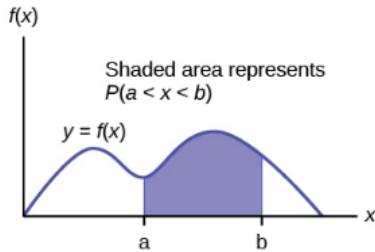# Statistics Ch 5 and 6

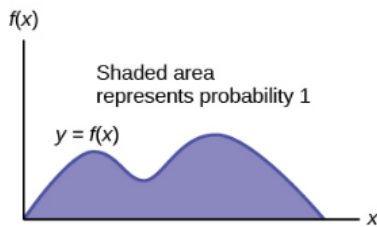## Ch 5.1 Continuous random variable

### A) Density Curve

Probability of a Continuous Random Variable X is defined by its Probability Density Function(pdf) or density curve: $f(x)$ so that
- Area under the density curve corresponds to probability or relative frequency (percent).



- Total area under the density curve is equal to 1
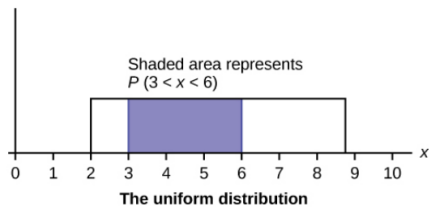


- the graph is always above x-axis.

> Probability = Area = Percent

Two important continuous Probability Distributions
1) Uniform Distribution – The probability of X is equally likely to occur. Histogram of sample data usually bars of similar heights.  There is lowest and highest value of X.
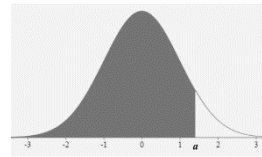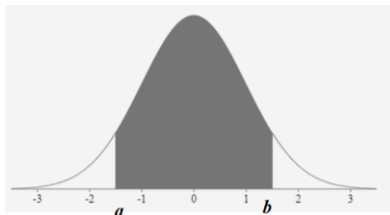Ex1.  X is modeled by Uniform Distribution for
  lowest 2 and highest 8.8.
Probability that X is between 3 and 6 is the shaded area under the density curve.
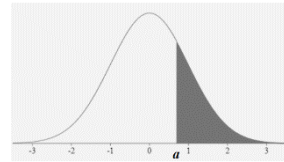


### b) Normal Distribution

Probability that X is between a and b is the area under the bell curve for x = a and x = b.





Shaded left area = probability that x is less than a.



Shaded right area  = probability that x is greater than a.

Notation and property of probability of Continuous random variable X.
Probability that X = a:  P(X = a) = 0
Probability that X is between a and b:
P( a < X < b) or  P(a ≤ X ≤ b)
Probability that X is less than a: P( X < a) = P( X ≤ a)
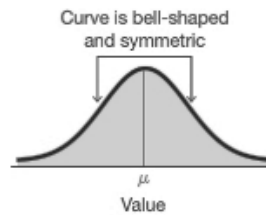Probability that X is greater than a: P(X > a) = P(x ≥ a)

## Ch 6.1 Standard Normal distribution

A random variable X has a distribution with a graph that is symmetric and bell-shaped, and it can be described by the equation given by

$$y = \frac{e^{\wedge}(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)}{\sigma\sqrt{2\pi}}$$ , then it has a "Normal distribution"

Normal Density Curve:



Note: the distribution is determined by µ and σ.
Z-score of x = $\frac{x-\mu}{\sigma}$ is the standardized value of x.
z-score tells the number of standard deviation X is above or below the mean. Positive z implies X is above the mean, negative z implies X is below the mean.

### A) Standard Normal

A standard Normal (also known as z distribution) is a normal distribution with parameters:
Mean µ = 0 and standard deviation σ = 1.
 The total area under its density curve is equal to 1.



Properties of standard normal (z-normal):
- area left of z of 0 = 0.5
- area right of z of 0 = 0.5
- area left of z =  area right of  - z
- area right of z = 1 – area left of z

## B) Empirical Rule (68-95-99.7)

If X is normally distributed, 68% are within 1 sd from the mean. 95% are within 2 sd from the mean, 99.7% are within 3 sd from the mean.

P( Z-score between -1 and 1 ) = 68%

P( Z-score between -2 and 2 ) = 95%

P( Z-score between -3 and 3 ) = 99.7%

Ex1.  Weight of a certain type of dog is normally distributed with mean = 18lb and standard deviation of 2 lb.

a) What is the z-score of 14lb and 22 lb? What is the probability that a dog weigh between 14 lb and 22lb?

b) What is the z-score of 20lb and 16lb? What is the probability that a dog weigh between 16lb and 20 lb?

c) What is the z-score of 24 lb and 12 lb? What is the probability that a dog weigh between 12 lb and 24lb?

## C) Probability of z-score in standard normal

Use online Normal distribution calculator

http://onlinestatbook.com/2/calculators/normal_dist.html

Specify mean $\mu$ =0  standard deviation $\sigma$=1

-For left area or P(x < a) click below

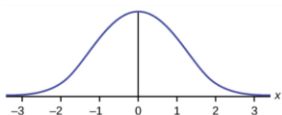-For right area or P(x > a) click above

-For area between two values a and b P( a < x < b), click between

-For area outside of a and b, P(x <a or x > b), click outside

-Click "Recalculate"

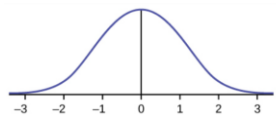Ex1: Find probability that z is between -1.8  and 1.8 . Sketch the area.



Use online Normal calculator Mean = 0, SD = 1

Click between, enter -1.8, 1.8

Recalculate: P( -1.8 < z < 1.8 ) =

Ex2. Find the probability that z is less than 0.44 Sketch the area.



Use online Normal calculator $\mu$ =0 , $\sigma$=1

Click below , enter 0.44

Recalculate: P( z < 0.44 ) =

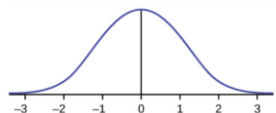Ex3. Find the probability that z is greater than 1.8 Sketch the area.



Use online Normal calculator $\mu$ =0 , $\sigma$=1

Click above , enter 1.8

Recalculate: P( z >1.8) =

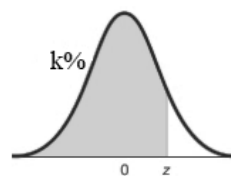Ex 4. Find the probability that z is less than – 1.2 Sketch the area.



Use online Normal calculator $\mu$ =0 , $\sigma$=1

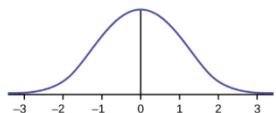Click below , enter -1.2

Recalculate: P( z < -1.2) =

## D) Find percentile of a z-score



k Percentile corresponds to a value that is higher thank% of all values. Or k% of data are less than the k percentile value. This corresponds to left area of k%.
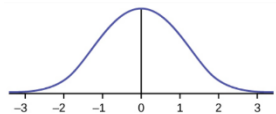
Ex5. What percentile is the z-score 2.2?



Use online Normal calculator Mean =0 , SD=1 Click below , enter 2.2

Recalculate: P( z < 2.2) =
Round to whole percent.

Ex 6. What percentile is the z-score -1.35

Use online Normal calculator $\mu$ =0 , $\sigma$=1



Click below , enter -1.35
Recalculate: P( z < -1.35) =
Round to whole percent.

## E) Find z-score given area or percentile.

Online Inverse Normal calculator is use to find the z-score that is the cut-off for the left area, right area or percentile.
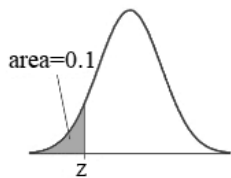
http://onlinestatbook.com/2/calculators/inverse_normal_dist.html

Specify area, mean= 0 and SD= 1
select if area is below, above, between or outside.
Click "Recalculate"

Ex1. Find the z-score that corresponds to bottom 10% of all values.
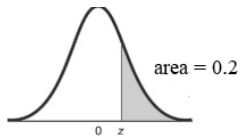


Use Inverse Normal calculator.
Specify area = 0.1
Mean = 0, sd =1,
Click below.
Recalculate.

P( z < _____ ) = 0.1,   cut-off z =

Ex2.  Find the cutoff for <u>top</u> 20%



Use Inverse Normal calculator.
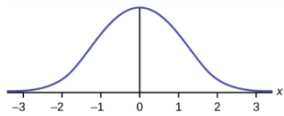Specify area = 0.2
Mean = 0, sd =1,
Click <u>above.</u> Recalculate.

P( z > _____) = 0.2  z-cutoff =

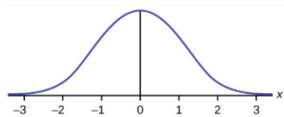Ex3. Find $P_{91}$ , 91th percentile of all z.



Use Inverse Normal calculator.
Specify area = 0.91

Mean = 0, sd =1,
Click below. Recalculate
P( z < _____) = 0.91  z-cutoff =

Ex4. Find $P_{15,}$ 15$^{th}$ percentile of all z.



Use Inverse Normal calculator.
Specify area = 0.15
Mean = 0, sd =1,

Click below. Recalculate

P( z < _____) = 0.15  z-cutoff =
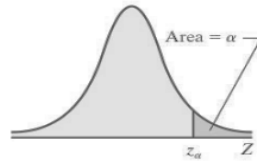
## F) Find Critical value $Z_\alpha$

α = <u>significant level</u>. The probability of unlikely, default is 0.05 if not specify.
**Critical value $z_\alpha$** : the positive z-value that separates significantly high values of  z with not significant z.
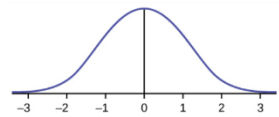Note: Significantly low critical value = - $z_\alpha$



To find $z_\alpha$:
Use Inverse Normal calculator

Specify area = α, Mean = 0, sd =1,
Click above. Recalculate.
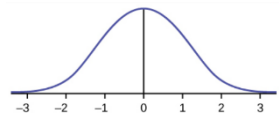
Ex1. Given α = 0.02, find the critical value $Z_{0.02}$



Use Inverse Normal calculator
Specify area = 0.02, Mean = 0, sd =1,

Click above. Recalculate.
$Z_{0.02}$ =

Ex2. Given α = 0.06, find the critical value $Z_{0.06}$



Use Inverse Normal calculator
Specify area = 0.06, Mean = 0, sd =1,

Click above. Recalculate.
$Z_{0.06}$ =

## Ch 6.2 Using the Normal distribution

When X is normally distributed with mean μ and standard deviation σ,  N( μ,σ) probability of   range of X can be represented by the area under normal

density curve $y = \dfrac{e^{\wedge(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$   (no need memorize.)



The same properties of bell shape density curve apply:
Probability = area = relative frequency (percentage)
 a) P( X > μ ) = 0.5,  P(X < μ) = 0.5
 b) P(  a < X < b) = area between a and b under the density bell curve.

## A) Find Probability of X in Normal distribution

Use online Normal distribution calculator to find prob.
http://onlinestatbook.com/2/calculators/normal_dist.html

Specify mean = μ, SD= σ
-For left area or P(x < a) click below
-For right area or P(x > a) click above
-For area between two values a and b P( a < x < b), click between
-For area outside of a and b, P(x <a or x > b), click outside
-Click "Recalculate"

## B) Find X given probability in Normal Distribution

Use Inverse Normal Calculator to find the cut-off given a left area, right area, between area or outside area of two tails.
http://onlinestatbook.com/2/calculators/inverse_normal_dist.html
Specify area, mean= μ, SD= σ
select if area is below, above, between or outside.
Click "Recalculate"
Note: left area = bottom percentage = percentile.
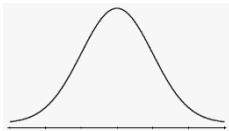Right area = top percent
 a) Find cutoff X for Top k percentage:
Use Inverse online calculator, above
 b) Find cutoff X for Bottom k percentage or k percentile:  Use Inverse online calculator, below

Ex1. Final exam of a standardized test is normally distributed with mean 63 and standard deviation 5.
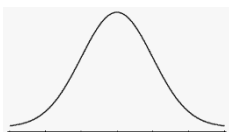a) Find the probability that a randomly selected student scored more than 65 on the exam.

Use Online Normal calculator.
Mean = 63, SD =5,
Click above, enter 65.
Recalculate
P( X > 65 ) = 0.3446.

b) Find the probability that a score less than 85

Use Online Normal calculator.
Mean = 63, SD =5,
Click below, enter 85.
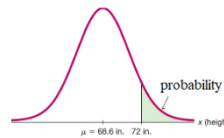Recalculate
P( X < 85 ) = 1

c) Find the 90th percentile score. ( top 10%)

Use Inverse Normal Calculator
Enter Area = 0.9,  Mean = 63, SD = 5. Click below
The cut-off X is at the below box.
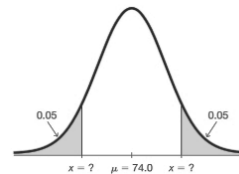
 P(X < _____) = 0.9

Ex2: Heights of men are normally distributed with mean of 68.6 in. and a standard deviation of 2.8 in. Find the probability that a randomly selected man has a height greater than 72 in.

 Use Online Normal calculator.
Mean = 68.6, sd =2.8,
Click above, enter 72.
Recalculate
P(X > 72) = 0.1123

Ex3. Given pulse rate of women is normally distributed with a mean of 74 bpm and a standard deviation of 12.5 bpm. Find the pulse rates of lowest 5% and highest 5% of women.

Total area = 5% + 5% =0.1
Find cutoff pulse rate for lowest 5% and highest 5%.
Use Inverse Normal calculator.
Area = 0.10, Mean = 74, SD=12.5. Click outside. Recalculate.
lowest 5% cutoff = 53.4 bpm highest 5%=94.6 bpm

Ex4. A circus farmer grows mandarin oranges finds that the diameters of mandarin oranges harvested on his farm follow a normal distribution with a mean diameter of 5.85 cm and standard deviation of 0.24cm.
a) Find the probability that a normally selected mandarin orange from his farm has a diameter larger than 6 cm.
Use Online Normal calculator.

Mean = 5.85, sd =0.24,
Click above, enter 6.
Recalculate

P( X > 6 ) =
b) Find the middle 20% diameter of mandarin oranges from his farm.

Use Inverse Normal Calculator
Area = 0.2, Mean = 5.85, sd=0.24
Click Between.

P( _____< X < _____) = 0.2

Ex5. A TV has a life that is normally distributed with a mean of 6.5 years and a standard deviation of 2.3 years. If the company offers a warranty to replace any TV within 3 years. What percent of TV will need to be replaced?

Mean = 6.5 , sd =2.3,

Click below, enter 3.

Recalculate

percent =?

x=3

P (X < 3) = 0.064 = 6.4%

# Ch 7 Notes

## Ch 7.1 Central Limit Theorem for Sample Means

### A) Sample distribution of sample mean:

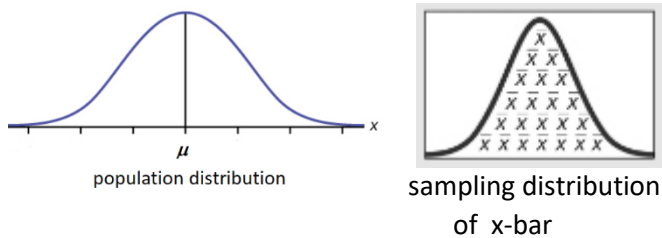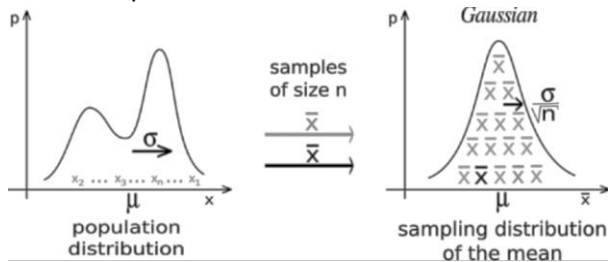When sample means ( $\bar{x}$ ) of same size n taken from the same population, the Sample means has the following behavior:

1) If the population distribution of X is normal, the distribution of $\bar{x}$ is always normal for all sample size n.



population distribution

sampling distribution of x-bar

2) When population distribution of X is not normal, The sampling distribution x-bar tends to be a normal distribution. The distribution become closer to normal when sample size increase.



population distribution

sampling distribution of the mean

### Central Limit Theorem:

For all sample of the same size n with n > 30, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ .

Note:

-This applies to all distribution of x. If X is normally distributed, n > 30 is not needed. Any n will work.

-The sample should be a Simple Random Sample.

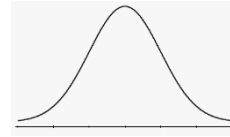| **Central limit Theorem** |
| :---: |
| Notation: $\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ |

Ex1  A standardized test with scores that are normally distributed with mean μ = 150 and standard deviation σ = 18.  A class of 20 students take the test. The mean score $\bar{x}$ of the 20 students are calculated.

a) Is the distribution of mean score  $(\bar{x})$ of 20 students
   Normally distributed?
   Ans: Yes because the original score is normal.

b) What is the mean and standard deviation of $\bar{x}$?

c) Find the probability that a student's score is greater than 160.



d) Find the probability that the mean score $\bar{x}$ of 20 students is greater than 160.



Ex2: Coke cans are filled so that the actual amounts have a mean of 12 oz and a standard deviation of 0.11 oz. The distribution of amount of coke is unknown.

a) Is the distribution of mean amount of coke in 36 cans normally distributed?
   Yes, because n > 30, according to CLT, $\bar{x}$  will be normally distributed.

b)  What is the mean and standard deviation of $\bar{x}$?
   Ans: $\mu_{\bar{x}} = 12$, $\sigma_{\bar{x}} = 0.11/\sqrt{36}$.

c)  Find the percent of individual coke with amount between 11.9 to 12.1 oz.



d) Find the percent of mean amount of 36 coke with between 11.9 and 12.1 oz.



Ex3. Annual incomes are known to have a distribution that is skewed to the right. Assume that 20 workers' mean incomes $\bar{x}$   are collected.

a)  Will the distribution of mean income $\bar{x}$  be normally distributed?
   Ans: No, since X is not normal and n < 30, CLT does not apply, $\bar{x}$  may not be normally distributed.

## Ch 7.2 Central Limit Theorem for Sums

For all sample of the same size n with n > 30, the sampling distribution of $\sum x$ can be approximated by a normal distribution with mean $(n)(\mu)$ and standard deviation $\sigma\sqrt{n}$ .
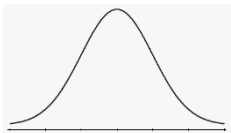
> **Central limit Theorem for Sum $\sum x$**
> Notation: $\mu_{sum} = (n)(\mu)$, $\sigma_{sum} = \sigma(\sqrt{n})$

-This applies to all distribution of x. If X is normally distributed, n > 30 is not needed. Any n will work.
-The sample should be a Simple Random Sample.

Ex1. An unknown distribution has a mean of 45 and a standard deviation of 8. A sample size of 50- is drawn randomly from the population.  .
a) Can Central Limit for Sum be used to model the distribution of the sum of 50 sample? Explain.

b) Find the probability that the sum of the 50 values is more than 2400.



Ex2: An elevator has a maximum weight limit of 5000 lb or 27 passengers. Assume adult males have weights that are normally distributed with a mean of 189 lb. and a standard deviation of 39 lb.
a) Can the sum of 27 passengers' weight be modeled by Normal distribution? If yes, what is the mean and standard deviation of sum of weight of 27 passengers?

b) Does this elevator appear to be safe? Find the probability that the total weight of 27 passengers exceed the maximum weight limit.



P (sum of 27 > 5000)  185 )

Ex3. The passenger load for a water-taxi is 3500 lb. Assume weights of passenger are normally distributed with mean 174 lb. with a standard deviation of 21 lb. Is it safe have a passenger limit of up to 20?
a) Can the sum weight of 20 passengers be modelled by Central limit Theorem of sum? If yes, what is the mean and standard deviation of the weights?
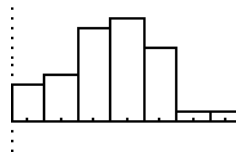
b) Determine the probability that 20 passenger's total weight will exceed the passenger load of 3500 lb.  Is the 20-passenger limit safe?
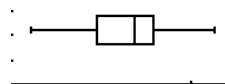


## Normality assessment:

The following procedure can be used to determine if sample data are from a population having a Normal distribution.

1) Histogram:  If the histogram departs dramatically from a bell shape, conclude data do not come from a normal distribution.
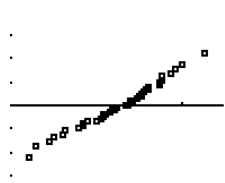


If Histogram from the sample does not departs from a bell shape distribution, conclude the Population may be Normal.

2) Outliers: If there is more than one outlier, conclude data do not come from a normal distribution.



3) Normal Quantile plot: If patterns of the points is reasonably close to a straight line and the points does not show some systematic pattern that is not a straight line pattern, conclude Normal distribution

**Normal quantile plot (or Normal probability plot)** is a plot of (x, y) where x is the original data and y is the corresponding z-score that is expected from a normal distribution.
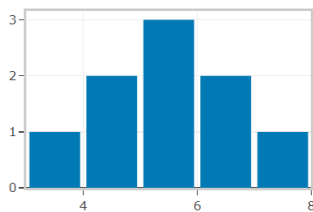
Normal probability plot does not show pattern different from a straight line.

Steps:

1) input data to Sample editor in Statdisk.

2) Data/Normality Assessment/

3) Check that number of outliers are at most 1. Check that Histogram has a approximately bell shape distribution. Check that points on Normal Quantile plot are reasonably close to a straight line.

Ex1. Sample question: Do the following values come from a population with normal distribution?
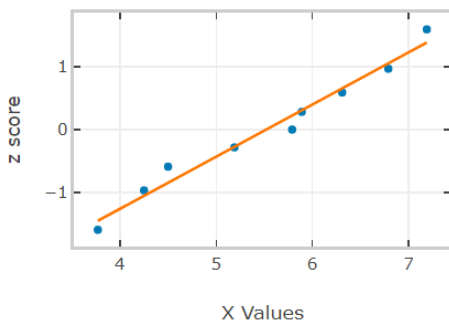7.19, 6.31, 5.89, 4.5, 3.77, 4.25, 5.19, 5.79, 6.79.

Input data to statdisk.
-number of outliers is 0



Histogram shows relatively symmetrical

-Normal quantile plot shows plot along a linear pattern.

**Normal Quantile Plot of Column 1 (n=9)**



X Values

Final conclusion: the sample are from a population that is Normal.