

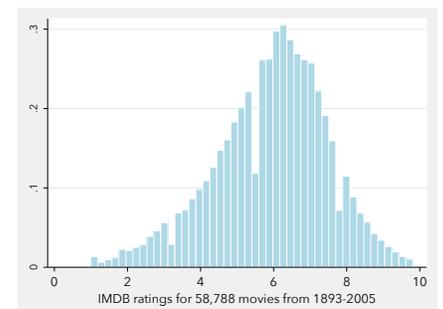
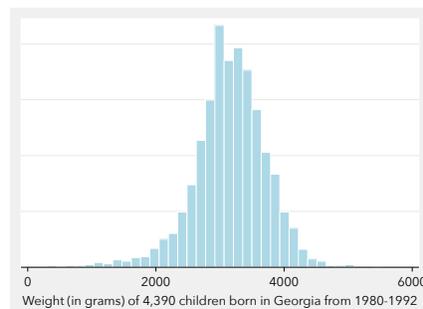
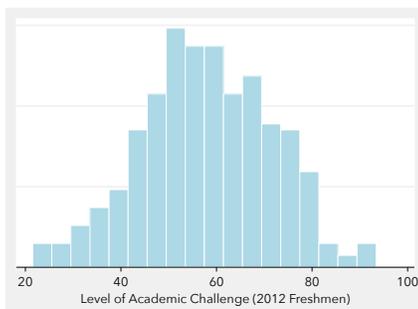
So far, we've investigated the binomial, hypergeometric, geometric, negative binomial, Poisson, uniform, exponential, and gamma distributions. In this activity, we're going to investigate the family of *normal* (or *Gaussian*) distributions.

Normal distributions can be useful in modeling measurement error or physical measurements of organisms, but they're also extremely useful for approximating other distributions (such as the binomial and Poisson distributions). Even when we have a population that is not well approximated by a normal distribution, averages of samples from the population can approximate a normal distribution (as we'll see when we learn about the Central Limit Theorem).

By the end of this activity, we'll need to be able to do the following:

1. Sketch a normal distribution with a given μ (mean) and σ (standard deviation)
2. Convert a normal distribution to a standard normal distribution (calculating a z-score)
3. Calculate probabilities and percentiles under a normal distribution with given μ and σ .

1. Identify similarities and differences among the following three histograms:



These unimodal, symmetric distributions can be approximated by **normal distributions**.

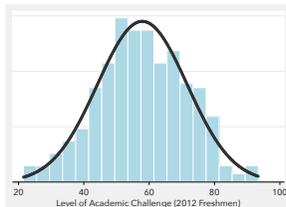
A continuous variable has a normal distribution with parameters μ and σ if its probability density function is:

$$\text{pdf: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty, \text{ where } \mu = E[X] \text{ and } \sigma = \sqrt{\text{var}(X)}$$

If variable X follows a normal distribution, we can write $X \sim N(\mu, \sigma)$

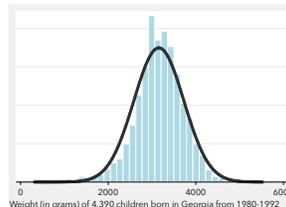
Below, I've sketched normal distributions on top of the histograms. To do this, I only needed to know μ and σ .

Notice that the normal pdf does not perfectly fit any of the datasets; it fits some better than others.



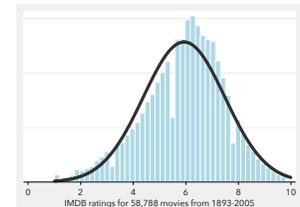
$$\mu = 57.89$$

$$\sigma = 13.75$$



$$\mu = 3156.30$$

$$\sigma = 570.44$$



$$\mu = 5.933$$

$$\sigma = 1.553$$

2. Normal distributions are theoretical, so no data will follow a normal distribution perfectly. That doesn't mean a normal distribution can't be useful, though. Identify which of the following random variables might be adequately modeled by a normal distribution. Explain.
- a) The weights of randomly selected American adults.
 - b) The income of randomly selected American families.
 - c) The shoe size of Olympic athletes.
3. All normal distributions have the same shape; they differ only in terms of where they are centered (μ) and how spread out they are (σ). When you want to sketch a normal curve, you should center it at μ and put the inflection points at $(\mu - \sigma)$ and $(\mu + \sigma)$.

IQ scores follow a normal distribution with $\mu = 100$ and $\sigma = 16$. That means IQ scores are distributed as

$$f(x) = \frac{1}{16\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{16}\right)^2}$$

To the right, sketch the distribution of IQ scores.

How could we demonstrate that the inflection points on a normal distribution are located at $\mu \pm \sigma$?

4. If we randomly choose an individual, what's the probability we choose someone with an IQ less than 90? Estimate the answer using the distribution you sketched above.

Estimated probability = _____

Because IQ scores are continuous, we need to find: $P(x < 90) = \int_0^{90} \frac{1}{16\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{16}\right)^2} dx$

That doesn't look like something I want to integrate, so I'm going to hand it off to wolframalpha.com.

Before I get the answer, let's have wolframalpha.com find the indefinite integral.



integrate (1/(16 * (2 * pi)^.5)) * e^(-.5 * ((x-100)/16)^2) with respect to x



Here's the output:

Indefinite integral:

$$\int \frac{1}{(16\sqrt{2\pi}) e^{0.5\left(\frac{x-100}{16}\right)^2}} dx = -0.5 \operatorname{erf}(4.41942 - 0.0441942x) + \text{constant}$$

erf(x) is the error function

Even if you know what the error function is*, you won't want to integrate to find normal distribution probabilities. Instead, we'll use technology.

5. Remember, we're trying to find $P(X < 90)$ given $X \sim N(100, 16)$. Choose the method you prefer...

1 StatKey: http://lock5stat.com/statkey/theoretical_distribution/theoretical_distribution.html#normal

a) Click Edit Parameters to enter in the mean and std. dev.

b) Check the appropriate box:

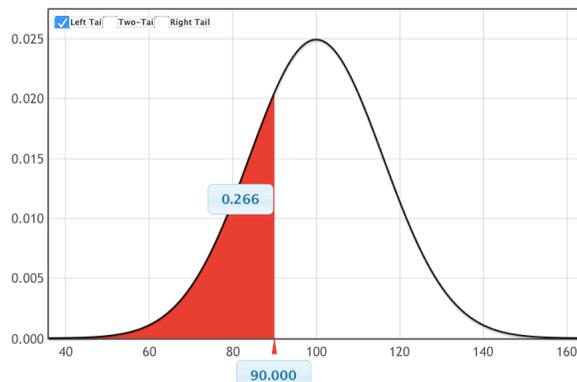
$P(X < \#)$ = left-tail

$P(\# < X < \#)$ = two-tail

$P(X > \#)$ = right-tail

c) Probability: Change the bottom cut-off number.

d) Percentile: Change the middle proportion



Use this applet to find: $P(X > 110) =$ _____ $P(90 < X < 110) =$ _____

2 Rossman/Chance: <http://www.rossmanchance.com/applets/NormCalc.html>

a) Optional: Give a name to the variable.

b) Enter the mean and std. dev; click Scale to Fit

c) Probability: Check the box; click to toggle < and >
Enter the appropriate x-value.

d) Percentile: Check the box; click to toggle < and >
Enter the appropriate probability.

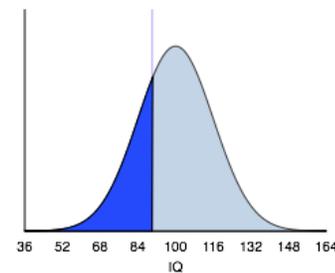
Variable:

Mean: SD:

Mean: SD:

Scale to Fit

	x	z	Probability
<input checked="" type="checkbox"/>	< 90	-0.6	0.2660
<input type="checkbox"/>	< 100	0.0	0.5000



Use this applet to find: $P(X > 110) =$ _____ $P(90 < X < 110) =$ _____

Compute error function: <http://math.stackexchange.com/questions/145087/how-to-calculate-the-integral-in-normal-distribution>
 What is the error function? http://en.wikipedia.org/wiki/Error_function

3 Wolfram Alpha: <http://www.wolframalpha.com>

P(X<90) in normal distribution with mean=100 and standard deviation=16

Input interpretation:
P(X < 90) where
 $X \approx$ normal distribution

mean	$\mu = 100$
standard deviation	$\sigma = 16$

P(A) is the probability of the event A

Result:
 $\frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{5}{8\sqrt{2}} \right) \right) \approx 0.265986$

erf(x) is the error function

Note that you could also type:

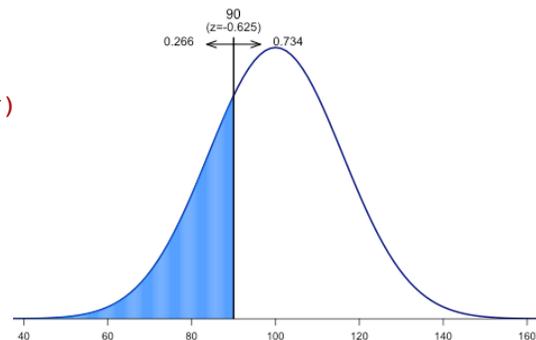
P(X<90) given X~N(100, 16^2) ← Notice you need to specify the variance; not the std. dev!

Use WA to find: $P(X > 110) =$ _____ $P(90 < X < 110) =$ _____

4 R: <http://cran.r-project.org> (installed on lab computers)

- a) Base R: $P(X < \#)$: `pnorm(##, mean = ##, sd = ##, lower.tail = TRUE)`
or `pnorm(##, mean, sd)`
- $P(X > \#)$: `pnorm(##, mean = ##, sd = ##, lower.tail = FALSE)`
or `1 - pnorm(##, mean, sd)`
- Percentile: `qnorm(p, mean = ##, sd = ##, lower.tail = FALSE)`

- b) mosaic package: `library(mosaic)`
Probability: `xpnorm(##, mean = ##, sd = ##)`
 $P(a < X < b)$: `xpnorm(c(a,b), mean = ##, sd = ##)`
- Percentile: `xqnorm(##, mean = ##, sd = ##)`



Use R to find: $P(X > 110) =$ _____ $P(90 < X < 110) =$ _____

6. Make sure you can use at least one of those methods to calculate normal distribution probabilities!

Let's look at one more method to calculate probabilities under normal distributions. It's the method I could have used as a sophomore in high school (before StatKey, the Rossman/Chance applets, WolframAlpha, or even R* were around). While I don't care that you ever use this method, it does introduce an important concept: z-scores.

As we've already discovered, all normal distributions look the same – they only have different centers (μ) and phatnesses (σ). This means we could take any X value on a normal distribution and think of it in terms of μ and σ .

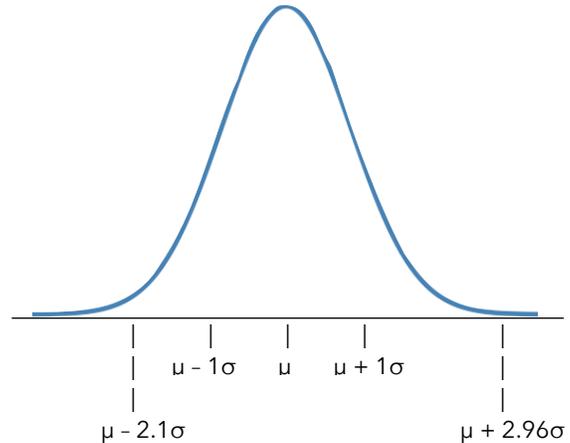
Every X-value can be represented by the number of standard deviations it is away from the mean.

Value = mean + (a number of) (standard deviations)

$$X = \mu + (z) (\sigma)$$

$$\mathbf{X = \mu + z\sigma}$$

where $z = \frac{x - \mu}{\sigma}$ = the number of standard deviation units X is away from the mean.



This means an IQ score of 116 is equivalent to a z-score of +1.

Convert these IQ scores to z-scores. Remember, $\mu = 100$ and $\sigma = 16$.

a) $X = 68$; $Z =$ _____

b) $X = 124$; $Z =$ _____

c) $X = 113$; $Z =$ _____

7. Suppose I hand back the next exam and you see your test score was $z = +2$. Should you be pleased? What does this mean? What does it mean to get a z-score of -0.3 ?

8. Suppose we convert every IQ score to a z-scores. How would the distribution change?

* While R wasn't introduced until 1993, it's an implementation of the S programming language which first appeared in 1976 (the year I was born). Coincidence? Of course.

9. If we take any normal distribution and standardize the values (convert all the X-values to z-scores), we're left with a **standard normal distribution with $\mu = 0$ and $\sigma = 1$** .

That's right... every normal distribution can be converted to a standard normal distribution with $\mu = 0$ and $\sigma = 1$.

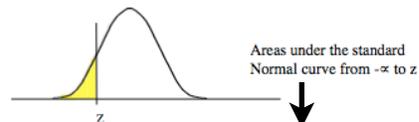
If we can estimate probabilities under a standard normal distribution, we can estimate probabilities under any normal distribution. That's the reasoning behind the standard normal table.

Standard normal table: <http://www.bradthiessen.com/html5/stats/m300/11c.pdf>

Let's try to use this table to find the probability of selecting an individual with an IQ less than 72.

- a) First, quickly sketch the IQ distribution and label the mean and standard deviation.

- b) Then, convert the IQ score of 72 into a z-score: $z = \underline{\hspace{2cm}}$



- c) Now, take a look at our standard normal table. The picture at the top indicates the numbers in the body of the table represent probabilities on the left-side of the normal distribution.

How do we find $P(z < -1.75)$?

z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
-4.00	0.00002	0.00002	0.00002	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003
-3.90	0.00003	0.00003	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00005	0.00005
-3.80	0.00005	0.00005	0.00005	0.00006	0.00006	0.00006	0.00006	0.00007	0.00007	0.00007
-3.70	0.00008	0.00008	0.00008	0.00009	0.00009	0.00009	0.00010	0.00010	0.00010	0.00011
-3.60	0.00011	0.00012	0.00012	0.00013	0.00013	0.00014	0.00014	0.00015	0.00015	0.00016
-3.50	0.00017	0.00017	0.00018	0.00019	0.00019	0.00020	0.00021	0.00022	0.00022	0.00023
-3.40	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.00030	0.00031	0.00033	0.00034
-3.30	0.00035	0.00036	0.00038	0.00039	0.00040	0.00042	0.00043	0.00045	0.00047	0.00048
-3.20	0.00050	0.00052	0.00054	0.00056	0.00058	0.00060	0.00062	0.00064	0.00066	0.00069
-3.10	0.00071	0.00074	0.00076	0.00079	0.00082	0.00085	0.00087	0.00090	0.00094	0.00097
-3.00	0.00100	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135
-2.90	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.80	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.70	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.60	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.50	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.40	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.30	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.20	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.10	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.00	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.90	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.80	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.70	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.60	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.50	0.0550	0.0561	0.0571	0.0581	0.0591	0.0601	0.0611	0.0621	0.0631	0.0641

10. If you have a TI calculator, you can also use it to get probabilities under normal distributions.

- a) To find $P(\text{IQ} < 72)$, go to the **DISTR** menu and select **normalcdf**(.
- b) You'll need to enter: **normalcdf**(left endpoint, right endpoint, mean, standard deviation). In this example:

normalcdf(-999, 72, 100, 16) = _____

11. Using whatever method you prefer, find the following probabilities. Check your answers with quick sketches!

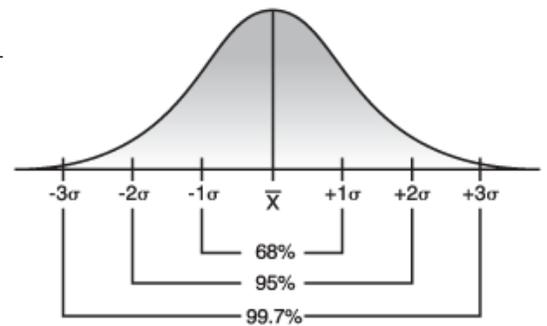
i) $P(122 < IQ < 137) =$ _____

ii) $P(84 < IQ < 116) = P(-1 < Z < 1) =$ _____

iii) $P(68 < IQ < 132) = P(-2 < Z < 2) =$ _____

iv) $P(52 < IQ < 148) = P(-3 < Z < 3) =$ _____

The previous answers should convince you of the **empirical rule**:



12. So far, we've calculated probabilities under normal distributions. We may also want to calculate percentiles (scores at which a certain proportion of values fall below).

Suppose you're elected president of Mensa, the high IQ society. You need to set an IQ cut-score for membership ("all Mensa members must have an IQ above ???"). Assuming you want **no more than the top 10%** of individuals to be Mensa members, **what IQ score should you set for your cut-score?**

Notice that in this question, you're given a probability and asked for a score.

13. Suppose you work for a university admissions office. Two students apply for the last open spot at your school:

Student A, who scored 1180 on the SAT (approximate normal distribution with $\mu = 1000$; $\sigma = 200$)

Student B, who scored 28 on the ACT (approximate normal distribution with $\mu = 20.6$; $\sigma = 6.2$)

Considering only their test scores, which applicant should you accept?
Justify your answer and defend your methodology.