R Lab #8c - Critical values for a normal distribution

Contents

The standard normal distribution

Recall from Chapter 10 that the standard normal distribution is simply a normal distribution with a mean of 0 and a standard deviation of 1.

Also, note that the textbook describes that about 67% and 95% of the area under the curve for a normal distribution will be within the mean +/- one and two standard deviations, respectively.

Thus, for the standard normal distribution about 95% of the area under the curve will be between -2 and 2. Since normal distributions are symmetrical, this means that about 2.5% of the area under the curve will be less than -2 and about 2.5% of the area under the curve will be greater than 2.

Why do we say "about" 95% and "about" 2.5%? This is because this is an approximation. You can see this by using the pnorm() function to get the exact area under the standard normal curve that is greater than 2 (see R Lab #8b for walkthrough of pnorm function):

```r
pnorm(2, 0, 1, lower.tail = F)
## [1] 0.02275013
```

We see that the area under the curve to the right of 2 is actually 0.02275013 rather than the expected 0.025.

Critical values

Rather than starting with a value and asking for the area under the curve to the right or left of it, we can instead start with the desired area under the curve and ask for a value.

For example, what if we want the *exact* value for which 2.5% of the area under a standard normal curve lies to the right?

This value would be a "critical value" for the right tail probability of 0.025 for a standard normal distribution (or a critical value for the left tail probability of 0.975).

Calculating critical values with qnorm() function

We can calculate a critical value with the function qnorm(). It takes four arguments: the (desired) probability, mean, standard deviation, and lower.tail (T or F).

For example, if we want the critical value for a right tail probability of 0.025 on a standard normal distribution:

```
qnorm(0.025, 0, 1, lower.tail = F)
## [1] 1.959964
```

Notice that we get the same value if we ask for the critical value for a left tail probability of 0.975.

```
qnorm(0.975, 0, 1, lower.tail = T)
## [1] 1.959964
```

This is where the 1.96 value that is used for calculating a 95% confidence interval comes from.

The q in qnorm is for quantile, and you can use the function to ask for any quantile/percentile/probability. For example, what is the value for standard normal distribution for with 30% of the probability lies to the right?

```
qnorm(0.3, 0, 1, lower.tail = F)
## [1] 0.5244005
```

Using qnorm() for any normal distribution

The qnorm() function works with any normal distribution. Let's illustrate its utility with another distribution.

In the NASA height example it was stated that in the heights of males aged 20-29 years in the USA has a normal distribution with $\mu = 177.6$ cm and $\sigma = 9.7$ cm.

What height is the 90th quantile? Another way to phrase this is what value would have a 0.9 probability to the left and 0.1 probability to the right?

Using qnorm():

```
qnorm(0.9, 177.6, 9.7, lower.tail = T)
## [1] 190.0311
```

or

```
qnorm(0.1, 177.6, 9.7, lower.tail = F)
## [1] 190.0311
```

This means that a male in this population that is 190.0311 cm tall would be taller than 90% of other males in the population.

Using qnorm() for sampling distributions of the mean

Back to the heights of males aged 20-29, with a normal distribution with $\mu = 177.6$ cm and $\sigma = 9.7$ cm.

If we repeatedly and randomly sample 50 males, measure the height of each male, and calculate the mean height of the sample then we can get a sampling distribution of the mean with $n = 50$.

The mean of this sampling distribution is 177.6 cm (same as population). The standard deviation of the sampling distribution of the mean is:

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{n}}$$

```r
sdm <- 9.7/sqrt(50)
sdm
## [1] 1.371787
```

So now we could answer a question like this:

For random samples of $n = 50$ from this population, what is the mean height that is greater than the mean height of 90% of samples?

```r
qnorm(0.9, 177.6, sdm, lower.tail = T)
## [1] 179.358
```

See that if the sample size is increased to 500 then our uncertainty in the mean discreases (i.e., sampling distribution of the mean becomes more narrow), and this value decreases:

```r
sdm <- 9.7/sqrt(500)
sdm
## [1] 0.4337972
qnorm(0.9, 177.6, sdm, lower.tail = T)
## [1] 178.1559
```

Comparing critical values for male heights example

Notice how these values change in the different normal distributions.

The population has the widest distribution ($\mu = 177.6$, $\sigma = 9.7$ cm, 90% left tail value = 190.0311).

The sampling distribution of the mean has a more narrow distribution because we go from $\sigma = 9.7$ cm to $\sigma_{\text{mean}} = 9.7$ / sqrt(n). So as n increases the standard deviation of the sampling distribution of the mean decreases, as does the 90% left tail value. At $n = 50$ the value is 179.358 and at $n = 500$ the value is 178.1559.

See figures 10.5-1 and 10.5-2 for a visual of this.

R commands summary

- **Critical value for normal distribution (q value for a particular lower (TRUE) or upper tail (FALSE))**
  - qnorm(p,mean,sd,lower.tail)