

Using RStudio with the Normal Distribution

Finding proportions

We can use Rstudio to find proportions, intervals, and percentiles from any normal distribution using some simple code and rules. First, the way we ask for a proportion of area under the curve in RStudio is by typing the following into the console window:

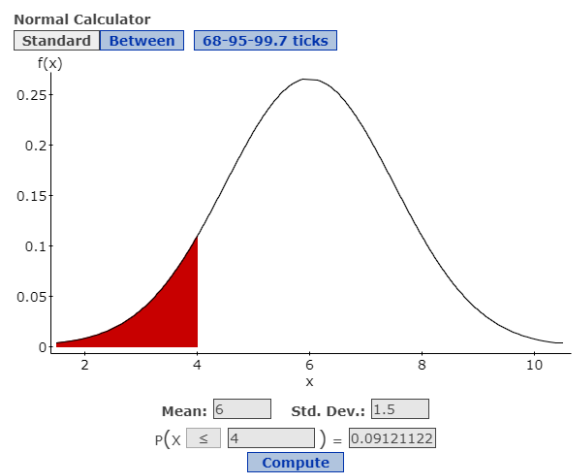
```
> pnorm(q = _____, mean = _____, sd = _____)
```

Just like when using the z-score formula, our cutoff (or observation of interest) needs to be specified as “q”. We can also put in any values for mean and standard deviation that we would like.

NOTE: RStudio will always give **cumulative proportions**. This means it will give the “less than” or left tail proportion of the curve. For example:

```
> pnorm(q = 4, mean = 6, sd = 1.5)
[1] 0.09121122
```

In StatCrunch this would look like:



What can we do if we want an upper tail proportion? Since the whole curve proportions is equal to 1, if we want $P(X \geq 4)$, we type:

```
> 1 - pnorm(q = 4, mean = 6, sd = 1.5)
[1] 0.9087888
```

You can check in StatCrunch.

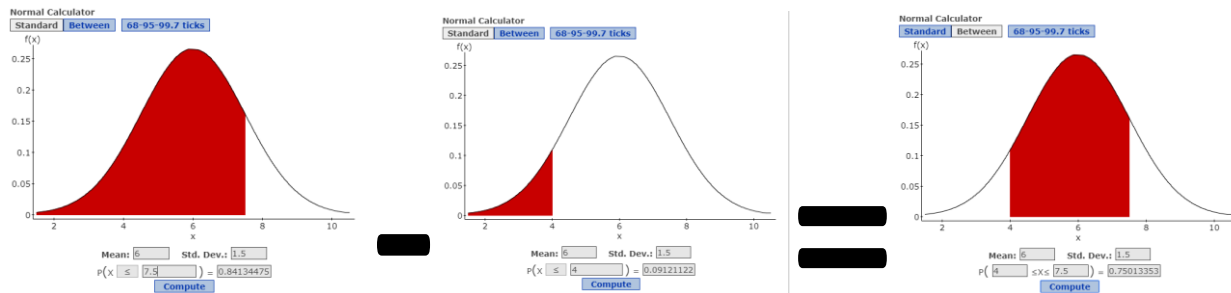
NOTE: By default, RStudio will assume the mean is 0 and the standard deviation is 1. So if we don't specify, it will use the standard normal distribution:

```
> pnorm(q = 0)
[1] 0.5
```

To find the proportion between an interval, always use proportion for the upper limit minus the proportion for the lower limit. For $P(4 \leq X \leq 7.5)$ on the $N(6, 1.5)$ distribution:

```
> pnorm(q = 7.5, mean = 6, sd = 1.5) - pnorm(q = 4, mean = 6, sd = 1.5)
[1] 0.7501335
```

Again, to visualize this in StatCrunch:



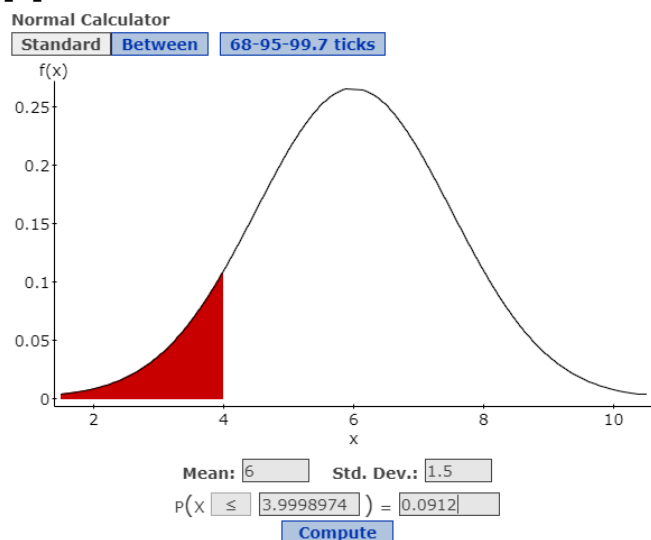
Finding cutoffs/endpoints/percentiles

When we have the proportion, but need to find the corresponding value on the number line we simply change the first letter in the RStudio code to *q* (the technical term for cutoff/endpoint is *quantile* which is where the *q* comes from):

```
> qnorm(p = _____, mean = _____, sd = _____)
```

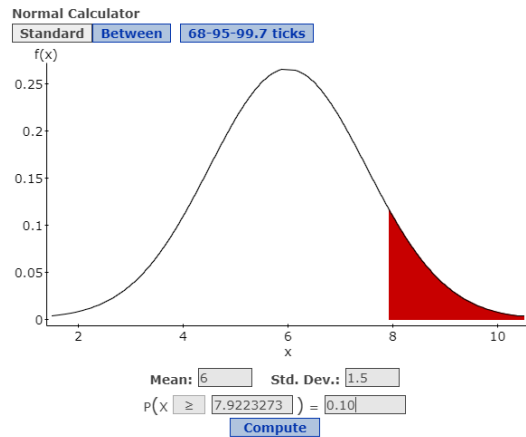
In this case we know the proportion “*p*”, but need the value. Again, in RStudio, the values calculated for the lower proportion. So, when we enter our value for “*p*”, the program will assume you mean the lower tail, or left side.

```
> qnorm(p = 0.0912, mean = 6, sd = 1.5)
[1] 3.999897
```



If you are trying to find the top percentage, or value where some % is above, think about it in the equivalent lower proportion. For example, the cutoff for the top 10% in our example above is the same as the cutoff for the lower 90%. $P(X \geq ?) = 0.10$:

```
> qnorm(p = 0.90, mean = 6, sd = 1.5)
[1] 7.922327
```



For the endpoints of an interval, consider the cutoffs for the percentage of each side. For example, what is the IQR for the normal distribution we have been using? We need Q1 and Q3, which are the 25th percentile and 75th percentile so:

```
> qnorm(p = 0.25, mean = 6, sd = 1.5)
[1] 4.988265
> qnorm(p = 0.75, mean = 6, sd = 1.5)
[1] 7.011735
```

So IQR is:

```
> 7.011735 - 4.988265
[1] 2.02347
```

To find the cutoffs for the exact middle 90%, think about what is left in each tail. Exactly 90% is in the middle, so we must have 5% in each tail. We want the 5th and 95th percentile and we would write the code:

```
> qnorm(p = 0.05, mean = 6, sd = 1.5)
[1] 3.53272
> qnorm(p = 0.95, mean = 6, sd = 1.5)
[1] 8.46728
```

