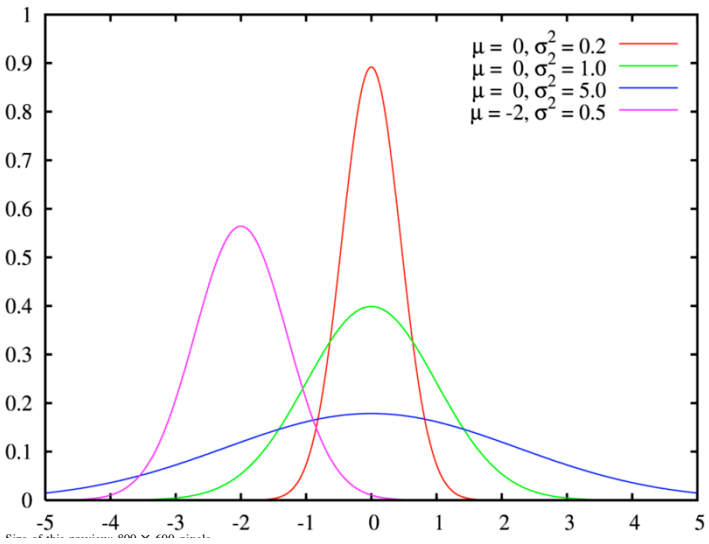
Section 8.7: Normal Distributions

So far we have dealt with random variables with a finite number of possible values. For example; if X is the number of heads that will appear, when you flip a coin 5 times, X can only take the values $0, 1, 2, 3, 4$, or 5.

Some variables can take a continuous range of values, for example a variable such as the height of 2 year old children in the U.S. population or the lifetime of an electronic component. For a continuous random variable X, the analogue of a histogram is a continuous curve (the probability density function) and it is our primary tool in finding probabilities related to the variable. As with the histogram for a random variable with a finite number of values, the total area under the curve equals 1. Probabilities correspond to areas under the curve and are calculated over intervals rather than for specific values of the random variable.

Although many types of probability density functions commonly occur for everyday random variables, we will restrict our interest to random variables with Normal Distributions and the probabilities will correspond to areas under a **Normal Curve** (or normal density function). It is important in that given any random variable with any distribution, the average of that variable over a fixed number of trials of the experiment will have a normal distribution. This applies in particular to sample means from a population. The shape of a Normal curve depends entirely on two parameters, μ and σ, which correspond, respectively, to the mean and standard deviation of the population for the associated random variable. Below we have a picture of a selection of Normal curves, for various values of μ and σ. The curve is always bell shaped. It is always centered (balanced) at the mean μ and larger values of σ give a curve that is more spread out. The area beneath the curve is always 1.



Properties of a Normal Curve

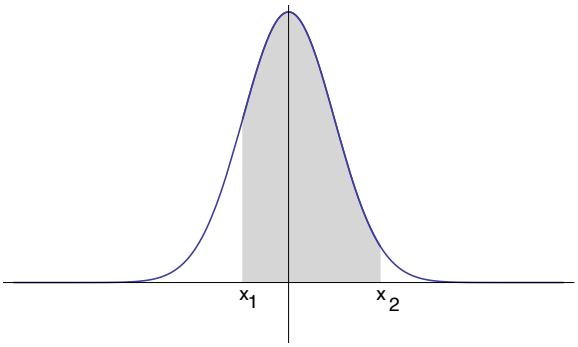
1. All Normal Curves have the same general bell shape.
2. The curve is symmetric with respect to a vertical line that passes through the peak of the curve.
3. The curve is centered at the mean μ which coincides with the median and the mode and is located at the point beneath the peak of the curve.
4. The area under the curve is always 1.

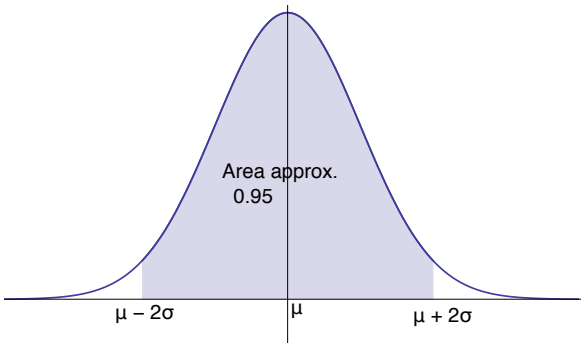
5. The curve is completely determined by the mean μ and the standard deviation σ. For the same mean, μ, a smaller value of σ gives a taller and narrower curve, whereas a larger value of σ gives a flatter curve.

6. The area under the curve to the right of the mean is 0.5 and the area under the curve to the left of the mean is 0.5.

7. The ical rule for mound shaped data applies to variables with normal distributions:
Approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval $(\mu - 2\sigma, \mu + 2\sigma)$ etc....

8. If a random variable X associated to an experiment has a normal probability distribution, the probability that the value of X derived from a single trial of the experiment is between two given values x_1 and x_2 ($P(x_1 \leq X \leq x_2)$) is the area under the associated normal curve between x_1 and x_2. For any given value x_1, $P(X = x_1) = 0$, so $P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)$.





The **standard Normal curve** is the normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$. We have included tables for the standard normal curve at the end of this set of notes. These tables give us the areas beneath the curve to the left of particular values of the Standard normal variable Z. As we will see below, this allows us to calculate probabilities that a range of values will occur for such a random variable Z. We can then standardize the values of any any normal random variable X and calculate the probabilities of events concerning X, using the standard tables.

Calculating Probabilities For a Standard Normal Random Variable

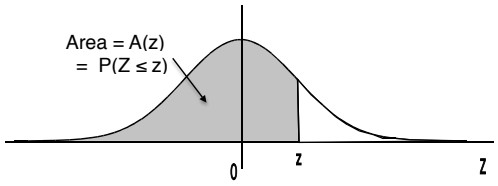
The **Table** shown at the end of your lecture consist of two columns, one gives a value for the variable z, and next to it, the table gives a value $A(z)$, which can be interpreted in either of two ways:

z	$A(z)$
1	.8413

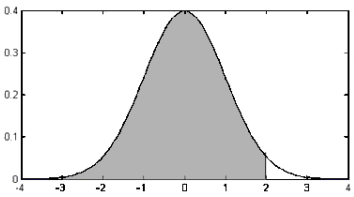
$A(z)$ = the area under the standard normal curve ($\mu = 0$ and $\sigma = 1$) to the left of this value of z, shown as the shaded region in the diagram below.

$A(z)$ = the probability that the value of the random variable Z observed for an individual chosen at random from the population is less than or equal to z. $A(z) = \mathrm{P}(Z \leq z)$.

The section of the table shown above tells us that the area under the standard normal curve to the left of the value $z = 1$ is 0.8413. It also tells us that if Z is normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$, then $\mathrm{P}(Z \leq 1) = .8413$.

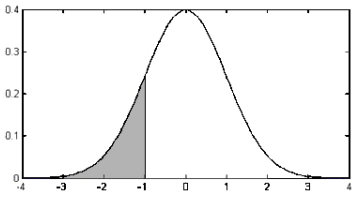


Example If Z is a standard normal random variable, what is $\mathrm{P}(Z \leq 2)$. Sketch the region under the standard normal curve whose area is equal to $\mathrm{P}(Z \leq 2)$. Use the tables at the end of this lecture to find $P(Z \leq 2)$.



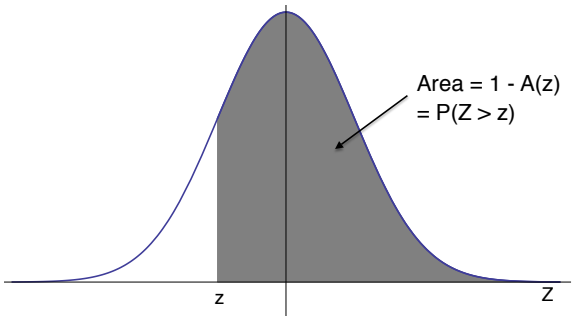
$P(Z \leqslant 2) = 0.9772.$

Example If Z is a standard normal random variable, what is $\mathrm{P}(Z \leq -1)$. Sketch the region under the standard normal curve whose area is equal to $\mathrm{P}(Z \leq -1)$.



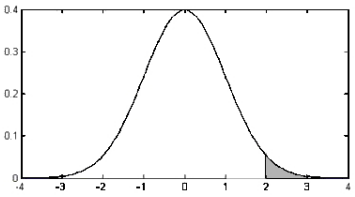
$P(Z \leqslant -1) = 0.1587.$

Recall now that the total area under the standard normal curve is equal to 1. Therefore the area under the curve to the right of a given value z is $1 - A(z)$. By the complement rule, this is also equal to $P(Z > z)$.

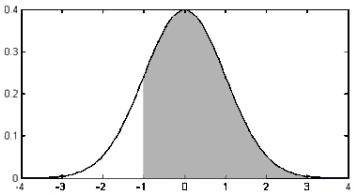


Example If Z is a standard normal random variable, use the above principle to find $P(Z \geq 2)$. Sketch the region under the standard normal curve whose area is equal to $\mathrm{P}(Z \geq 2)$.

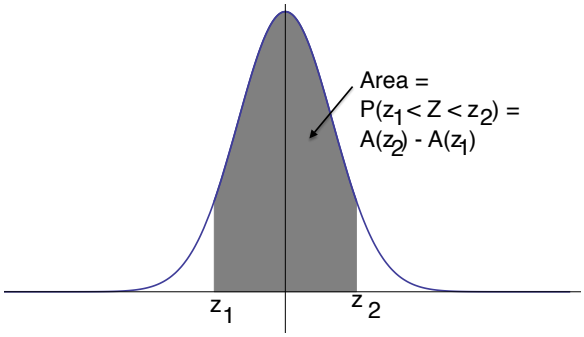
$P(Z \leqslant 2) = 0.9772$ so $P(Z \geqslant 2) = 1 - 0.9772 = 0.0228$.



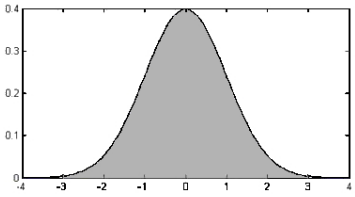
Example If Z is a standard normal random variable, find $P(Z \geq -1)$. Sketch the region under the standard normal curve whose area is equal to $\mathrm{P}(Z \geq -1)$.



$P(Z \leqslant -1) = 0.1587$ so $P(Z \geqslant -1) = 1 - 0.1587 = 0.8413$.



Example If Z is a standard normal random variable, find $\mathrm{P}(-3 \leq Z \leq 3)$. Sketch the region under the standard normal curve whose area is equal to $\mathrm{P}(-3 \leq Z \leq 3)$.



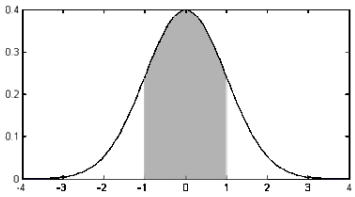
$\mathrm{P}(-3 \leq Z \leq 3) = \mathrm{P}(Z \leq 3) - \mathrm{P}(Z \leq -3) = 0.9987 - 0.0013 = 0.9973$.

Empirical Rule for the standard normal distribution ($\mu = 0$, $\sigma = 1$.) If the data has a normal distribution with $\mu = 0$, $\sigma = 1$, we have the following empirical rule:

- Approximately 68% of the measurements will fall within 1 standard deviation of the mean or equivalently in the interval $(-1, 1)$.
- Approximately 95% of the measurements will fall within 2 standard deviations of the mean or equivalently in the interval $(-2, 2)$.
- Approximately 99.7% of the measurements(essentially all) will fall within 3 standard deviations of the mean, or equivalently in the interval $(-3, 3)$.

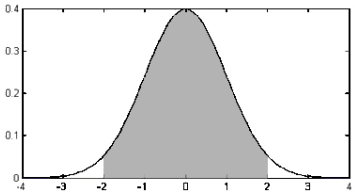
Verify the empirical rule: If Z is a standard normal random variable

(a) What is $P(-1 \leq Z \leq 1)$?



$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.8413 - 0.1587 = 0.6827.$

(b) What is $P(-2 \leq Z \leq 2)$?



$P(-2 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -2) = 0.9772 - 0.0228 = 0.9545.$

(c) What is $P(-3 \leq Z \leq 3)$(see above)?

Using your TI Eighty-Something calculator for Standard Normal Distribution

You can use your calculator to calculate the above probabilities for the standard normal distribution.

1. Bring up the distribution menu, using $\boxed{2nd}$ $\boxed{vars}$. Then select **normalcdf**.
2. To calculate $P(a \leq Z \leq b)$ where Z is a standard normal random variable (with mean 0 and standard deviation 1) we calculate **normalcdf(a, b, 0, 1)**
3. When the lower bound of our interval is $a = -\infty$, we use -E99 to represent a (keys on calculator; $\boxed{(-)}$ $\boxed{2nd}$ $\boxed{,}$ $\boxed{9}$ $\boxed{9}$)
4. When the upper bound of our interval is $b = \infty$, we use E99 to represent a (keys on calculator; $\boxed{2nd}$ $\boxed{,}$ $\boxed{9}$ $\boxed{9}$)

Example Let Z be a standard normal random variable.

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $P(-1.53 \leq Z \leq 2.16)$ and find the area using your calculator.

$P(-1.53 \leq Z \leq 2.16) = P(Z \leq 2.16) - P(Z \leq -1.53) = 0.9846 - 0.0630 = 0.9216.$

(b) Sketch the area beneath the density function of the standard normal random variable, corresponding to $P(-\infty \leq Z \leq 1.23)$ and find the area using your calculator.

$P(-\infty \leq Z \leq 1.23) = 0.8907.$

(c) Sketch the area beneath the density function of the standard normal random variable, corresponding to $P(1.12 \leq Z \leq \infty)$ and find the area using your calculator.

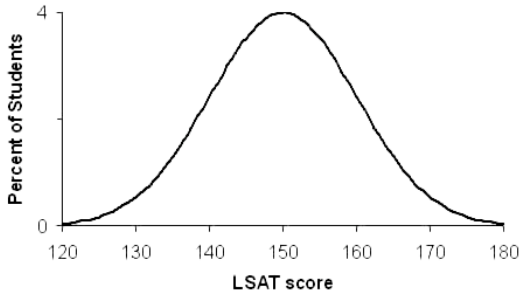
$P(1.12 \leq Z \leq \infty) = 1 - \big(P(Z \leq 1.12)\big) = 1 - 0.8686 = 0.1314.$

Non-Standard Normal Random Variables

We have already introduced the empirical rule for mound shaped distributions and used it to solve the problem shown below. We will now expand the same general principle to all Normal distributions.

Example Recall how we used the empirical rule to solve the following problem earlier in the course:

The scores on the LSAT exam, for a particular year, are normally distributed with mean $\mu = 150$ points and standard deviation $\sigma = 10$ points. What percentage of students got a score between 130 and 170 points in that year (or what percentage of students got a Z-score between -2 and 2 on the exam)?



We can use the same principle to calculate probabilities for any variable with a normal distribution. We do not have a table for every normal random variable, otherwise we would have infinitely many tables to store. In order to use the tables to calculate probabilities for a non-standard normal random variable, X, we first standardize the relevant values of X, by calculating their Z-scores. We then use the tables for the standard normal random variable Z, given above, to calculate the probability.

Property If X is a normal random variable with mean μ and standard deviation σ, then the random variable Z defined by the formula (Note this is the Z-score of X):

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

has a standard normal distribution. The value of Z gives the number of standard deviations between X and the mean μ.

To calculate $P(a \leq X \leq b)$, where X is a normal random variable with mean μ and standard deviation σ;

- We calculate the Z-scores for a and b:

  $$P(a \leq X \leq b) = P\Big(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\Big) = P\Big(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\Big)$$

  where Z is a standard normal random variable.
- We then use the table for standard normal probability distribution to calculate the probability.
- If $a = -\infty$, then $\frac{a-\mu}{\sigma} = -\infty$ and similarly if $b = \infty$, then $\frac{b-\mu}{\sigma} = \infty$.

Example If the length of newborn alligators, X, is normally distributed with mean $\mu = 6$ inches and standard deviation $\sigma = 1.5$ inches, what is the probability that an alligator egg about to hatch, will deliver a baby alligator between 4.5 inches and 7.5 inches?

$\mathrm{P}(4.5 \leq X \leq 7.5) = \mathrm{P}\left(\frac{4.5-6}{1.5} \leq Z \leq \frac{7.5-6}{1.5}\right) = \mathrm{P}(-1 \leq z \leq 1) = 0.6827$ or about 68%.

Example Time to failure of a particular brand of lightbulb is normally distributed with mean $\mu = 400$ hours and standard deviation $\sigma = 20$ hours.

(a) What percenatge of the bulbs will last longer than 438 hours?

$\mathrm{P}(438 \leq X < \infty) = \mathrm{P}\left(\frac{438-400}{20} \leq Z \leq \infty\right) = \mathrm{P}(1.9 \leq z) = 1 - \mathrm{P}(Z \leqslant 1.9) = 1 - 0.9713 = 0.0287$ or about 2.9%.

(b) What percentage of the bulbs will fail before 360 hours?

$\mathrm{P}(-\infty < X \leq 438) = \mathrm{P}\left(-\infty \leq Z \leq \dfrac{360-400}{20}\right) = \mathrm{P}(Z \leq -2) = 0.0228$ or about 2.9%.

Using your calculator for non-standard normal distributions

You can use your calculator to calculate the above probabilities for a normal distribution.

1. Bring up the distribution menu, using $\boxed{2nd}$ $\boxed{vars}$. Then select **normalcdf**.
2. To calculate $P(a \leq X \leq b)$ where X is a normal random variable with mean μ and standard deviation σ we calculate **normalcdf(a, b, μ, σ)**
3. When the lower bound of our interval is $a = -\infty$, we use -E99 to represent a (keys on calculator; $\boxed{(-)}$ $\boxed{2nd}$ $\boxed{\ ,\ }$ $\boxed{9}$ $\boxed{9}$)
4. When the upper bound of our interval is $b = \infty$, we use E99 to represent a (keys on calculator; $\boxed{2nd}$ $\boxed{\ ,\ }$ $\boxed{9}$ $\boxed{9}$)

Example Let X be a normal random variable with mean $\mu = 100$ and standard deviation $\sigma = 15$, what is the probability that the value of x falls between 80 and 105; $P(80 \leq X \leq 105)$.

$$\mathrm{P}(80 \leq X \leq 105) = \mathrm{P}\left(\frac{80-100}{15} \leq Z \leq \frac{105-100}{15}\right) = \mathrm{P}(-1.3333 \leq Z \leq 0.3333) = 0.6305 - 0.0912 = 0.5393.$$

Example Dental Anxiety Assume that scores on a Dental anxiety scale (ranging from 0 to 20) are normal for the general population, with mean $\mu = 11$ and standard deviation $\sigma = 3.5$.

(a) What is the probability that a person chosen at random will score between 10 and 15 on this scale?

$$\mathrm{P}(80 \leq X \leq 105) = \mathrm{P}\left(\frac{10-11}{3.5} \leq Z \leq \frac{15-11}{3.5}\right) = \mathrm{P}(-0.2857 \leq Z \leq 1.1429) = 0.8735 - 0.3875 = 0.4859.$$

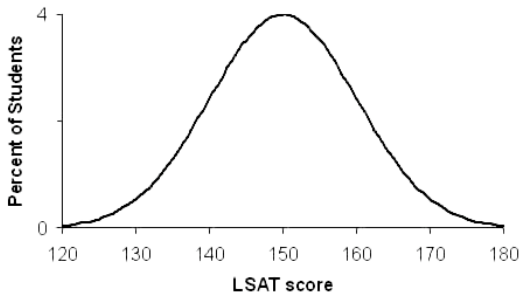
(b) What is the probability that a person chosen at random will have a score larger then 10 on this scale?

$$\mathrm{P}(10 \leq X < \infty) = \mathrm{P}\left(\frac{10-11}{3.5} \leq Z < \infty\right) = \mathrm{P}(-0.2857 \leq Z < \infty) = 1-(0.3875) = 0.6125.$$

(c) What is the probability that a person chosen at random will have a score less than 5 on this scale?

$$\mathrm{P}(-\infty < X \leq 5) = \mathrm{P}\left(\infty < Z \leq \frac{5-11}{3.5}\right) = \mathrm{P}(Z \leq -1.7143) = 0.0432.$$

Example Let X denote the scores on the LSAT for a particular year. The mean is $\mu = 150$ and the standard deviation is $\sigma = 10$. The "histogram" or density function for the scores looks like:



Although, technically, the variable X is not continuous, the histogram is very closely approximated by the above curve and the probabilities can be calculated from it. What percentage of students had a score of 165 or higher on this LSAT exam?

$$\mathrm{P}(165 \leq X < \infty) = \mathrm{P}\left(\frac{165-150}{10} \leq Z < \infty\right) = \mathrm{P}(1.5 \leq Z < \infty) = 1 - \mathrm{P}(Z \leq 1.5) = 1 - (0.9332) = 0.0668.$$

Example Let X denote the weight of newborn babies at Memorial Hospital. The weights are normally distributed with mean $\mu = 8$ lbs and standard deviation $\sigma = 2$ lbs.

(a) What is the probability that the weight of a newborn, chosen at random from the records at Memorial Hospital, is less than or equal to 9 lbs?

$$\mathrm{P}(X \leq 9) = \mathrm{P}\left(Z \leq \frac{9-8}{2}\right) = \mathrm{P}(Z \leq 0.5) = 0.6915.$$

(b) What is the probability that the weight of a newborn baby, selected at random from the records of Memorial Hospital, will be between 6 lbs and 8 lbs?

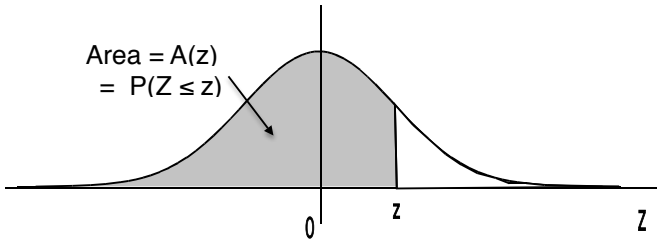
$$\mathrm{P}(6 \leq X \leq 8) = \mathrm{P}\left(\frac{6-8}{2} \leq Z < \frac{8-8}{2}\right) = \mathrm{P}(1 \leq Z < 0) = 0.5 - 0.1587 = 0.3413.$$

Example Let X denote Miriam's monthly living expenses. X is normally distributed with mean $\mu = \$1,000$ and standard deviation $\sigma = \$150$. On Jan. 1, Miriam finds out that her money supply for

January is \$1,150. What is the probability that Miriam's money supply will run out before the end of January?

If Miriam's monthly expenses exceed $\$1,150$ she will run out of money before the end of the month. Hence we want $\mathrm{P}(1,150 \leq X)$: $\mathrm{P}\left(\frac{1150-1000}{150} \leq X\right) = \mathrm{P}(1 \leq Z) = 1 - \mathrm{P}(Z \leq 1) = 1 - (0.8413) = 0.1587$.

Areas under the Standard Normal Curve



z	$A(z)$	z	$A(z)$	z	$A(z)$	z	$A(z)$	z	$A(z)$
−3.50	.0002	−2.00	.0228	−.50	.3085	1.00	.8413	2.50	.9938
−3.45	.0003	−1.95	.0256	−.45	.3264	1.05	.8531	2.55	.9946
−3.40	.0003	−1.90	.0287	−.40	.3446	1.10	.8643	2.60	.9953
−3.35	.0004	−1.85	.0322	−.35	.3632	1.15	.8749	2.65	.9960
−3.30	.0005	−1.80	.0359	−.30	.3821	1.20	.8849	2.70	.9965
−3.25	.0006	−1.75	.0401	−.25	.4013	1.25	.8944	2.75	.9970
−3.20	.0007	−1.70	.0446	−.20	.4207	1.30	.9032	2.80	.9974
−3.15	.0008	−1.65	.0495	−.15	.4404	1.35	.9115	2.85	.9978
−3.10	.0010	−1.60	.0548	−.10	.4602	1.40	.9192	2.90	.9981
−3.05	.0011	−1.55	.0606	−.05	.4801	1.45	.9265	2.95	.9984
−3.00	.0013	−1.50	.0668	.00	.5000	1.50	.9332	3.00	.9987
−2.95	.0016	−1.45	.0735	.05	.5199	1.55	.9394	3.05	.9989
−2.90	.0019	−1.40	.0808	.10	.5398	1.60	.9452	3.10	.9990
−2.85	.0022	−1.35	.0885	.15	.5596	1.65	.9505	3.15	.9992
−2.80	.0026	−1.30	.0968	.20	.5793	1.70	.9554	3.20	.9993
−2.75	.0030	−1.25	.1056	.25	.5987	1.75	.9599	3.25	.9994
−2.70	.0035	−1.20	.1151	.30	.6179	1.80	.9641	3.30	.9995
−2.65	.0040	−1.15	.1251	.35	.6368	1.85	.9678	3.35	.9996
−2.60	.0047	−1.10	.1357	.40	.6554	1.90	.9713	3.40	.9997
−2.55	.0054	−1.05	.1469	.45	.6736	1.95	.9744	3.45	.9997
−2.50	.0062	−1.00	.1587	.50	.6915	2.00	.9772	3.50	.9998
−2.45	.0071	−.95	.1711	.55	.7088	2.05	.9798		
−2.40	.0082	−.90	.1841	.60	.7257	2.10	.9821		
−2.35	.0094	−.85	.1977	.65	.7422	2.15	.9842		
−2.30	.0107	−.80	.2119	.70	.7580	2.20	.9861		
−2.25	.0122	−.75	.2266	.75	.7734	2.25	.9878		
−2.20	.0139	−.70	.2420	.80	.7881	2.30	.9893		
−2.15	.0158	−.65	.2578	.85	.8023	2.35	.9906		
−2.10	.0179	−.60	.2743	.90	.8159	2.40	.9918		
−2.05	.0202	−.55	.2912	.95	.8289	2.45	.9929		

Extras: Calculating Percentiles: Using the tables in reverse

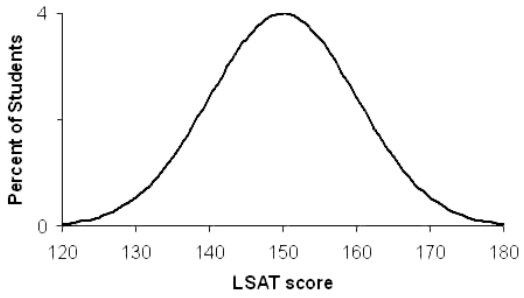
Percentiles for a Normal Distribution Recall that x_p is the pth percentile for the random variable X if $p\%$ of the population have values of X which are at or lower than x_p and $(100-p)\%$ have values of X at or greater than x_p. To find the pth percentile of a normal distribution with mean μ and standard deviation σ, we can use the tables in reverse or use the invNorm function on our calculator.

1. We are looking for the value of a such that $\mathrm{P}(a \leq X) = p$.
2. To evaluate the value of the pth percentile of a random variable X with a normal distribution with mean μ and standard deviation σ, we we pull up the distributions menu on our calculator using the buttons $\boxed{2nd}$ $\boxed{VARS}$. We then select the *invNorm* function and calculate **invNorm(** $p/100,\ \mu,\ \sigma$**)**
3. For example to calculate the 25th percentile of a normal random variable X, with mean $\mu = 75$ and standard deviation $\sigma = 15$ we calculate **invNorm(0.25, 75, 15)** which should give an answer of 64.883. This means that 25% of the population of interest has a value of X less than or equal to 64.883.

Example Calculate the 95th, 97.5th and 60th percentile of a normal random variable X, with mean $\mu = 400$ and standard deviation $\sigma = 35$.

- 95^{th}-percentile: $b = \dfrac{a-400}{35}$ and looking in the table gives $b = 1.65$ so $a = 35 \cdot 1.65 + 400 = 457.75$.
- 97.5^{th}-percentile. Looking at the table gives $b = 1.97$ so $a = 35 \cdot 1.95 + 400 = 468.25$.
- 60^{th}-percentile. Looking at the table gives $b = 1.97$ so $a = 35 \cdot 0.27 + 400 = 409.45$.

Example : The scores on the LSAT for a particular year, X, have a normal distribution with mean, $\mu = 150$, and the standard deviation, $\sigma = 10$. The distribution is shown below.



(a) Find the 90th percentile of the distribution of scores.

90^{th}-percentile $a = 162.8155$.

Extras: Old Exam Questions

1 The lifetime of Didjeridoos is normally distributed with mean $\mu = 150$ years and standard deviation $\sigma = 50$ years. What proportion of Didjeridoos have a lifetime longer than 225 years?

(a) 0.0668 (b) 0.5668 (c) 0.9332 (d) 0.5 (e) 0.4332

$$\mathrm{P}(225 \leq X) = \mathrm{P}\left(\frac{225-150}{50} \leq Z\right) = \mathrm{P}(1.5 \leq Z) = 1 - \mathrm{P}(Z \leq 1.5) = 1 - 0.9332 = 0.0668.$$

2 Test scores on the OWLs at Hogwarts are normally distributed with mean $\mu = 250$ and standard deviation $\sigma = 30$. Only the top 5% of students will qualify to become an Auror. What is the minimum score that Harry Potter must get in order to qualify?

(a) 200.65 (b) 299.35 (c) 280 (d) 310 (e) 275.5

We need to find a so that $\mathrm{P}(a \leq X) = 0.05$. Let $\alpha = \frac{a-\mu}{\sigma}$. Then $\mathrm{P}(a \leq X) = \mathrm{P}(\alpha \leq Z) = 0.05$ so $\mathrm{P}(\alpha \leq Z) = 1 - \mathrm{P}(Z \leq \alpha)$ so $\mathrm{P}(Z \leq \alpha) \leq 1 - 0.05 = 0.95$. From the table $\mathrm{P}(\alpha \leq Z) = 0.95$ so $\alpha \approx 1.65$. Hence $a = 250 + 30 \cdot 1.65 = 299.3456$ to four decimal places so (b) is the correct answer.

3 Find the area under the standard normal curve between $z = -2$ and $z = 3$.

(a) 0.9759 (b) 0.9987 (c) 0.0241 (d) 0.9785 (e) 0.9772

$$\mathrm{P}(-2 \leq Z \leq 3) = \mathrm{P}(Z \leq 3) - \mathrm{P}(Z \leq -2) = 0.9987 - 0.0228 = 0.9759.$$

4 The number of pints of Guinness sold at "The Fiddler's Hearth" on a Saturday night chosen at random is Normally Distributed with mean $\mu = 50$ and standard deviation $\sigma = 10$. What is the probability that the number of pints of Guinness sold on a Saturday night chosen at random is greater than 55.

(a) .6915 (b) .3085 (c) .8413 (d) .1587 (e) .5

$$\mathrm{P}(55 \leq X) = \mathrm{P}\left(\frac{55-50}{10} \leq Z < \infty\right) = \mathrm{P}(0.5 \leq Z) = 1 - \mathrm{P}(Z \leq 0.5) = 1 - (0.6915) = 0.3085.$$

Extras: Approximating a Binomial Distribution using A Normal Distribution

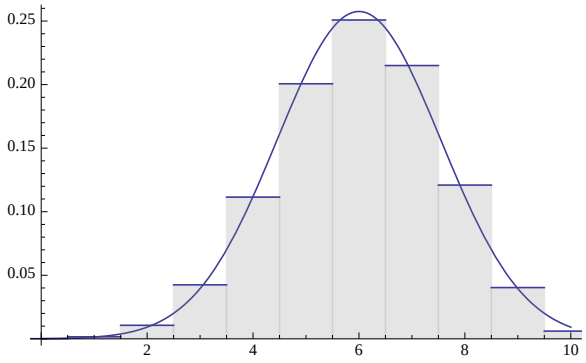
Recall that a **binomial random variable**, X, counts the number of success' in n independent trials of an experiment with two outcomes, success and failure.

Below we show the histograms for a binomial random variable, with $p = 0.6$, $q = 0.4$, as the value of n (= the number of trials) varies from $n = 10$ to $n = 30$ to $n = 100$ to $n = 200$. We have superimposed the density function for a normal random variable with mean $\mu = E(X) = np$ and standard deviation $\sigma = \sigma(X) = \sqrt{npq}$ on each histogram for the binomial distribution. We can see that even with $n = 10$, areas from the histogram are already well approximated by areas under the corresponding normal curve. As n increases, the approximation gets better and better and the Normal Distribution with the appropriate mean and standard deviation gives a very good approximation to the probabilities for the binomial distribution.

n = 10 Let X denote the number of success' in $n = 10$ independent trials of a binomial experiment, with $p = 0.6$ and $q = 0.4$. The random variable X can take on any of the values, 0, 1, 2, ..., 10. We have a formula for the probability that X takes the value k, namely

$$P(X = k) = \binom{10}{k} p^k (1-p)^{10-k} = \binom{10}{k} (0.6)^k (0.4)^{10-k}, \qquad k = 0, 1, \ldots, 10.$$

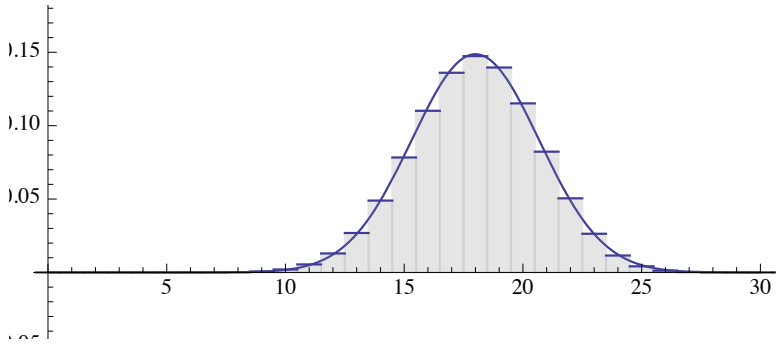
The probabilities are shown in the histogram below along with the Normal curve with $\mu = 6 = E(X)$ and $\sigma = \sqrt{npq} = \sigma(X)$.



n = 30: Here we have the probability distribution for X, where X is the number of success' in 30 trials of a binomial experiment, with $p = 0.6$ and $q = 0.4$. The random variable X can take on any of the values, 0, 1, 2, ..., 30.

$$P(X = k) = \binom{30}{k} p^k (1-p)^{30-k} = \binom{30}{k} (0.6)^k (0.4)^{30-k}, \qquad k = 0, 1, \ldots, 30.$$

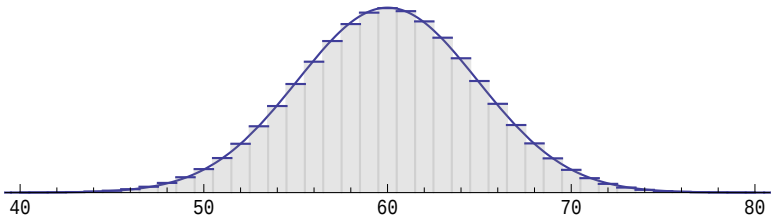
The probabilities are shown in the histogram below along with the Normal curve with $\mu = 18 = E(X)$ and $\sigma = \sqrt{npq} = \sigma(X)$.



n = 100: Here we have the probability distribution for X, where X is the number of success' in 100 trials of a binomial experiment, with $p = 0.6$ and $q = 0.4$. The random variable X can take on any of the values, 0, 1, 2, ..., 100.

$$P(X = k) = \binom{100}{k} p^k (1-p)^{100-k} = \binom{100}{k} (0.6)^k (0.4)^{100-k}, \qquad k = 0, 1, \ldots, 100.$$

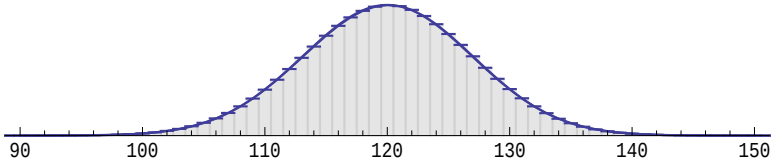
The probabilities are shown in the histogram below along with the Normal curve with $\mu = 60 = E(X)$ and $\sigma = \sqrt{npq} = \sigma(X)$.



n = 200: Here we have the probability distribution for X, where X is the number of success' in 200 trials of a binomial experiment, with $p = 0.6$ and $q = 0.4$. The random variable X can take on any of the values, 0, 1, 2, ..., 200.

$$P(X = k) = \binom{200}{k} p^k (1-p)^{200-k} = \binom{200}{k} (0.6)^k (0.4)^{200-k}, \qquad k = 0, 1, \ldots, 200.$$

The probabilities are shown in the histogram below along with the Normal curve with $\mu = 120 = E(X)$ and $\sigma = \sqrt{npq} = \sigma(X)$.



Example: Will Melinda McNulty win the election?

Suppose that Melinda has just one opponent, Mark Reckless, then she needs to get more than 50% of the votes to win. Lets assume that the population is very large and I take a random sample of 100 people and ask if they will vote for Melinda or not in the upcoming election. Now because the population is very large, the variable X = number of people who say yes has a distribution which is almost identical to a binomial distribution with $n = 100$. We do not know what p is but we would like for p to be greater than or at the very least equal to 0.5 (we're on Melinda's side).

Now suppose that in our poll, we had only 40% of the sample say that they will vote for Melinda. This is not good news, but we know that it may be just due to variation in sample statistics. We can use our normal approximation to the binomial to check the likelihood of getting a sample with this result in the most conservative winning/drawing scenario where 50% of the population will vote for Melinda.

So, assuming that $p = 0.5$, the distribution of X is approximately normal with mean $\mu = np = 50$ and standard deviation $\sigma = \sqrt{npq} = \sqrt{25} = 5$. Use the normal distribution to estimate the likelihood that we would get a sample where $X \leq 40$ i.e. estimate $P(X \leq 40)$.

Using the binomial distribution, $\mathrm{P}(X \leq 40) = \mathrm{P}(X = 40) + \mathrm{P}(X = 39) + \cdots + \mathrm{P}(X = 0) = C(100, 40)(0.5)^{40}(0.5)^{60} + \cdots C(100, 0)(0.5)^{0}(0.5)^{100} \approx 0.0284439668.$

Using the normal distribution, $\mathrm{P}(X \leq 40) = \mathrm{P}\left(Z \leq \dfrac{40 - 50}{5}\right) = \mathrm{P}(Z \leq -2) \approx 0.0228.$