Econ 5023: Statistics for Decision Making

Univariate Statistics (XII): What is a Large/Unlikely Value in Statistics?

Le Wang

December 3, 2018

Question: What are large values?

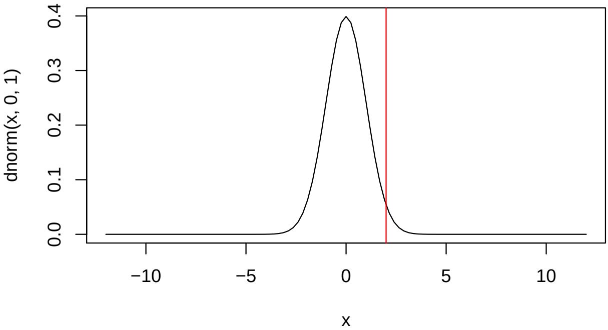
For example, is 2 a large value?

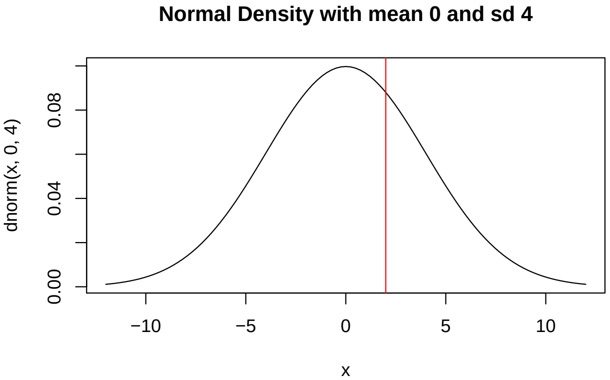
Question: What are unlikely values?

For example, is 2 an unlikely value?

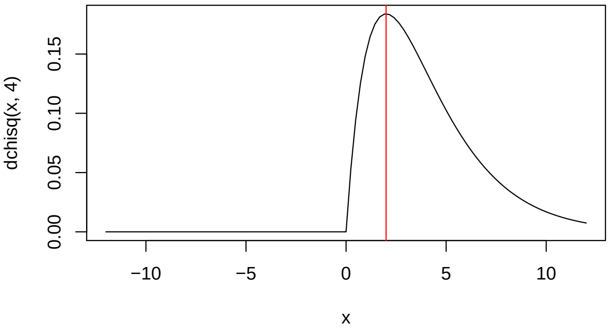
The answer depends on the distribution from which a number is drawn.

Normal Density with mean 0 and sd 1

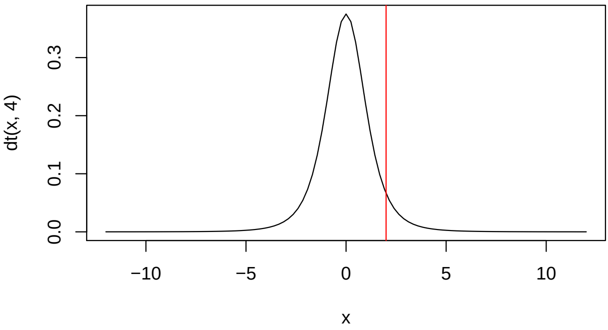



Normal Density with mean 0 and sd 4
dnorm(x, 0, 4)
0.00
0.04
0.08
−10
−5
0
5
10
x

Chi−square Density with 4 degrees of freedom



t Density with 4 degrees of freedom



Question: How do we quantify the “largeness” or “unlikeliness” of a value?

For example, is 2 a large or unlikely value?

Two ways

1. Compare to the pre-defined "large" or "unlikely" value **from the underlying distribution**
2. Calculate how many values **from the underlying distribution** are larger than this value.

In order to know whether a number is large or unlikely, we need to know about two things

1. What is the underlying distribution (or benchmark)?
2. How do we define large? How large is large?
   2.1 In terms of magnitude, or
   2.2 Large positive or Negative values?

What do you need to know about the underlying distribution

1. Normal: $N(\mu, \sigma^2)$ (mean and variance)
2. χ^2_k: degrees of freedom
3. t_k: degrees of freedom
4. F_{k_1,k_2}: **TWO** degrees of freedom

Remember that every distribution is associated with different probabilities assigned to values.

Large Values $\iff$ **Unlikely Values**

Approach 1:

Compare to the pre-defined “large” value **from the underlying distribution**

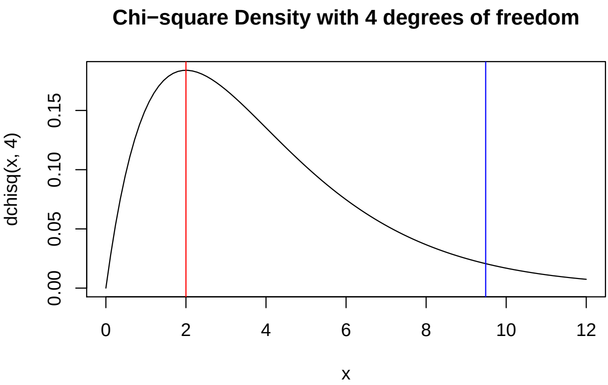
Let's look at two "easier" distributions first.

1. χ^2_k
2. F_{k_1,k_2}

They are easy because all the values are **positive**. The definition of "large" is more straightforward.

Let's assume that

1. The random variable is drawn from a χ^2_4 distribution. The value is 2.
2. We pre-define a "large" value as the value greater than top 5 percent (95^{th} percentile) of the distribution.



$\Longrightarrow$ 2 is a small number!

```
qchisq(.95, 4)

## [1] 9.487729
```

Exercise: Now, let's assume that

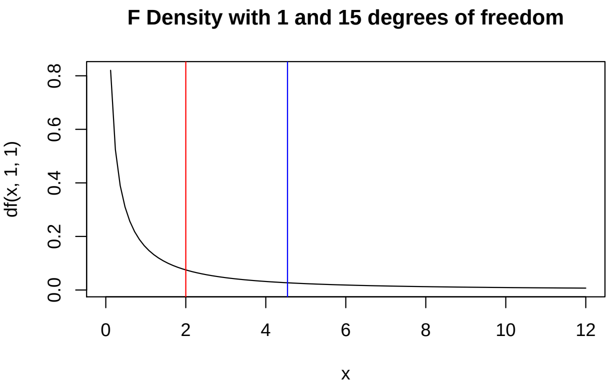
1. The random variable is drawn from a χ^2_4 distribution. The value is 2.
2. We pre-define a "large" value as the value among the upper half of the distribution (top 50%) [i.e., greater than 50^{th} percentile of the distribution].

Exercise: Now, let's assume that

1. The random variable is drawn from a χ_4^2 distribution. The value is 10.
2. We pre-define a "large" value as the top 1% value, i.e., the value greater than 99^{th} percentile of the distribution.

Let's assume that

1. The random variable is drawn from a $F_{1,15}$ distribution. The value is still 2.
2. We pre-define a "large" value as the top 5% value, i.e., the value greater than 95^{th} percentile of the distribution.



$\Longrightarrow$ 2 is a small number!

```r
qf(.95, 1,15)

## [1] 4.543077
```

Exercise: Now, let's assume that

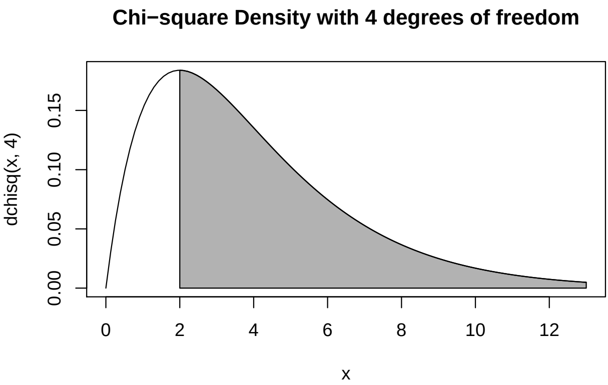
1. The random variable is drawn from a $F_{1,15}$ distribution. The value is 5.
2. We pre-define a "large" value as the top 1% value, i.e., the value greater than 99^{th} percentile of the distribution.

Approach 2:

Calculate the percentage of values **from the underlying distribution** that are greater than this particular value

Let's assume that

1. The random variable is drawn from a χ^2_4 distribution. The value is 2.



How to calculate this area

```
1-pchisq(2, 4)

## [1] 0.7357589
```

```r
1-pchisq(2, 4)

## [1] 0.7357589
```

How to interpret this? There are 73.57589 percent of the values greater than 2. Is it large?

Suppose that the large value is defined as those greater than 95 percent of the values, or among the top 5% of the distribution.

Answer: It is not!

Now, let's look at both normal and t distributions. They are trickier because they cover both negative or positive values.

Then, the question is: do we care about the direction of the magnitude?

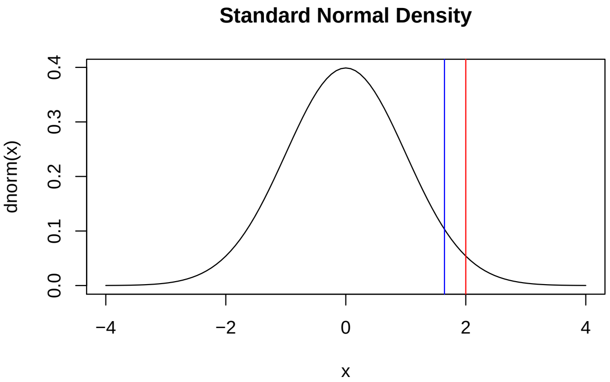
Let's say, we care about the "largeness" in terms of both magnitude and direction.

Approach 1:

Question: Is this a large, positive value?

Let's assume that

1. The random variable is drawn from a $N(0,1)$ distribution. The value is still 2.
2. We pre-define a "large", positive value as the top 5% the value greater than 95^{th} percentile of the distribution.



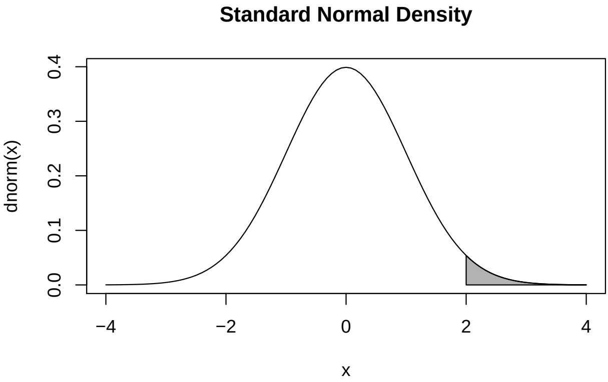
$\Longrightarrow$ 2 is a large, positive number!

```
qnorm(.95)

## [1] 1.644854
```

Exercise: Now, let's assume that

1. The random variable is drawn from a $N(0,1)$ distribution. The value is 2.
2. We pre-define a "large", positive value as the top 1% value, i.e., the value greater than 99^{th} percentile of the distribution.

We can also use the second approach, similar to above



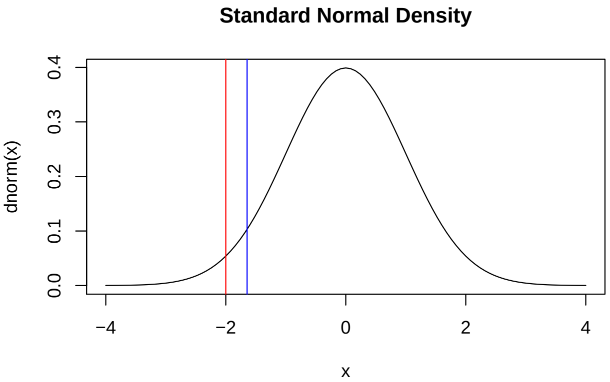
How to calculate this area

```
pnorm(-2)

## [1] 0.02275013

1-pnorm(2)

## [1] 0.02275013
```

Question: Is this a <u>large</u>, <u>negative</u> value?

Let's assume that

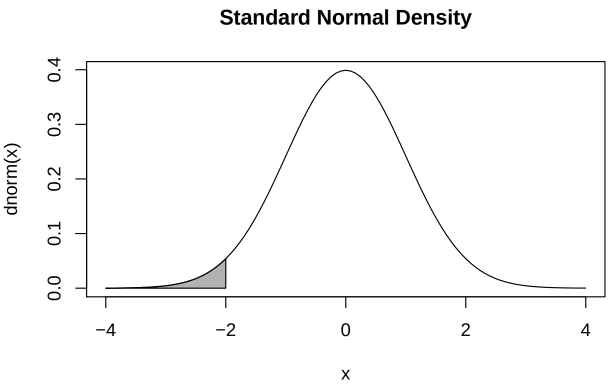
1. The random variable is drawn from a $N(0,1)$ distribution. The value is now -2.
2. We pre-define a "large" negative value as the **bottom** 5%, i.e., the value **smaller** than 5^{th} percentile of the distribution.



$\Longrightarrow$ -2 is a large, negative number!

```
qnorm(.05)

## [1] -1.644854
```

We can also use the second approach, similar to above



How to calculate this area

```
pnorm(-2)

## [1] 0.02275013
```

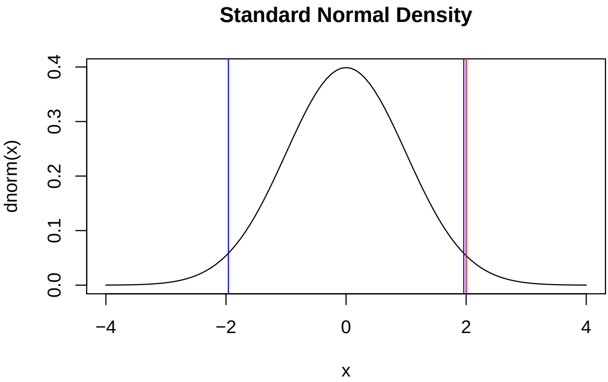
Tricky case: we care ONLY about the magnitude of the value!

What is the large number?

Both large, positive and large, negative values are considered as large!

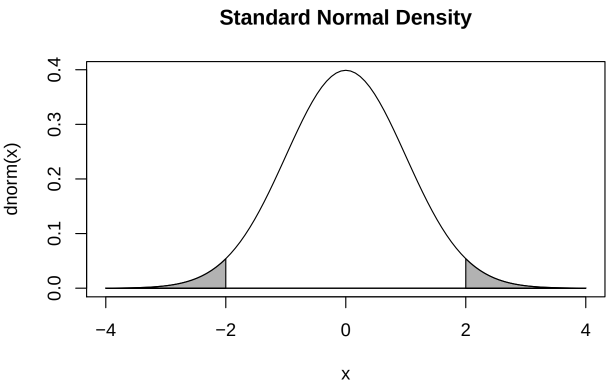
Top 5 percent large number is split between

1. top 2.5 percent positive numbers
2. and top 2.5 percent negative number



$\Longrightarrow$ 2 is a large number!

```
qnorm(.975)

## [1] 1.959964

qnorm(.025)

## [1] -1.959964
```



How to calculate this area

```
pnorm(-2)*2
```

```
## [1] 0.04550026
```

Notice that it is the same with a random variable distributed from a t-distribution with k degrees of freedom since it is also symmetric.

Even if a distribution is not symmetric, it is very straightforward to follow the same procedure, other than the way to calculate the percentage greater than a particular value when we consider only magnitudes. For the latter, we cannot simply just multiply the percentage smaller than the bottom, say, 2.5%, by 2. We will leave this for you to figure out on your own.