

Worksheet 1

Prep-Work (Distributions)

1) Let X be the random variable whose *c.d.f.* is given below.

$$F_X(x) = \begin{cases} 0 & \text{if } x < 5 \\ 0.3 & \text{if } 5 \leq x < 10 \\ 0.5 & \text{if } 10 \leq x < 15 \\ 0.8 & \text{if } 15 \leq x < 20 \\ 1.0 & \text{if } 20 \leq x \end{cases}$$

Compute the mean, μ_X . (Hint: First identify all possible values of X , then compute values for the *p.m.f.*, $f_X(x)$).

12

2) Let X be binomial random variable with $n = 40$ and $p = 0.15$. Use *Excel* to compute (i) $f_X(8) = \mathbf{0.1086}$

and $F_X(8) = \mathbf{0.8645}$.

3) Let X be a continuous random variable that is uniform on the interval $[0,10]$. (i) What is the probability that X is **at most** 8.75? $= \mathbf{.875}$

(ii) What is the probability that X is **no less than** 4.25? $= \mathbf{.575}$

4) Let W be the working lifetime, measured in years, of the microchip in your new digital watch. Suppose that W has an exponential distribution with mean 4 years. Use *Integrating.xls* and the probability density function f_W to compute the probabilities that the chip lasts for (i) **at least** 8 years $= \mathbf{.1348}$ and (ii) **at most** 2 years $= \mathbf{.3935}$

5) Let X be an **exponential** random variable with $\mu_X = 9.2$. Compute the following. (i) $f_X(6) = \mathbf{.0566}$

(ii) $P(X = 6) = \mathbf{0}$ (iii) $F_X(6) = \mathbf{.4791}$ (iv) $P(X \geq 6) = \mathbf{.5209}$ (v) $E(X) = \mathbf{9.2}$

6) Use *Integrating.xls* to determine whether or not the function given below could be a *p.d.f.* for some continuous random variable.

$$f_X(x) = \begin{cases} 1.2 \cdot x^2 + 1.2 \cdot x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Definition		Computation		Plot Interval		Integration Interval		$\int_a^b f(x) dx$
Formula for $f(x)$		x	$f(x)$	A	B	a	b	
0			0	0	1	0	1	1.0000

Constants

s	
t	
u	
v	
w	

Integrate

Hide

Unhide

Worksheet 2

Prep-Work (Variance)

Part 1-Variance(Dispersion) and Standard Deviation

1)Discrete Random Variable: Example 1(MBD Proj2.ppt) –from Text from Variance Section What is similar and what is different about the two random variables, X and Y in the text Example 1?

a) What is the mean of each random variable, X and Y?

4

4

b) Looking at the values of X and Y, which random variable has the larger variance?

Y

c) From the tables, what is the variance of X? .7 And of Y? 3.3

d) From the tables, what is the standard deviation of X? .84 And of Y? 1.82

e) Look at the calculation of the variance of X and Y . From this, write down the formula for **the variance of a discrete random variable.**

$$V(X) = \sum_{\text{all } x} (x - \mu_X)^2 \cdot f_X(x)$$

2)The *p.m.f.* of a finite random variable Y is given below.

y	-2	-1	0	1	2
$f_Y(y)$	0.10	0.15	0.20	0.25	0.30

Compute $V(Y) = 1.75$ and $\sigma_Y = 1.32$

3)Continuous Random Variable: Example 4(MBD Proj2.ppt) –from Text from Variance Section

- a. Write down the formula for the variance of a continuous random variable.

$$V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x) dx$$

- b. The random variable giving the time between computer breakdowns is an exponential random variable(a continuous random variable) with $\alpha = 16.8$.
c. What is the formula for the pdf of this random variable?

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{16.8} \cdot e^{-x/16.8} & \text{if } 0 \leq x \end{cases}$$

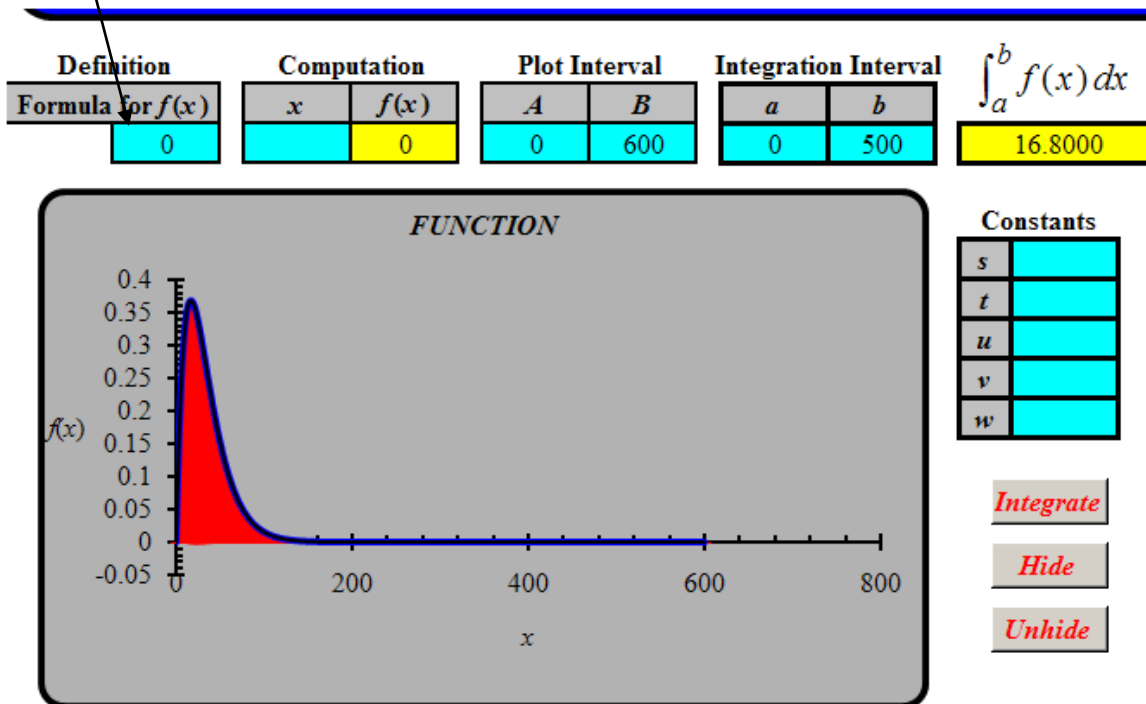
- d. What is the formula for the mean of this random variable?

$$E(X) = \alpha$$

e. Find the mean using Integrating.xls.

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx .$$

=x*(1/16.8)*EXP(-x/16.8)



f. What is the formula for the variance of this random variable?

$$V(X) = \int_0^{\infty} (x-16.8)^2 * \frac{1}{16.8} e^{-x/16.8} dx \quad (\text{When using Excel})$$

$$V(X) = \alpha^2$$

g. Find the variance.

$$V(X) = (16.8)^2 = 282.24$$

h. What is the standard deviation of this random variable?

$$\sigma_x = \sqrt{V(X)} = \sqrt{\alpha^2} = \alpha = 16.8$$

i. Sketch a graph of the pdf of this random variable.

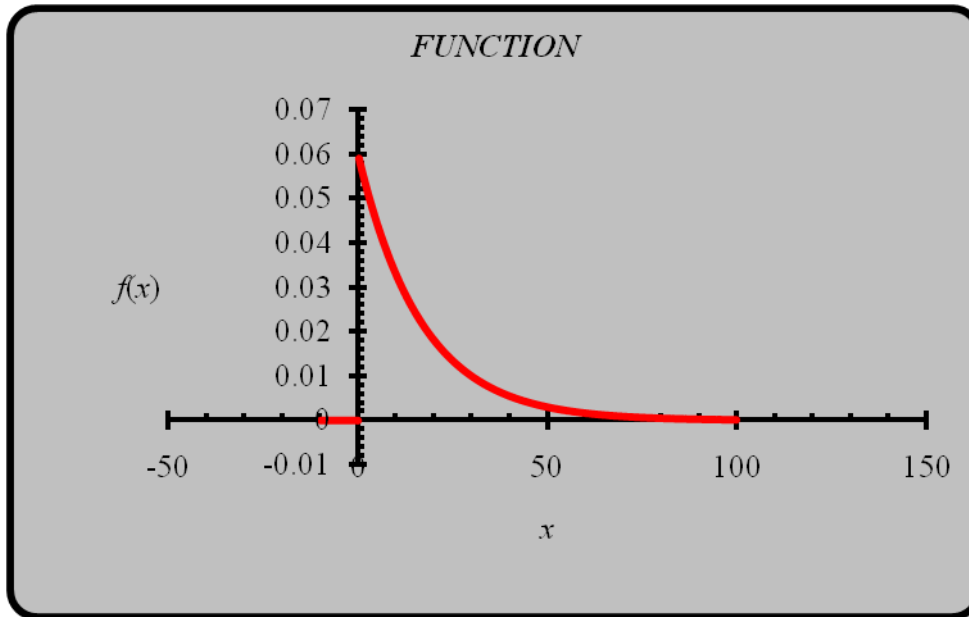
$$=IF(x<0,0,(1/16.8*EXP(-x/16.8)))$$

Definition	
Formula for $f(x)$	
	0.05952
	4

Computation	
x	$f(x)$
	0.05952
	4

Plot Interval	
a	b
-10	100

Constants	
s	
t	
u	
v	
w	



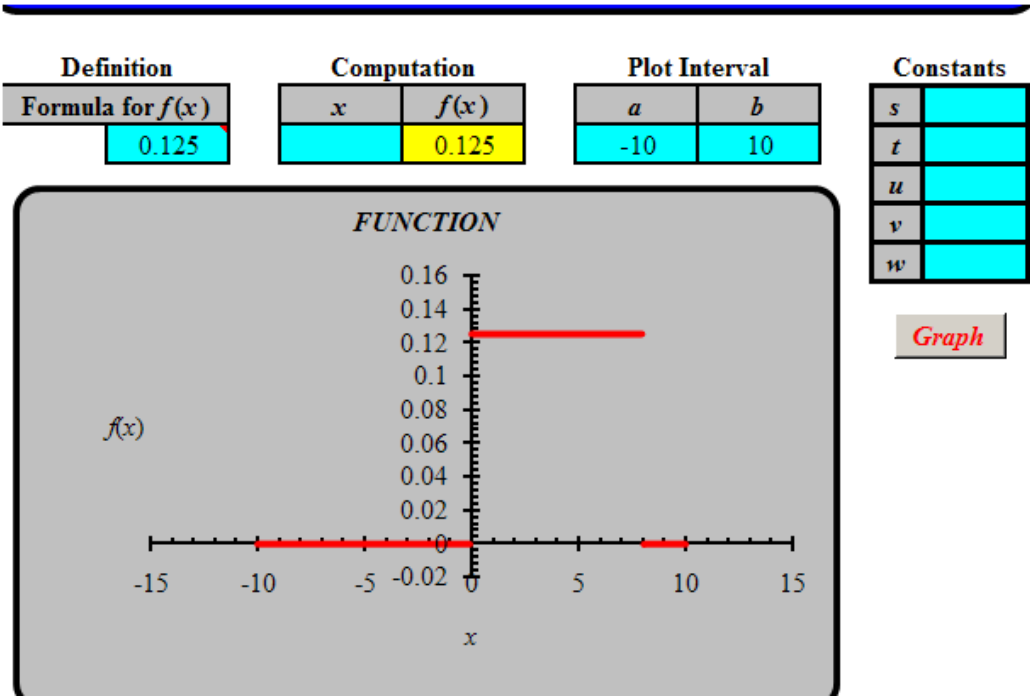
j. Guess the standard deviation of a general exponential random variable.

$$\sigma_x = \sqrt{V(X)} = \sqrt{\alpha^2} = \alpha$$

4)Uniform Distribution

A uniformly distributed random variable has a pdf with the same value for all values of the variable. Suppose X is uniform random variable taking all values between 0 and 8.

a) Sketch a graph of the pdf..



b)

What must be true of the area under the graph?

1

C)What is the formula for the pdf?

$$f_x(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x \leq 8 \\ 0 & \text{if } 8 < x \end{cases}$$

D)What is the mean of the random variable X ? (Excel not needed.)

$$(0+8)/2=4$$

E)Find the variance of X . (Excel needed.)

$$(8-0)^2 / 12=5.33$$

F)Find the standard deviation of X .

$$2.309$$

Part 2-Variance of Distributions; Sample Statistics

1)Variance of Binomial Distribution: Use Bionomial2.xls

- a) The Excel file contains the calculation to find the expected value, variance, and standard deviation of the Binomial distribution with $n = 28$ and $p = 0.2$. Note down the answers.

expected value(5.6) , variance(4.48), and standard deviation(2.1166)

- b) Now adapt the file to find the expected value, variance, and standard deviation for $n = 50$ and $p = 0.2$. Note down the answers.

the expected value(10), variance(8), and standard deviation(2.824)

- c) Adapt the file again for $n = 50$ and $p = 0.4$. Write down the expected value, variance, and standard deviation. Similar to part (b)

the expected value(20), variance(12), and standard deviation(3.46)

- d) In some order, the formulas for the expected value, variance, and standard deviation of the Binomial distribution with n trial and probability p are the following: $\sqrt{np(1-p)}$; $np(1-p)$; np . Match them up by checking the formulas against the values you found in Questions #1-3.

Binomial Distribution	
Expected value	np
Variance	$np(1-p)$
Standard deviation	$\sqrt{np(1-p)}$

2)What if we have a sample instead of a whole distribution? (Think about the errors of the historical signals; these are a sample.) How do you find the mean, variance and standard deviation of the sample? We need new formulas, which follow:

<u>For a Sample:</u> Mean	Variance	Standard deviation
$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$	$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$	$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$

=average(.....)	=var(.....)	=stdev(.....)
-----------------	-------------	---------------

3) Example 8 from text: Let X be the number of days that a heart transplant recipient stays in the hospital after a transplant. An insurance executive wanted to estimate the mean, μ_X , and standard deviation, σ_X . To do this, she took a random sample of 12 transplant recipients. The numbers of days for which these people were hospitalized are: 8, 7, 9, 10, 9, 10, 6, 7, 6, 8, 10, 8.

a. Calculate sample standard deviation.

1.46

b. Use **VAR** and **STDEV** to compute s^2 and s for the following random sample of values of a random variable X .

s^2 (2.15) and s (1.46)

4) Let X be the continuous random variable with *p.d.f.*

$$f_X(x) = \begin{cases} 1.2 \cdot x^2 + 1.2 \cdot x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Use **Integrating.xls** to compute $V(X)$ and σ_X .

$V(X) = .05$ and $\sigma_X = .223$

5)

Let X be the exponential random variable with parameter $\alpha = 4$. Recall that both the mean and standard deviation of X are equal to 4. Let S be the standardization of X . Compute $P(S \leq 1)$. (Hint: First express $P(S \leq 1)$ in terms of a probability for X , then use the formula for the cumulative distribution function of X to finish the exercise.)

0.8647

6)

In the future we want to learn about a whole population from a sample. For example, if you sample shoppers to see how much they will pay for a new item, what can you conclude?

In order to draw conclusions from the sample (referred to as “making a statistical inference”), we have to know how the mean of a sample varies as we take new samples. This is what the Central Limit Theorem tells us and this is what we will do today.

Central Limit Theorem says that as sample size, n , gets larger, the distribution of sample means is approximately

-Normal, and has

-Same mean as original distribution; that is, Mean = μ

-Standard deviation = original standard deviation over square root sample size; that is, Standard Deviation = $\frac{\sigma}{\sqrt{n}}$

a)

Let X be a random variable with a mean of 15.9 and a standard deviation of 0.24. Let \bar{x} be the sample mean for random samples of size $n = 180$. Compute the expected value, variance, and standard deviation of \bar{x} .

expected value(15.9), variance(0.00032), and standard deviation of \bar{x} (0.0178)

b) CLT game (We will do this in class)

Worksheet 3

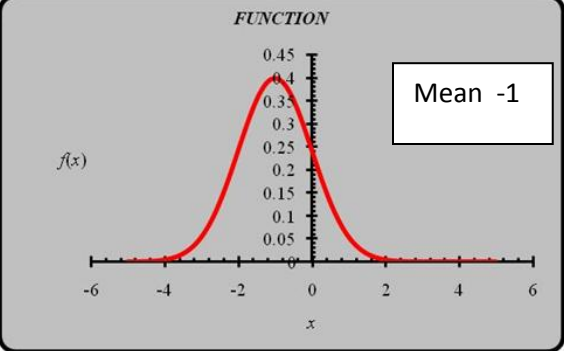
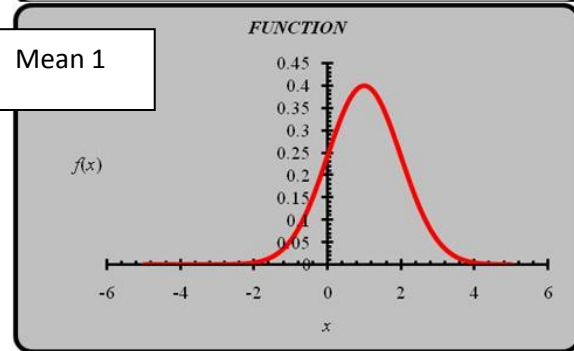
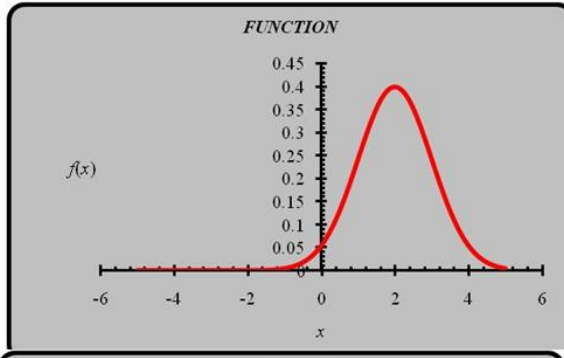
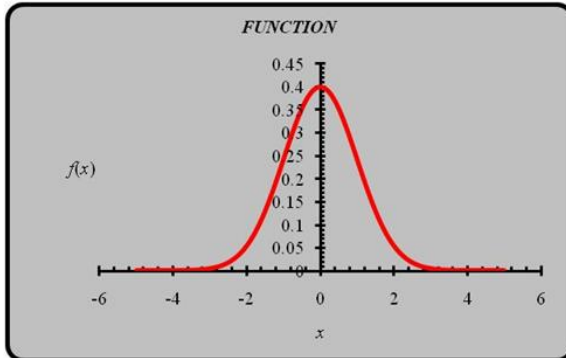
Prep-Work (Normal Distributions)

The Normal Distribution

1. Using = NORMDIST(x , μ , σ , false), graph the pdf for $\sigma = 1$ and $\mu = 0, 1, 2, 3, -1$, Use the interval $[-5, 5]$.

Mean 0

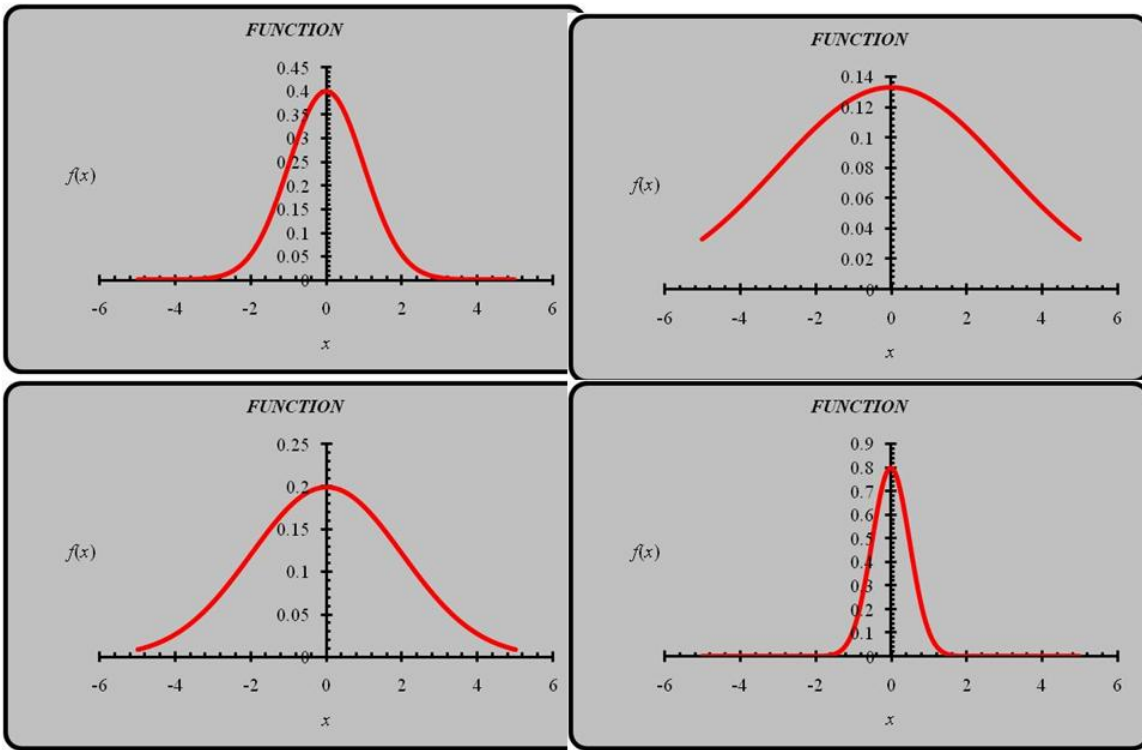
Mean 2



2. What does the value of μ tell you? What does changing μ do?

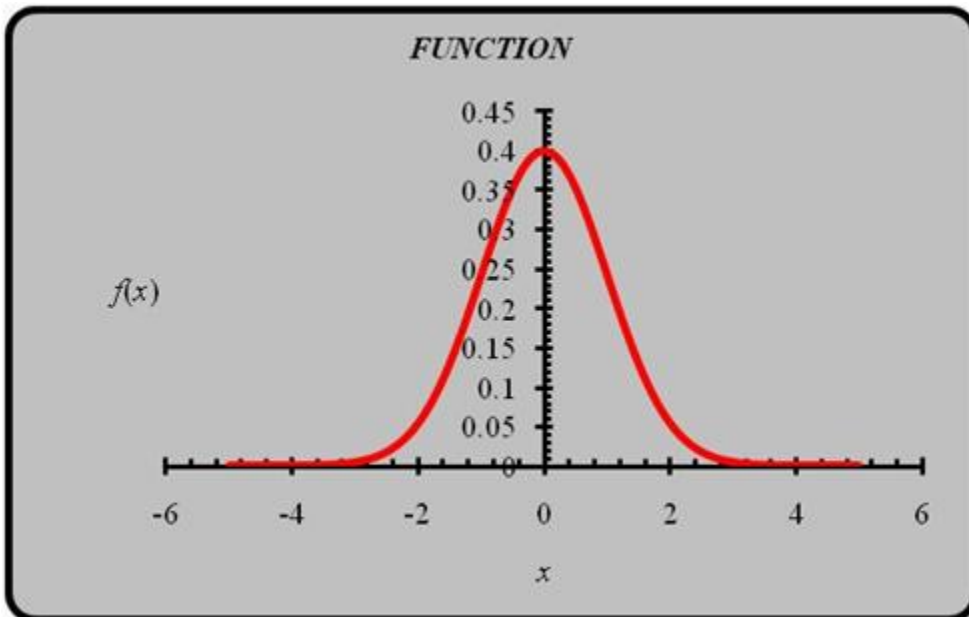
The x-value of the peak(Typical value), The location peak changes

3. Using = NORMDIST(x, μ, σ , false), graph the pdf for $\sigma = 1$ and $\mu = 0$ and $\sigma = 1, 2, 3, 0.5$, Use the interval $[-5, 5]$.



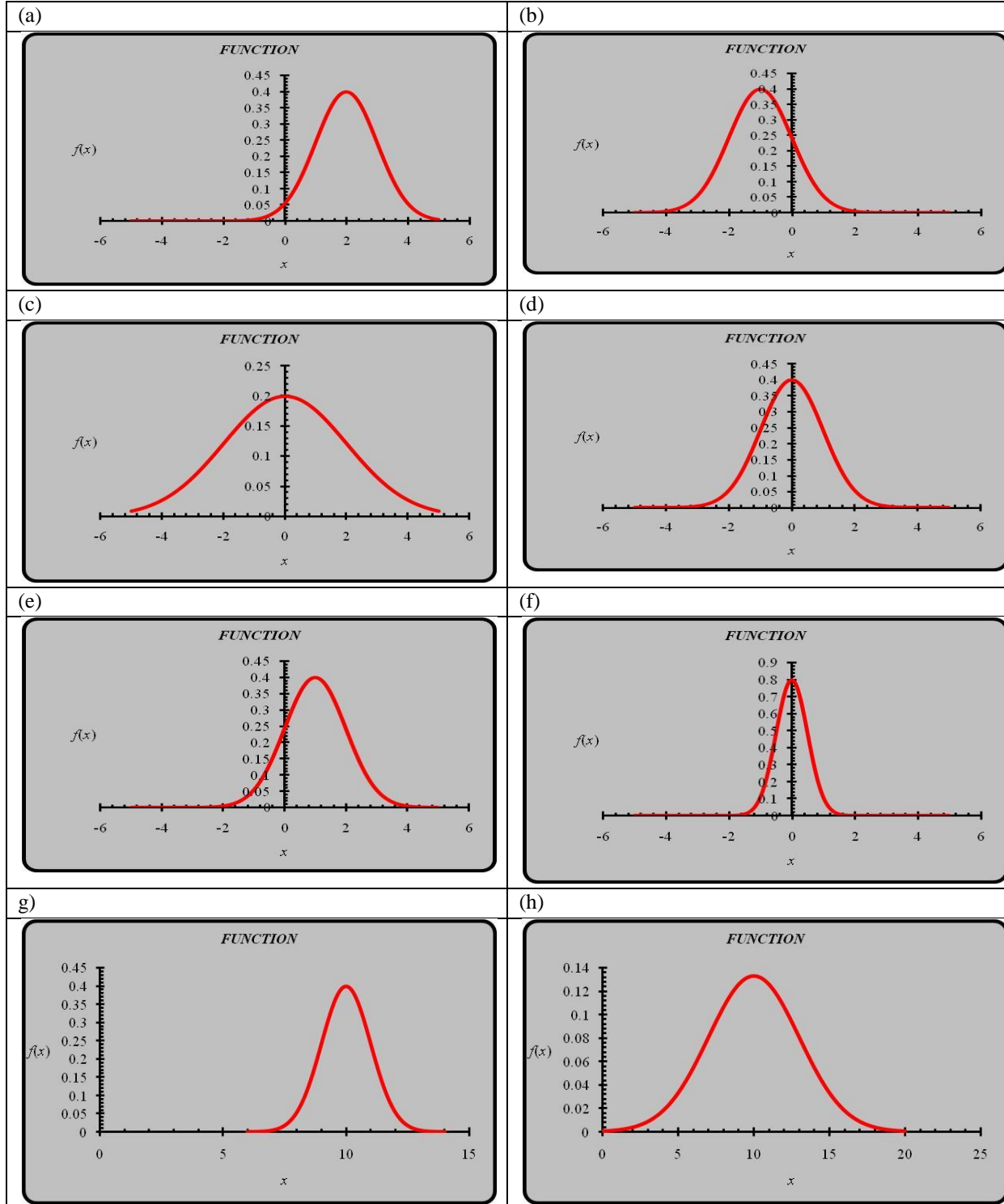
4. What does the value of σ tell you? The average distance from the average value
 What does changing σ do? When it is larger the graph gets wider

5. Standard normal distribution has mean of zero and standard deviation of 1. Which is its graph



6. Match the following graphs of normal pdfs with the one of the value of the parameters μ and σ . You will not use all the values of the parameters.

μ, σ	(0,1)	(1,0)	(1, 1)	(2,1)	(-1,1)	(0, 2)	(0, 0.5)	(10, 1)	(10,3)	(10,10)
answers	d	none	e	a	b	c	f	g	h	none



The normal distribution with mean μ and standard deviation σ has pdf

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

though we use = NORMDIST(x, μ , σ , false) for computation. The standard normal has $\mu = 0$, $\sigma = 1$.

Probabilities and the standard normal distribution. Let X have the standard normal distribution.

7. Using the pdf, write an expression for the probability that X is within one standard deviation of the mean. (Use the formula at the top of the page.)

$$P(-1 \leq X \leq 1) = \int_{-1}^1 \frac{1}{1\sqrt{2\pi}} e^{-0.5\left(\frac{x-0}{1}\right)^2} dx$$

8. Using the pdf, calculate the probability that X is within one standard deviation of the mean.
Using integrating.xls & answer in 2 we get .6827
9. Using the cdf, calculate the probability that X is within one standard deviation of the mean. **To find cdf at 1 use NORMDIST(1, 0, 1, true)**

$$P(-1 \leq X \leq 1) = F_X(1) - F_X(-1) = 0.6827$$

Probabilities for any normal distribution: "Rule of Thumb"

10. The results in #2-10 are true for all normal distributions. Summarize your results in the following table

<u>Distance from Mean in Normal Distribution</u>	<u>Probability</u>
Within one standard deviation of mean	0.6827
Within two standard deviations of mean	.9545
Within three standard deviations of mean	.9973

Standardization of Normal Random Variables. If X is normally distributed, its standardization is

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

- 11
What is the distribution of Z ?
Standard Normal

Suppose that X is normally distributed, with a mean X of 30 and standard deviation of 5.

- 12 What is the Z -value (that is, the standardized value) of $X = 35$?
1
13 What is the standardized value of $X = 40$?
2
14 if a value of X is three standard deviations above the mean, what is its Z value? 3 What is the X value? 45

Finding the Z value corresponding to particular probabilities

1. Using Excel, find the value of z_0 such that $P(Z < z_0) = 0.975$. Give two decimal places. Use NORMDIST and trial and error.
1.96

Definition	Computation	Plot Interval		Integration Interval			
Formula for $f(x)$	x	$f(x)$	A	B	a	b	$\int_a^b f(x) dx$
0.39894	0.39894	-5	5	-5	1.95996	0.9750	

FUNCTION

Constants	
s	
t	
u	
v	
w	

Integrate

Hide

Unhide

2. Find the value of z_0 such that $P(-z_0 < Z < z_0) = 0.99$.

Definition	Computation	Plot Interval		Integration Interval			
2.575							$\int_a^b f(x) dx$
Formula for $f(x)$	x	$f(x)$	A	B	a	b	$\int_a^b f(x) dx$
0.39894	0.39894	-5	5	-5	2.57589	0.9950	

FUNCTION

Constants	
s	
t	
u	
v	
w	

Integrate

Hide

A 50 kg sack of flour contains a weight of flour that is normally distributed with mean 51 kg and standard deviation 2 kg.

3. What is the Z-value of a weight of 50 kg?

-0.5

Standardization of Mean from Samples of Size n . By the Central Limit Theorem, the sample means \bar{x} is normally distributed with mean μ and standard deviation σ/\sqrt{n} . Thus the standardization, has the standard normal distribution, where

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

This is true no matter what the distribution of X provided the samples are random and n is large enough (usually above 30). (Quite remarkable!)

4. A sample of 4 sacks of flour has mean 50 kg. What is the Z-value of this mean?

-1

5. A sample of 25 sacks of flour has mean 50 kg. What is the Z-value of this mean?

-2.5

6. A sample of 100 sacks of flour has mean 50 kg. What is the Z-value of this mean?

-5

Confidence Intervals

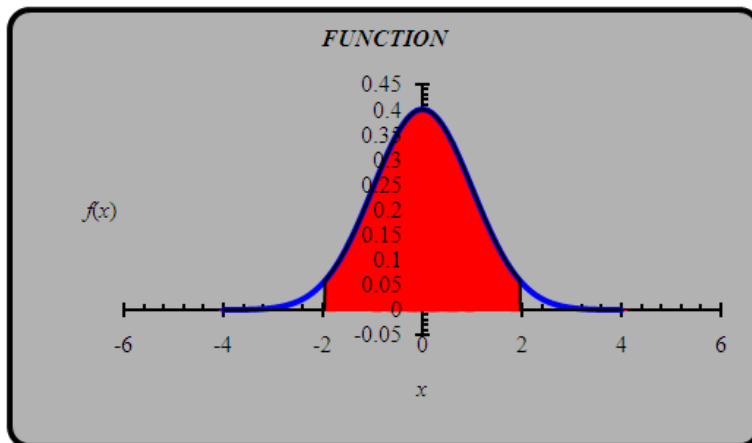
Last time we showed that $P(-1.96 < Z < 1.96) = 0.95$, where Z is the standard normal variable.

The variable Z represents the standard normal variable.

$$\underline{P(-1.96 < Z < 1.96) = 0.95}$$

1. Represent this on a diagram.

Definition	Computation		Plot Interval		Integration Interval		$\int_a^b f(x) dx$
Formula for $f(x)$	x	$f(x)$	A	B	a	b	
0.39894		0.39894	-4	4	-1.96	1.96	0.9500



Constants

s	
t	
u	
v	
w	

Integrate

Hide

Unhide

2. Explain what this result means in words.

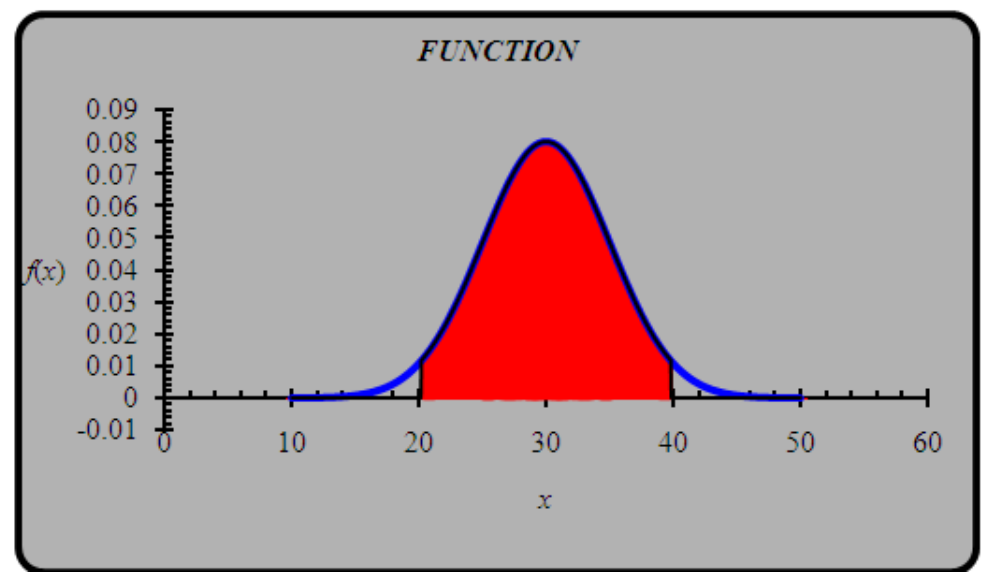
We are 95 % confident the Z value will fall between -1.96 and 1.96

Suppose that X is normally distributed, with a mean X of 30 and standard deviation of 5. Let

$$Z = \frac{x-30}{5}$$

3. What is the value of $P(-1.96 < \frac{x-30}{5} < 1.96)$? Illustrate on a diagram.

Definition	Computation		Plot Interval		Integration Interval		$\int_a^b f(x) dx$
Formula for $f(x)$	x	$f(x)$	A	B	a	b	
1.2E-09		1.2E-09	10	50	20.2	39.8	0.9500



Constants

s	30
t	5
u	
v	
w	

Integrate

Hide

Unhide

$$P(-1.96 < \frac{x-30}{5} < 1.96) = P(-1.96 < Z < 1.96) = 0.95$$

4. What is the value of $P(30-1.96*5 < x < 30+1.96*5)$? Illustrate on a diagram.

$$P(30-1.96*5 < x < 30+1.96*5) = P(-1.96 < Z < 1.96) = 0.95$$

Same illustration as in part 3.

5. What is the value of $P(20.2 < X < 39.8)$? Illustrate on a diagram.

This is the same probability; just evaluate the values of X at the end points as in the graph in part 3.

Standardization of Mean from Samples of Size n . By the Central Limit Theorem, for a sample of size n , the sample means \bar{x} are normally distributed with mean μ and standard deviation σ/\sqrt{n} . Thus the standardization, Z , has the standard normal distribution, where

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

This is true no matter what the distribution of X provided the samples are random and n is large enough (usually above 30). (Quite remarkable!)

Continuing the example where the random variable X has a mean 30 and standard deviation 5. Let's take a sample of 100 and find the mean \bar{x}

6. What is the mean of all the possible \bar{x} s?

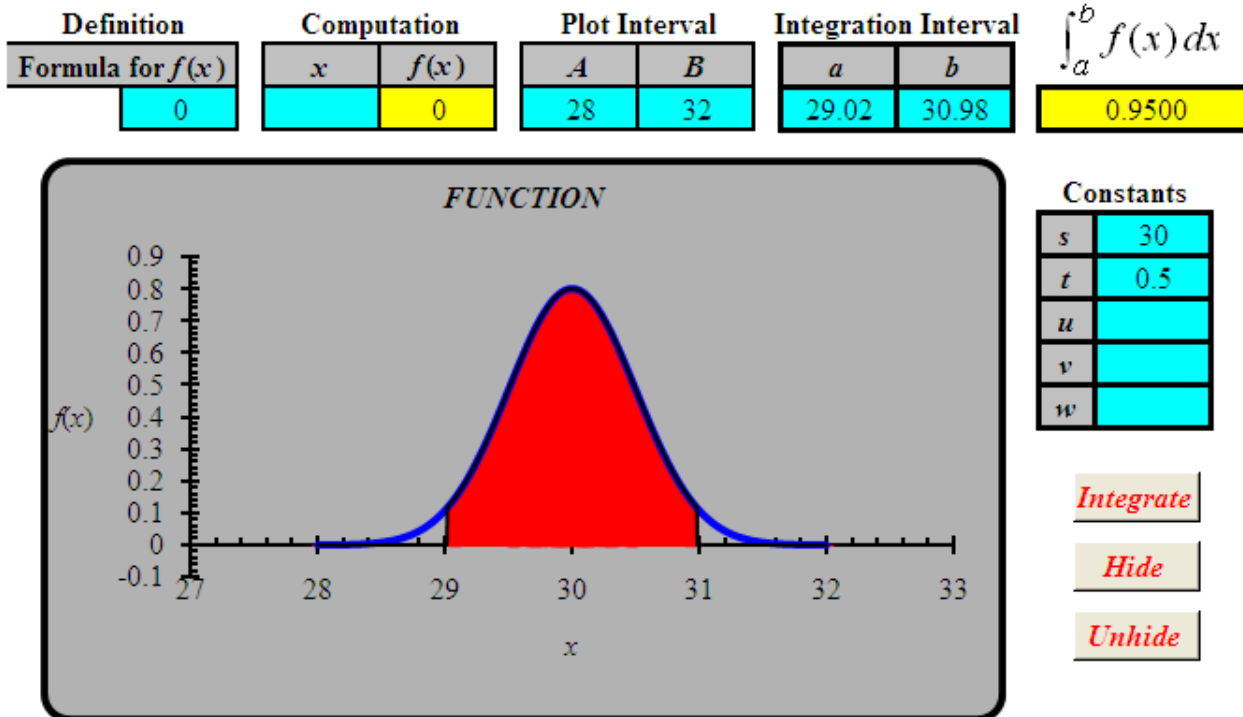
The mean value is still the same, i.e. 30.

7. What is the standard deviation of all the possible \bar{x} s?

The standard deviation is $5/\sqrt{100} = 0.5$

8. What is the value of $P(30 - 1.96 * 0.5 < \bar{x} < 30 + 1.96 * 0.5)$? Illustrate on a diagram.

This is the 95% confidence interval. So the probability is 0.95

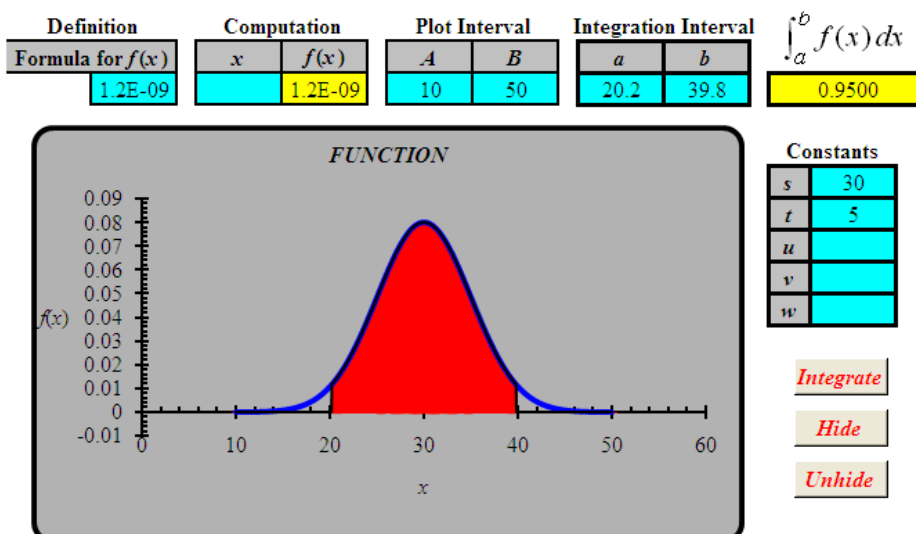


9. What is the value of $P(29.02 < \bar{x} < 30.98)$? Illustrate on a diagram.

Its value is 0.95, just evaluate the endpoints of the interval. We are 95% sure the values of \bar{x} will fall between 29.02 and 30.98. The graph is the same as in part 8.

10. What is the interval in which there is a 95% chance of finding an x value?

For X , the confidence interval is from $30 - 1.96 * 5$, to $30 + 1.96 * 5$. That is 20.2 to 39.8 as illustrated below.



11. Give an intuitive explanation of why the interval for \bar{x} is shorter than the interval for X .

The reason is that there is more concentration around the mean since we are dividing by the square root of the sample size, i.e. by 10.

12. What would happen to the length of the interval if the size of the sample (now 100) was increased? Would it get longer or shorter? Why?

The size of the interval will be squeezed further. It is inversely proportional to the square root of the size of the sample.

Now suppose that the mean of a distribution is NOT known, but that the standard deviation is known. Suppose we take a sample of size n and find that it has mean \bar{x} . Then it can be shown that there is a 95% chance that the mean of the population lies in the interval given by the formula

$$\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

This is called the 95% confidence interval for the mean.

Example 3, Normal Distributions. An administrator samples 50 other administrators' salaries and find the mean of the sample to be $\bar{x} = \$88,989$ and the standard deviation of the sample to be $s = \$22,358$. The standard deviation of the sample is a good approximation to σ , the standard deviation of the population.

13. Find the 95% confidence interval for the mean of all such administrators' salaries.

Using

$$\bar{x} - 1.96 \frac{\sigma_x}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma_x}{\sqrt{n}}$$

(\$82,791, \$ 95187)

14. What does the interval in #13 tell you?

95% chance that the mean national mean salary for his counter parts is between \$82,791 & 95187

The reason the administrator took the sample was to show that he was paid less than the mean.

15. If the administrator's own salary is \$83,500, can he claim with 95% certainty that he is paid less than the mean?

No . Because his salary is within the 95% confidence interval

16. If the administrator's own salary is \$81,500, can he claim with 95% certainty that he is paid less than the mean?

Yes . Because his salary is outside the 95% confidence interval

Suppose the sample size had been 100 instead of 50.

The 95% confidence interval is NOW (\$84,607, \$93371)

17. With a salary of \$83,500 would he have been able to claim he was paid less than the mean?

Yes . Because his salary is outside the 95% confidence interval

18. With a salary of \$81,500 would he have been able to claim that he was paid less than the mean?

Yes . Because his salary is outside the 95% confidence interval