Stat 250 Gunderson Lecture Notes
4: Random Variables

All models are wrong; some models are useful. -- *George Box*

Patterns make life easier to understand and decisions easier to make. Earlier we discussed the different types of data or variables and how to turn the data into useful information with graphs and numerical summaries. Having some notion of probability from the previous chapter, we can now view the variables as "random variables" – the numerical outcomes of a random circumstance. We will look at the pattern of the distribution of the values of a random variable and we will see how to use the pattern to find probabilities. These patterns will serve as models in our inference methods to come.

What is a Random Variable?

Recall in our discussion on probability we started out with some random circumstance or experiment that gave rise to our set of all possible outcomes *S*. We developed some rules for calculating probabilities about various events. Often the events can be expressed in terms of a "random variable" taking on certain outcomes. Loosely, this random variable will represent the value of the variable or characteristic of interest, but *before we look*. Before we look, the value of the variable is not known and could be any of the possible values with various probabilities, hence the name of a "random" variable.

> ***Definition:***
> A ***random variable*** assigns a number to each outcome of a random circumstance, or, equivalently, a random variable assigns a number to each unit in a population.

We will consider ***two broad classes*** of random variables: **discrete** random variables and **continuous** random variables.

> ***Definitions:***
> A ***discrete random variable*** can take one of a countable list of distinct values.
> A ***continuous random variable*** can take any value in an interval or collection of intervals.

Try It! Discrete or Continuous

A car is selected at random from a used car dealership lot. For each of the following characteristics of the car, decide whether the characteristic is a continuous or a discrete random variable.

a. Weight of the car (in pounds).

b. Number of seats (maximum passenger capacity).

c. Overall condition of car (1 = good, 2 = very good, 3 = excellent).

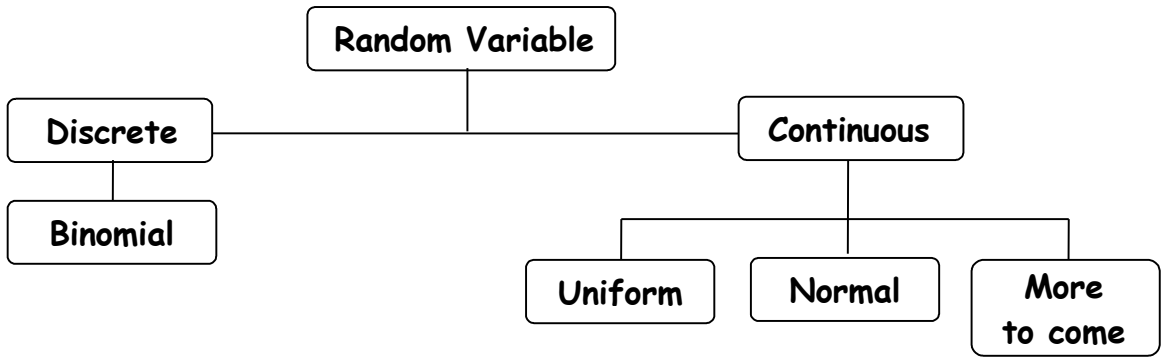
d. Length of car (in feet).

In statistics, we are interested in the **distribution of a random variable** and we will use the distribution to compute various probabilities. The probabilities we compute (for example, *p*-values in testing theories) will help us make reasonable decisions.

So just what is the distribution of a random variable? Loosely, it is a model that shows us what values are possible for that particular random variable and how often those values are expected to occur (i.e. their probabilities). The model can be expressed as a function, table, picture, depending on the type of variable it is.

We will first discuss discrete random variables and their models. We will work with the broad class of general discrete random variables and then focus on a **particular family of discrete random variables** called the **Binomial**. The Binomial random variable arises in situations where you are counting the number of *successes* that occur in a sample.

Next we look at properties for continuous random variables and spend more time studying the **family of uniform random variables and normal random variables**. Later in this class you will be introduced to more models for continuous random variables that are primarily used in statistical testing problems. Below is a summary of the types of random variables we will work with in this course.



Technical Note: Sometimes a random variable fits the technical definition of a discrete random variable but it is more convenient to treat it, that is, model it, as if it were continuous. We will learn when it is reasonable to model a discrete binomial random variable as being approximately normal. Finally we will also learn how to model sums and differences of random variables.

Some general notes about random variables are:

- random variables will be denoted by capital letters (*X, Y, Z*);
- outcomes of random variables are represented with small letters (*x, y, z*).

So when we express probabilities about the possible value of a random variable we use the capital letter. For example, the probability that a random variable takes on the value of 2 would be expressed as $P(X = 2)$.

General Discrete Random Variables

A **discrete** random variable, X, is a random variable with a finite or countable number of possible outcomes. The probability notation your text uses for a Discrete Random Variable is given next:

> ***Discrete Random Variable:***
> *X* = the random variable.
> *k* = a number that the discrete random variable could assume.
> *P(X = k)* is the probability that the random variable *X* equals *k*.

The **probability distribution function (pdf) for a discrete random variable *X*** is a table or rule that assigns probabilities to the possible values of the *X.*
One way to show the distribution is through a table that lists the possible values and their corresponding probabilities:

Value of *X*	x_1	x_2	x_3	…
Probability	p_1	p_2	p_3	…

- **Two conditions** that must apply to the probabilities for a discrete random variable are:
  Condition 1: The sum of all of the individual probabilities must equal 1.
  Condition 2: The individual probabilities must be between 0 and 1.
- A ***probability histogram*** or better yet, a ***probability stick graph***, can be used to display the distribution for a discrete random variable.
  The *x*-axis represents the values or outcomes.
  The *y*-axis represents the probabilities of the values or outcomes.
- The **cumulative distribution function (cdf) for a discrete random variable *X*** is a table or rule that provides the probabilities $P(X \le k)$ for any real number *k*. Generally, the term cumulative probability refers to the probability that *X* is ***less than or equal to*** a particular value.

Try It! Psychology Experiment

A psychology experiment on the behavior of young children involves placing a child in a designated area with five different toys. Over a fixed time period various observations are made. One response measured is the number of toys the child plays with.
Based on many results, the (partial) probability distribution below was determined for the discrete random variable *X* = number of toys played with by children (during a fixed time period).

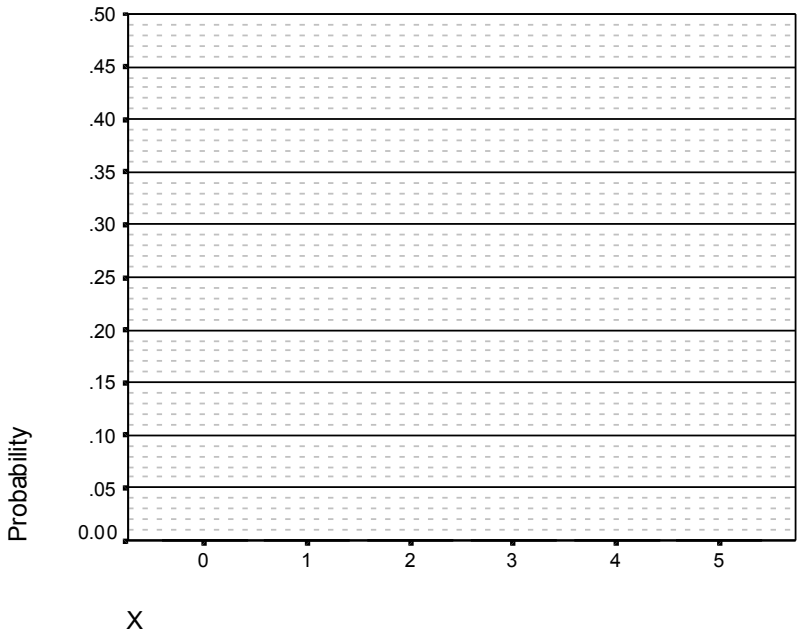
***X* = # toys**	0	1	2	3	4	5
Probability	0.03	0.16	0.30	0.23	0.17	

a. What is the missing probability *P(X = 5)*?

Psychology Experiment *continued*

***X* = # toys**	0	1	2	3	4	5
Probability	0.03	0.16	0.30	0.23	0.17	

b. Graph this discrete probability distribution function for *X*.



c. What is the probability that a child will play with ***at least*** 3 toys?

d. Given the child has played with at least 3 toys, what is the probability that he/she will play with all 5 toys?

e. Finish the table below to provide the cumulative distribution function of *X*.

***X* = # toys**	0	1	2	3	4	5
Cum Probability ***P(X ≤ k)***	0.03	0.03+0.16 = 0.19	0.03+0.16+0.30 = 0.49			

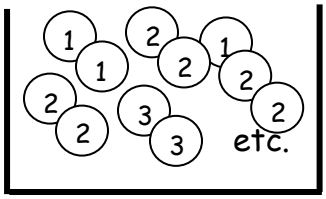
Expectations for Random Variables

Just as we moved from summarizing a set of data with a graph to numerical summaries, we next consider computing the mean and the standard deviation of a random variable. The mean can be viewed as the expected value over the long run (in many repetitions of the random circumstance) and the standard deviation can be viewed is approximately the average distance of the possible values of X from its mean.

> ***Definition:***
> The ***expected value*** of a random variable is the mean value of the variable *X* in the sample space, or population, of possible outcomes. *Expected value*, denoted by E(*X*), can also be interpreted as the mean value that would be obtained from an infinite number of observations on the random variable.

Motivation for the expected value formula …

Consider a population consisting of 100 families in a community. Suppose that 30 families have just 1 child, 50 families have 2 children, and 20 families have 3 children. What is the mean or average number of children per family for this population?



Population of 100 families

Mean = (sum of all values)/100
= [1(30) + 2(50) + 3(20)]/100
= **1**(30/100) + **2**(50/100) + **3**(20/100)
= 1(0.30) + 2(0.50) + 3(0.20)
= 1.9 children per family

Mean = Sum of (value x **probability of that value)**

> ***Definitions:***
> **Mean and standard deviation of a discrete random variable**
> Suppose that X is a discrete random variable with possible values x_1, x_2, x_3, … occurring with probabilities p_1, p_2, p_3, …, then
> the expected value (or mean) of *X* is given by $\mu = E(X) = \sum x_i p_i$
> the *variance* of *X* is given by $V(X) = \sigma = \sum (x_i - \mu)^2 p_i$
> and so the *standard deviation* of *X* is given by $\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$
> The sums are taken over all possible values of the random variable *X*.

Try It! Psychology Experiment

Recall the probability distribution for the discrete random variable X = number of toys played with by children.

***X* = # toys**	0	1	2	3	4	5
Probability	0.03	0.16	0.30	0.23	0.17	0.11

a. What is the expected number of toys played with?

***Note**: The expected value may not be a value that is ever expected on a single random outcome. Instead, it is the average over the long run.*

b. What is the standard deviation for the number of toys played with?

c. Complete the interpretation of this standard deviation (in terms of an average distance):

On average, the number of toys played with vary by about _____

from the mean number of toys played with of _____.

Binomial Random Variables

An important class of discrete random variables is called the **Binomial Random Variables**.

A binomial random variable is that it ***COUNTS*** the number of times a certain event occurs out of a particular number of observations or trials of a random experiment.

Examples of Binomial Random Variables:

- The number of girls in six independent births.
- The number of tall men (over 6 feet) in a random sample of 30 men from a large male population.

A **binomial experiment** is defined by the following conditions:

1. There are *n* "trials", *n* is determined in advance and is not a random value.
2. There are two possible outcomes on each trial,
   called "success" (S) and "failure" (F).
3. The outcomes are independent from one trial to the next.
4. The probability of a "success" remains the same from one trial to the next, and this probability is denoted by *p*. The probability of a "failure" is $1 - p$ for every trial.

A **binomial random variable** is defined as
***X* = number of successes in the *n* trials of a binomial experiment**.

Try It! Are the Conditions Right for Binomial?

a. Observe the sex of the next 50 children born at a local hospital.
   X = number of girls

b. A ten-question quiz has five True-False questions and five multiple-choice questions, each with four possible choices. A student randomly picks an answer for every question.
   X = number of answers that are correct.

c. Four students are randomly picked without replacement from large student body listing of 1000 women and 1000 men.
   X = number of women among the four selected students.

   What if the student body listing consisted only of 10 women and 10 men?

***Rule of Thumb*: population at least 10 times as large as the sample → ok!**

The Binomial Formula

We will develop the formula together using our probability knowledge. Suppose that of the online shoppers for a particular website that start filling a shopping cart with items, 25% actually make a purchase (complete a transaction). We have a random sample of 10 such online shoppers.

If the stated rate is true, what is the probability that ...

... all 10 shoppers will actually make a purchase?

... none of the shoppers will make a purchase?

... just 1 shopper will make a purchase?

With only the basic probability knowledge, you just calculated three binomial probabilities that are based on the following formula.

The *binomial distribution:*

Probability of exactly *k* successes in *n* trials …

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0,1,2,\ldots,n$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ *(this represents the # of ways to select k items from n)*

Try it! The first part …

You can think of the computation of $\binom{n}{k}$ in the following way …

Suppose you had n friends, how many ways could you invite k to dinner? The ones "at the ends" are easy to do without even using the formula or a calculator. Your calculator is likely to have this complete function or at least a factorial ! option. On many calculators this combinations function is found under the math → probability menu and expressed as nCr.

1. $\binom{10}{0} =$

2. $\binom{10}{10} =$

3. $\binom{10}{1} =$

4. $\binom{10}{9} =$

5. $\binom{10}{2} =$

Try it! Finding Binomial Probabilities

Recall we have a random sample of n = 10 online shoppers from a large population of such shoppers and that p = 0.25 is the population proportion who actually make a purchase.

a. What is the probability of selecting ***exactly one*** shopper who actually makes a purchase?

b. What is the probability of selecting ***exactly two*** shoppers who actually make a purchase?

c. What is the probability of selecting ***at least one*** shopper who actually makes a purchase?

d. How many shoppers in your random sample of size 10 would you ***expect*** to actually make a purchase?

In the previous question (part d), you just computed the **mean of a binomial distribution**.

If X has the **binomial distribution Bin(n, p)** then

Mean of X is $\mu = E(X) = np$

Standard Deviation of X, is $\sigma = \sqrt{np(1-p)}$

Try it! More Work with the Binomial

Suppose that about 10% of Americans are left-handed. Let X represent the number of left-handed Americans in a random sample of 12 Americans.

Then X has a _____distribution (*be as specific as you can).*

Note that the mean or expected number of left-handed Americans in such a random sample would be $\mu = np = 12(0.10) = 1.2$. The standard deviation (reflecting the variability in the results from the mean across many such random samples) is $\sigma = \sqrt{np(1-p)} = \sqrt{12(0.10)(0.90)} = 1.04$.

a. What is the probability that the sample contains 2 or fewer left-handed Americans?

b. Suppose a random sample of 120 Americans had been taken instead of just 12. So now X has a Binomial(n = 120, p = 0.10) model. The mean or expected number of left-handed Americans in a random sample of 120 will be $\mu = np = 120(0.10) = 12$. The standard deviation for the number of left-handed Americans will be $\sigma = \sqrt{np(1-p)} = \sqrt{120(0.10)(0.90)} = 3.29$.

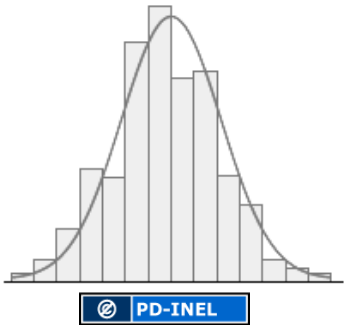
   So how might you try to find the probability that a random sample of 120 Americans would result in 20 or fewer left-handed Americans? Note that 2 out of n = 12 is 16.67% and that 20 out of n = 120 is also equal to 16.67%.

General Continuous Random Variables

A **continuous random variable**, X, takes on all possible values in an interval (or a collection of intervals). The way that we determine probabilities for continuous random variables differs in one important respect from how we determine probabilities for discrete random variables. For a discrete random variable, we can find the probability that the variable *X* exactly equals a specified value. We can't do this for a continuous random variable. Instead, we are only able to find the probability that *X* could take on values in an interval. We do this by determining the corresponding area under a ***curve*** called the probability density function of the random variable.

We have already summarized the general shapes of distributions of a quantitative response that often arise with real data. The shape of a distribution was found by drawing a smooth ***curve*** that traces out the overall pattern that is displayed in a histogram. With a histogram, the area of each rectangle is proportional to the frequency or count for each class. The curve also provides a visual image of proportion through its area. If we could get the equation of this smoothed curve, we would have a simple and somewhat accurate summary of the distribution of the response.

The picture at the right shows a smoothed curve that is symmetric and bell shaped, even though the underlying histogram is only approximately symmetric. If the data came from a representative sample, the smooth curve could serve as a model, that is, as the probability distribution for the continuous response for the population.



So the **probability distribution of a continuous random variable** is described by a **density curve**. The probability of an event is the area under the curve for the values of X that make up the event.

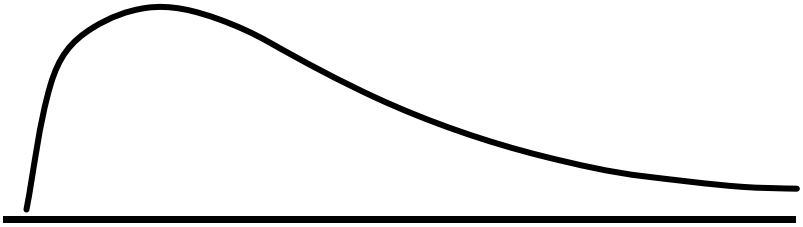
The probability model for a continuous random variable assigns probabilities to intervals.

> ***Definition:***
> A curve (or function) is called a ***Probability Density Curve*** if:
> 1. It lies on or above the horizontal axis.
> 2. Total area under the curve is equal to 1.

KEY IDEA: AREA under a density curve over a range of values corresponds to the PROBABILITY that the random variable *X* takes on a value in that range.

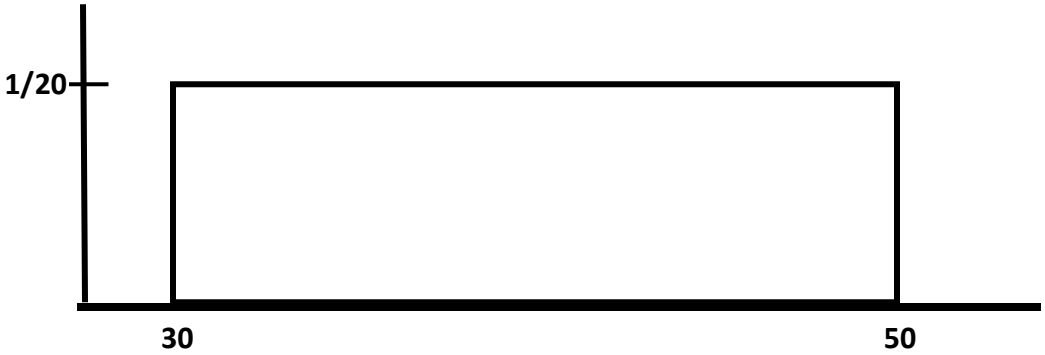
Try It! Some Probability Density Curves

I. A density curve for modeling income for employed adults (in $1000s) for a city.



How would you use the above density curve to estimate the probability of a randomly selected employed adult from this city having an income between $30,000 and $40,000?

II. Consider the following curve:



a. Is this a density curve? Why?

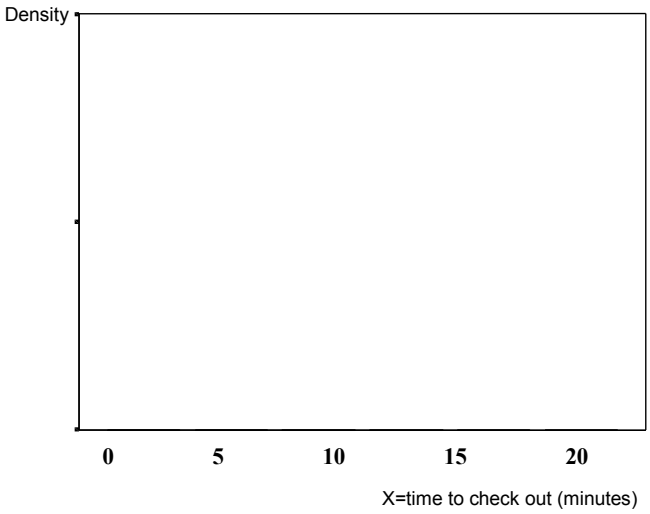
b. If yes, find the probability of observing a response that is less than 35.

c. What does the value of 35 correspond to for this distribution?

Try it! Checkout time at a store

Let *X* be the checkout time at a store, which is a random variable that is uniformly distributed on values between 5 and 20 minutes. That is, *X* is *U*(5, 20).

a. What does the density look like? Sketch it and include a value on the y-axis.



b. What is the probability a person will take more than 10 minutes to check out?

c. Given a person has already spent 10 minutes checking out, what is the probability they will take no more than 5 additional minutes to check out?

d. What is the expected time to check out at this store?

Definition: **Mean of a continuous random variable.**

Expected Value or Mean = Balancing point of the density curve **E(X) = μ**

(Sometimes one would need calculus/integration to find it -- integral instead of sums)

There are many density curves that can be used as models. Next we focus on an important family of densities called the **NORMAL DISTRIBUTIONS.**

Normal Random Variables

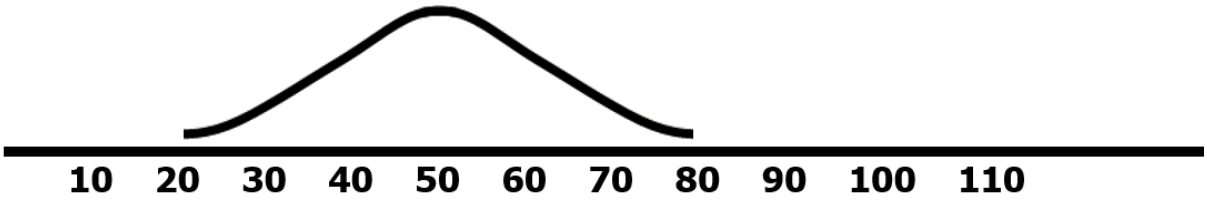
We had our first introduction to normal random variables back in our summarizing data section as a special case of bell-shaped distributions. The family of normal distributions is very important because many variables have this shape and form approximately and many statistics that we use in our inference methods are based on sums or averages which generally have (approximately) a normal distribution.

A Normal Curve: Symmetric, bell-shaped, centered at the mean μ and its spread is determined by the standard deviation σ. In fact, the points of inflection on each side of the mean mark the values which are one standard deviation away from the mean.

Notation: If a population of measurements follows a normal curve, and if *X* is the measurement for a randomly selected individual from the population, then *X* is said to be a **normal random variable**. *X* is also said to have a **normal distribution**. Any normal random variable can be completely characterized by its mean and its standard deviation.

The variable *X* is normally distributed with mean μ and standard deviation σ is denoted by:

A *N*(50,10) curve is sketched below. Add a *N*(80,5) curve to this picture.
Keep in mind the features of the empirical rule (68-95-99.7) which applies to a normal curve.



Standardized Scores:

A normal distribution is indexed by its population mean μ, and its population standard deviation σ, and denoted by $N(\mu,\sigma)$. Recall that the standard deviation is a useful "**yardstick**" for measuring how far an individual value falls from the mean. The **standardized score** or ***z*-score** is the distance between the observed value and the mean, measured in terms of number of standard deviations. Values that are above the mean have positive *z*-scores, and values that are below the mean have negative *z*-scores.

Standardized score or *z*-score: $z = \dfrac{\text{observed value - mean}}{\text{Standard deviation}} = \dfrac{x-\mu}{\sigma}$

Finding Probabilities for z-Scores:

Standard scores play a role in how we will find areas (and thus probabilities) under a normal curve. We simply convert the endpoints of the interval of interest to the corresponding standardized scores and then use a table (computer/calculator) to find probabilities associated with these standardized scores. When we convert to standardized scores, the variable *X* is converted to the **Standard Normal Random Variable**, *Z*, which has the $N(0,1)$ distribution.

Try It! Finding Probabilities for *Z*

1. Find $P(Z \le 1.22)$.

   Think about it: What is $P(Z < 1.22)$?

2. Find $P(Z > 1.22)$.

3. Find $P(-1.58 < Z < 2.24)$

4. What is the probability that a standard normal variable Z is within 2 standard deviations of the mean? That is, find $P(-2 \le Z \le 2)$.

In the Extreme (for $z > 0$)

z	3.09	3.72	4.26	4.75	5.20	5.61	6.00
Probability	.999	.9999	.99999	.999999	.9999999	.99999999	.999999999

5. What is $P(Z \le 4.75)$? $P(Z > 10.20)$?

7. What is the 90th percentile of the standard normal $N(0,1)$ distribution?

Table A.1 provides the areas to the left for various values of Z

From Utts, Jessica M. and Robert F. Heckard. Mind on Statistics, Fourth Edition. 2012. Used with permission

Table entry for z is the area to the left of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

How to Solve General Normal Curve Problems

One way to find areas and thus probabilities for any general $N(\mu,\sigma)$ distribution involves the idea of standardization.

If the variable X has the $N(\mu,\sigma)$ distribution,

then the standardized variable $Z = \dfrac{X-\mu}{\sigma}$ will have the $N(0,1)$ distribution.

We will see how this standardization idea works through our next example.

Try It! Scholastic Scores

A proposal before the Board of Education has specified certain designations for elementary schools depending on how their students do on the Evaluation of Scholastic Scoring. This test is given to all students in all public schools in grades 2 through 4. Schools that score in the top 20% are labeled excellent. Schools in the bottom 25% are labeled "in danger" and schools in the bottom 5% are designated as failing. Previous data suggests that scores on this test are approximately normal with a mean of 75 and a standard deviation of 5.

a. What is the probability that a randomly selected school will score below 70?

b. What is the score cut-off required for schools to be labeled excellent? Show all work.

Notes:

1. The normal distribution is a density for continuous random variables.
2. The normal distribution is not the only continuous distribution (recall you worked with a family of uniform continuous distributions a few pages back).
3. Computers and calculators often have the ability to find areas or percentiles for many density curves built right in to a function. You might be introduced to some of these in your lab sessions and are welcome to use them. But drawing a picture of what you are trying to find will be beneficial and serve as one way to show your work.

Approximating Binomial Distribution Probabilities

Recall Our Left-Handed Problem

In an earlier problem it was stated that about 10% of Americans are left-handed. Let *X* = the number of left-handed Americans in a random sample of 120 Americans (part c had us think about a sample size being 120 instead of 12).

Then *X* has an exact Binomial distribution with *n* = 120 and *p* = 0.10. The mean or expected number of left-handed Americans in the sample is $\mu = np = 120(0.10) = 12$. The standard deviation of *X* is $\sigma = \sqrt{np(1-p)} = \sqrt{120(0.10)(0.90)} = 3.29$.

Suppose we want to find the probability that a random sample of 120 will contain 20 or fewer left-handed Americans. Using the exact binomial distribution we would start with:

$$P(X \le 20) = P(X = 0) + P(X = 1) + \cdots + P(X = 19) + P(X = 20)$$

This would not be too much fun to compute by hand since each of the probabilities for *X* = 0 up to *X* = 20 would be found using the binomial probability formula: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

But there is ***an easier way*** that will give us approximately the probability of having 20 or fewer. The easier way involves using a normal distribution. The normal distribution can be used to approximate probabilities for other types of random variables, one being binomial random variables when the sample size *n* is large.

Normal Approximation to the Binomial Distribution

If ***X* is a binomial** random variable based on *n* trials with success probability *p*,
and ***n* is large**, then the random variable ***X* is also approximately** …

Conditions:
The approximation works well when both *np* and $n(1-p)$ are at least 10.

Try It! Returning to our Left-Handed Problem

About 10% of Americans are left-handed. Let *X* = the number of left-handed Americans in a random sample of 120 Americans. Then *X* has an exact Binomial distribution with *n* = 120 and *p* = 0.10. The **mean** number of left-handed Americans in the sample $\mu = np = 120(0.10) = 12$. And the **standard deviation** of *X* = $\sigma = \sqrt{np(1-p)} = \sqrt{120(0.10)(0.90)} = 3.29$.

a. We want to find the **probability that a random sample of 120 will contain 20 or fewer left-handed Americans**. Since *np* = 120(0.10) = 12 and *n*(1 – *p*) = 120(0.90) = 108 are both at least 10, we can use the normal approximation for the distribution of *X*.

***X* has approximately a Normal distribution: *N*(_____, _____)**

$P(X \le 20) =$

b. **How likely is it that more than 20% of the sample will be left-handed Americans**?

Sums, Differences, and Combinations of Random Variables

There are many instances where we want information about combinations of random variables. One type of combination of variables is a linear combination. Two primary linear combinations that arise are sums and differences.

Sum = *X* + *Y* **Difference = *X* – *Y***

The next two summary boxes present the rules for finding the mean and the variance (and thus standard deviation) of a sum and of a difference. We will see the results for a difference when we study learning about the difference between two proportions and about the difference between two means.

Rules for Means:

Mean(*X* + *Y*) = Mean(*X*) + Mean(*Y*)
Mean(*X* – *Y*) = Mean(*X*) – Mean(*Y*)

Rules for Variances (if *X* and *Y* are independent)**:**

Variance(*X* + *Y*) = Variance(*X*) + Variance(*Y*)
Variance(*X* – *Y*) = Variance(*X*) + Variance(*Y*)

Think about it: *why is the variance of a <u>difference</u> found by taking the <u>sum</u> of the variances?*

Additional Notes

A place to … jot down questions you may have and ask during office hours, take a few extra notes, write out an extra problem or summary completed in lecture, create your own summary about these concepts.