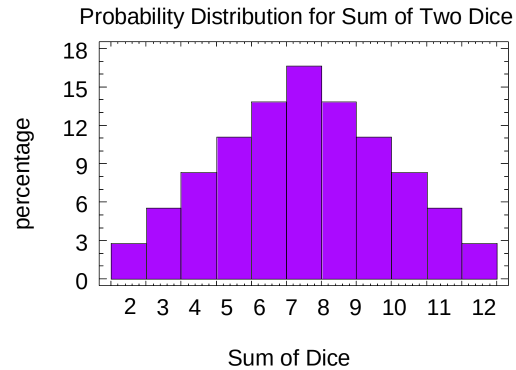
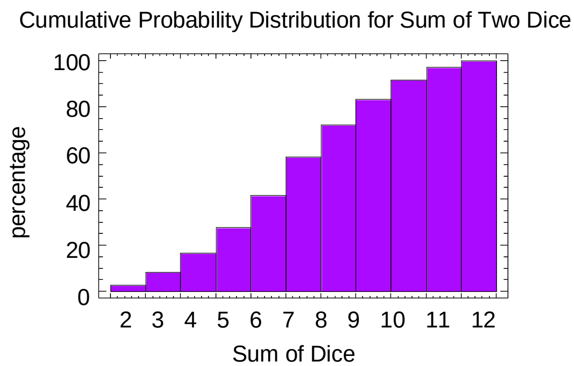
Sum of 2 Dice	Probability
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36
Total Probability is:	1





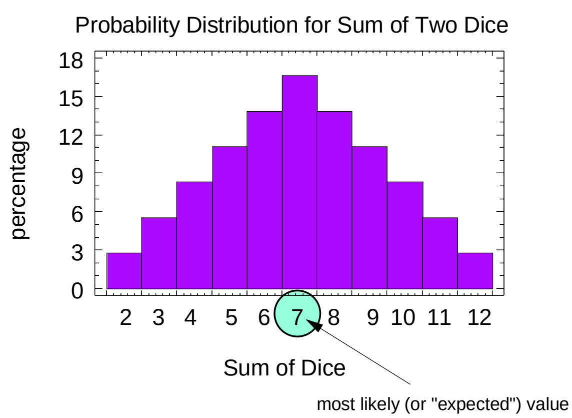
Properties of a probability function:

$$0 \leq P(\tilde{x} = x) \leq 1$$

The probability of any value of a random variable is in the range (0,1) inclusive

$$\sum_{\text{all x}} P(\tilde{x}) = 1.0$$

The probabilities of all values of a random variable in the sample space sum to 1.



Features of Probability Functions:

Measure of Central Location:

$$E(\tilde{x}) = \sum_{\text{all x}} x \bullet P(x)$$

The **expected value** of a random variable is the weighted sum of all its possible values, with the weights being the probability of each value.

Sum of 2 Dice	Probability	Products
2	1/36	2 x 0.0278
3	2/36	3 x 0.056
4	3/36	4 x 0.083
5	4/36	5 x 0.111
6	5/36	6 x 0.139
7	6/36	7 x 0.167
8	5/36	8 x 0.139
9	4/36	9 x 0.111
10	3/36	10 x 0.083
11	2/36	11 x 0.056
12	1/36	12 x 0.0278
Total Probability is:	1	

$$\text{Expected Value} = \sum_{\text{all x}} P(\tilde{x}_i) \bullet x_i = 7.0$$

Measure of the Spread of the Distribution:

$$V(\tilde{x}) = \sigma^2 = E\left[(\tilde{x}-\mu)^2\right] = \sum_{\text{all x}} (x-\mu)^2 \bullet P(x)$$

The **variance** of a random variable is the expected value of the weighted sum of the squared deviations from the mean of the probability distribution, with the weights being the probability of each value.

Tschebysheff's Theorem: Given any probability distribution with mean μ and standard deviation σ the probability of obtaining a value within *k* standard deviations of the mean is at least 1-1/k^2.

$$\text{i.e.,}\quad \Pr\left\{|\tilde{x}-\mu| \le k\sigma\right\} \ge 1 - \frac{1}{k^2}.$$

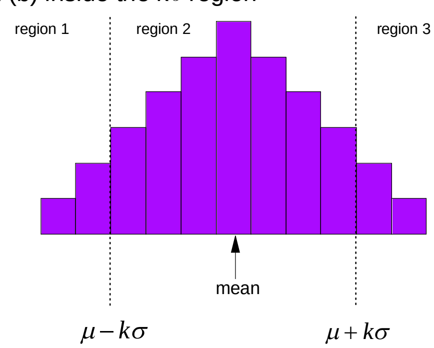
Note that Tschebysheff's theorem also implies that the probability that we get a value more than *k* standard deviations away from the mean is 1/k^2.

Proof:

1. We know that from our definition of the variance of a p.d.

$$V(\tilde{x}) = \sigma^2 = E\left[(\tilde{x}-\mu)^2\right] = \sum_{\text{all x}} (x-\mu)^2 \bullet P(x)$$

2. We can split our probability distribution into 3 parts corresponding to the areas (a) outside the k σ region and (b) inside the kσ region



3. Then we can decompose the variance as follows:

$$\sigma^2 = \sum_{\text{region 1}} (x-\mu)^2 \bullet P(x) + \sum_{\text{region 2}} (x-\mu)^2 \bullet P(x) + \sum_{\text{region 3}} (x-\mu)^2 \bullet P(x)$$

4. Since all three of these terms are greater than zero, we can drop the second term, leaving us with:

$$\sigma^2 \ge \sum_{\text{region 1}} (x-\mu)^2 \bullet P(x) + \sum_{\text{region 3}} (x-\mu)^2 \bullet P(x)$$

5. Now, since the absolute value of the deviation of *x* from the mean,μ is at least $k\sigma$ for all terms, we can write:

$$\sigma^2 \ge \sum_{\text{region 1}} k^2\sigma^2 \bullet P(x) + \sum_{\text{region 3}} k^2\sigma^2 \bullet P(x)$$

or, dividing through by $k^2\sigma^2$

$$\frac{1}{k^2} \ge \sum_{\text{region 1}} P(x) + \sum_{\text{region 3}} P(x)$$

probability that $x > \mu + k\sigma$

probability that $x < \mu - k\sigma$

Therefore, the probability that a random variable is greater than *k* standard deviations away from the mean of its probability distribution is less than or equal to *1/k^2*.

More Rules of Expectations

$E[k] = k$ where k is a constant

$V[k] = 0$ the variance of a constant is zero (duh!)

$E[k\tilde{x}] = k \bullet E[\tilde{x}]$

$V[k\tilde{x}] = k^2 \bullet V[\tilde{x}]$

$E[\tilde{x} \pm \tilde{y}] = E[\tilde{x}] \pm E[\tilde{y}]$

$E[\tilde{x} \times \tilde{y}] = E[\tilde{x}] \times E[\tilde{y}]$ if x & y are independent.

$V[\tilde{x} \pm \tilde{y}] = V[\tilde{x}] + V[\tilde{y}]$ if x & y are independent.

always positive

The Expectation of a linear transformation of a random variable.
The linear transformation is:

$$\tilde{y} = a + b \bullet \tilde{x}$$

This implies that the expectation of the linear transformation of *x* is:

$$\mathrm{E}[\tilde{y}] = E[a] + b \bullet [E(\tilde{x})] = a + bE[\tilde{x}]$$

A particularly useful linear transformation:
Take any random variable *x* with mean μ and standard deviation σ.
Form the new random variable:

$$\tilde{z} = \frac{(\tilde{x} - \mu)}{\sigma}$$

or, rearranging, we have,

$$\tilde{z} = \frac{1}{\sigma}\tilde{x} - \frac{\mu}{\sigma}.$$

Now, let's find the expectation and variance of our new random variable *z:*

$$E[\tilde{z}] = \frac{1}{\sigma}E[\tilde{x}] - \frac{\mu}{\sigma}.$$

$$= \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0.$$

The expectation of our new random variable *z* is zero!

The **variance** of our transformed random variable is:

$$V[\tilde{z}] = V\left[\frac{1}{\sigma}\tilde{x} - \frac{\mu}{\sigma}\right]$$

$$= V\left[\frac{1}{\sigma}\tilde{x}\right] + V\left[\frac{\mu}{\sigma}\right] \quad \text{(Expectations Rule)}$$

$$= \frac{1}{\sigma^2}V[\tilde{x}] + 0 \quad \text{(Expectations Rule)}$$

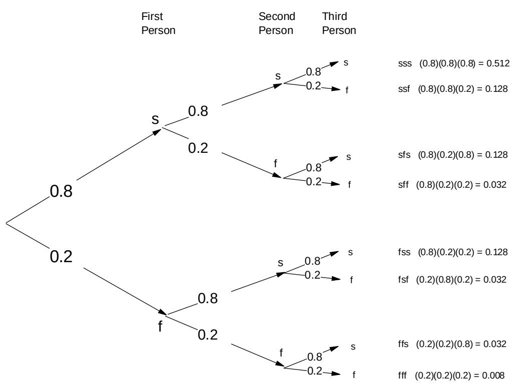
$$= \frac{1}{\sigma^2} \times \sigma^2 = 1$$

So, our new random variable has a variance of 1.

Notice that this **standardized random variable, *z*,** has a mean = 0 and a variance/standard deviation = 1 *irrespective of the type of distribution from which the random variable comes.*

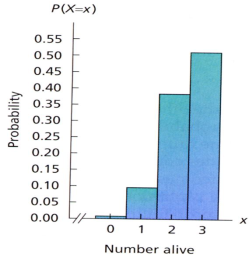
sss	ssf	sfs	sff
fss	fsf	ffs	fff

Outcome	Probability
sss	(0.8)(0.8)(0.8) = 0.512
ssf	(0.8)(0.8)(0.2) = 0.128
sfs	(0.8) (0.2) (0.8) = 0.128
sff	(0.8)(0.2)(0.2) = 0.032
fss	(0.2)(0.8)(0.8) = 0.128
fsf	(0.2)(0.8)(0.2) = 0.032
ffs	(0.2)(0.2)(0.8) = 0.032
fff	(0.2)(0.2)(0.2) = 0.008



Number alive (x)	Probability P(X = x)
0	0.008
1	0.096
2	0.384
3	0.512

Probability histogram for the random variable X, the number of people out of three that are alive at age 65



sss	ssf	sfs	sff
fss	fsf	ffs	fff

Two alive

One dead

$$(0.8)^2 \bullet (0.2)^1 = 0.64 \times 0.2 = 0.128$$

Probability one alive

Probability one dead

Number of trials

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = 3$$

Number alive

$$P(\tilde{x} = 2) = \binom{3}{2} \bullet (0.8)^2 \bullet (0.2)^1 = 3 \bullet 0.128 = 0.384$$

Let $\tilde{x}$ denote the total number of successes in n Bernoulli trials with success probability p. Then the probability distribution of the random variable $\tilde{x}$ is given by

$$P(\tilde{x} = \mathbf{x}) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The random variable $\tilde{x}$ is called a **binomial random variable** and is said to have the **binomial distribution** with parameters *n* and *p.*

To Find a Binomial Probability Formula for a Particular Problem:

Assumptions
1. n identical trials are to be performed.
2. Two outcomes, success or failure, are possible for each trial.
3. The trials are independent.
4. The success probability, p, remains the same from trial to trial.
   - Step 1 Identify a success.
   - Step 2 Determine p, the success probability.
   - Step 3 Determine n, the number of trials.
   - Step 4 The binomial probability formula for the number of successes, $\tilde{x}$, is

$$P(\tilde{x} = \mathbf{x}) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Given that the probability of a 20-year old's surviving until age 65 is 0.8:
a) What is the probability that exactly 2 of 3 randomly selected people will be alive?

$$P(\tilde{x} = 2) = \binom{3}{2}(0.8)^2(0.2)^{3-2} = \frac{3!}{2!(3-2)!}(0.8)^2(0.2)^1 = 0.384$$

Given that the probability of a 20-year old's surviving until age 65 is 0.8:
b) What is the probability that at most 1 of 3 randomly selected people will be alive?

$$P(\tilde{x} \le 1) = P(\tilde{x} = 0) + P(\tilde{x} = 1)$$

$$= \binom{3}{0}(0.8)^0(0.2)^{3-0} + \binom{3}{1}(0.8)^1(0.2)^{3-1}$$

$$= 0.008 + 0.096 = \mathbf{0.104}$$

Given that the probability of a 20-year old's surviving until age 65 is 0.8:
c) What is the probability that at least 1 of 3 randomly selected people will be alive?

$$P(\tilde{x} \ge 1) = P(\tilde{x} = 1) + P(\tilde{x} = 2) + P(\tilde{x} = 3)$$

But note that, because of the complementation rule:

$$P(\tilde{x} \ge 1) = 1 - P(\tilde{x} < 1) = 1 - P(\tilde{x} = 0)$$

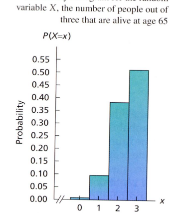
Therefore, we can solve the easier, related problem:

$$= 1 - \binom{3}{0}(0.8)^0(0.2)^{3-0} = 1 - 0.008 = 0.992$$

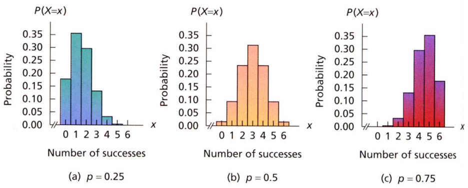
$$P(\tilde{x} = 3) = \binom{3}{3}(0.8)^3(0.2)^{3-3} = (1)(0.8)^3(0.2)^0 = 0.512$$

Number alive (x)	Probability P(X = x)
0	0.008
1	0.096
2	0.384
3	0.512

Probability histogram for the random variable X, the number of people out of three that are alive at age 65



Probability histograms for binomial distributions with parameters $n = 6$ and a) $p = 0.25$, (b) $p = 0.5$, and (c) $p = 0.75$



Measure of Central Location:

$$E(\tilde{x}) = \sum_{\text{all x}} x \bullet P(x)$$

The **expected value** of a random variable is the weighted sum of all its possible values, with the weights being the probability of each value.

Measure of the Spread of the Distribution:

$$V(\tilde{x}) = \sigma^2 = E\left[(\tilde{x} - \mu)^2\right] = \sum_{\text{all x}} (x - \mu)^2 \bullet P(x)$$

The **variance** of a random variable is the expected value of the weighted sum of the squared deviations from the mean of the probability distribution, with the weights being the probability of each value.

The expected value (mean) of a binomial distribution

$$E(\tilde{x}) = \sum_{\text{all x}} x \bullet P(x)$$

The probability that a binomial rv equals a given value is:

$$P(\tilde{x} = \mathbf{x}) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Substitute this probability into the expected value formula:

$$E(\tilde{x}) = \sum_{\text{all x}} x \bullet \binom{n}{x} p^x (1-p)^{n-x}$$

Let's simplify this for a small number of trials, say *n=3*

$$E(\tilde{x}) = \sum_{\text{all x}} x \bullet \binom{3}{x} p^x (1-p)^{3-x}$$

This gives us:

$$E(\tilde{x}) = 0 \bullet \binom{3}{0} p^0 (1-p)^{3-0} + 1 \bullet \binom{3}{1} p^1 (1-p)^{3-1} + 2 \bullet \binom{3}{2} p^2 (1-p)^{3-2} + 3 \bullet \binom{3}{3} p^3 (1-p)^{3-3}$$

Simplifying, we have:

$$E(\tilde{x}) = 0 + 1 \bullet 3p(1-p)^2 + 2 \bullet 3p^2(1-p)^1 + 3 \bullet 1p^3(1-p)^0$$

$$= 0 + 3(p - 2p^2 + p^3) + 6(p^2 - p^3) + 3p^3$$

$$= 3p, \text{ or, more generally,}$$

$$E(\tilde{x}) = np$$

The mean of a binomial random variable equals the probability of a success times the number of trials.

The variance and standard deviation of a binomial distribution:

$\sigma^2 = np(1-p)$ variance of a binomial distribution

$\sigma = \sqrt{np(1-p)} = (np(1-p))^{0.5}$ standard deviation of a binomial distribution

N = number of elements in a population

r = number of elements possessing a specific characteristic
(Call that characteristic a "success")

n = number of elements in the sample

Then, the probability of finding *x* successes in a sample of size *n* when sampling *without replacement* is:

$$p(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N} = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \text{ Hypergeometric Probability Distribution}$$

Mean: $\mu = n\left(\frac{r}{N}\right)$ Variance: $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

Problem using the hypergeometric distribution: Suppose that Airport & Intown Taxi Company (Chapel Hill's finest) has ten cabs, three of which have defective radios. The company's dispatcher, unaware that any of the radios is defective, chooses four of the company's cabs at random and assigns the company's four drivers to them.

(a) What is the probability that none of these four cabs has a defective radio?

N = 10, r = 3, n = 4, x = 0

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{3}{0}\binom{7}{4}}{\binom{10}{4}} = \frac{\frac{3!}{0!3!}\frac{7!}{3!4!}}{\frac{10!}{6!4!}} = \frac{1}{6} = 0.16667$$

(b) What is the probability that one of these four cabs has a defective radio?

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{3}{1}\binom{7}{3}}{\binom{10}{4}} = \frac{\frac{3!}{1!2!}\frac{7!}{3!4!}}{\frac{10!}{6!4!}} = \frac{1}{2} = 0.5$$

(c) What is the probability that 2 of these four cabs have a defective radio?

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}} = \frac{\frac{3!}{1!2!}\frac{7!}{2!5!}}{\frac{10!}{6!4!}} = \frac{3}{10} = 0.3$$

(d) What is the expected value of the number of cabs chosen by the dispatcher that have defective radios?

$$\text{Mean: } \mu = n\left(\frac{r}{N}\right) = 4\left(\frac{3}{10}\right) = \frac{6}{5} = 1.2 \text{ cabs}$$

(e) What is the standard deviation of the number of cabs chosen by the dispatcher that have defective radios?

$$\sigma = \sqrt{\sigma^2} = \sqrt{n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)} = \sqrt{4\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)\left(\frac{6}{9}\right)} = \sqrt{0.56} = 0.75$$

Sampling and the Binomial Distribution

Suppose that a simple random sample of size *n* is taken from a finite population in which the proportion of members that have a specified attribute is *p.* Then the number of members sampled that have the specified attribute has exactly a binomial distribution with parameters *n* and *p* if the sampling is done with replacement and *approximately* a binomial distribution with parameters *n* and *p* if the sampling is done without replacement and the sample size does not exceed 5% of the population size.

Assumptions necessary to apply the Poisson Distribution

Assumption	Bank Example
1.We know the average number of events (lambda) occuring in a given time segment.	1. Say the time segment of interest is one hour..
2. Probability of an occurrence remains constant throughout the time segment	2. The hour is one in which there is a steady flow of customers
3. Divide the time segment into small subsegments (length t) where the probability of two or more occurances in a subsegment is small enought to be ignored.	3.Impossible for two people to enter the bank simultaneously (i.e., in the same second (1/3,600th of an hour) (t=1/3,600 = 0.000278 hour)
4. Independence of occurrences between any two non-overlapping segments	4. Arrivals at the bank are not influenced by the length of the lines.

The Poisson Distribution

$$P(x) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!}, & \text{for } x = 0,1,2,,.\infty, \quad \lambda > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where

$\lambda =$ the mean number of events in a given segment of time $(t = 1)$

$t =$ the length of a particular subsegment $(t \le 1)$

$E[x] = \mu_x = \lambda t =$ the expected number of events in one subsegment length t

Suppose a bank knows that it averages 60 customers per hour over lunch hour:

$$\lambda = 60$$

If the customers are evenly distributed during the hour, the expected number of customers every 5 minutes is:

$$\lambda \bullet t = 60 \bullet 0.0833 = 5$$

What is the probability that a bank will have 3 or more customers in a 5 -minute period?

$$P(x \ge 3) = 1 - [P(x=0) + P(x=1) + P(x=2)]$$

$$= 1 - \left[\frac{e^{-\lambda t}((\lambda t)^0)}{0!} + \frac{e^{-\lambda t}((\lambda t)^1)}{1!} + \frac{e^{-\lambda t}((\lambda t)^2)}{2!}\right]$$

$$= 1 - \left[\frac{e^{-5}(5^0)}{0!} + \frac{e^{-5}(5^1)}{1!} + \frac{e^{-5}(5^2)}{2!}\right]$$

$$P(x \ge 3) = 1 - \left[\frac{0.00674 \bullet (1)}{1} + \frac{0.00674 \bullet (5)}{1} + \frac{0.00674 \bullet (25)}{2}\right] = 0.8753$$

More easily, you could look in Appendix C of the text to find the probability given:

$$x \le 2 \text{ and } \lambda \bullet t = 60 \bullet 0.0833 = 5$$

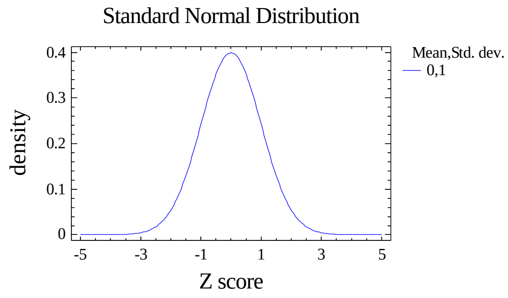
Using the Poisson Distribution to find the probability of an event:

- Define the length of the segment unit (t=1)
- Determine the length of the sub-segment of interest (t=?)
- Determine the mean of the random variable x $E[x] = \mu_x = \lambda t$
- Define the event of interest, x, and use the Poisson formula or the Poisson tables to find the probability

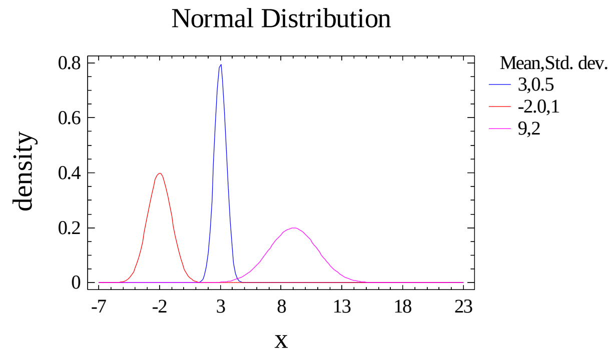
Finally, the mean and variance of the Poisson are the same:

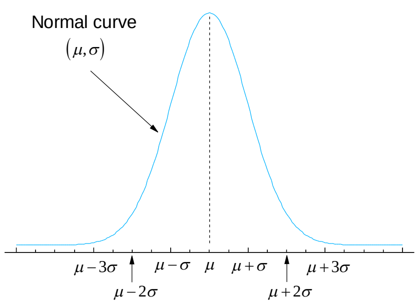
$$E[x] = \mu_x = \lambda \bullet t$$

$$\sigma_x^2 = \lambda \bullet t$$



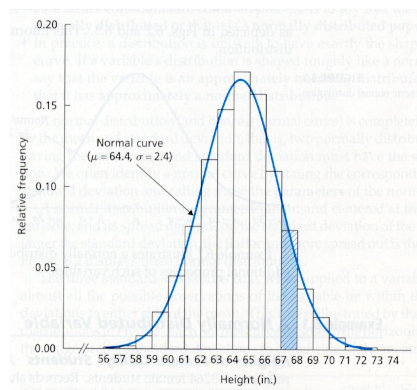
$$f(\tilde{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$





Frequency and relative-frequency distributions for heights

Height (inches)	Freq. f	Rel. freq.
56 < 57	3	0.0009
57 < 58	6	0.0018
58 < 59	26	0.0080
59 < 60	74	0.0227
60 < 61	147	0.0450
61 < 62	247	0.0757
62 < 63	382	0.1170
63 < 64	483	0.1480
64 < 65	559	0.1713
65 < 66	514	0.1575
66 < 67	359	0.1100
67 < 68	240	0.0735
68 < 69	122	0.0374
69 < 70	65	0.0199
70 < 71	24	0.0074
71 < 72	7	0.0021
72 < 73	5	0.0015
73 < 74	1	0.0003
	3264	1.0000



Major Fact about normally distributed variables and normal-curve areas:

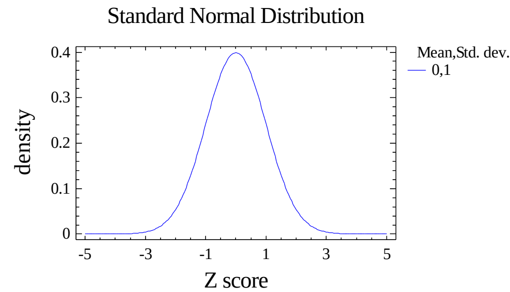
For a normally distributed variable, the percentage of all possible observations that lie within any specified range equals the corresponding area under its associated normal curve, expressed as a percentage. This result holds approximately for a variable that is approximately normally distributed.

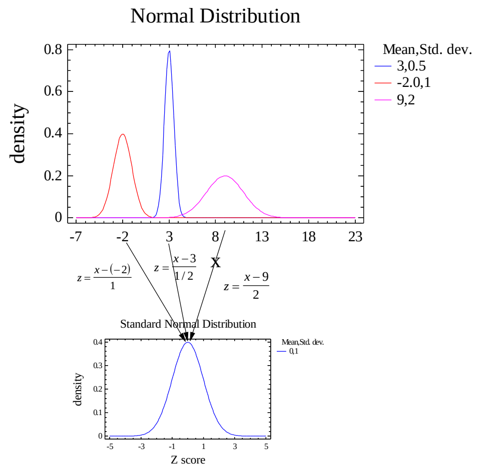
To Summarize:

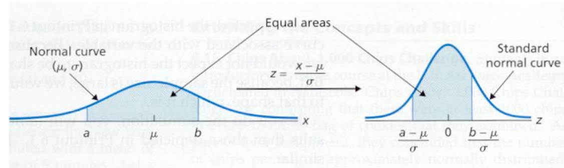
- Once we know the **mean** and the **standard deviation** of a normally distributed variable, we know its distribution and associated normal curve.
- Percentages for a normally distributed variable are equal to areas under its associated normal curve.
- The mean and standard deviation of a normal distribution are its **sufficient statistics;** they completely define the variable's distribution.

Given a random variable, $\tilde{x}$, with a normal distribution, the "standardized random variable, $\tilde{z}$, will have a normal distribution with mean zero and standard deviation = 1.

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



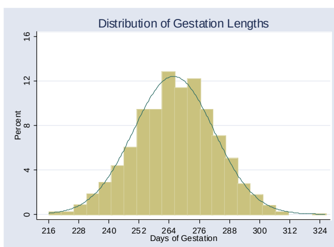




Basic Properties of the Standard Normal Curve

- Property 1: The total area under the standard normal curve is 1.
- Property 2: The standard normal curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
- Property 3: The standard normal curve is symmetric about 0; that is, the part of the curve to the left of the dashed line in Fig. 6.8 is the mirror image of the part of the curve to the right of it.
- Property 4: Almost all the area under the standard normal curve lies between -3 and 3.

histogram gestatio, bin(22) percent title("Distribution of Gestation Lengths") ylabel(0 4 to 16) xlabel(216 228 to 324) norm



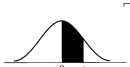

```
summarize gestatio, detail
                   Days of Gestation
-------------------------------------------------------------
      Percentiles      Smallest
 1%          228            216
 5%          238            217
10%          245            219       Obs                1000
25%          255            221       Sum of Wgt.        1000

50%          266                      Mean              265.8
                        Largest       Std. Dev.      16.03487
75%          277            306
90%          286            306       Variance       257.1171
95%        291.5            306       Skewness      -.0854922
99%        301.5            326       Kurtosis       2.989899
```

Normal Curve Areas

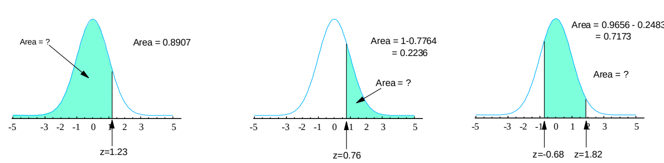
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: This table is abridged from Table 1 of *Statistical Tables and Formulas*, by A. Hald (New York: John Wiley & Sons, Inc., 1952). Reproduced by permission of A. Hald and the publishers, John Wiley & Sons, Inc.

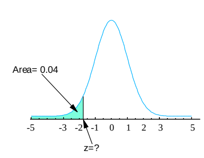


Normal Curve Areas

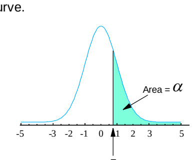
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990



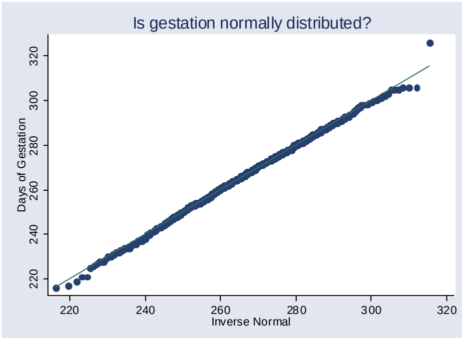
Finding the z-score for a specified area:



The z_α notation:
The symbol z_α is used to denote the z-score having an area of α (alpha) to its right under the standard normal curve.



qnorm gestatio, title("Is gestation normally distributed?")



Binomial Coefficient for 30 Successes in 50 Trials

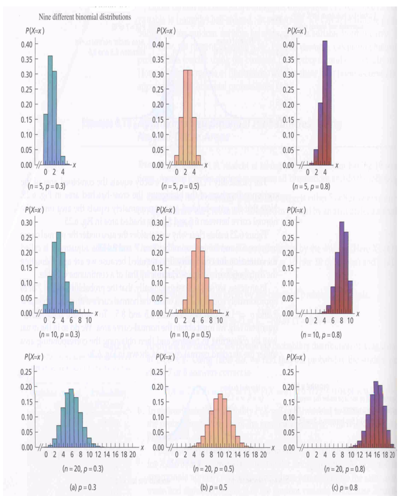
$$\binom{50}{30} = \frac{50!}{30!20!} = \frac{3.041 \times 10^{64}}{\left(2.653 \times 10^{32}\right)\left(2.433 \times 10^{18}\right)} = \frac{3.041 \times 10^{64}}{6.455 \times 10^{50}} = 4.713 \times 10^{13}$$

Do this 51 times to generate the true binomial distribution!

Binomial Distributions for varying probabilities of success and sample size.

Note that when *p* is far from 0.5, the binomial distribution is skewed. However, as sample size (i.e., number of trials) rises even these skewed distributions become more symmetrical and normal.

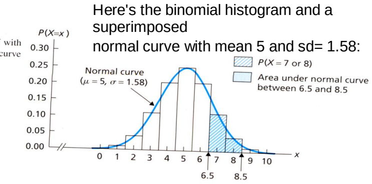
Sample Size Rises



Assume that a student taking a true-false statistics exam guesses at all 10 questions with a 50% chance of being correct on each question. What is the binomial distribution?

Here's the binomial histogram and a superimposed
normal curve with mean 5 and sd= 1.58:

Probability histogram for X with superimposed normal curve



Probability distribution of the number of correct answers out of 10 by the student

Number correct x	Probability $P(X = x)$
0	0.0010
1	0.0098
2	0.0439
3	0.1172
4	0.2051
5	0.2461
6	0.2051
7	0.1172
8	0.0439
9	0.0098
10	0.0010

$$\mu = np = 5$$

$$\sigma = \sqrt{np(1-p)} = 1.58$$

Because the normal curve only approximates the binomial histogram, we must correct for the continuity of the normal distribution. We do this by subtracting one-half unit from the low side of the range we're interested in (6.5 instead of 7.0) and adding one-half unit on the high side (8.5 instead of 8.0)

The area under the normal curve from 6.5 to 8.5 approximates the area contained in the histogram between 7 and 8.

NB: Be sure to make the continuity correction before converting to standard normal scores.

Procedure for using Normal Aproximiation to the Binomial Distribution:

Step 1: Determine, *n,*the number of trials, and *p,* the success probability.

Step 2: Determine whether both *np* and *n(1-p)* are 5 or greater. If they are not, do not use the normal approximation. (Rule of Thumb)

Step 3: Find μ and σ, using the formulas $\mu = np$ and $\sigma = \sqrt{np(1-p)}$

Step 4: Make the correction for continuity and find the required area under the normal curve with parameters μ and σ.

Mean and Variance of Continuous Probability Functions

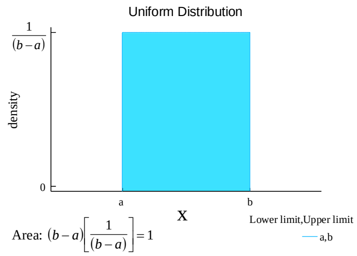
The Mean of Random Variable *x* is: $E[\tilde{x}] = \mu_x = \int_{-\infty}^{\infty} x f(x)\,d\tilde{x}$

The Expectation of the function of *x, g(x)* is: $E[g(\tilde{x})] = \int_{-\infty}^{\infty} g(x) f(x)\,d\tilde{x}$

The Expectation of x^2 is: $E[\tilde{x}^2] = \int_{-\infty}^{\infty} x^2 f(x)\,d\tilde{x}$

The Variance of *x* is: $V[\tilde{x}] = E\left[(\tilde{x} - \mu_x)^2\right] = E[\tilde{x}^2] - (E[\tilde{x}])^2$

You should verify that the formulas for mean and variance of a continuous distribution are the same as those given earlier for discrete distributions except that integral signs are substituted for summation signs and *f(x)dx* is substituted for *P(x).*



Area: $(b-a)\left[\frac{1}{(b-a)}\right]=1$

The Uniform Distribution

$$f(x)=\begin{cases}\frac{1}{(b-a)}, & a\le x\le b\\ 0, & \text{otherwise}\end{cases}$$

with

$$\mu=\frac{1}{2}(b+a) \text{ and } \sigma=\frac{(b-a)}{\sqrt{12}}$$

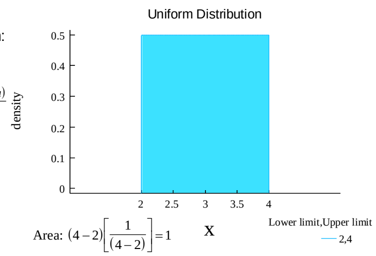
Deriving the mean of the uniform distribution:

$$\mu_x=\int_a^b xf(x)dx, \quad =\int_a^b x\left(\frac{1}{b-a}\right)dx, \quad =\frac{1}{b-a}\int_a^b xdx,$$

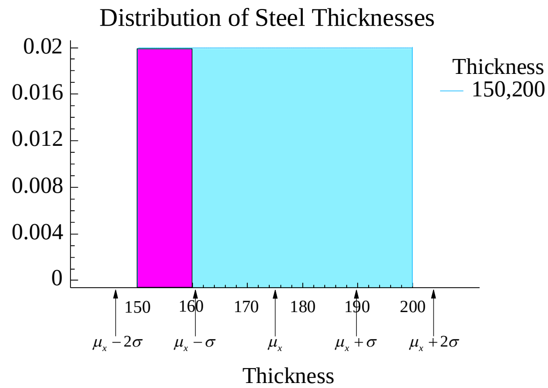
$$=\frac{1}{b-a}\Big|_a^b F(b)-F(a) \quad =\frac{1}{b-a}\left[\frac{1}{2}\left(b^2-a^2\right)\right] \quad =\frac{1}{2}\frac{(b-a)(b+a)}{(b-a)}$$

$$=\frac{1}{2}(b+a)$$

Can you derive the variance of the uniform distribution from the formula for the variance of a continuous distribution?



Area: $(4-2)\left[\frac{1}{(4-2)}\right]=1$



$$P\left(\tilde{x}<160=\left([160-150]\times 0.02\right)=10\times 0.02=0.2\right.$$

The Exponential Distribution: $f(x)=\begin{cases}\lambda e^{-\lambda x}, & \lambda>0, x\ge 0\\ 0, & \text{otherwise}\end{cases}$

with: $\mu=\frac{1}{\lambda}$ and $\sigma=\frac{1}{\lambda}$

Right-tailed probabilities: $P(x\ge a)=e^{-\lambda a}, a\ge 0 \text{ and } \lambda>0$