Chapter 6 Lecture Notes

Matt Warner

6.1 Probability Distributions for a Continuous Random Variable

2 Types of Random Variables

- **Discrete**
  - **possible values** - a finite set or a countably infinite set
    (usually associated with counting, e.g. "the number of...")
  - **Probability distribution** - specified by giving a
    - Table

x	1	2	3	4	5	6
$p(x)$	0.40	0.20	0.12	0.10	0.14	0.04

    - Probability Mass Function

$$e.g.\;\; p(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \qquad \text{for x = 0,1,2..., n}$$

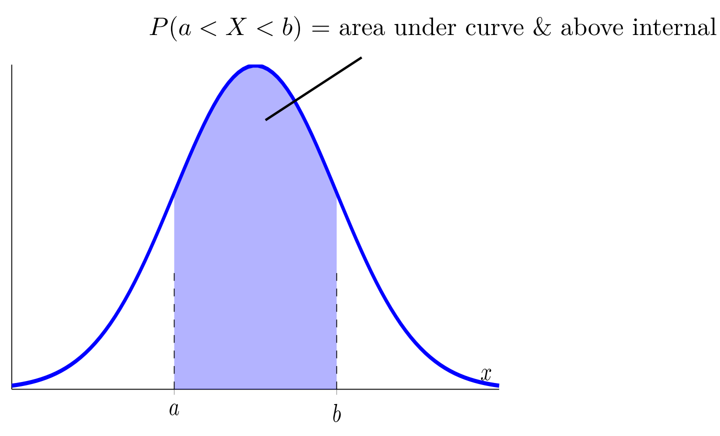
    - **Example** - Binomial distribution

- **Continuous**
  - **Possible values** - an infinite set that forms an interval on the number line
    (usually associated with measuring, e.g. heights, weights, volumes, times, etc.)
  - **Probability distribution** - specified by giving a **probability density function**
  - **Example** - Normal distribution

Probability Density Function

A Probability Density Function is a curve that

- Is on or above the x-axis
- Has total area = 1.0



The Normal Distribution

- The **Normal Distribution** is defined by a probability density function that is a symmetric, bell shaped curve.

- It is used to suggest or hypothesize a distribution for an ***entire population***

- There are many normal distributions. Each is identified by giving values of

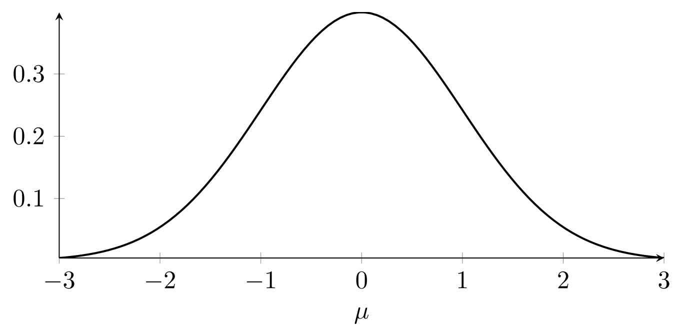
  - μ = population mean (Note $\bar{x}$ = sample mean)

  - σ = population standard deviation (Note: s = samplestd. dev.)

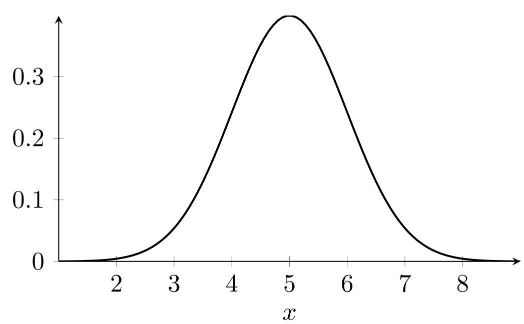
Question 1

Sketch the normal distribution corresponding to each of the following

a) $Z \sim \mathrm{N}(\mu = 0, \sigma = 1)$



b) X $\sim \mathrm{N}(\mu = 5, \sigma = 1)$



Note:-

- The Standard normal distribution is the most important one to understand.
- Any generic or **non-standard normal** X∼N(μ, σ) can be converted into the standard normal using the formula $Z = \dfrac{x-\mu}{\sigma}$
- Any desired probability about the random variable X is found by **standardizing X** to produce Z and then using the table of standard normal cumulative probabilites (commonly called the **Z table**)

Question 2

Let the random variable X represent the SAT score of a randomly chosen student. Suppose that the SAT scores of all students in the population vary according to a normal distribution with mean $\mu = 1050$ and standard deviation $\sigma = 150$

What does the empirical Rule tell us about the following probabilities

a) $P(900 < X < 1200) = .68 \rightarrow 1$ standard deviation from the mean

b) $P(750 < X < 1350) = .95 \rightarrow 2$ standard deviation from the mean

c) $P(600 < X < 1500) = .997 \rightarrow 3$ standard deviation from the mean

---

Find the probability that a randomly selected student scores less than 1275

That is,

$$P(X < 1275)$$

To find this we can use the formula

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

So, we get

$$Z = \frac{1275-1050}{150}$$

So,

$$Z = 1.50$$

Now we can find the the probability by using the Z table.

$$P(X < 1275) = .9394$$

Find the probability that the student scores more than 800

That is,

$$P(X > 800)$$

So,

$$P(\frac{x-800}{150} > \frac{800-1050}{150})$$

Now we have,

$$P(Z > -1.67)$$

which gives us,

$$P(X > 800) = 0.0475$$

Find $\mathbf{P(1100 \leq X \leq 1270)}$

That is,

$$P(\frac{1100 - 1050}{150} \leq Z \leq \frac{1270 - 1050}{150})$$

Solving that leaves us with,

$$P(.33 \leq Z \leq 1.47)$$

Lets get all the area to the left of 1.47

$$P(Z = 1.47) = .9292$$

Now, lets get everything to the left of .33

$$P(Z = .33) = .6293$$

Now, we subtract them

$$.9292 - .6293 = \boxed{.2999}$$

25% of students score <u>less</u> than ____________.

We need to find the Z score in the table find by finding the closest value to .25

$$P(X < .25) = -.67$$

Using the equation

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

We can solve for x

$$-0.67 = \frac{x - 1050}{150}$$

$$x = 949.5$$

The top 10% of students score more than __________.

Note:-

This is the same as asking for the value b so that $P(X > b) = .10$, which could also be called the 90th percentile.

Find the value in the table closest to .90

$$Z = 1.28$$

using the formula for Z, we can find b,

$$1.28 = \frac{b - 1050}{150}$$

So,

$$b = 1242.2$$

Therefore, the top 10% of students score more than 1242.2

find a symmetric interval about the mean that will capture the central 90% of SAT scores. That is, find the values L and U (where L and U are equally distant from the mena), so that

$$P(L < X < U) = .90$$

If we are asked for the middle 90%, then our left and right tails are both .05

So our z scores are ± 1.6449

Now we solve

$$1.6449 = \frac{u - 1050}{150} = 1296.7$$

$$-1.6449 = \frac{l - 1050}{150} = 808.3$$

So our interval is,

$$(1296.7, 808.3)$$

Question 3

The virginia Cooperative Extension reports that the mean weight of yearling Angus steers is 1152 pounds. Suppose that weights of all such animals can be described by a Normal model with a standard deviation of 84 pounds

What percent of steers weigh

---

a) Less than 120 Pounds

That is,

$$P(X < 1200)$$

So,

$$P(\frac{x - 1152}{84} < \frac{1200 - 1152}{84})$$

$$P(Z < .57)$$

Now we look up the value 0.57 in the table to find the probability.

$$P(Z < .57) = .7157$$

---

b) Less than 950 pounds

That is,

$$P(X < 950)$$

c) Greater than 1250 pounds

That is,

$$P(X > 1250)$$

So,

$$P(X > 1250) = \frac{1250 - 1152}{84}$$

So we have,

$$Z = 1.17$$

Looking in the table we find,

$$P(Z = 1.17) = .8790$$

Now we subtract 1 from that giving us,

$$\begin{aligned} P(X > 1250) &= 1 - .8790 \\ &= .1210 \end{aligned}$$

Note:-

Always round the Z scores to 2 decimal places