

## Simple Statistical Distributions on the TI-83

### Math 211 – ASU

In this course we look at basic statistical distributions for both the discrete and continuous cases. We start with a random variable  $X$  for some scenario.

#### *Discrete Case*

If  $p(X = x) \geq 0$  for just a finite set of  $x$  values, then we would say this is a *discrete* situation. For example, if we flip a coin ten times and let  $X$  = number of times heads appears, then the possible values for  $X$  are the integers between 0 and 10 inclusive. For these values,  $p(X = x)$  will be positive, and we can actually calculate these probabilities using the Binomial Probability Model, which we'll discuss in a moment. It wouldn't make sense to let  $X = 4.5$  (*i.e.* we got 4.5 heads in ten flips). From a probability standpoint, we'd say  $p(X = 4.5) = 0$ . We could summarize this scenario by saying that:

$$\begin{aligned} p(X = x) &\geq 0, & x = 0,1,2,3,4,5,6,7,8,9,10 \\ p(X = x) &= 0, & \text{all other } x \end{aligned}$$

Let's perform this example. We will use the Binomial Probability Model since this scenario meets the following conditions: it is independent from trial to trial, the probability of success does not change, and we are performing multiple trials.

The formula for a binomial probability is  $B(p, n, k) = {}_n C_k \cdot (p)^k \cdot (1 - p)^{n-k}$ , where  $p$  = probability of success,  $k$  = number of trials and  $n$  = number of successes.. However, we will use a built-in feature on the TI-83 (and higher) models called “**binompdf**” to perform these calculations. Hit 2<sup>nd</sup>-DISTR on your calculator to find this choice. Please note there is also a choice for **binomcdf** which we'll discuss in a moment.

The call sequence for binompdf is **binompdf(k,p,n)**, where the variables are as described in the preceding paragraph. The ‘pdf’ stands for *probability density function*. For our purposes it calculates the probability at a value  $n$ .

Say we flip our coin 10 times. The probability of heads is  $p = 0.5$ . Say we want 5 heads to appear. The call sequence would be **binompdf(10,0.5,5)**, which gives 0.246. Hence we have about a 24.6% likelihood of getting 5 heads in 10 flips.

Suppose we want to know the likelihood of getting *at most* 5 heads. Now we use the cdf feature, which stands for *cumulative density function*. This calculates the probability of the event occurring for all values from 0 up to and including  $n$ . We'd type **binomcdf(10,0.5,5)** = 0.623, which says there's a 62.3% likelihood of us getting up to and including 5 heads in 10 flips.

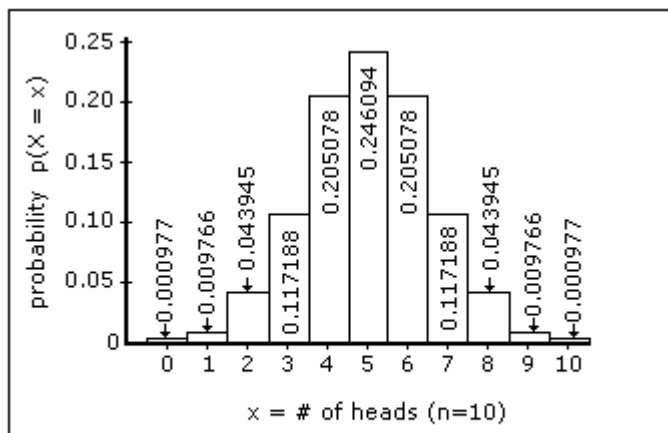
**binompdf** = probability **at** a value for  $n$

**binomcdf** = probability for the **interval** 0 up to  $n$ .

We can set up each probability and sketch a distribution model using the binompdf feature. For illustrative purposes, the following table also shows the cumulative (binomcdf) values as  $n$  progresses from 0 to 10.

$n$ = number of heads	Probability of $n$ heads	Cumulative probability
0	$\text{binompdf}(10,0.5,0) = 0.000977$	$\text{binomcdf}(10,0.5,0) = 0.000977$
1	$\text{binompdf}(10,0.5,1) = 0.009766$	$\text{binomcdf}(10,0.5,1) = 0.010742$
2	$\text{binompdf}(10,0.5,2) = 0.043945$	$\text{binomcdf}(10,0.5,2) = 0.054688$
3	$\text{binompdf}(10,0.5,3) = 0.117188$	$\text{binomcdf}(10,0.5,3) = 0.171875$
4	$\text{binompdf}(10,0.5,4) = 0.205078$	$\text{binomcdf}(10,0.5,4) = 0.376953$
5	$\text{binompdf}(10,0.5,5) = 0.246094$	$\text{binomcdf}(10,0.5,5) = 0.623047$
6	$\text{binompdf}(10,0.5,6) = 0.205078$	$\text{binomcdf}(10,0.5,6) = 0.828125$
7	$\text{binompdf}(10,0.5,7) = 0.117188$	$\text{binomcdf}(10,0.5,7) = 0.945313$
8	$\text{binompdf}(10,0.5,8) = 0.043945$	$\text{binomcdf}(10,0.5,8) = 0.989258$
9	$\text{binompdf}(10,0.5,9) = 0.009766$	$\text{binomcdf}(10,0.5,9) = 0.999023$
10	$\text{binompdf}(10,0.5,10) = 0.000977$	$\text{binomcdf}(10,0.5,10) = 1.000000$

Notice the cumulative values in the third column increase and eventually sum to 1. This should not be a surprise: if you flip a coin 10 times, it is certain ( $p = 1$ ) that you'll get at most 10 heads. Cumulative probabilities always increase in value and top out at 1. A histogram graph of the individual probabilities is given below.



Histogram graph of the probabilities for getting  $x$  heads in 10 flips of a fair coin. Since  $p = 0.5$  for heads, note the symmetry of this graph. Each bar is  $\text{binompdf}(10,0.5,x)$ . The sum of the bars = 1.

Suppose we want to know the probability of getting *at least* 5 shots, or  $p(X \geq 5)$ . The binomcdf works by accumulating probabilities from the left (from zero), so we need to reword this problem so that the cdf features can be used. The complement of making *at least* 5 is to make *at most* 4. You would add up the bars from 0 to 4, but leave 5 open. We get  $\text{binomcdf}(10,0.5,4) = 0.376953$ , from above. Hence, the probability of getting at least 5 heads is  $1 - 0.377$  (rounded) = 0.623, or about 62.3%.

Lastly, suppose we want to know the probability of getting between 3 and 5 heads, or  $p(3 \leq X \leq 5)$ . We can simply add them up:  $p(X = 3) + p(X = 4) + p(X = 5) = 0.568$  (rounded). We can also recognize that this would be the same as  $p(X \leq 5) - p(X \leq 2) = \text{binomcdf}(10,0.5,5) - \text{binomcdf}(10,0.5,2) = 0.623047 - 0.054688 = 0.568$  (rounded, same as before). In the second method,

we first add up all bars from 0 to 5, then add up all bars from 0 to 2, leaving just the bars from 3 to 5. This second method works best for higher quantities, where the direct application of the formula would be tedious.

### *Continuous Case*

Some scenarios allow for  $X$  to roam between all values within an interval. For example, the heights of men might range from 60 inches to 84 inches (5 feet to 7 feet). It is reasonable to assume that all heights in between, including all non-integer values, are possible to achieve. Hence, the random variable  $X$  would range continuously from 60 inches to 84 inches. There are also cases where the discrete case contains so many possible values that we can model these multitudes of individual cases with one smooth curve, thereby transforming it into a continuous case. For example, daily rainfall is measured in hundredths of an inch, and the usual range might be anywhere from zero inches up to 5 inches. This would be over 500 individual possibilities! (0.00, 0.01, 0.02, *etc.*, on up through 4.98, 4.99 and 5.00). We would probably want to model this situation as a continuous model, even though technically it would be discrete.

Suppose we use some function  $f(x)$  to be a probability density function (pdf, again). The basic requirement is that  $f(x) \geq 0$  and that the area under the curve equals one, or in calculus terms:

$$\int_R f(x) dx = 1$$

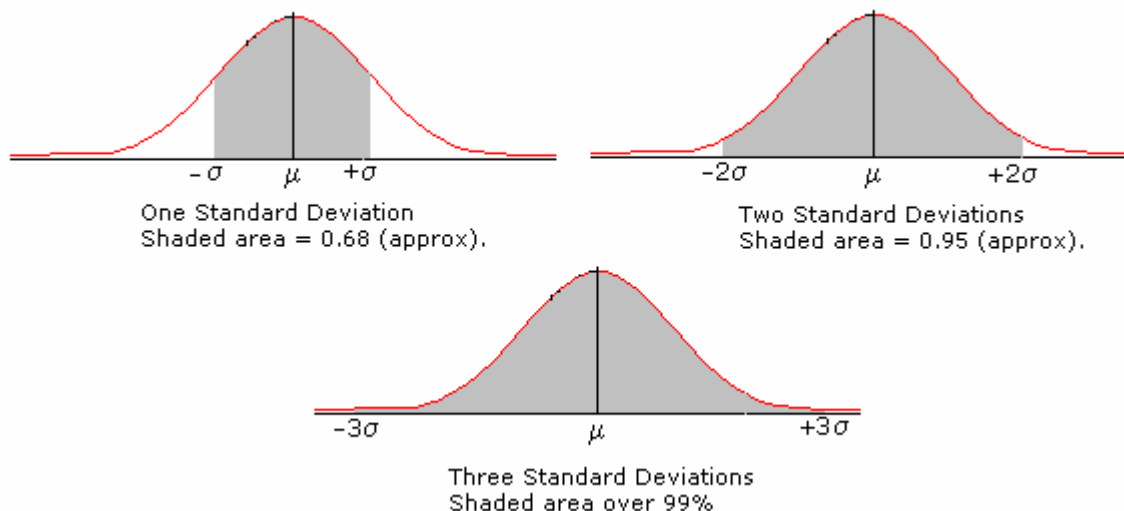
where  $R$  is some interval corresponding to a scenario.

In statistics, the normal distribution curve (also known as the bell curve) is a useful model for situations where the data tends to collect near the mean, and thin out further away from the mean. In a large class, the distribution of test scores often results in a normal distribution, for example. Most data is near the mean, a few are way above or way below it. The typical shape of a normal curve is given below.



*Normal "Bell" Curve*

The normal curve is designed to have the mean be at  $\mu = 0$ , and in such a way that about 68% of the data is captured within 1 standard deviation ( $\sigma$ ) of the mean ( $\mu$ ), 95% of the data captured within 2 standard deviations, and over 99% within 3 standard deviations. We view these figures as 'areas under the curve', which implies we would use calculus to determine these values. However, the function used to create this graph is difficult to integrate and instead we use tables or the calculator to determine these areas.



**Example:** Suppose we are told that the mean IQ of a population is  $\mu = 100$  and the standard deviation is  $\sigma = 10$ . This means that roughly 68% of the population has an IQ between 90 and 110, 95% of the population has an IQ between 80 and 120, and over 99% has an IQ between 70 and 130.

In random variable notation, we'd say:

$$p(90 \leq X \leq 110) = 0.68, \quad p(80 \leq X \leq 120) = 0.95, \quad p(70 \leq X \leq 130) = 0.99 +$$

We can use symmetries and simple arithmetic to determine other probabilities. Consider these examples:

- a) What is the probability a randomly chosen person has an IQ between 100 and 110?

Answer: This is just half the region from 90 to 110, which we know covers an area (has a probability) of 0.68, so the answer is  $p(100 \leq X \leq 110) = 0.34$ , half the 0.68.

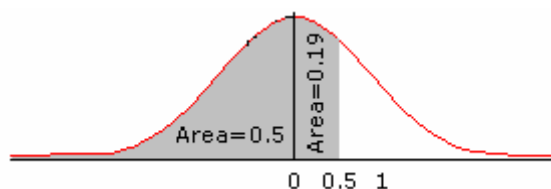
- b) What is the probability some one has an IQ above 110?

Answer: We know the right half of the distribution has a probability 0.5 (this means half the population is above the average – no surprise there). We know that between 100 and 110 the area covers 0.34. This leaves  $0.5 - 0.34 = 0.16$  as our answer.

Suppose we want to know the likelihood of a randomly chosen person having an IQ below 105. In random variables, this is  $p(X \leq 105)$ . To answer this question, we need to re-center the graph so that the center (mean) is at zero and so that each standard deviation equals 1 unit. All values then get altered accordingly – this process is called *normalization*:

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

where  $x$  is the original value in question and  $z$  is its equivalent on the normalized normal curve (called the  $z$ -score). So, in this case,  $z = (105 - 100)/10 = 0.5$ . We just need to determine the area below 0.5. See the picture for assistance:

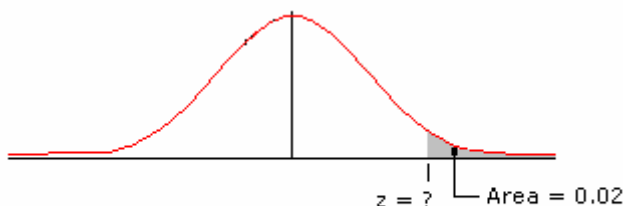


The left half of the region has an area of 0.5. We just need to find the area between 0 and 0.5. Your calculator has a nifty feature that will calculate this. Go to 2<sup>nd</sup>-DISTR (above the VARS key) and select “**normalcdf**” and hit enter. You have to tell it to run from 0 to 0.5, so your command will look like this:

**normalcdf(0,0.5)**

This gives 0.19, which added to 0.5 gives 0.69. Hence, the probability a randomly chosen person has an IQ below 105 is  $p(X \leq 105) = 0.69$ . A common trick that works well with the calculator for open-ended intervals is to use -5 and 5 as default left and right bounds when needed (in place of negative and positive infinity). In the above example, we could have typed **normalcdf(-5,0.5)** and have arrived at the same result. Please note: there is nothing magical with -5 and 5 other than they’re far enough away from the mean to act reasonably well as ‘open-ended’ endpoints. The accuracy is still well within what we want.

Consider a scenario where we want to work backwards. Consider a club that only accepts those who score in the top 2% of IQ. What would be the minimum IQ to get into this club? We treat this as an area problem: we want the  $z$ -score that gives an area of 0.02 to the right (see figure).



What  $z$ -score has an area of 0.02 to the right and 0.98 to the left? That's what we need to determine.

Your calculator has a nifty feature called **InvNorm** (2<sup>nd</sup>-DISTR, then choice #3). InvNorm works by returning the  $z$ -score for a given area from the left, so in this case we want the  $z$ -score that gives an area of 0.98 to the left, which will be the same as the  $z$ -score corresponding to the area of 0.02 to the right. **InvNorm(0.98)** gives us  $z = 2.05$ .

Now that we have  $z = 2.05$ , we can use the renormalization formula to determine  $x$ :  $2.05 = (x - 100)/10$ , which with a little algebra, gives  $x = 120.5$ .

**Practice problems (Discrete Cases)**

1. A salesman has a 40% likelihood of making a sale. One day 25 people visited his shop. Find the probability he...
  - a. Made exactly 12 sales.
  - b. Made at most 12 sales.
  - c. Made at least 13 sales.
  - d. Made between 7 and 12 sales.
  
2. Jeff target shoots. He has a probability of 0.85 of making his shot. He makes 30 attempts. Find the following probabilities:
  - a.  $p(X = 20)$
  - b.  $p(X \leq 22)$
  - c.  $p(X \geq 24)$
  - d.  $p(20 \leq X \leq 26)$

**Practice problems (Continuous Cases)**

1. Using your calculator, determine these probabilities of a normal distribution in the scenario where the mean  $\mu = 0$  and the standard deviation  $\sigma = 1$ . This distribution has already been normalized. Use symmetry and complements if needed.

$$p(-1 \leq X \leq 1) =$$

$$p(-1 \leq X \leq 0) =$$

$$p(0 \leq X \leq 1) =$$

Notice anything consistent about the above three?

$$p(-2 \leq X \leq 2) =$$

$$p(-3 \leq X \leq 3) =$$

$$p(1 \leq X \leq 3) =$$

$$p(0.5 \leq X \leq 1.9) =$$

$$p(X \leq -1.2) =$$

$$p(X \geq -1.5) =$$

$$p(X \geq 0.75) =$$

2. In this scenario the mean  $\mu = 12$  and the standard deviation  $\sigma = 2$ . Renormalize and find the following probabilities:

$$p(9 \leq X \leq 13) =$$

$$p(X \leq 14) =$$

$$p(7 \leq X \leq 10.5) =$$

$$p(X \geq 8.5) =$$

3. Suppose the average man's height is 70 inches (5-10) and the standard deviation is 2.5 inches.
- What is the probability a randomly chosen man stands between 5-10 and 6 feet?
  - What is the probability a randomly chosen man stands less than 5-7?
  - What is the probability a randomly chosen man stands greater than 5-9?
  - What is the probability a man stands above 6-3?
  - The tall man's club admits members in the top 1% of heights. What is the minimum height requirement to get into this club?