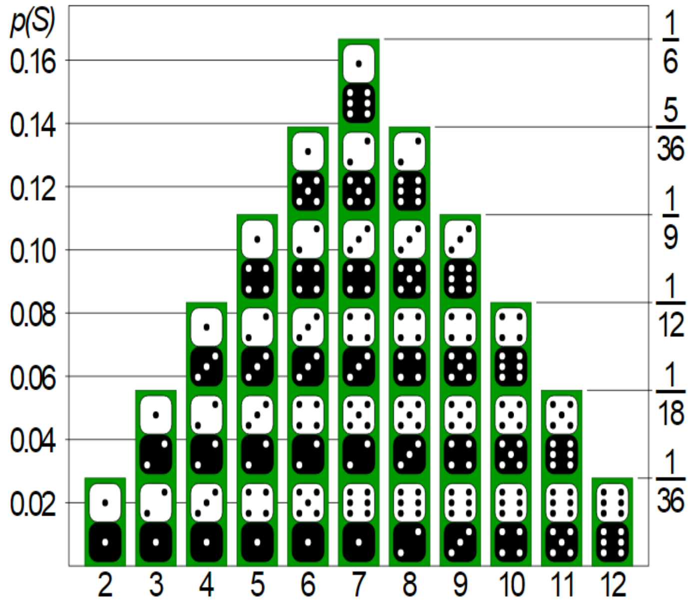
Sample Sum	Sample Mean	Occurrence	Probability
2	1	1	1/36
3	1.5	2	2/36
4	2	3	3/36
5	2.5	4	4/36
6	3	5	5/36
7	3.5	6	6/36
8	4	5	5/36
9	4.5	4	4/36
10	5	3	3/36
11	5.5	2	2/36
12	6	1	1/36

The Sampling Distribution of the Sample Mean



$\bar{x}$	P($\bar{x}$)
1.0	1/36
1.5	2/36
2.0	3/36
2.5	4/36
3.0	5/36
3.5	6/36
4.0	5/36
4.5	4/36
5.0	3/36
5.5	2/36
6.0	1/36

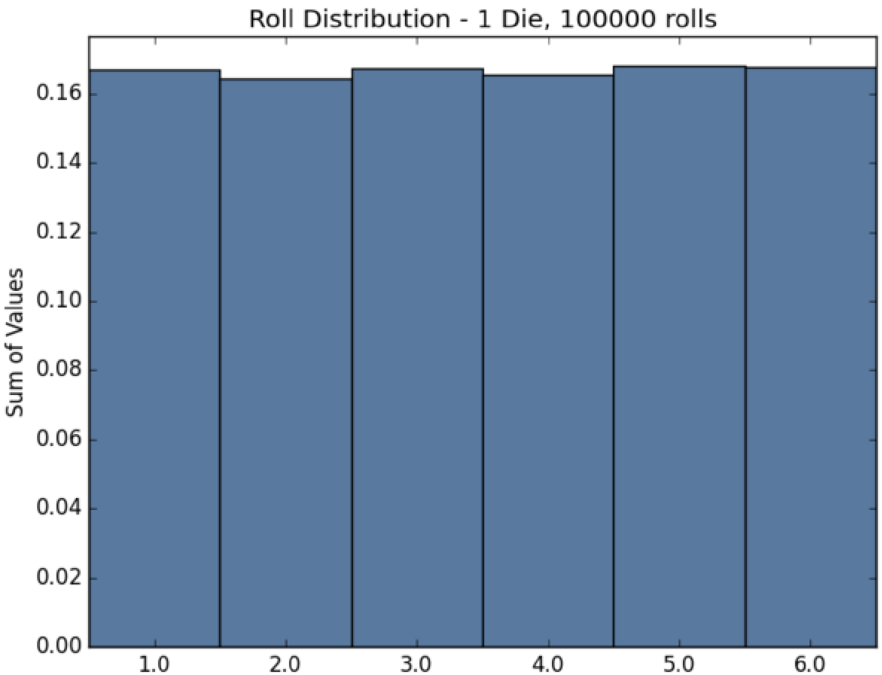
The Sampling Distribution of the Sample Mean



$\bar{x}$	$P(\bar{x})$
1.0	1/36
1.5	2/36
2.0	3/36
2.5	4/36
3.0	5/36
3.5	6/36
4.0	5/36
4.5	4/36
5.0	3/36
5.5	2/36
6.0	1/36

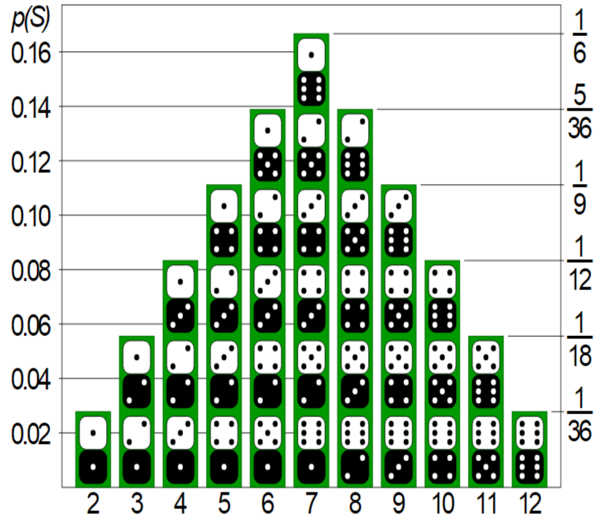
- $\mu_{\bar{x}} = 1*\frac{1}{36} + 1.5*\frac{2}{36} + 2*\frac{3}{36} + 2.5*\frac{4}{36} + 3*\frac{5}{36} + 3.5*\frac{6}{36} + 4*\frac{5}{36} + 4.5*\frac{4}{36} + 5*\frac{3}{36} + 5.5*\frac{2}{36} + 6*\frac{1}{36} = 3.5$
- $\sigma_{\bar{x}}^2 = (1-3.5)^2*\frac{1}{36} + (1.5-3.5)^2*\frac{2}{36} + (2-3.5)^2*\frac{3}{36} + (2.5-3.5)^2*\frac{4}{36} + (3-3.5)^2*\frac{5}{36} + (3.5-3.5)^2*\frac{6}{36} + (4-3.5)^2*\frac{5}{36} + (4.5-3.5)^2*\frac{4}{36} + (5-3.5)^2*\frac{3}{36} + (5.5-3.5)^2*\frac{2}{36} + (6-3.5)^2*\frac{1}{36} = 1.46$
- $\sigma_{\bar{x}} = \sqrt{1.46} = 1.208$

Population Mean



$$\mu = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$$

$$\sigma^2 = (1 - 3.5)^2 * \frac{1}{6} + (2 -$$

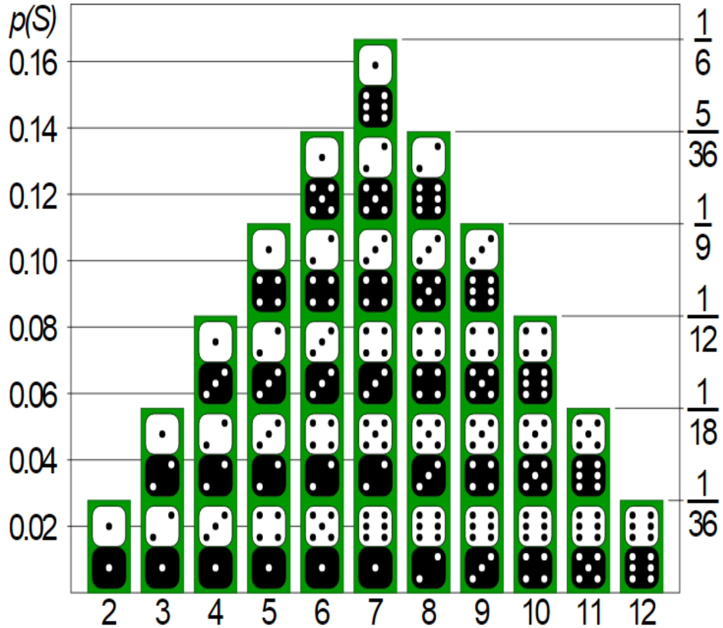


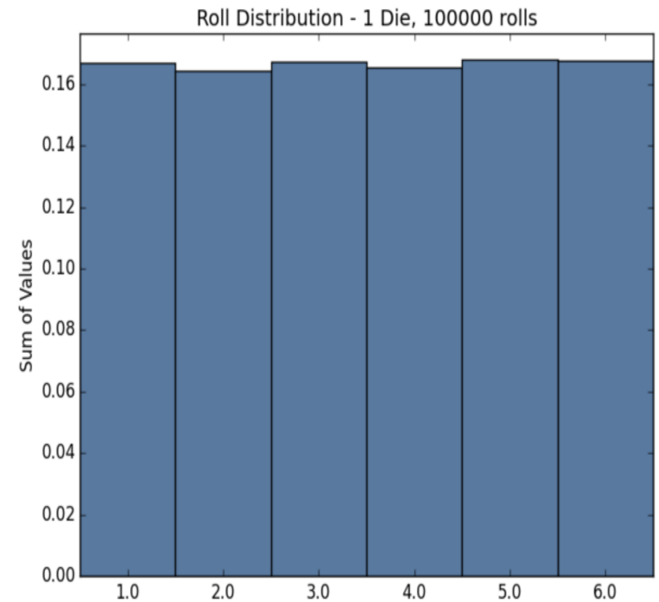
- Sample
  - $\mu_{\bar{x}}$ =3.5
  - $\sigma_{\bar{x}}^2$ = 1.46
  - $\sigma_{\bar{x}} = 1.208$

- Sample
  - $\mu_{\bar{x}}$ =3.5 $\quad E(\bar{x}) = \mu_{\bar{x}} = \mu$
  - $\sigma_{\bar{x}}^2$ = 1.46 $\quad \frac{\sigma^2}{n} = \frac{2.92}{2} = 1.46$
  - $\sigma_{\bar{x}} = 1.208$ $\quad \frac{\sigma}{\sqrt{n}} = \frac{1.709}{\sqrt{2}} = 1.208$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Population
  - μ=3.5
  - σ^2 = 2.92
  - σ = 1.709





$$E(\bar{x}) = \mu_{\bar{x}} = \mu$$

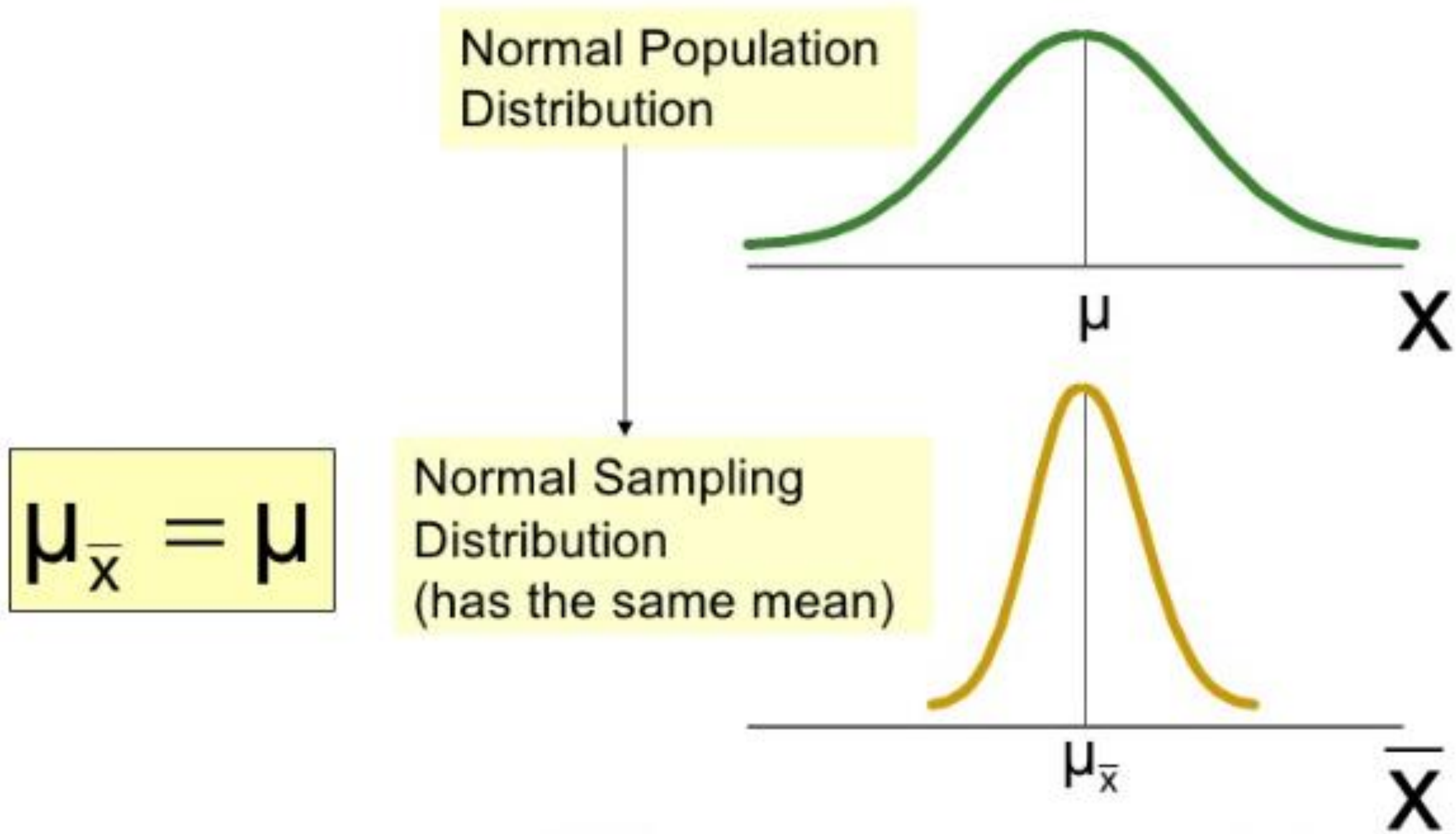
The mean of the sampling distribution is equal to the mean of the population (μ).

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The variance of the sampling distribution is determined by the standard deviation of the population (σ), and the sample size (n).

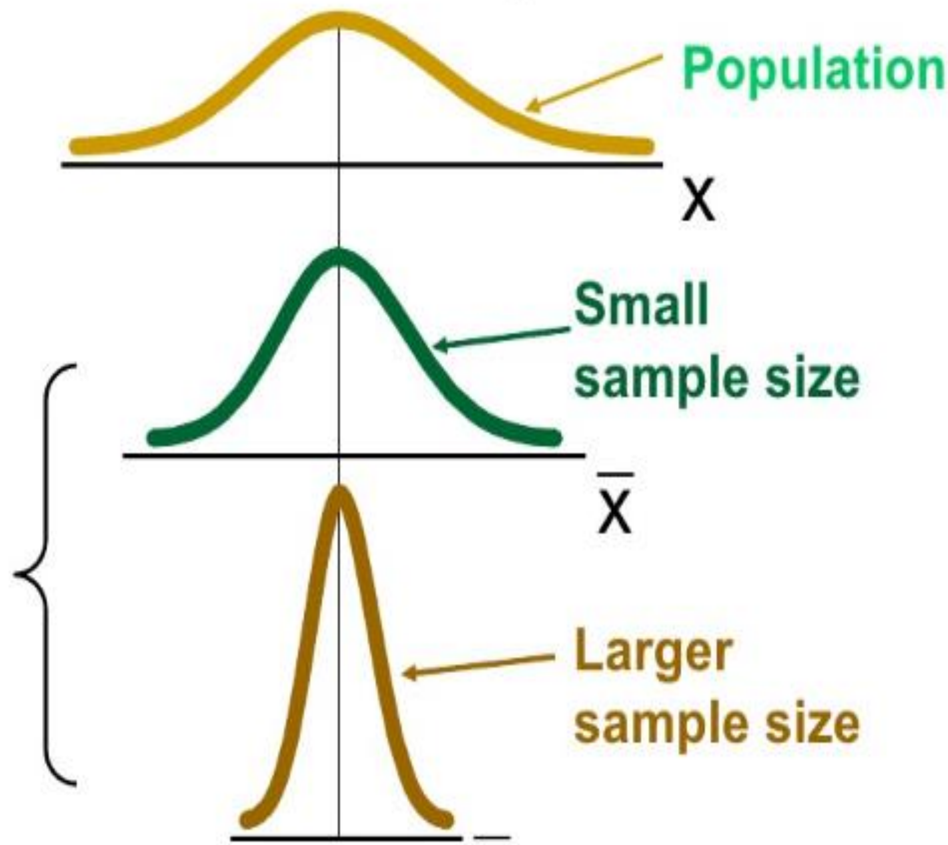
The standard error of the mean

The mean of the sampling distribution



The standard error of the sampling distribution

(the value of $\bar{x}$ becomes closer to μ as n increases):



As n increases,

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$

decreases

Calculating Z-Scores with the Sampling Distribution of the Sample Mean

$$Z = \frac{\boxed{X} - \mu}{\sigma}$$

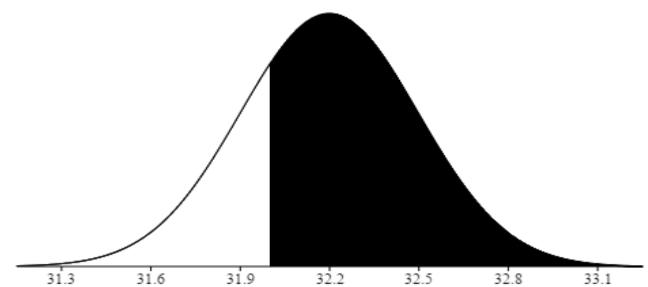
$$Z = \frac{\boxed{\bar{X}} - \mu}{\sigma / \sqrt{n}}$$

You are looking at one random variable x

You are looking at sample mean $\bar{x}$

Problem

- The manager of a bottling plant has observed that the amount of soda in each "32-ounce" bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce.



- If a customer buys one bottle, what is the probability that the bottle will contain more than 32 ounces?

```
> 1-pnorm((32-32.2)/0.3)
[1] 0.7475075
```

- If a customer buys a carton of four bottles, what is the probability that the mean amount of the four bottles will be greater than 32 ounces?

```
> a <-((32-32.2)/(0.3/sqrt(4)))
> 1-pnorm(a)
[1] 0.9087888
```

Additional exercise- sampling distribution

- Suppose you take a sample of 25 high-school students, and measure their IQ. Assuming that IQ is normally distributed with μ = 100 and σ = 15, what is the probability that your sample's IQ will be 105 or greater?

```
> a <-((105-100)/(15/sqrt(25)))
> 1-pnorm(a)
[1] 0.04779035
```

Additional exercise- sampling distribution

- The mathematics scores of a class of students are normal distribution, with an average of 72 and a standard deviation of 9. When 10 students are randomly selected, what is the probability average score of the 10 students more then 80?

```
> a<-((80-72)/(9/sqrt(10)))
> a
[1] 2.810913
> 1-pnorm(a)
[1] 0.002470053
```

Sampling distribution of a proportion

- The estimator of a population proportion of successes is the sample proportion.

- We count the number of successes in a sample and compute:

$$\hat{P} = \frac{X}{n}$$

  - X is the number of successes, n is the sample size.

- If we assume that 40 students are female in the 100 NYUST students. What is the proportion of the female students in this sample?

Sampling distribution of a proportion

- If samples are repeated drawn from a population, the distribution of $\hat{p}$ will be approximately normally distributed

- Sample proportions can be standardized

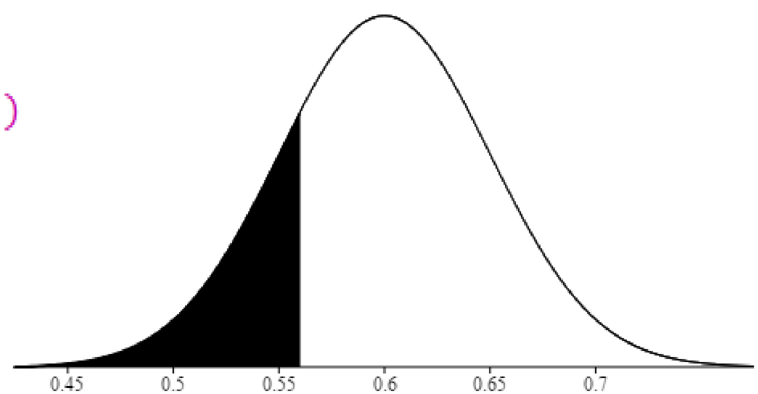
$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}}$$

Problem

- A random sample of 100 students is taken from the population of all part-time students in the United States, for which the overall proportion of females is assumed as 0.6.

- What is the probability that sample proportion p-hat is less than or equal to 0.56?

```
> a <-(0.56-0.6)/(sqrt((0.6*(1-0.6))/100))
> pnorm(a)
[1] 0.2071081
```



Additional exercise- sampling distribution

- Suppose the proportion of all college students who have used marijuana in the past 6 months is about 40 percent. For a class of 100 sample student that is representative of the population of all students on marijuana use. What is the probability that the proportion of students who have used marijuana in the past 6 months is more than 32 students? Do you think it is a serious problem in the campus?

```
> phat <- 0.32
> p <- 0.4
> n <- 100
> z <- (phat-p)/sqrt((p*(1-p)/n))
> pnorm(z, lower.tail = FALSE)
[1] 0.9487648
```

Additional exercise- sampling distribution

- Suppose that 23% of patients with hypertension have adverse effects after taking a certain drug. To find 160 patients with hypertension after taking the drug.

- What is the probability that the patients who have adverse effects after taking a certain drug is less than 32 patients? What is the total number of estimated patients?

```
> phat <- 32/160
> p <- 0.23
> n <- 160
> z <- (phat-p)/sqrt((p*(1-p)/n))
> estp<-pnorm(z)
> num<-estp*n
> ceiling(num)
[1] 30
```

Where are we and where are we going ?

Getting a grasp on data

Populations and Samples

Making use of data (inference)

- Estimation
- Hypotheses

Additional exercise

- Household size in the United States has a mean of 2.6 people and standard deviation of 1.4 people. It should be clear that this distribution is skewed right as the smallest possible value is a household of 1 person but the largest households can be very large indeed. Then what is the probability that the mean size of a random sample of 100 households is more than 3?

```
> a<-(3-2.6)/(1.4/sqrt(100))
> pnorm(a,lower.tail = FALSE)
[1] 0.002137367
```

Additional exercise

- The Swanson Auto Body business repaints cars that have been in an accident or which are in need of a new paint job. Its quality standards call for an average of 1.2 paint defects per door panel. What is the probability of finding exactly 2 defects when 2 doors are inspected.

Poisson Distribution Formula

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

```
> x<-2
> lambda<-2.4
> dpois(x,lambda)
[1] 0.2612677
```

Additional exercise

- A mid-management team consists of 10 people, 6 males and 4 females. Recently top management selected 4 people from this team for promotion. It was stated that the selections were based on random selection. All 4 people selected were males. The females are upset and believe that there may have been more than random selection involved here. What probability distribution should be used to analyze this situation and what is the probability that all 4 promotions would go to males if the selections were random?

$$\frac{C_x^m C_{k-x}^n}{C_k^{m+n}}$$

```
> m<-6
> x<-4
> n<-4
> k<-4
> dhyper(x,m,n,k)
[1] 0.07142857
```

Additional exercise

- It is assumed that the scores of students in a certain school are normal, with an average score of 70 points and a standard deviation of 5 points. Assuming that the passing criteria is 60 points, how many students will fail the exam if 1000 students take that exam?

```
> x<- 60
> mu<-70
> sigma<- 5
> z <- (x - mu)/sigma
> pass<-pnorm(z)*1000
> ?"ceiling"
> ceiling(pass)
[1] 23
```

Additional exercise

- According to the US Census Bureau's American Community Survey, 87% of Americans over the age of 25 have earned a high school diploma. Suppose we are going to take a random sample of 200 Americans in this age group and calculate what proportion of the sample has a high school diploma. What is the probability that number of people with a high school diploma is less than 170?

```
> phat <- 170/200
> p <- 0.87
> n <- 200
> z <- (phat-p)/sqrt((p*(1-p)/n))
> pnorm(z)
[1] 0.2001644
```

Additional exercise

- Suppose a contract calls for, at most, 10 percent of the items in a shipment to be red. To check this without looking at every item in the large shipment, a sample of n = 10 items is selected. If 1 or fewer are red, the shipment is accepted; otherwise it is rejected. Then what is the probability that the shipment is accepted?

$\mathrm{P}(X \leq 1)$

```
> n<-10
> p<-0.1
> k<-1
> pbinom(k,n,p,lower.tail = TRUE)
[1] 0.7360989
```