Continuous Probability Distributions

Huining Kang

HuKang@salud.unm.edu

August 5, 2020

Overview

- Content
    - Review discrete probability distribution
    - Probability distributions of continuous variables
    - The Normal distribution
- Objective
    - Consolidate the understanding of the concepts related to probability distribution
    - Understand the concepts related to the continuous probability distribution
    - Understand the normal distribution and standard normal distribution. know how to calculate the probabilities of the events based on the standard normal distribution

Review of discrete probability distributions

- Example
  - 10% of a certain population is color blind
  - Draw a random sample of 5 people from the population, and let X be the number of people who are color blind among this sample.
  - Questions
    - What are the possible values that X assumes?
    - What is the probability that X assumes each of the above possible values
- Solution
  - X follows Binomial distribution $Binomial(n, p)$, where $n = 5, p = 0.1$, and $q = 1 - p = 0.9$.

Posible values of X	x					
	0	1	2	3	4	5
Probability density function $f(x) = P(X = x)$	q^5	$5q^4p$	$10q^3p^2$	$10q^2p^3$	$5qp^4$	p^5
	.5905	.3281	.0729	.0081	.0005	.0001
Cumulative distribution function $F(x) = P(X \leq x)$	$F(0)$	$F(1)$	$F(2)$	$F(3)$	$F(4)$	$F(5)$
	.5905	.9185	.9914	.9995	.9999	1

- Probability density function

$$f(x) = P(X = x) = \frac{n \cdot (n-1) \cdots (n-x+1)}{x!} q^{n-x} p^x, x = 0, 1, 2, 3, 4, 5.$$

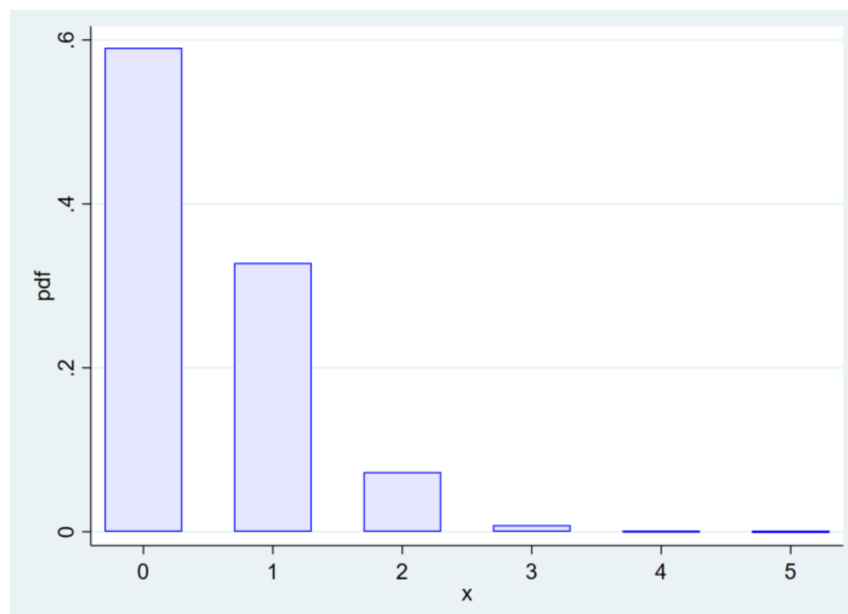
Review of discrete probability distributions

- Solution
  - X follows Binomial distribution $Binomial(n, p)$, where $n = 5, p = 0.1$, and $q = 1 - p = 0.9$.

Posible values of X	x					
	0	1	2	3	4	5
Probability density function $f(x) = P(X = x)$	q^5	$5q^4p$	$10q^3p^2$	$10q^2p^3$	$5qp^4$	p^5
	.5905	.3281	.0729	.0081	.0005	.0001
Cumulative distribution function $F(x) = P(X \le x)$	$F(0)$	$F(1)$	$F(2)$	$F(3)$	$F(4)$	$F(5)$
	.5905	.9185	.9914	.9995	.9999	1

$$f(x) = P(X = x) = \frac{n \cdot (n-1) \cdots (n-x+1)}{x!} q^{n-x} p^x,$$

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



Review of discrete probability distributions

- What is the probability distribution of a discrete random variable?
  - (From the textbook) is a table, graph, formula, or other device used to specify all possible values of a discrete random variable along with their respective probabilities.
  - (Also from the textbook) is a device that can be used to describe the relationship between the values of a random variable and the probabilities of their occurrence.
  - (From Wikipedia) is the mathematical function that gives the probabilities of different possible outcomes for an experiment.
- My definition
  - The relationship between the possible outcomes (values of a random variable) and the probability of the their occurrence is referred to as the probability distribution.
  - Probability distribution (of a random variable) may be expressed in the form of a table, graph or formula

Review of discrete probability distributions

- Why is it important?
  - It can help us to calculate the probability of an event under more complex conditions.
  - If you know the type of the probability distribution (e.g. binormial, Poisson, etc.), you can calculate the probability of an event using the tables or statistical software.
- Example
  - What is the probability that at least one is color blind?

$$P(X \geq 1) = 1 - P(X = 0) = 1 - q^5 = 1 - 0.9^5 = 0.4095$$

Stata command: disp 1 – binomial(5, 0, 0.1)

  - What is the probability that at least two are color blind?

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0.9185 = 0.0815$$

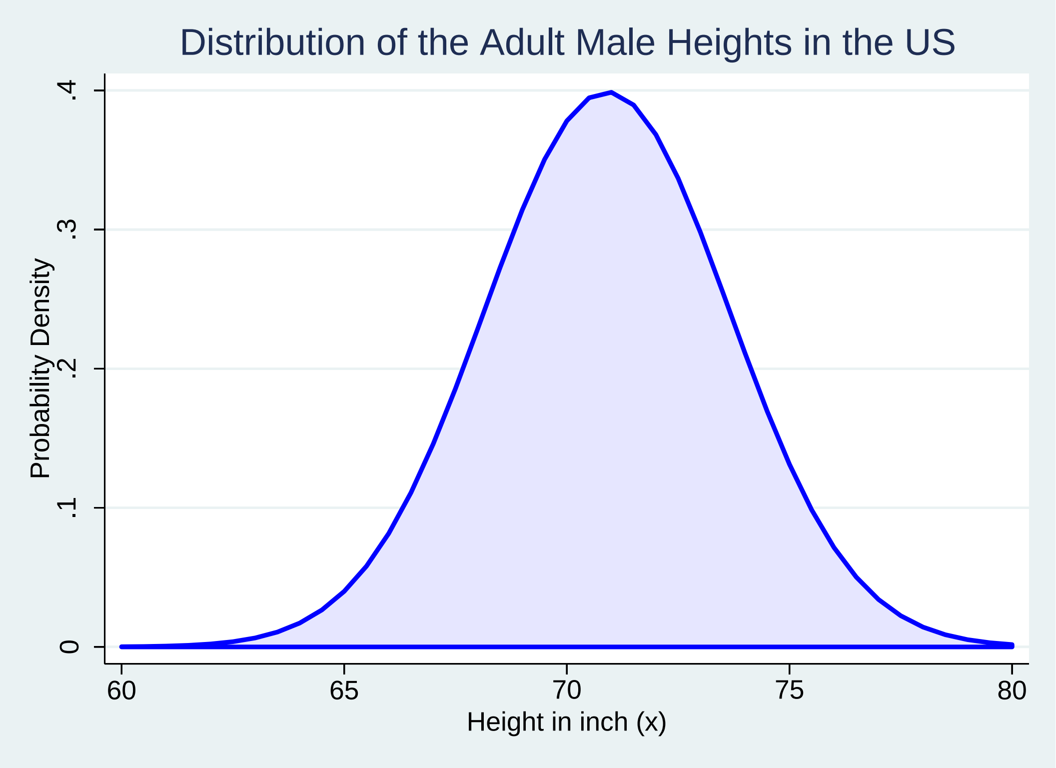
Stata command: disp 1 – binomial(5, 1, 0.1)

Probability distributions of continuous variables

- Examples of the continuous random variable
  - $X =$ the height of a randomly selected adult male from the US
  - $T =$ time from the diagnosis to the death of a woman randomly selected from the patients with ovarian cancer.
- Characteristics
  - Does not possess the gaps or interruptions
  - Can take on an infinite number of possible values, corresponding to every value in an interval.
    - X could be any value between 60 and 80 inches
    - T can assume any positive values
- Challenge in the theory
  - We cannot model the continuous random variables with the same methods as we used for the discrete random variables
    - Tables or Histogram won't work for a continuous random variables
  - There are some similarities, but we have to use different methods

Probability distributions of continuous variables

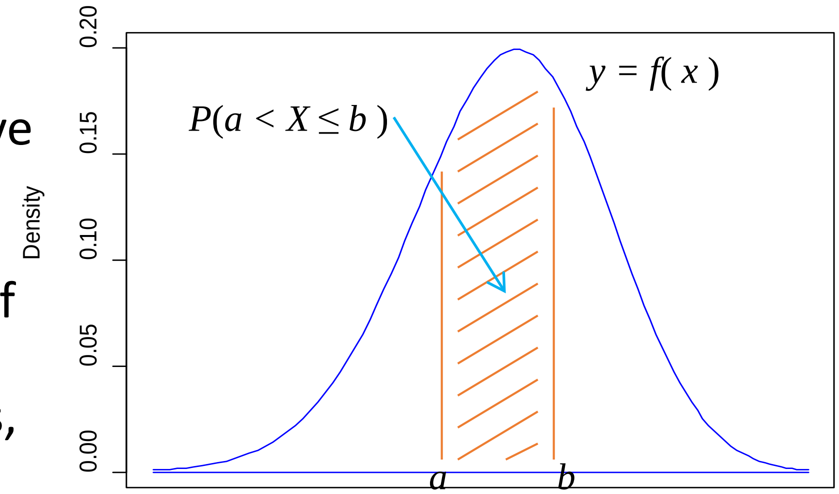
- If searching online with the key words distribution, height of US adult males, you may find something similar to the following graph
  - Average is $\mu = 70.9$ in.
  - Standard deviation is $\sigma = 2.75$ in



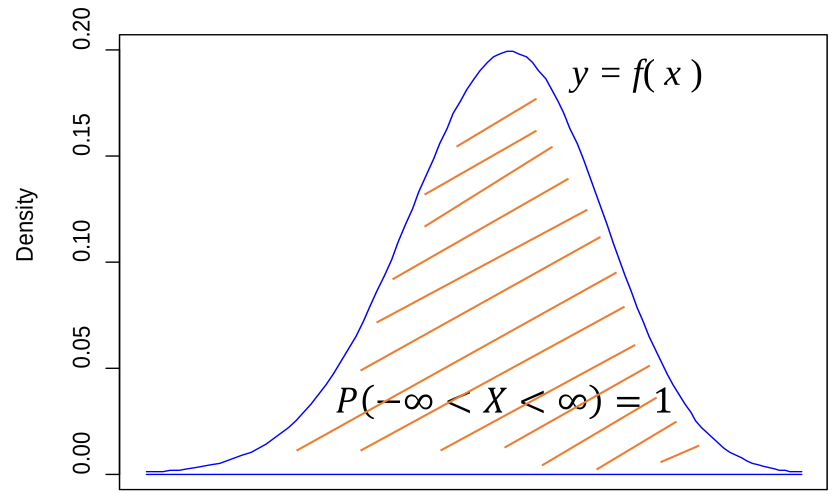
  - Impression
    - It looks like a smooth version of a histogram
    - The curve is a graph of certain function $y = f(x)$
    - The values of the height with the curve that is high are more likely to occur than where it is low

Probability distributions of continuous variables

- Probability density function (pdf)
  - Suppose X is a continuous random variable. If there exists a nonnegative function $f(x)$ such that for any interval $[a, b]$, $P(a \leq X \leq b)$ is equal to the area under the graph of $f(x)$ enclosed by the two vertical lines at the point a and b and x-axis, then $f(x)$ is called the probability density function (pdf) of X

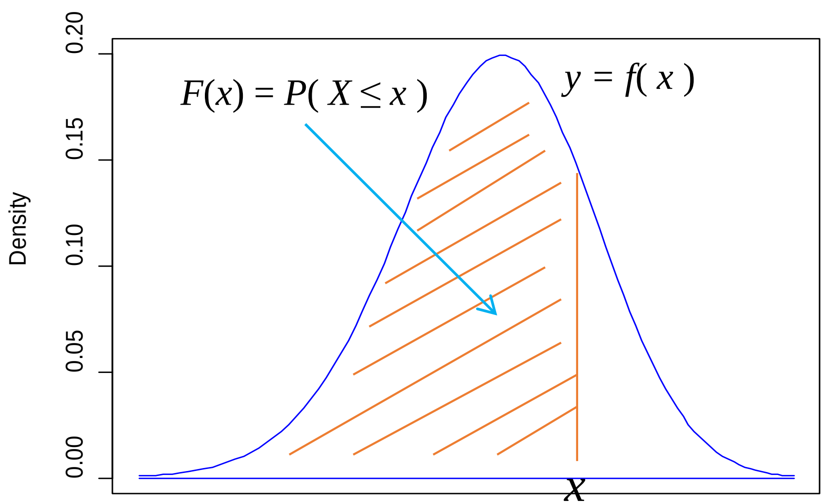


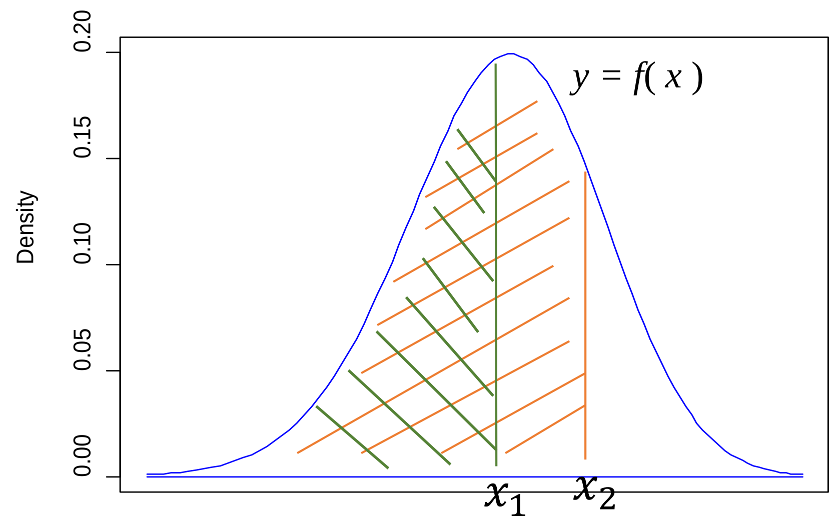
- Properties
  - $f(x) \geq 0$
  - Given a specific value c, $P(X = c) = 0$.
  - $P(a \leq X \leq b) = P(a < X < b)$
  - The entire area under the graph of $f(x)$ and above x-axis is 1
    - $P(-\infty < X < \infty) = 1$



Probability distributions of continuous variables

- Cumulative distribution function (cdf)
  - By definition the cumulative distribution function (cdf) $F(x) = P(X \leq x)$ is the area enclosed by the graph of $f(x)$, x-axis and the vertical line at point
    - $F(x) = P(X \leq x) = P(-\infty < X \leq x)$.



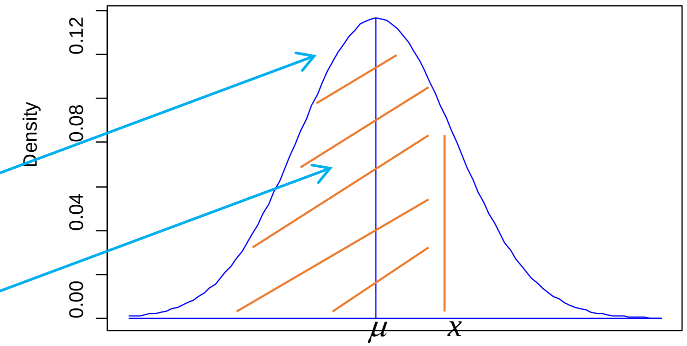


- Properties
  - $x_1 < x_2$, Then $F(x_1) \leq F(x_2)$
  - $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$

Probability distributions of continuous variables – Normal distribution

- Definition $X \sim N(\mu, \sigma^2)$

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

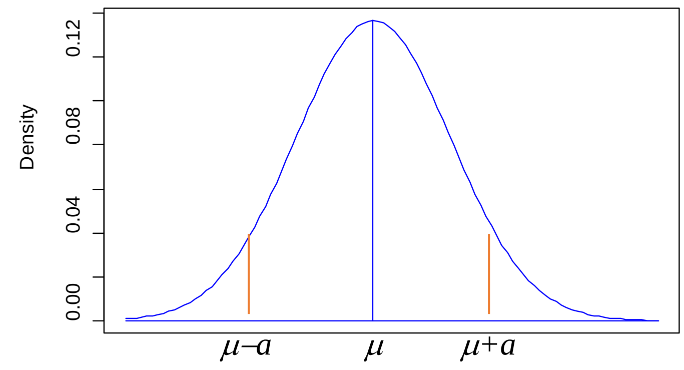


  - cdf $$F(x) = P(X \le x)$$
  - Mean = μ, Variance = σ^2

- Properties
  - Symmetric about its mean μ

    $$P(X \le \mu) = P(\ge \mu) = 0.5$$

    $$P(X \le \mu - a) = P(\ge \mu + a)$$



  - Mean = median = mode
  - Area under the curve = 1

Normal Distribution

- Properties
  - 68 – 95 – 99.7 rule (for how much of the distribution is within 1, 2 and 3σ's from the center μ)

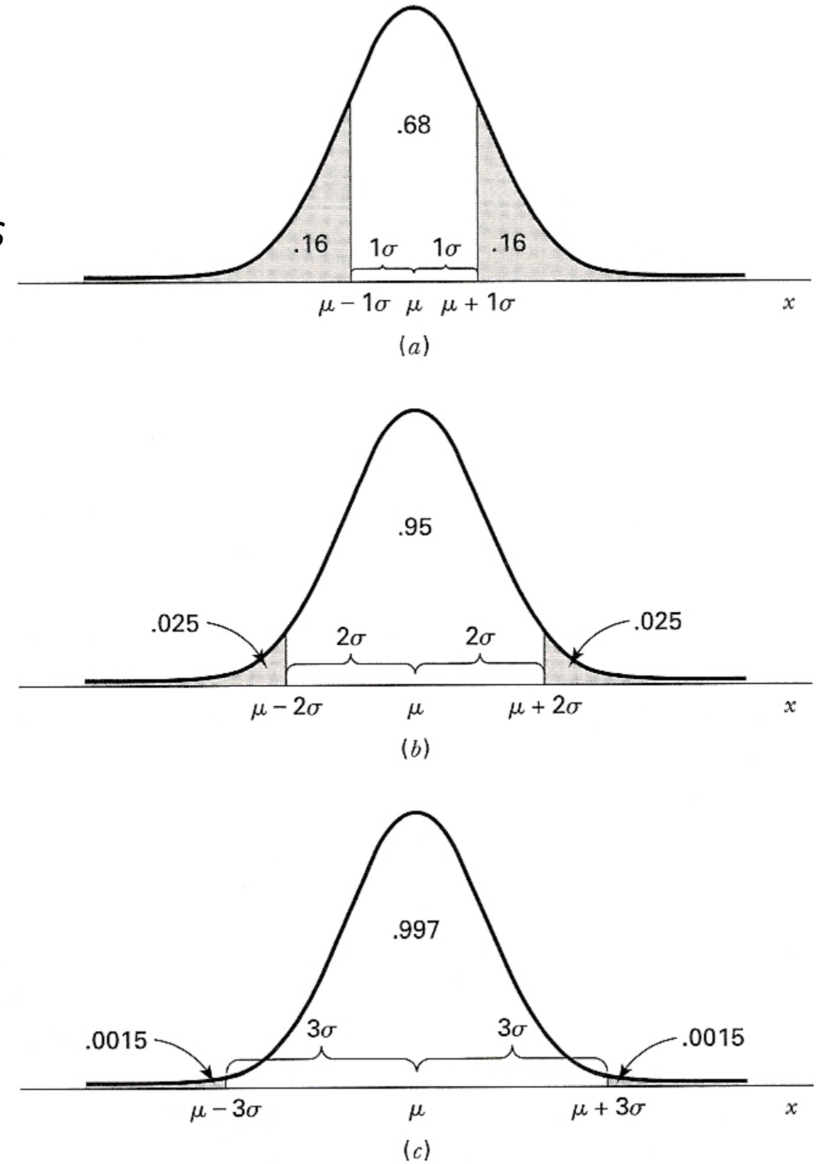


FIGURE 4.6.2 Subdivision of the area under the normal curve (areas are approximate).

Normal Distribution

- Properties
  - Normal distribution is completely determined by μ and σ.
    - μ = location, σ: shape (page 118, 10th Ed)

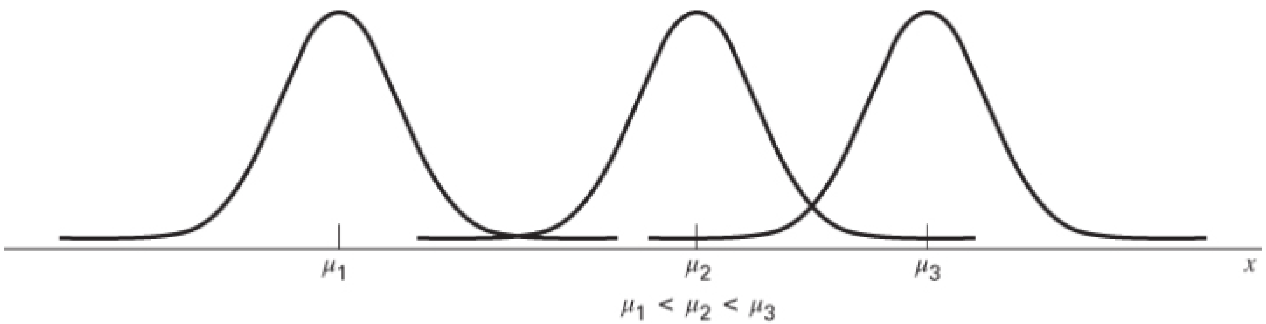


FIGURE 4.6.3 Three normal distributions with different means but the same amount of variability.

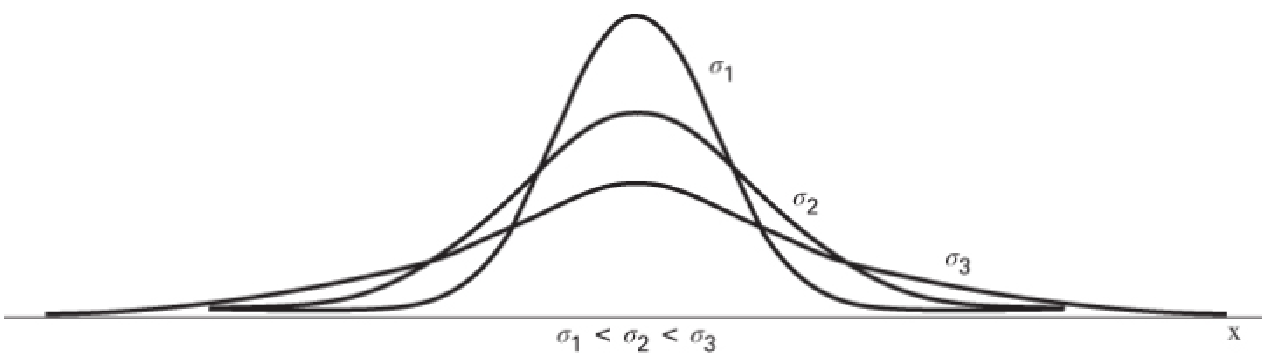
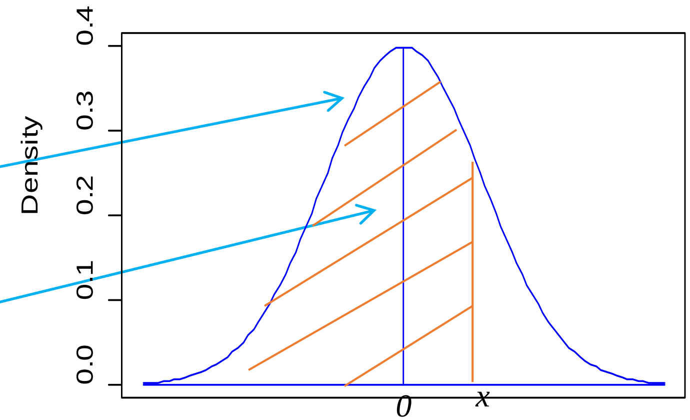


FIGURE 4.6.4 Three normal distributions with different standard deviations but the same mean.

Standard Normal distribution

- Standard normal distribution
  - $Z \sim N(0, 1)$, $\mu = 0$, $\sigma = 1$.
  - pdf $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - cdf $\Phi(z) = P(Z \le z)$



- Calculation of standard normal distribution
  - Table of standard normal distribution function (Page A38 - A39)
  - Stata function
    - cdf of standard normal distribution: normal(z)
    - pdf of standard normal distribution: normalden(z)

Calculation of Standard Normal Distribution

- Stata function
  - cdf of standard normal distribution: normal(*z*)
  - pdf of standard normal distribution: normalden(*z*)

Example	Stata calculation
4.6.1 $P(Z \le 2) = \Phi(2)$	. disp normal(2) .97724987
4.6.2 $P(-2.55 < Z \le 2.55) = \Phi(2.55) - \Phi(-2.55)$	. disp normal(2.55) - normal(-2.55) .98922771
4.6.3 $P(-2.74 < Z \le 1.53) = \Phi(1.53) - \Phi(-2.74)$	. disp normal(1.53) - normal(-2.74) .93391968
4.6.4 $P(Z \ge 2.71) = 1 - P(Z < 2.71) = 1 - \Phi(2.71)$	. disp 1- normal(2.71) .00336416
4.6.5 $P(0.84 < Z \le 2.45) = \Phi(2.45) - \Phi(0.84)$	. disp normal(2.45) - normal(.84) .19331138

Summary

- Reviewed
  - Discrete probability distribution
- Learned
  - Probability distributions of continuous variables
  - The Normal distribution
- Objective
  - Consolidate the understanding of the concepts related to probability distribution
  - Understand the concepts related to the continuous probability distribution
  - Understand the normal distribution and standard normal distribution know how to calculate the probabilities of the events based on the standard normal distribution

Homework

- Due Wednesday, August 19.
- Page 122 (10^{th} Ed)
    - Exercise 4.6.1
    - Exercises 4.6.3 – 4.6.6. What relationship between $\Phi(x)$ and $\Phi(-x)$ is implied by these results. Please explain why it is true. (hint: using graph)
    - Exercises 4.6.7 – 4.6.9.