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.

Normal Distributions

Probabilities correspond to areas under the curve and are calculated over intervals rather than for specific values of the random variable.

Normal Distributions

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, we will restrict our attention to random variables with Normal Distributions and the probabilities will correspond to areas under a **Normal Curve** (or normal density function).

Normal Distributions

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, we will restrict our attention to random variables with Normal Distributions and the probabilities will correspond to areas under a **Normal Curve** (or normal density function).

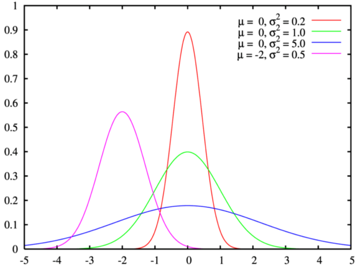
This is the most important example of a continuous random variable, because of something called the **Central Limit Theorem**: given *any* random variable with *any* distribution, the average (over many observations) of that variable will (essentially) have a normal distribution. This makes it possible, for example, to draw reliable information from opinion polls.

Normal Distributions

The shape of a Normal curve depends on two parameters, μ and σ, which correspond, respectively, to the mean and standard deviation of the population for the associated random variable.

Normal Distributions

The shape of a Normal curve depends on two parameters, μ and σ, which correspond, respectively, to the mean and standard deviation of the population for the associated random variable. The graph below shows a selection of Normal curves, for various values of μ and σ. The curve is always bell shaped, and always centered at the mean μ. 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.

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.

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.

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.

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.

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.

Properties of a Normal Curve

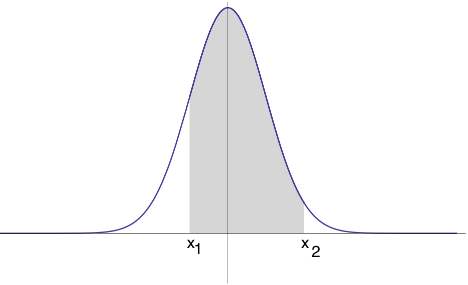
7. The empirical rule (68%, 95%, 99.7%) for mound shaped data applies to variables with normal distributions.
   For example, approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval $(\mu - 2\sigma, \mu + 2\sigma)$.

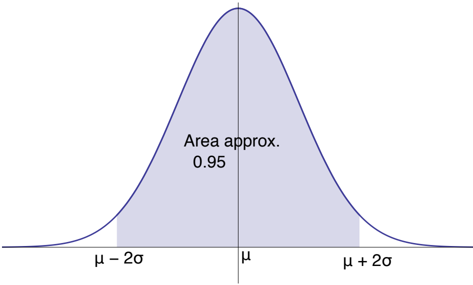
Properties of a Normal Curve

7. The empirical rule (68%, 95%, 99.7%) for mound shaped data applies to variables with normal distributions.
   For example, approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval $(\mu - 2\sigma, \mu + 2\sigma)$.
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 ($\mathbf{P}(x_1 \leqslant X \leqslant x_2)$) is the area under the associated normal curve between x_1 and x_2. For any given value x_1, $\mathbf{P}(X = x_1) = 0$, so
   $\mathbf{P}(x_1 \leqslant X \leqslant x_2) = \mathbf{P}(x_1 < X < x_2)$.

Properties of a Normal Curve

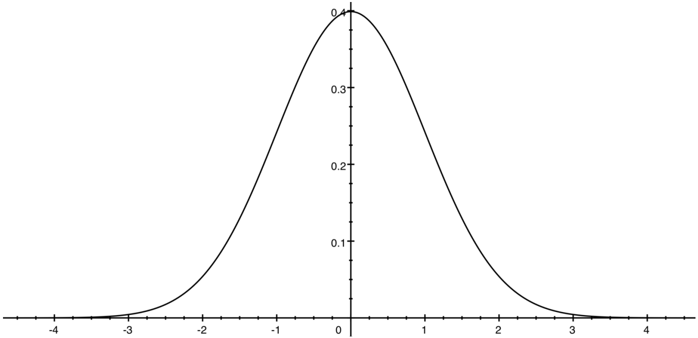
Here are a couple of pictures to illustrate items 7 and 8.





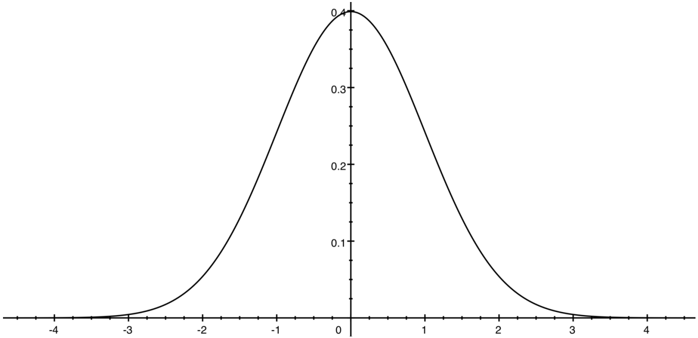
The standard Normal curve

The **standard Normal curve** is the normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.



The standard Normal curve

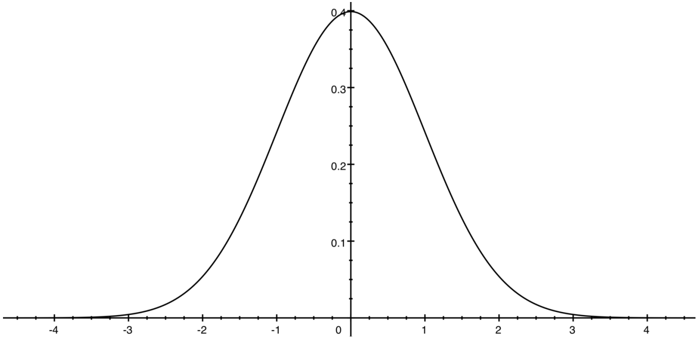
The **standard Normal curve** is the normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.



We will see later how probabilities for any normal curve can be recast as probabilities for the standard normal curve.

The standard Normal curve

The **standard Normal curve** is the normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.

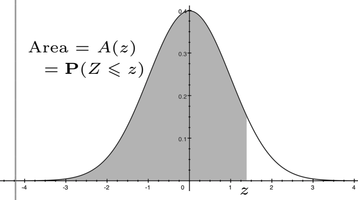


We will see later how probabilities for any normal curve can be recast as probabilities for the standard normal curve.

For the standard normal, probabilities are computed either by means of a computer/calculator or via a table.

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		



Probabilities for the standard Normal

The table consists of two columns. One (on the left) gives a value for the variable z, and one (on the right) gives a value $A(z)$, which can be interpreted in either of two ways:

z	$A(z)$
1	0.8413

Probabilities for the standard Normal

The table consists of two columns. One (on the left) gives a value for the variable z, and one (on the right) gives a value $A(z)$, which can be interpreted in either of two ways:

z	$A(z)$
1	0.8413

$A(z)$ = 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 on the next page.

Probabilities for the standard Normal

The table consists of two columns. One (on the left) gives a value for the variable z, and one (on the right) gives a value $A(z)$, which can be interpreted in either of two ways:

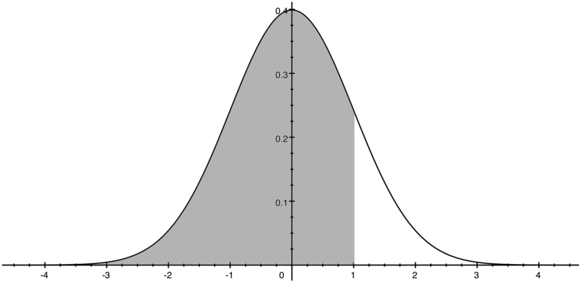
z	$A(z)$
1	0.8413

$A(z) =$ 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 on the next page.

$A(z) =$ 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.
$A(z) = \mathbf{P}(Z \leqslant z)$.

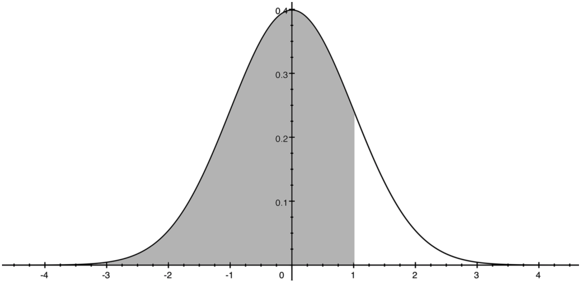
Probabilities for the standard Normal

The shaded area is $A(1) = 0.8413$, correct to 4 decimal places.



Probabilities for the standard Normal

The shaded area is $A(1) = 0.8413$, correct to 4 decimal places.



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 $\mathbf{P}(Z \leqslant 1) = .8413$.

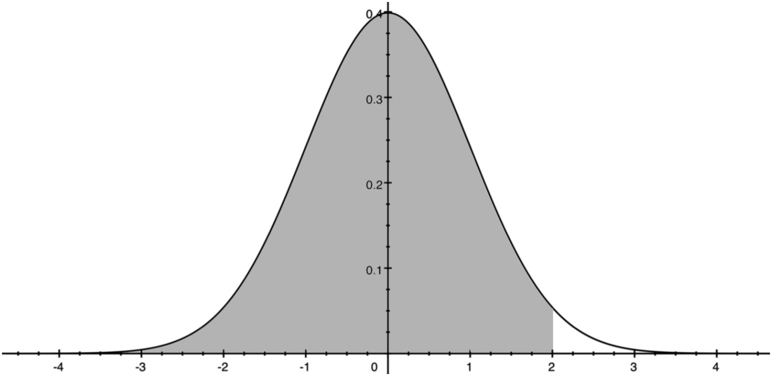
Examples

If Z is a standard normal random variable, what is $\mathbf{P}(Z \leqslant 2)$? Sketch the region under the standard normal curve whose area is equal to $\mathbf{P}(Z \leqslant 2)$. Use the table to find $\mathbf{P}(Z \leqslant 2)$.

Examples

If Z is a standard normal random variable, what is $\mathbf{P}(Z \leqslant 2)$? Sketch the region under the standard normal curve whose area is equal to $\mathbf{P}(Z \leqslant 2)$. Use the table to find $\mathbf{P}(Z \leqslant 2)$.

$\mathbf{P}(Z \leqslant 2) = 0.9772.$



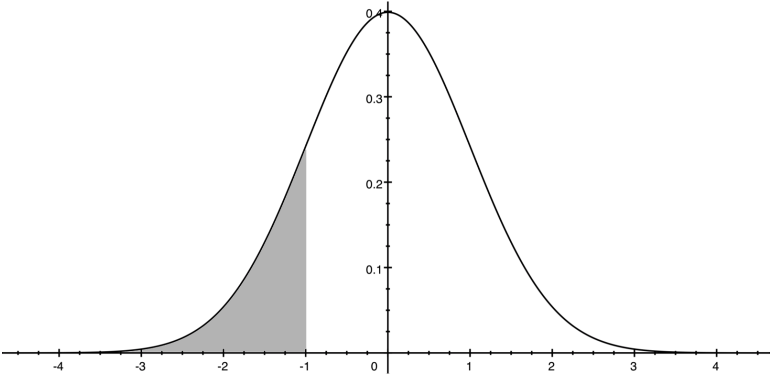
Examples

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

Examples

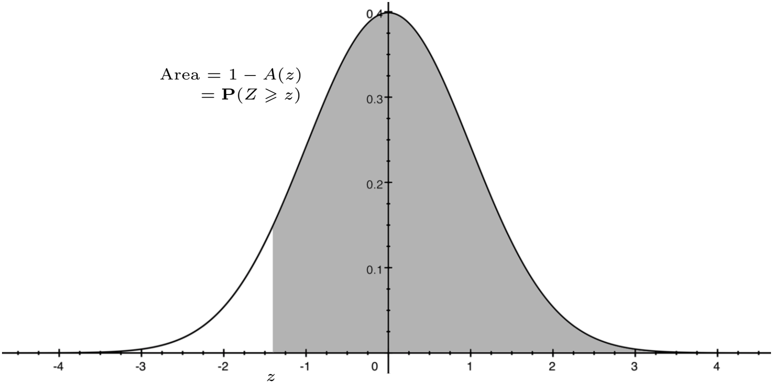
If Z is a standard normal random variable, what is $\mathbf{P}(Z \leqslant -1)$? Sketch the region under the standard normal curve whose area is equal to $\mathbf{P}(Z \leqslant -1)$.

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



Area to the right of a value

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 $\mathbf{P}(Z > z)$.



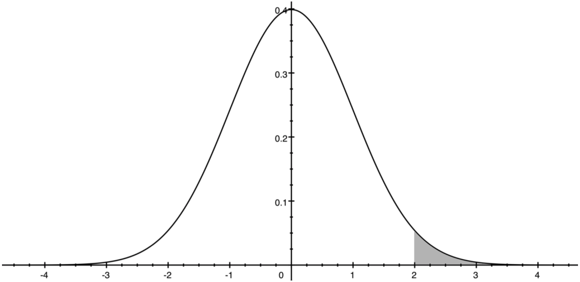
Examples

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

Examples

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

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



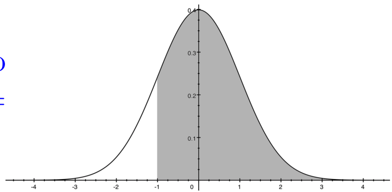
Examples

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

Examples

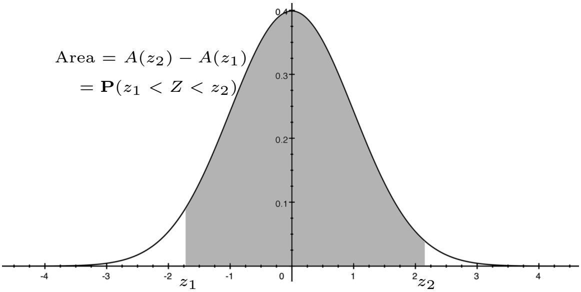
If Z is a standard normal random variable, find $\mathbf{P}(Z \geqslant -1)$. Sketch the region under the standard normal curve whose area is equal to $\mathbf{P}(Z \geqslant -1)$.

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



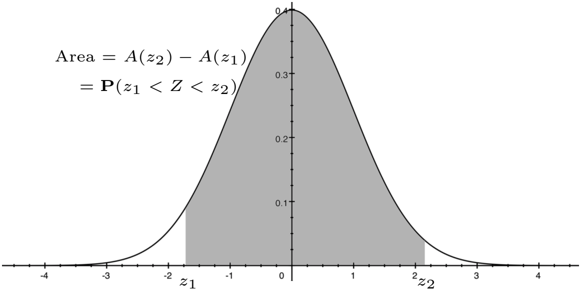
The area between two values

We can also use the table to compute
$\mathbf{P}(z_1 < Z < z_2) = \mathbf{P}(z_1 \leqslant Z < z_2) = \mathbf{P}(z_1 < Z \leqslant z_2) = \mathbf{P}(z_1 \leqslant Z \leqslant z_2) = A(z_2) - A(z_1)$.



The area between two values

We can also use the table to compute $\mathbf{P}(z_1 < Z < z_2) = \mathbf{P}(z_1 \leqslant Z < z_2) = \mathbf{P}(z_1 < Z \leqslant z_2) = \mathbf{P}(z_1 \leqslant Z \leqslant z_2) = A(z_2) - A(z_1)$.

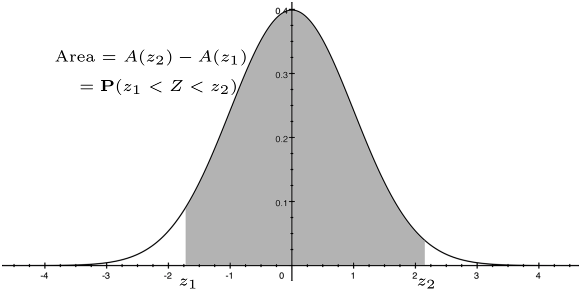


Our previous examples can be thought of like this:

$$\mathbf{P}(Z \leqslant z) = \mathbf{P}(-\infty < Z \leqslant z) = A(z) - A(-\infty) = A(z)$$

The area between two values

We can also use the table to compute $\mathbf{P}(z_1 < Z < z_2) = \mathbf{P}(z_1 \leqslant Z < z_2) = \mathbf{P}(z_1 < Z \leqslant z_2) = \mathbf{P}(z_1 \leqslant Z \leqslant z_2) = A(z_2) - A(z_1)$.



Our previous examples can be thought of like this:

$$\mathbf{P}(Z \leqslant z) = \mathbf{P}(-\infty < Z \leqslant z) = A(z) - A(-\infty) = A(z)$$

$$\mathbf{P}(z < Z) = \mathbf{P}(z < Z < \infty) = A(\infty) - A(z) = 1 - A(z)$$

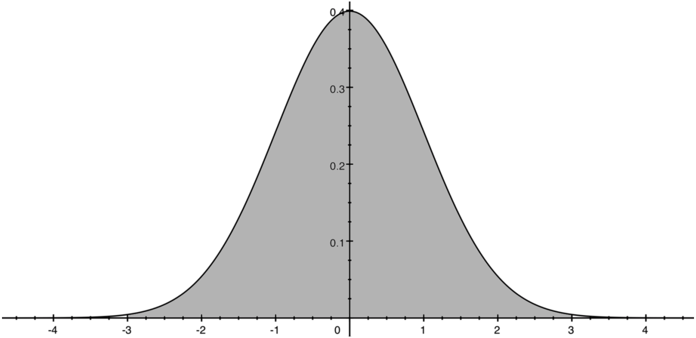
Example

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

Example

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

$\mathbf{P}(-3 \leqslant Z \leqslant 3) = \mathbf{P}(Z \leqslant 3) - \mathbf{P}(Z \leqslant -3) = 0.9987 - 0.0013 = 0.9973.$



Empirical Rule for the standard normal

If 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 $(-1, 1)$.

Empirical Rule for the standard normal

If 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 $(-2, 2)$.

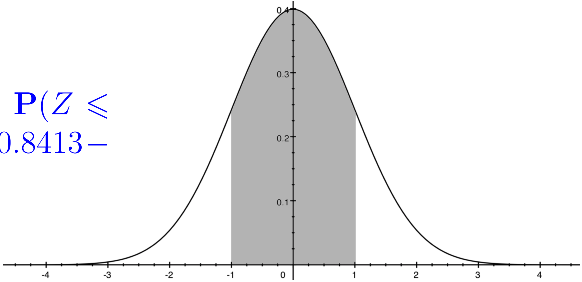
Empirical Rule for the standard normal

If 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 $(-3, 3)$.

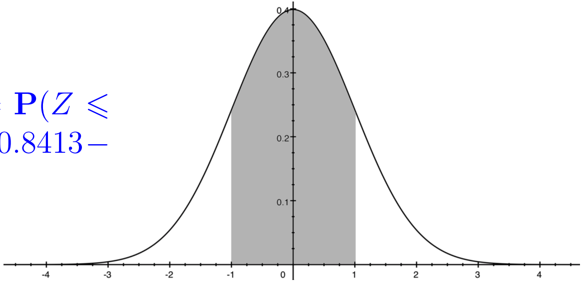
Verifying the empirical rule

$\mathbf{P}(-1 \leqslant Z \leqslant 1) = \mathbf{P}(Z \leqslant 1) - \mathbf{P}(Z \leqslant -1) = 0.8413 - 0.1587 = 0.6827.$

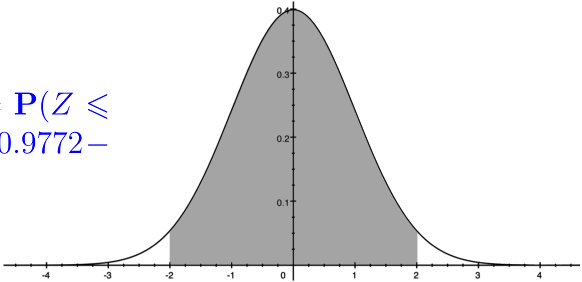


Verifying the empirical rule

$\mathbf{P}(-1 \leqslant Z \leqslant 1) = \mathbf{P}(Z \leqslant 1) - \mathbf{P}(Z \leqslant -1) = 0.8413 - 0.1587 = 0.6827.$



$\mathbf{P}(-2 \leqslant Z \leqslant 2) = \mathbf{P}(Z \leqslant 2) - \mathbf{P}(Z \leqslant -2) = 0.9772 - 0.0228 = 0.9545.$



Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

$\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16) = \mathbf{P}(Z \leqslant 2.16) - \mathbf{P}(Z \leqslant -1.53) = 0.9846 - 0.0630 = 0.9216.$

Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

$\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16) = \mathbf{P}(Z \leqslant 2.16) - \mathbf{P}(Z \leqslant -1.53) = 0.9846 - 0.0630 = 0.9216.$

(b) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-\infty \leqslant Z \leqslant 1.23)$ and find the area.

Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

$\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16) = \mathbf{P}(Z \leqslant 2.16) - \mathbf{P}(Z \leqslant -1.53) = 0.9846 - 0.0630 = 0.9216.$

(b) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-\infty \leqslant Z \leqslant 1.23)$ and find the area.

$\mathbf{P}(-\infty \leqslant Z \leqslant 1.23) = 0.8907.$

Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

$\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16) = \mathbf{P}(Z \leqslant 2.16) - \mathbf{P}(Z \leqslant -1.53) = 0.9846 - 0.0630 = 0.9216.$

(b) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-\infty \leqslant Z \leqslant 1.23)$ and find the area.

$\mathbf{P}(-\infty \leqslant Z \leqslant 1.23) = 0.8907.$

(c) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(1.12 \leqslant Z \leqslant \infty)$ and find the area.

Examples

(a) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16)$, and find the area.

$\mathbf{P}(-1.53 \leqslant Z \leqslant 2.16) = \mathbf{P}(Z \leqslant 2.16) - \mathbf{P}(Z \leqslant -1.53) = 0.9846 - 0.0630 = 0.9216.$

(b) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(-\infty \leqslant Z \leqslant 1.23)$ and find the area.

$\mathbf{P}(-\infty \leqslant Z \leqslant 1.23) = 0.8907.$

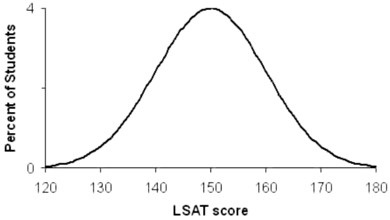
(c) Sketch the area beneath the density function of the standard normal random variable, corresponding to $\mathbf{P}(1.12 \leqslant Z \leqslant \infty)$ and find the area.

$\mathbf{P}(1.12 \leqslant Z \leqslant \infty) = 1 - \big(\mathbf{P}(Z \leqslant 1.12)\big) = 1 - 0.8686 = 0.1314.$

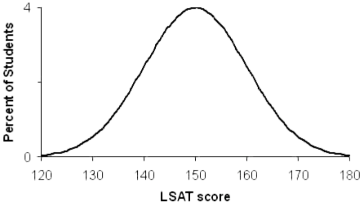
General Normal Random Variables

Recall how we used the empirical rule to solve the following problem:

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)?



General Normal Random Variables



We will now use normal distribution tables to solve this kind of problem. We do not have a table for every normal random variable (there are infinitely many of them!). So we will convert problems about general normal random to problems about the standard normal random variable, by **standardizing** — converting all relevant values of the general normal random variable to z-scores, and then calculating probabilities of these z-scores from a standard normal table (or using a calculator).

Standardizing

If X is a normal random variable with mean μ and standard deviation σ, then the random variable Z defined by

$$Z = \frac{X - \mu}{\sigma} \qquad \text{“}z\text{-score of } Z\text{”}$$

has a standard normal distribution. The value of Z gives the number of standard deviations between X and the mean μ (negative values are values below the mean, positive values are values above the mean).

Standardizing

To calculate $\mathbf{P}(a \leqslant X \leqslant b)$, where X is a normal random variable with mean μ and standard deviation σ:

Standardizing

To calculate $\mathbf{P}(a \leqslant X \leqslant b)$, where X is a normal random variable with mean μ and standard deviation σ:

- Calculate the z-scores for a and b, namely $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$

Standardizing

To calculate $\mathbf{P}(a \leqslant X \leqslant b)$, where X is a normal random variable with mean μ and standard deviation σ:

- Calculate the z-scores for a and b, namely $(a-\mu)/\sigma$ and $(b-\mu)/\sigma$
- $$\begin{aligned}\mathbf{P}(a \leqslant X \leqslant b) &= \mathbf{P}\Big(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\Big) \\ &= P\Big(\frac{a-\mu}{\sigma} \leqslant Z \leqslant \frac{b-\mu}{\sigma}\Big)\end{aligned}$$

  where Z is a standard normal random variable.

Standardizing

To calculate $\mathbf{P}(a \leqslant X \leqslant b)$, where X is a normal random variable with mean μ and standard deviation σ:

- Calculate the z-scores for a and b, namely $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$
- $$\begin{aligned}\mathbf{P}(a \leqslant X \leqslant b) =& \mathbf{P}\Big(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\Big)\\ &= P\Big(\frac{a-\mu}{\sigma} \leqslant Z \leqslant \frac{b-\mu}{\sigma}\Big)\end{aligned}$$

  where Z is a standard normal random variable.
- If $a = -\infty$, then $\frac{a-\mu}{\sigma} = -\infty$ and similarly if $b = \infty$, then $\frac{b-\mu}{\sigma} = \infty$.

Standardizing

To calculate $\mathbf{P}(a \leqslant X \leqslant b)$, where X is a normal random variable with mean μ and standard deviation σ:

- Calculate the z-scores for a and b, namely $(a-\mu)/\sigma$ and $(b-\mu)/\sigma$
- $$\begin{aligned}\mathbf{P}(a \leqslant X \leqslant b) =& \mathbf{P}\left(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\right) \\ =& P\left(\frac{a-\mu}{\sigma} \leqslant Z \leqslant \frac{b-\mu}{\sigma}\right)\end{aligned}$$

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

Examples

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?

Examples

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?

$$\mathbf{P}(4.5 \leqslant X \leqslant 7.5) = \mathbf{P}\left(\frac{4.5-6}{1.5} \leqslant Z \leqslant \frac{7.5-6}{1.5}\right) =$$

$\mathbf{P}(-1 \leqslant z \leqslant 1) = 0.6827$ or about 68%.

Examples

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

Examples

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

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

Examples

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

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

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

Examples

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

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

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

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

Examples

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

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

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

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

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

Examples

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; $\mathbf{P}(80 \leqslant X \leqslant 105)$?

Examples

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; $\mathbf{P}(80 \leqslant X \leqslant 105)$?

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

Examples

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; $\mathbf{P}(80 \leqslant X \leqslant 105)$?

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

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?

Examples

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; $\mathbf{P}(80 \leqslant X \leqslant 105)$?

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

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?

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

Examples

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

Examples

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

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

Examples

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

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

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

Examples

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

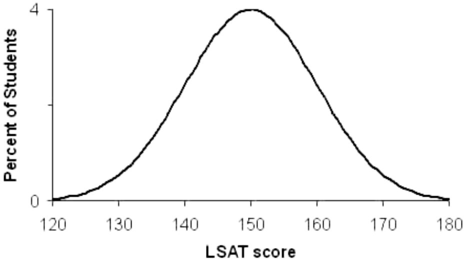
$\mathbf{P}(10 \leqslant X < \infty) = \mathbf{P}\left(\frac{10-11}{3.5} \leqslant Z < \infty\right) = \mathbf{P}(-0.2857 \leqslant Z < \infty) = 1 - (0.3875) = 0.6125.$

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

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

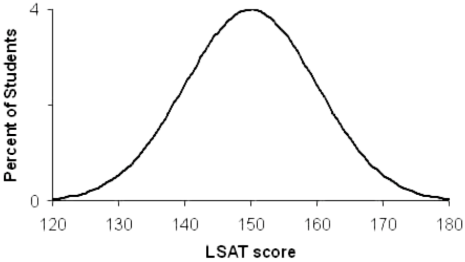
Examples

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



Examples

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



Although, technically, the variable X is not continuous, the histogram is very closely approximated by a normal curve and the probabilities can be calculated from it.

Examples

What percentage of students had a score of 165 or higher on this LSAT exam?

Examples

What percentage of students had a score of 165 or higher on this LSAT exam?

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

Examples

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.

Examples

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?

Examples

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?

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

Examples

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?

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

(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?

Examples

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?

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

(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?

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

Examples

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?

Examples

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 $\mathbf{P}(1,150 \leqslant X)$: $\mathbf{P}\left(\frac{1150-1000}{150} \leqslant X\right) = \mathbf{P}(1 \leqslant Z) =$ $1 - \mathbf{P}(Z \leqslant 1) = 1 - (0.8413) = 0.1587.$

Calculating Percentiles/Using the table in reverse

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 a function on a calculator).

Calculating Percentiles/Using the table in reverse

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

Calculating Percentiles/Using the table in reverse

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: From the table we see that 95% of the area under a standard normal curve is to the left of 1.65. Which reading x of X has z-score 1.65? Want $1.65 = (x - 140)/35$, so $x = 35 \cdot 1.65 + 400 = 457.75$. This is the 95^{th}-percentile of X; 95% of all readings of X give a value at or below 457.75.

Calculating Percentiles/Using the table in reverse

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: From the table we see that 95% of the area under a standard normal curve is to the left of 1.65. Which reading x of X has z-score 1.65? Want $1.65 = (x - 140)/35$, so $x = 35 \cdot 1.65 + 400 = 457.75$. This is the 95^{th}-percentile of X; 95% of all readings of X give a value at or below 457.75.
- 97.5^{th}-percentile: $35 \cdot 1.95 + 400 = 468.25$.

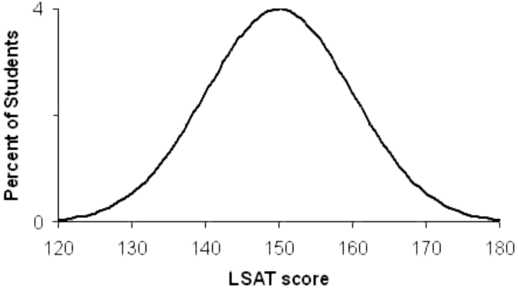
Calculating Percentiles/Using the table in reverse

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: From the table we see that 95% of the area under a standard normal curve is to the left of 1.65. Which reading x of X has z-score 1.65? Want $1.65 = (x - 140)/35$, so $x = 35 \cdot 1.65 + 400 = 457.75$. This is the 95^{th}-percentile of X; 95% of all readings of X give a value at or below 457.75.
- 97.5^{th}-percentile: $35 \cdot 1.95 + 400 = 468.25$.
- 60^{th}-percentile: $35 \cdot 0.27 + 400 = 409.45$.

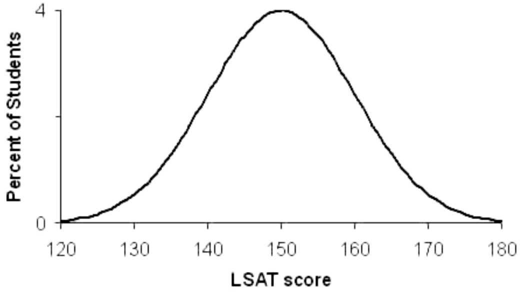
Calculating Percentiles/Using the table in reverse

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



Calculating Percentiles/Using the table in reverse

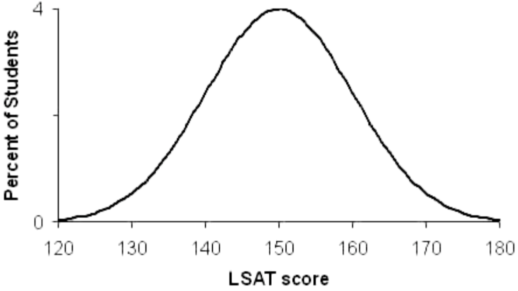
The scores on the LSAT for a particular year have a normal distribution with mean $\mu = 150$ and standard deviation $\sigma = 10$. The distribution is shown below.



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

Calculating Percentiles/Using the table in reverse

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



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

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

The table in the back of the book

In the back of the book there is a table like the one we have used. The z values run from 0 to 3.19 and look different to our values.

The table in the back of the book

In the back of the book there is a table like the one we have used. The z values run from 0 to 3.19 and look different to our values. The difference is that the function in the book is defined for positive z, and measures the area under the standard normal curve from 0 to z.

The table in the back of the book

In the back of the book there is a table like the one we have used. The z values run from 0 to 3.19 and look different to our values. The difference is that the function in the book is defined for positive z, and measures the area under the standard normal curve from 0 to z.

Let's see how the two tables are related. Let's use $B(z)$ to denote the values of the table in the book.

The table in the back of the book

In the back of the book there is a table like the one we have used. The z values run from 0 to 3.19 and look different to our values. The difference is that the function in the book is defined for positive z, and measures the area under the standard normal curve from 0 to z.

Let's see how the two tables are related. Let's use $B(z)$ to denote the values of the table in the book.

- If $0 \leqslant z < \infty$, $A(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(-\infty < Z < 0) + \mathbf{P}(0 \leq Z \leq z) = 0.5 + B(z)$
- So for $0 \leqslant z < \infty$, $A(z) = 0.5 + B(z)$

The table in the back of the book

In the back of the book there is a table like the one we have used. The z values run from 0 to 3.19 and look different to our values. The difference is that the function in the book is defined for positive z, and measures the area under the standard normal curve from 0 to z.

Let's see how the two tables are related. Let's use $B(z)$ to denote the values of the table in the book.

- If $0 \leqslant z < \infty$, $A(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(-\infty < Z < 0) + \mathbf{P}(0 \leq Z \leq z) = 0.5 + B(z)$
- So for $0 \leqslant z < \infty$, $A(z) = 0.5 + B(z)$
- If $-\infty < z < 0$, $A(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(Z \geq -z) = \mathbf{P}(0 < Z < \infty) - \mathbf{P}(0 \leq Z \leq -z) = 0.5 - B(-z)$
- So for $-\infty < z < 0$, $A(z) = 0.5 - B(-z)$

Old exam questions

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

Old exam questions

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

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

Old exam questions

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

Old exam questions

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

Old exam questions

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

Old exam questions

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

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

Old exam questions

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

Old exam questions

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

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

Approximating Binomial with Normal

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

Approximating Binomial with Normal

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

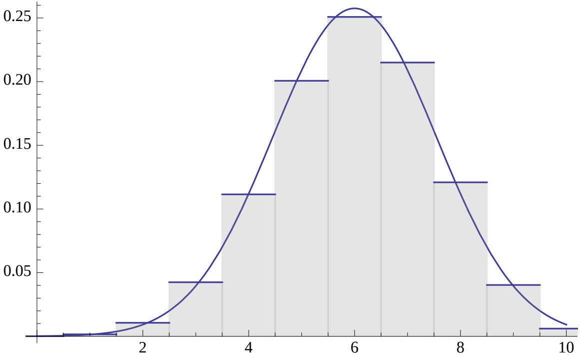
Below are histograms for a binomial random variable, with $p = 0.6$, $q = 0.4$, as the value of n (= the number of trials) varies from $n = 10$ to $n = 30$ to $n = 100$ to $n = 200$. Superimposed on each histogram is the density function for a normal random variable with mean $\mu = \mathbf{E}(X) = np$ and standard deviation $\boldsymbol{\sigma} = \boldsymbol{\sigma}(X) = \sqrt{npq}$. Even at $n = 10$, areas from the histogram are 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.

Approximating Binomial with Normal

$n = 10$: The histogram below shows the $n = 10$, $p = 0.6$ Binomial distribution histogram,

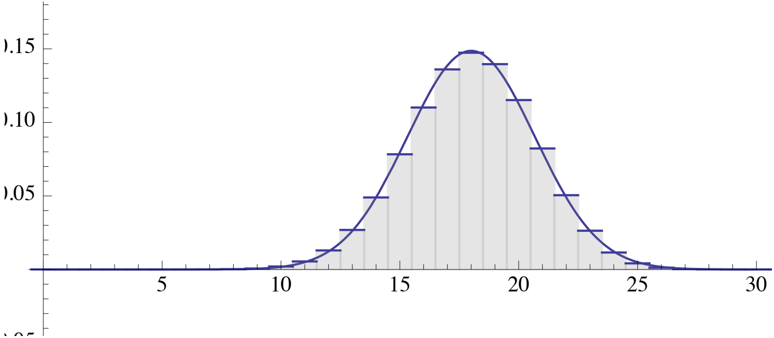
$$\mathbf{P}(X = k) = \binom{10}{k}(0.6)^k(0.4)^{10-k}$$

for $k = 0, 1, \ldots, 10$, along with a normal density curve with $\mu = 6 = np = \mathbf{E}(X)$ and $\sigma = 1.55 = \sqrt{npq} = \sigma(X)$.



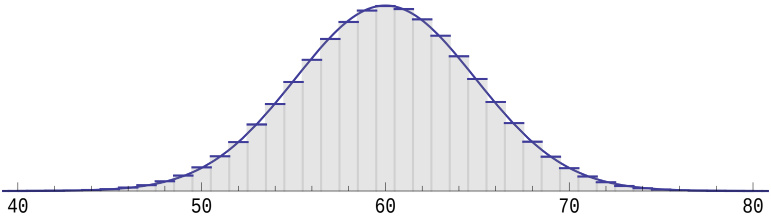
Approximating Binomial with Normal

$n = 30$: Here's the histogram of the $n = 30$, $p = 0.6$ Binomial distribution for $k = 0, 1, \ldots, 30$, along with a normal density curve with $\mu = 18 = \mathbf{E}(X)$ and $\sigma = 2.68 = \sqrt{npq} = \sigma(X)$.



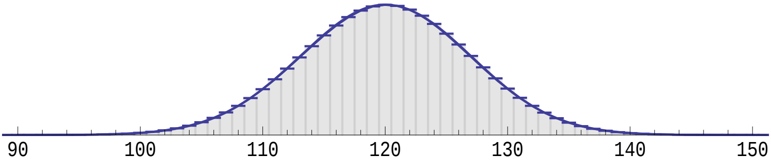
Approximating Binomial with Normal

$n = 100$: Here's the histogram of the $n = 100$, $p = 0.6$ Binomial distribution for $k = 0, 1, \ldots, 100$, along with a normal density curve with $\mu = 60 = \mathbf{E}(X)$ and $\sigma = 4.9 = \sqrt{npq} = \sigma(X)$.



Approximating Binomial with Normal

$n = 200$: Finally, here's the histogram of the $n = 200$, $p = 0.6$ Binomial distribution for $k = 0, 1, \ldots, 200$, along with a normal density curve with $\mu = 120 = \mathbf{E}(X)$ and $\sigma = 6.93 = \sqrt{npq} = \sigma(X)$.



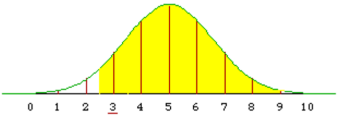
Using the approximation — continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

Using the approximation — continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

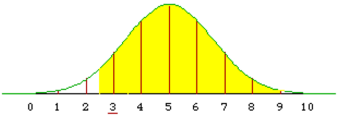
E.g., suppose $n = 10$, $p = 0.5$, and we want to know $\mathbf{P}(X \geq 3)$. It is tempting to estimate this by calculating $P(N \geq 3)$ where N is Normal, mean 5 and variance 2.5. But as the picture below shows, that will give us an answer that is too small.



Using the approximation — continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

E.g., suppose $n = 10$, $p = 0.5$, and we want to know $\mathbf{P}(X \geq 3)$. It is tempting to estimate this by calculating $P(N \geq 3)$ where N is Normal, mean 5 and variance 2.5. But as the picture below shows, that will give us an answer that is too small.

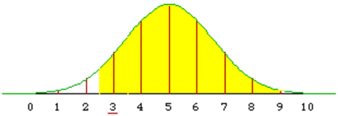


To best match up the Binomial histogram area and the Normal curve area, we should calculate $\mathbf{P}(N \geq 2.5)$. This is called the *continuity correction*.

Using the approximation — continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

E.g., suppose $n = 10$, $p = 0.5$, and we want to know $\mathbf{P}(X \geq 3)$. It is tempting to estimate this by calculating $P(N \geq 3)$ where N is Normal, mean 5 and variance 2.5. But as the picture below shows, that will give us an answer that is too small.



To best match up the Binomial histogram area and the Normal curve area, we should calculate $\mathbf{P}(N \geq 2.5)$. This is called the *continuity correction*.

$\mathbf{P}(X \geq 3) \approx .945$, $\mathbf{P}(N \geq 3) \approx .897$, $\mathbf{P}(N \geq 2.5) \approx .943$.

Continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

Continuity correction

Given a binomial distribution X with n trials, success probability p, we can approximate it using a Normal random variable N with mean np, variance $np(1-p)$.

The continuity correction tells us that when we move from X to N, we should make the following changes to the probabilities we are calculating:

- $X \geq a$ changes to $N \geq a - 0.5$
- $X > a$ changes to $N \geq a + 0.5$
- $X \leq a$ changes to $N \leq a + 0.5$
- $X < a$ changes to $N \leq a - 0.5$

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

We model the number of passengers who show up as a Binomial random variable X with $n = 205$, $p = 0.96$. We want to know that probability that $X > 200$.

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

We model the number of passengers who show up as a Binomial random variable X with $n = 205$, $p = 0.96$. We want to know that probability that $X > 200$.

We estimate X using a Normal random variable N with mean $205 \times 0.96 = 196.8$, variance $205 \times 0.96 \times 0.04 = 7.872$, standard deviation ≈ 2.8.

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

We model the number of passengers who show up as a Binomial random variable X with $n = 205$, $p = 0.96$. We want to know that probability that $X > 200$.

We estimate X using a Normal random variable N with mean $205 \times 0.96 = 196.8$, variance $205 \times 0.96 \times 0.04 = 7.872$, standard deviation ≈ 2.8.

The continuity correction says that we should estimate $\mathbf{P}(X > 200)$ by $\mathbf{P}(N \geq 200.5)$.

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

We model the number of passengers who show up as a Binomial random variable X with $n = 205$, $p = 0.96$. We want to know that probability that $X > 200$.

We estimate X using a Normal random variable N with mean $205 \times 0.96 = 196.8$, variance $205 \times 0.96 \times 0.04 = 7.872$, standard deviation ≈ 2.8.

The continuity correction says that we should estimate $\mathbf{P}(X > 200)$ by $\mathbf{P}(N \geq 200.5)$. The z-score of 200.5 is ≈ 1.32. So

$$\mathbf{P}(X > 200) \approx \mathbf{P}(Z \geq 1.32) \approx 0.09.$$

Example

An aeroplane has 200 seats. Knowing that passengers show up to flights with probability only 0.96, the airlines sells 205 seats for each flight. What is the probability that a given flight will be oversold (i.e., that more than 200 passengers will show up)?

We model the number of passengers who show up as a Binomial random variable X with $n = 205$, $p = 0.96$. We want to know that probability that $X > 200$.

We estimate X using a Normal random variable N with mean $205 \times 0.96 = 196.8$, variance $205 \times 0.96 \times 0.04 = 7.872$, standard deviation ≈ 2.8.

The continuity correction says that we should estimate $\mathbf{P}(X > 200)$ by $\mathbf{P}(N \geq 200.5)$. The z-score of 200.5 is ≈ 1.32. So

$$\mathbf{P}(X > 200) \approx \mathbf{P}(Z \geq 1.32) \approx 0.09.$$

From a Binomial calculator, the exact probability is ≈ 0.084.

Polling example I

Melinda McNulty is running for the city council this May, with one opponent, Mark Reckless. She needs to get more than 50% of the votes to win.

Polling example I

Melinda McNulty is running for the city council this May, with one opponent, Mark Reckless. She needs to get more than 50% of the votes to win.

I take a random sample of 100 people and ask them if they will vote for Melinda or not. Now assuming the population is large, the variable X = number of people who say "yes" has a distribution which is basically a binomial distribution with $n = 100$.

Polling example I

Melinda McNulty is running for the city council this May, with one opponent, Mark Reckless. She needs to get more than 50% of the votes to win.

I take a random sample of 100 people and ask them if they will vote for Melinda or not. Now assuming the population is large, the variable X = number of people who say "yes" has a distribution which is basically a binomial distribution with $n = 100$.

We do not know what p is. Suppose that in our poll, we found that 40% of the sample say that they will vote for Melinda. This is not good news, as it suggests $p \approx .4$, but this may be just due to variation in sample statistics.

Polling example I

We can use our normal approximation to the binomial to see how hopeless the situation is, by asking the question: suppose in reality 50% of the population will vote for Melinda. How likely is it that in a sample of 100 people, we find 40 or fewer people who support Melinda?

Polling example I

We can use our normal approximation to the binomial to see how hopeless the situation is, by asking the question: suppose in reality 50% of the population will vote for Melinda. How likely is it that in a sample of 100 people, we find 40 or fewer people who support Melinda?

Assuming $p = 0.5$, the distribution of X, the number of Melinda supporters we find in a sample of 100 is approximately normal with mean $\mu = np = 50$ and standard deviation $\sigma = \sqrt{npq} = \sqrt{25} = 5$.

Polling example I

We can use our normal approximation to the binomial to see how hopeless the situation is, by asking the question: suppose in reality 50% of the population will vote for Melinda. How likely is it that in a sample of 100 people, we find 40 or fewer people who support Melinda?

Assuming $p = 0.5$, the distribution of X, the number of Melinda supporters we find in a sample of 100 is approximately normal with mean $\mu = np = 50$ and standard deviation $\sigma = \sqrt{npq} = \sqrt{25} = 5$.

$$\mathbf{P}(X \leqslant 40) = \mathbf{P}\left(Z \leqslant \frac{40 - 50}{5}\right) = \mathbf{P}(Z \leqslant -2) \approx 0.0228$$

(so things don't look so good for Melinda...)

Polling example II

In a large population, some unknown proportion p of the people hold opinion o. A pollster, wanting to estimate p, polls 1000 people chosen at random, and asks each if they hold opinion o. She lets X be the number that say "yes".

Polling example II

In a large population, some unknown proportion p of the people hold opinion o. A pollster, wanting to estimate p, polls 1000 people chosen at random, and asks each if they hold opinion o. She lets X be the number that say "yes".

X is a Binomial random variable with $n = 1000$, some unknown mean $1000p$ and unknown variance $1000p(1-p)$. So it is very closely approximated by a normal random variable with mean $1000p$, variance $1000p(1-p)$.

Polling example II

In a large population, some unknown proportion p of the people hold opinion o. A pollster, wanting to estimate p, polls 1000 people chosen at random, and asks each if they hold opinion o. She lets X be the number that say "yes".

X is a Binomial random variable with $n = 1000$, some unknown mean $1000p$ and unknown variance $1000p(1-p)$. So it is very closely approximated by a normal random variable with mean $1000p$, variance $1000p(1-p)$.

Question: If the pollster uses the proportion $X/1000$ as an estimate for p, how likely is it that she gets an answer that within $\pm 3.1\%$ of the truth?

Polling example II

In a large population, some unknown proportion p of the people hold opinion o. A pollster, wanting to estimate p, polls 1000 people chosen at random, and asks each if they hold opinion o. She lets X be the number that say "yes".

X is a Binomial random variable with $n = 1000$, some unknown mean $1000p$ and unknown variance $1000p(1-p)$. So it is very closely approximated by a normal random variable with mean $1000p$, variance $1000p(1-p)$.

Question: If the pollster uses the proportion $X/1000$ as an estimate for p, how likely is it that she gets an answer that within $\pm 3.1\%$ of the truth?

I.e., what is

$$\mathbf{P}\left(-0.031 \leq \frac{X}{1000} - p \leq 0.031\right)?$$

Polling example II

$$\mathbf{P}(-0.031 \leq \frac{X}{1000} - p \leq 0.031) = \mathbf{P}(1000p - 31 \leq X \leq 1000p + 31)$$

Polling example II

$$\mathbf{P}(-0.031 \leq \frac{X}{1000} - p \leq 0.031) = \mathbf{P}(1000p - 31 \leq X \leq 1000p + 31)$$

z-score of $1000p - 31$ is $\frac{-31}{\sqrt{1000p(1-p)}} \approx \frac{-0.98}{\sqrt{p(1-p)}}$.

z-score of $1000p + 31$ is $\frac{31}{\sqrt{1000p(1-p)}} \approx \frac{0.98}{\sqrt{p(1-p)}}$.

Polling example II

$$\mathbf{P}(-0.031 \le \tfrac{X}{1000} - p \le 0.031) = \mathbf{P}(1000p - 31 \le X \le 1000p + 31)$$

z-score of $1000p - 31$ is $\frac{-31}{\sqrt{1000p(1-p)}} \approx \frac{-0.98}{\sqrt{p(1-p)}}$.

z-score of $1000p + 31$ is $\frac{31}{\sqrt{1000p(1-p)}} \approx \frac{0.98}{\sqrt{p(1-p)}}$.

So $\mathbf{P}(-0.031 \le \frac{X}{1000} - p \le 0.031) \approx \mathbf{P}(\frac{-0.98}{\sqrt{p(1-p)}} \le Z \le \frac{0.98}{\sqrt{p(1-p)}})$

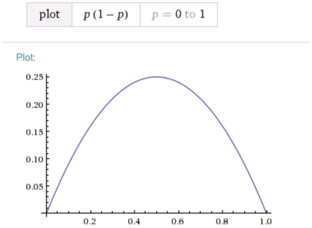
Polling example II

$$\mathbf{P}(-0.031 \leq \tfrac{X}{1000} - p \leq 0.031) = \mathbf{P}(1000p - 31 \leq X \leq 1000p + 31)$$

z-score of $1000p - 31$ is $\frac{-31}{\sqrt{1000p(1-p)}} \approx \frac{-0.98}{\sqrt{p(1-p)}}$.

z-score of $1000p + 31$ is $\frac{31}{\sqrt{1000p(1-p)}} \approx \frac{0.98}{\sqrt{p(1-p)}}$.

So $\mathbf{P}(-0.031 \leq \frac{X}{1000} - p \leq 0.031) \approx \mathbf{P}(\frac{-0.98}{\sqrt{p(1-p)}} \leq Z \leq \frac{0.98}{\sqrt{p(1-p)}})$



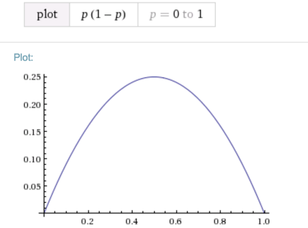
Polling example II

$$\mathbf{P}(-0.031 \le \tfrac{X}{1000} - p \le 0.031) = \mathbf{P}(1000p - 31 \le X \le 1000p + 31)$$

z-score of $1000p - 31$ is $\frac{-31}{\sqrt{1000p(1-p)}} \approx \frac{-0.98}{\sqrt{p(1-p)}}$.

z-score of $1000p + 31$ is $\frac{31}{\sqrt{1000p(1-p)}} \approx \frac{0.98}{\sqrt{p(1-p)}}$.

So $\mathbf{P}(-0.031 \le \frac{X}{1000} - p \le 0.031) \approx \mathbf{P}(\frac{-0.98}{\sqrt{p(1-p)}} \le Z \le \frac{0.98}{\sqrt{p(1-p)}})$



$\mathbf{P}(\frac{-0.98}{\sqrt{p(1-p)}} \le Z \le \frac{0.98}{\sqrt{p(1-p)}})$ is smallest when $p(1-p)$ is biggest, which is when $p = 0.5$ and $0.98/\sqrt{p(1-p)} = 1.96$

Polling example II

When it is at its smallest,

$$\mathbf{P}(-0.031 \le \frac{X}{1000} - p \le 0.031) \approx \mathbf{P}(-1.96 \le Z \le 1.96) \approx .95$$

Polling example II

When it is at its smallest,

$$\mathbf{P}(-0.031 \leq \frac{X}{1000} - p \leq 0.031) \approx \mathbf{P}(-1.96 \leq Z \leq 1.96) \approx .95$$

Conclusion: When using the results of a 1000-person opinion poll to estimate some unknown population proportion, we can be at least 95% confident that our estimate will be within ±3.1% of the true proportion, meaning that at least 95 out of every 100 (or 19 out of every 20) opinion polls conducted will result in an observed proportion that is within ±3.1% of the true proportion.

Polling example II

When it is at its smallest,

$$\mathbf{P}(-0.031 \le \frac{X}{1000} - p \le 0.031) \approx \mathbf{P}(-1.96 \le Z \le 1.96) \approx .95$$

Conclusion: When using the results of a 1000-person opinion poll to estimate some unknown population proportion, we can be at least 95% confident that our estimate will be within ±3.1% of the true proportion, meaning that at least 95 out of every 100 (or 19 out of every 20) opinion polls conducted will result in an observed proportion that is within ±3.1% of the true proportion.

- But 1 out of every 20 polls will be wrong!
- ±3.1% is called the "margin or error"
- All this assumes that the polling was done randomly
- Works regardless of the size of the population being polled