Basics of inference

E. Pastucha

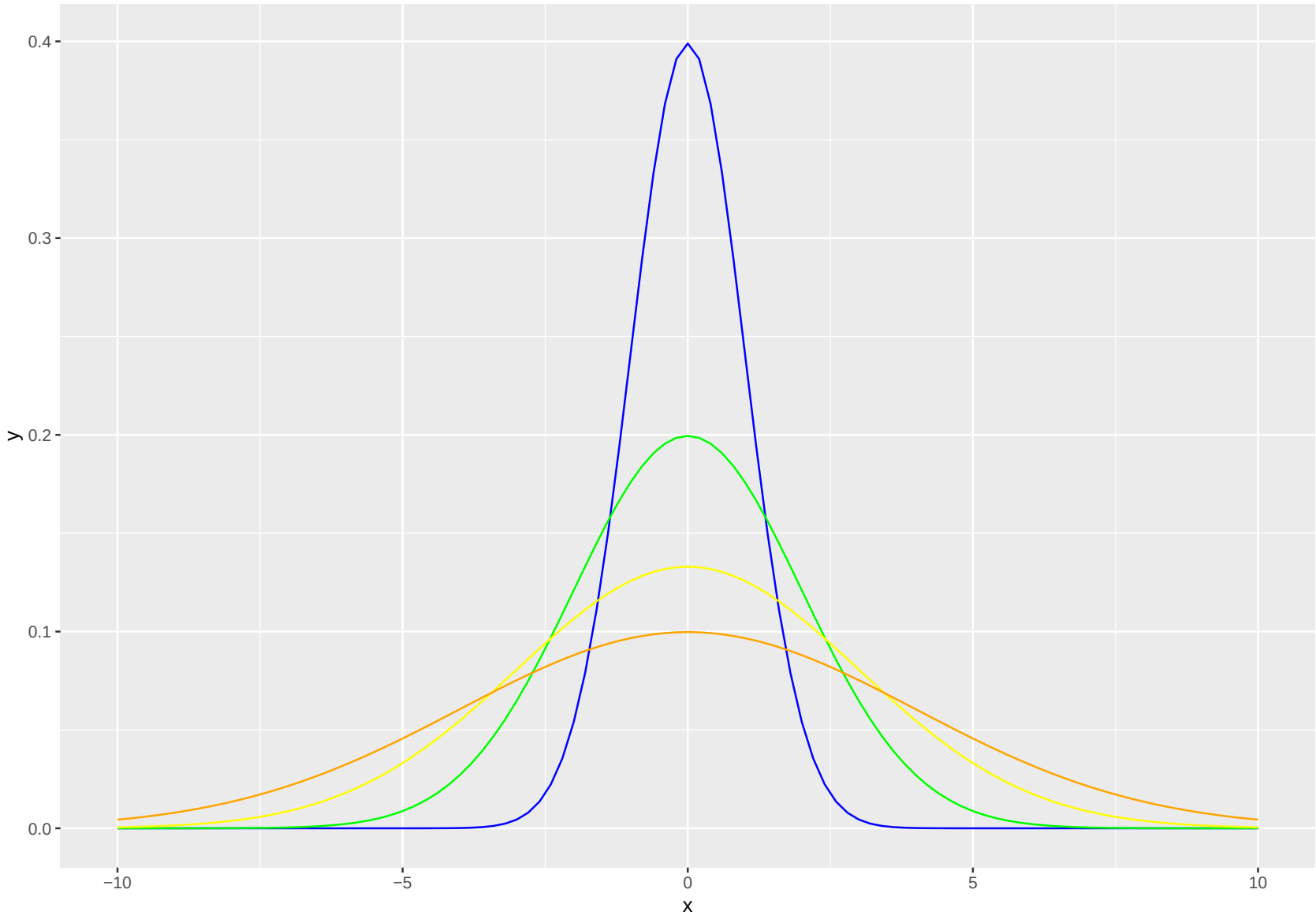
September 2024

Distributions

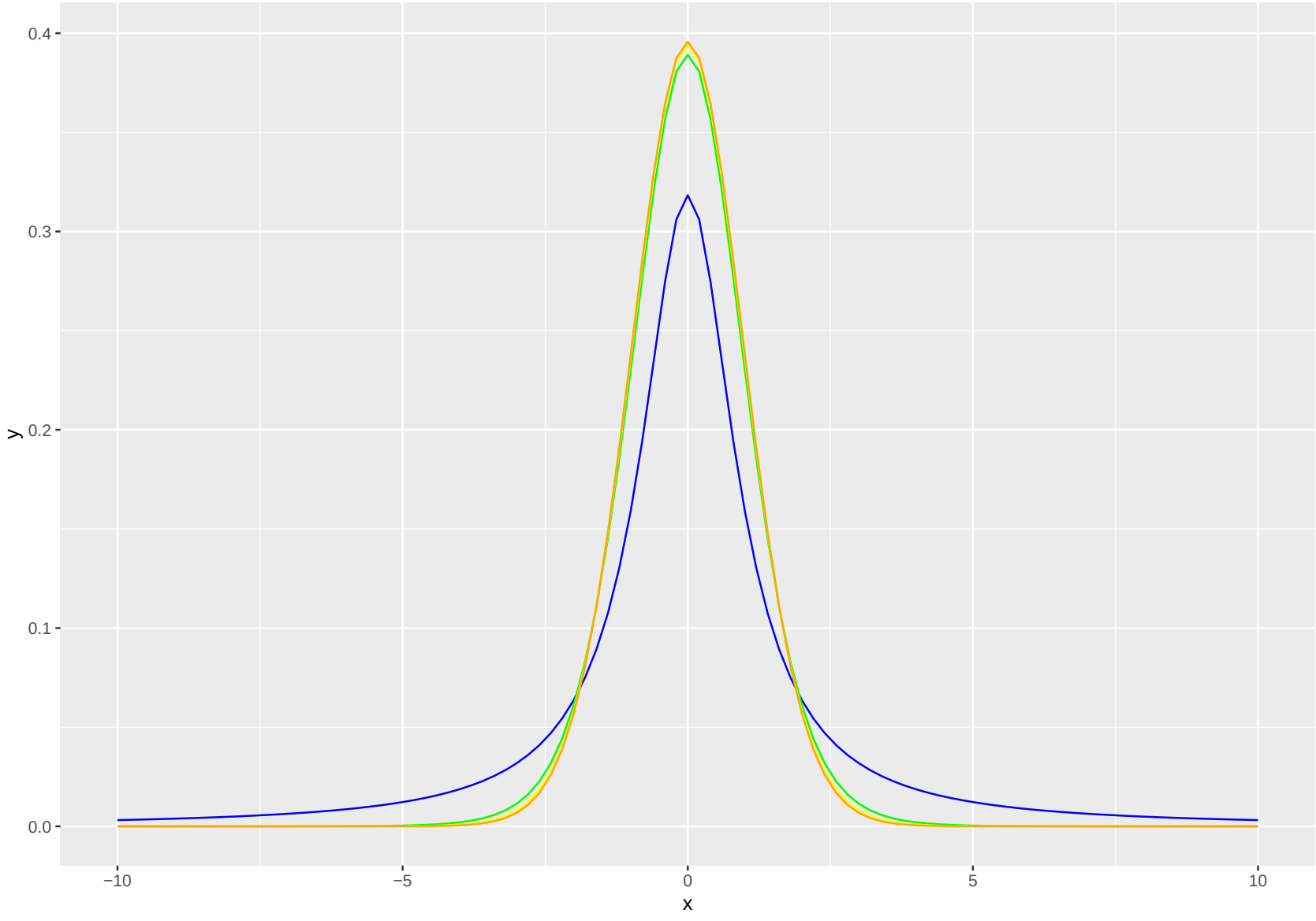
What is a distribution?

What does it describe?

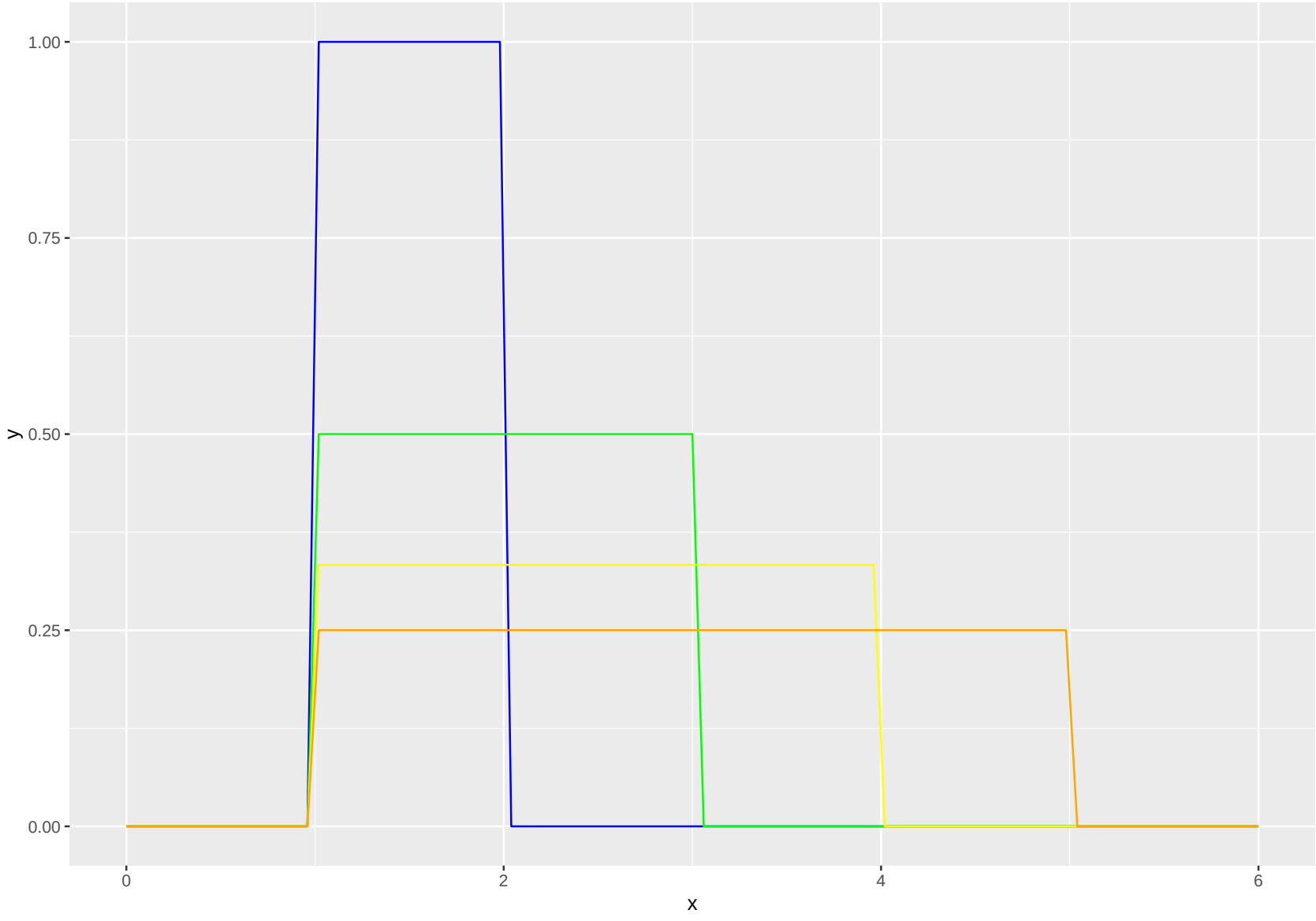
Normal Distribution



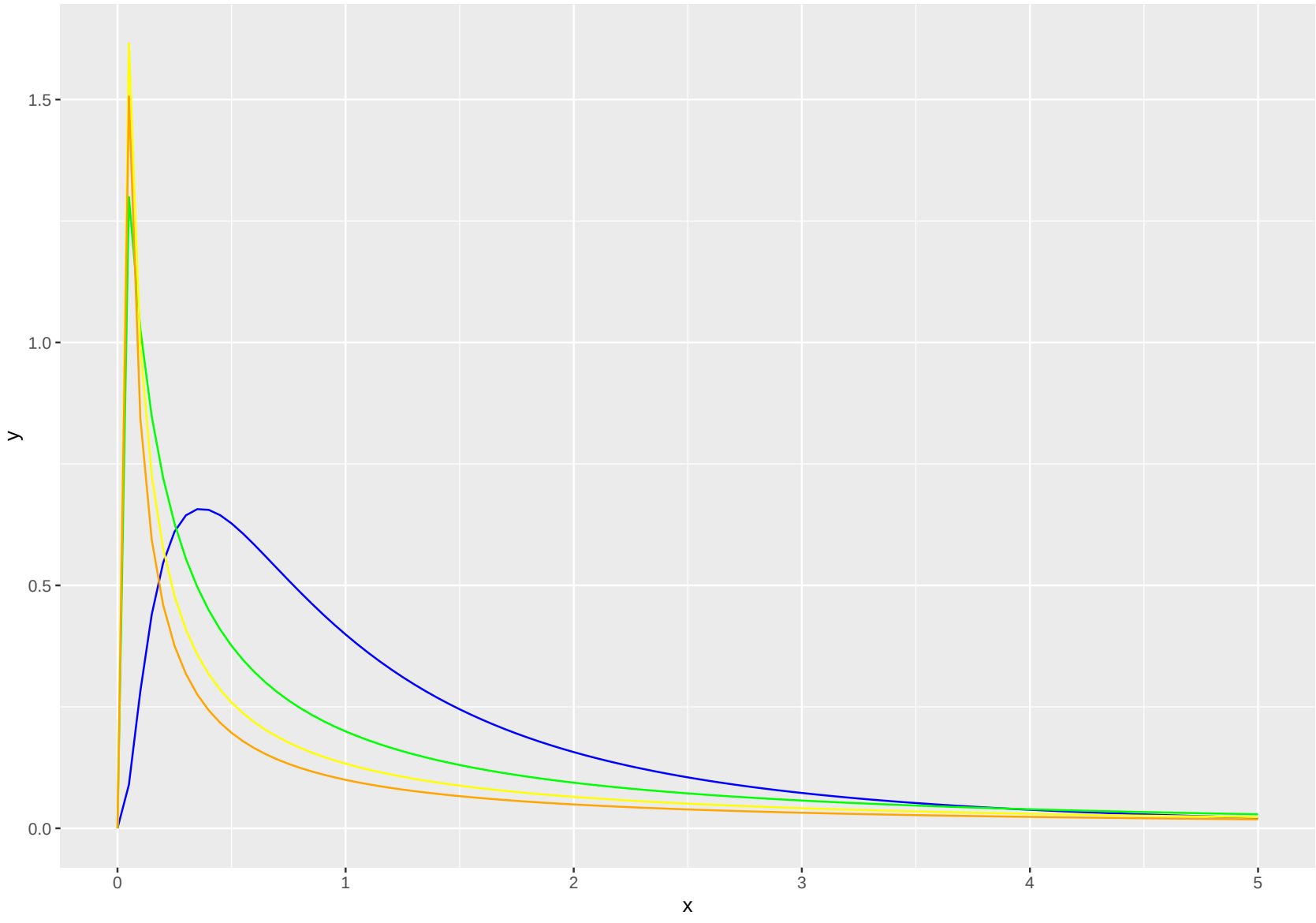
Student t-distribution



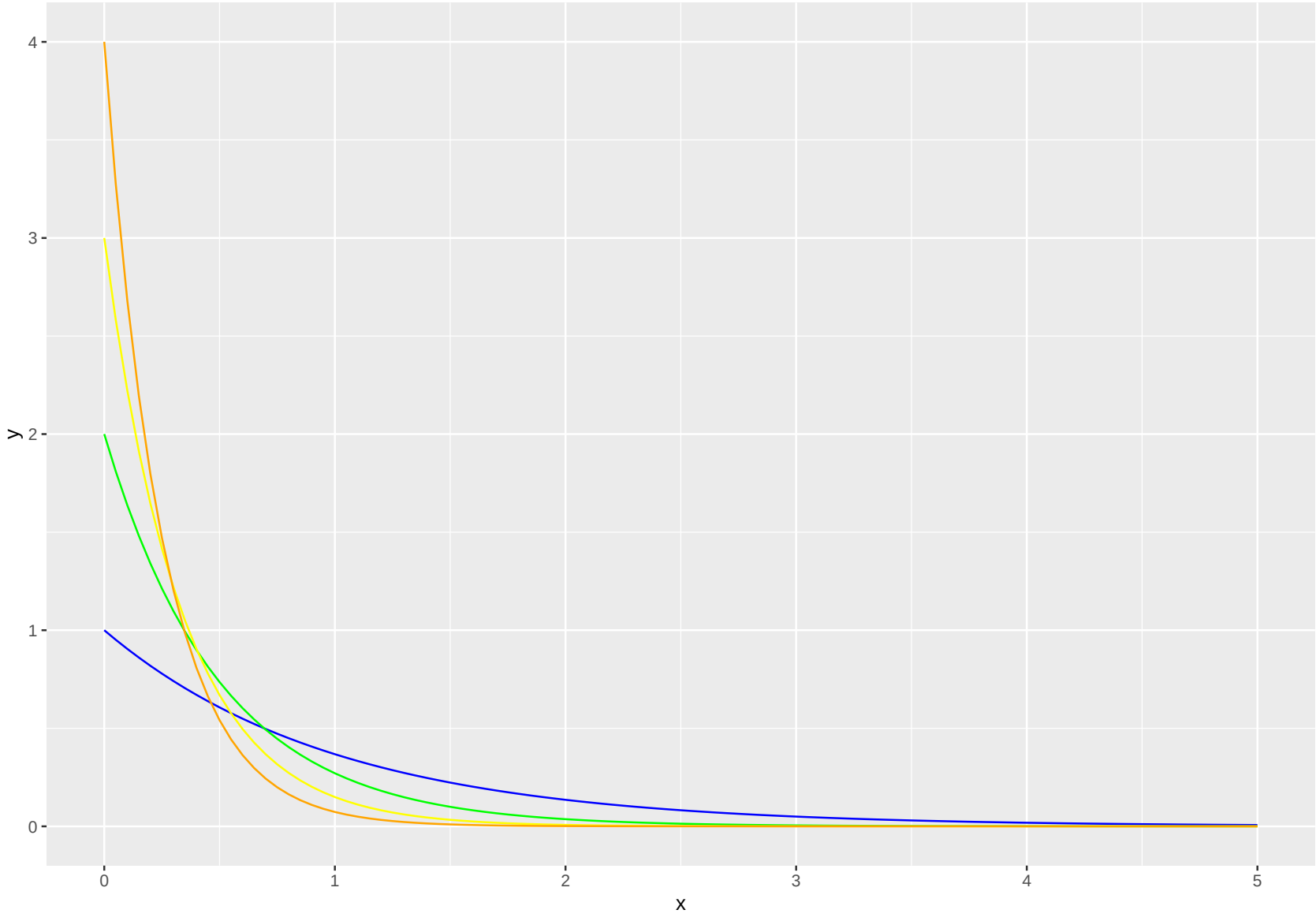
Uniform distribution



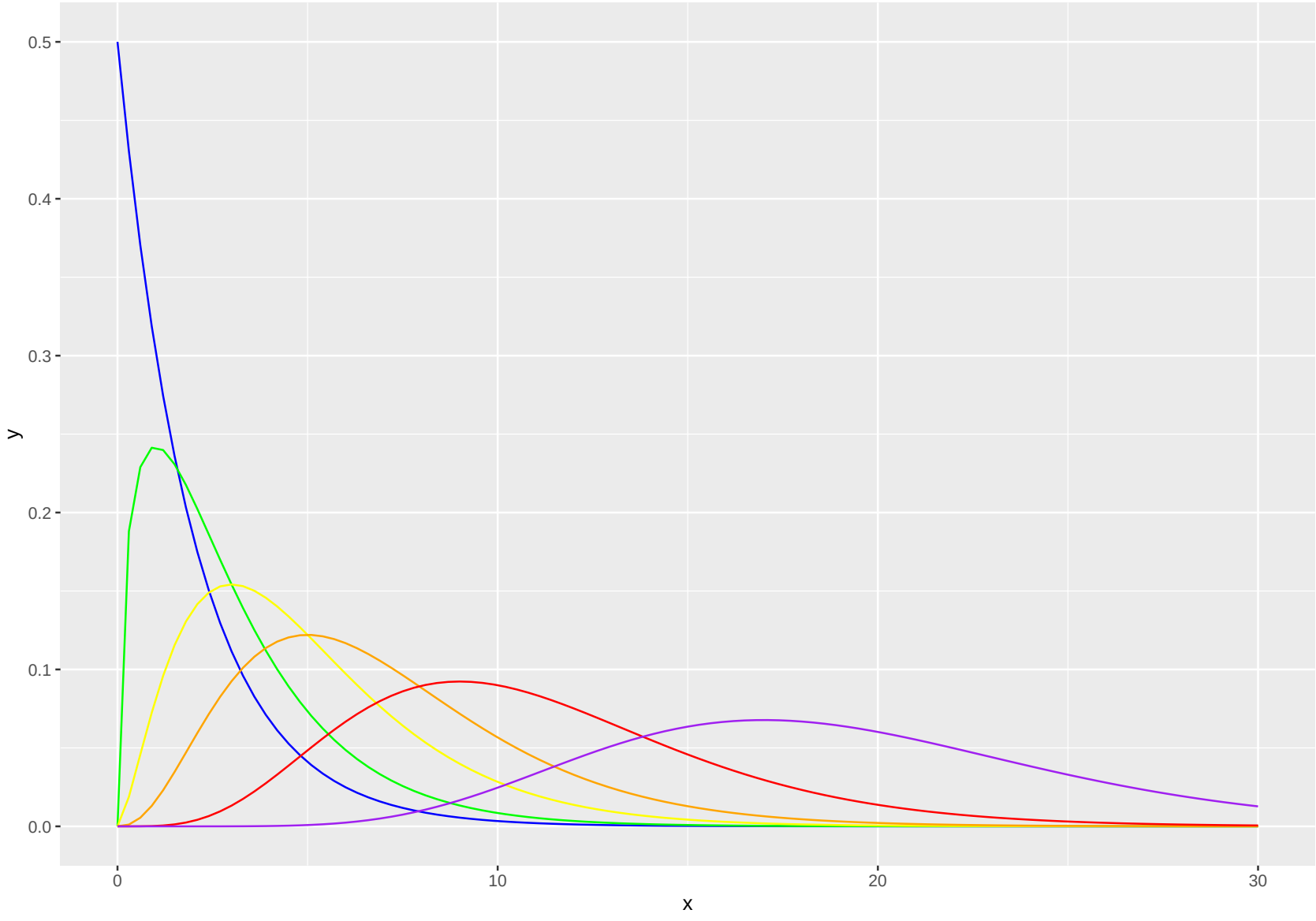
log-normal distribution



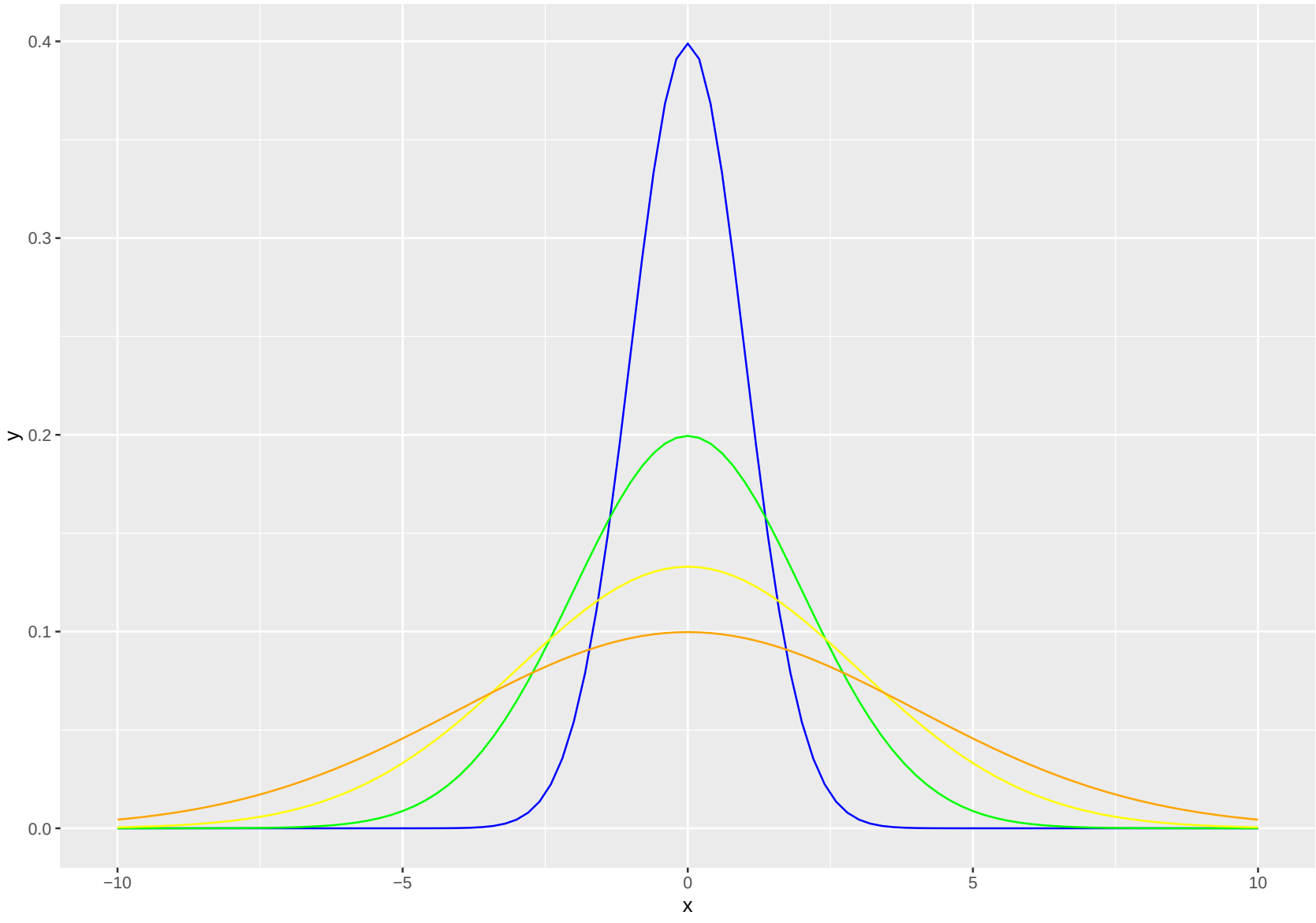
Exponential distribution



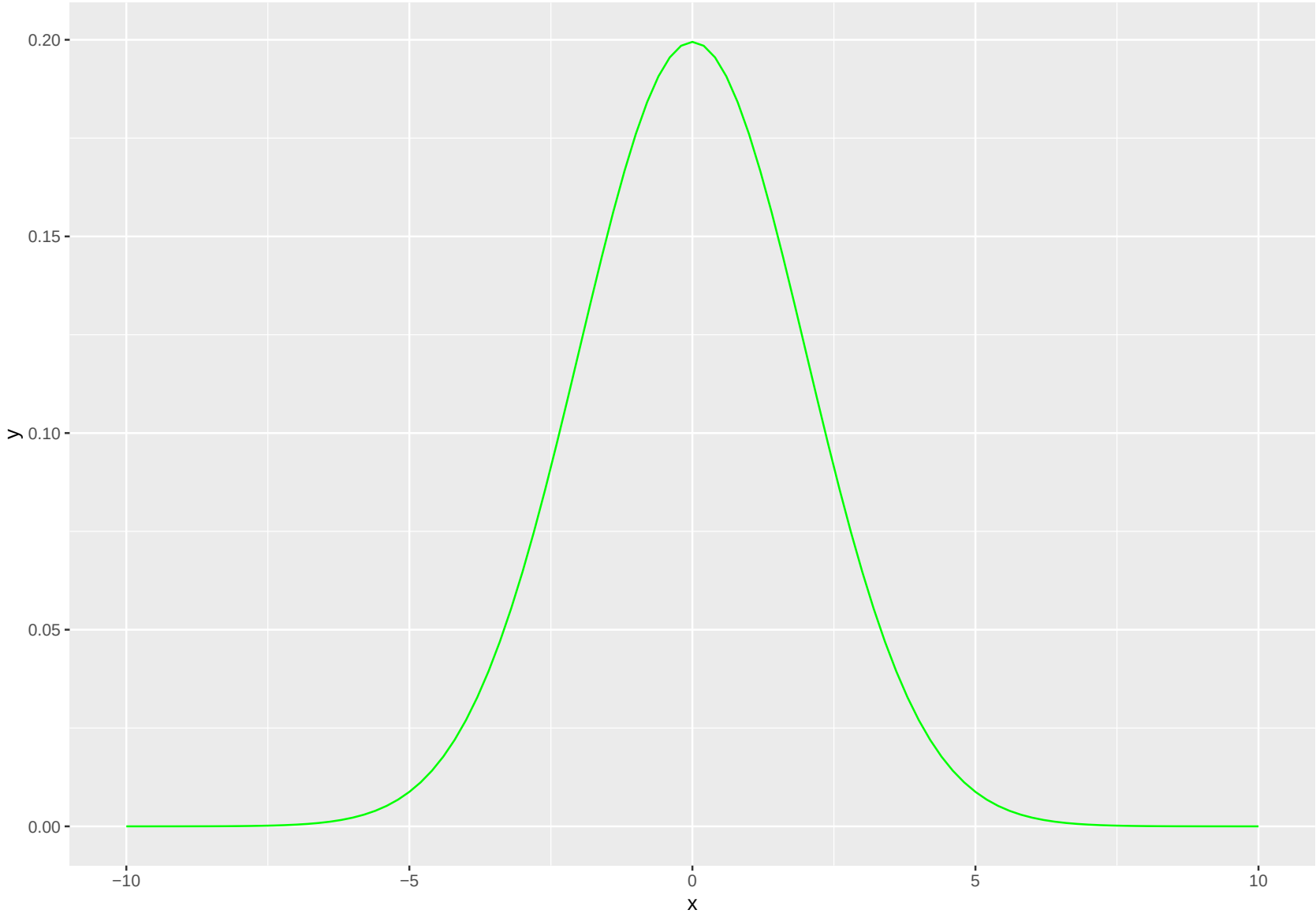
χ^2 distribution



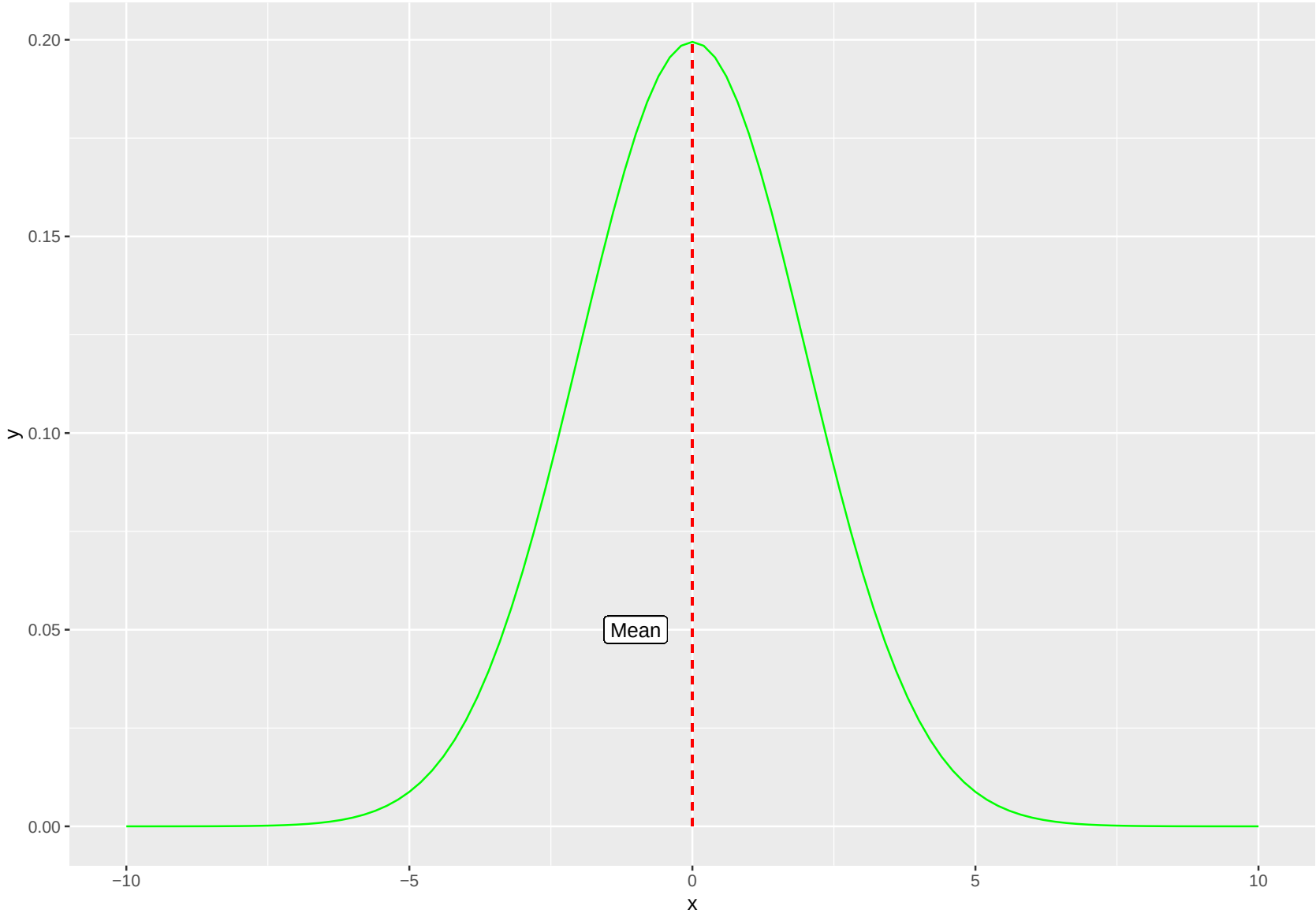
Normal Distribution



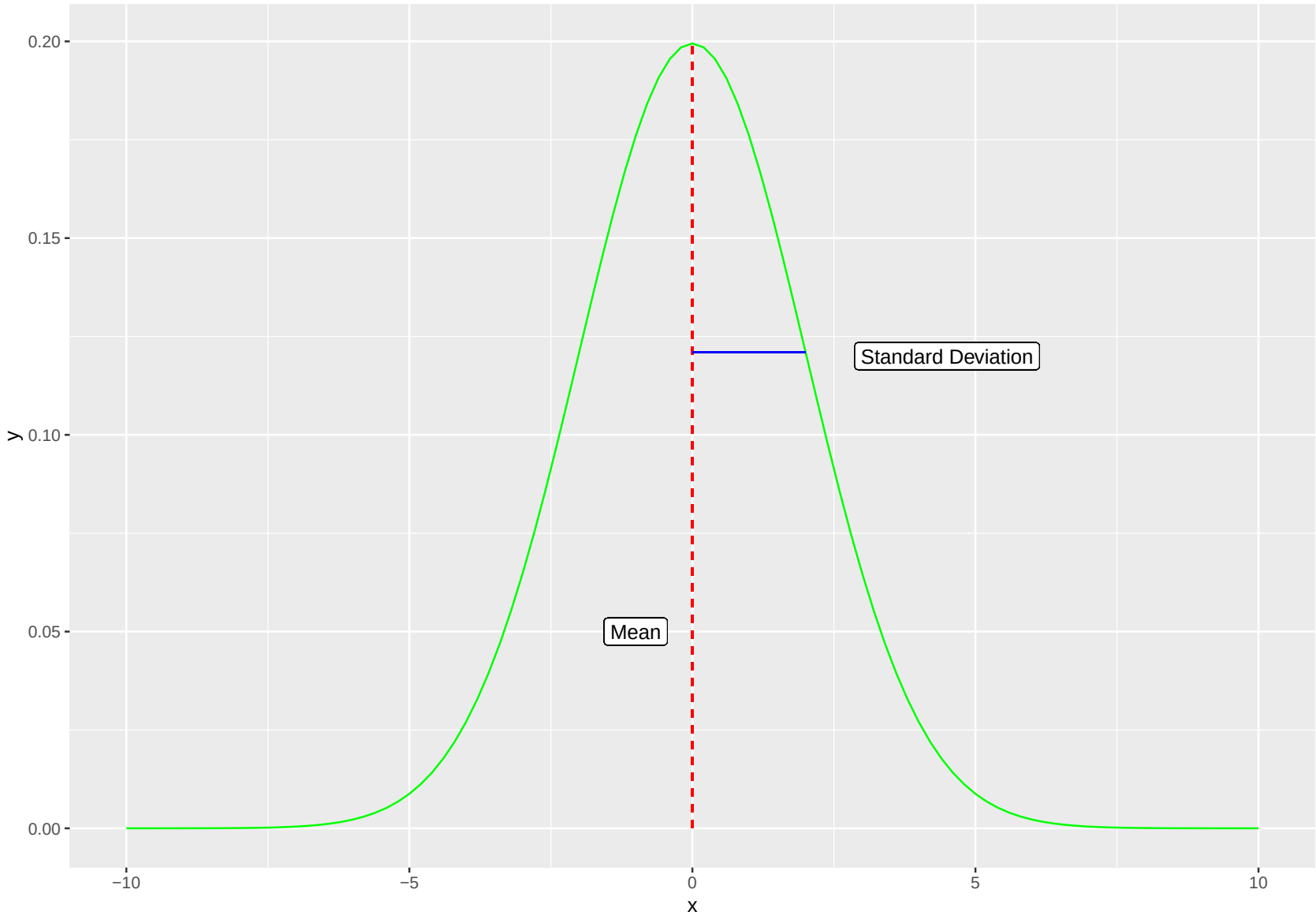
Normal Distribution



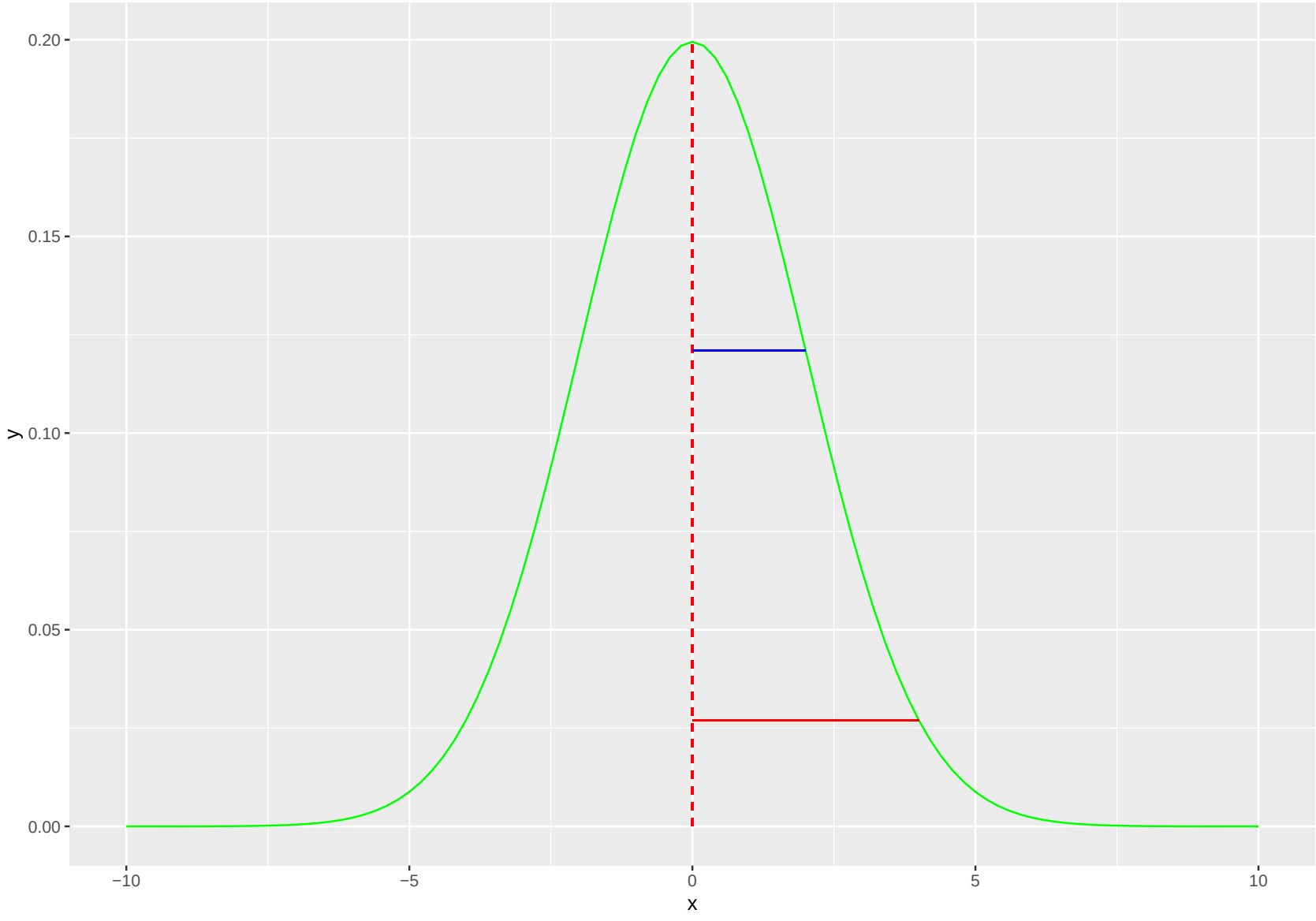
Normal Distribution



Normal Distribution



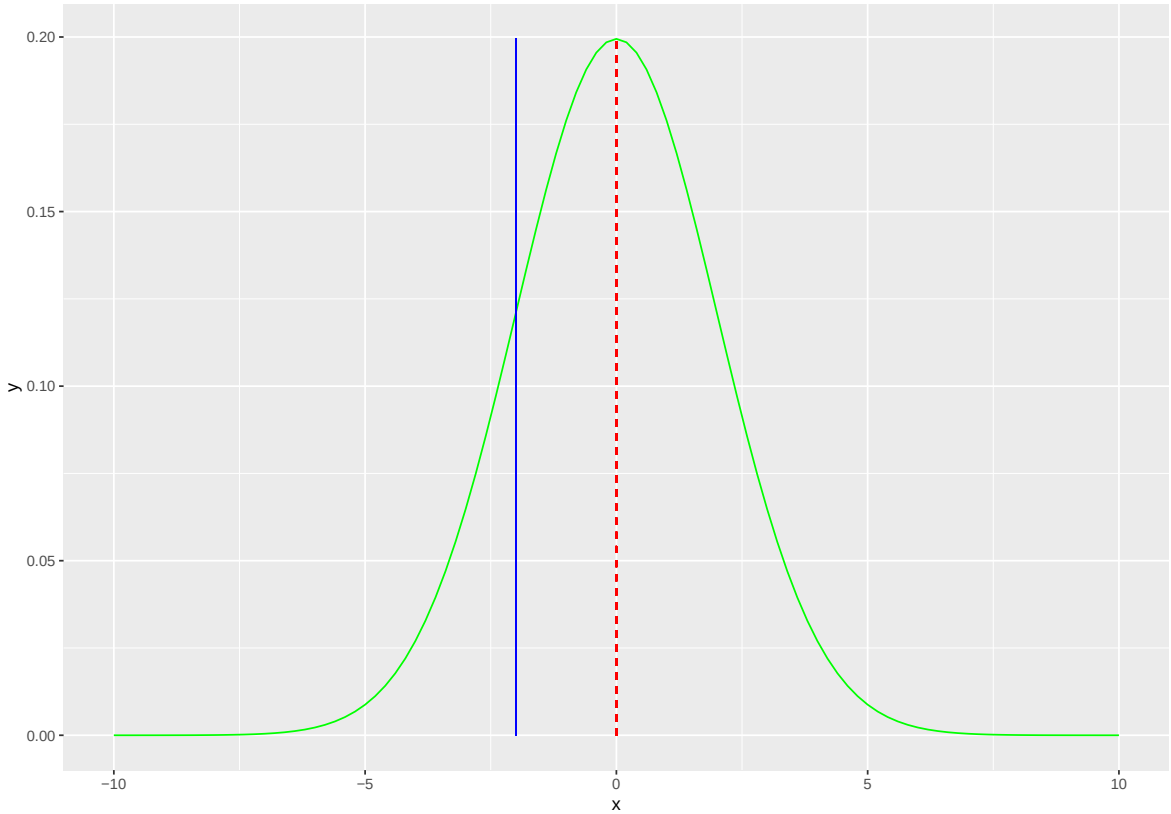
Normal Distribution



Normal Distribution Z-score

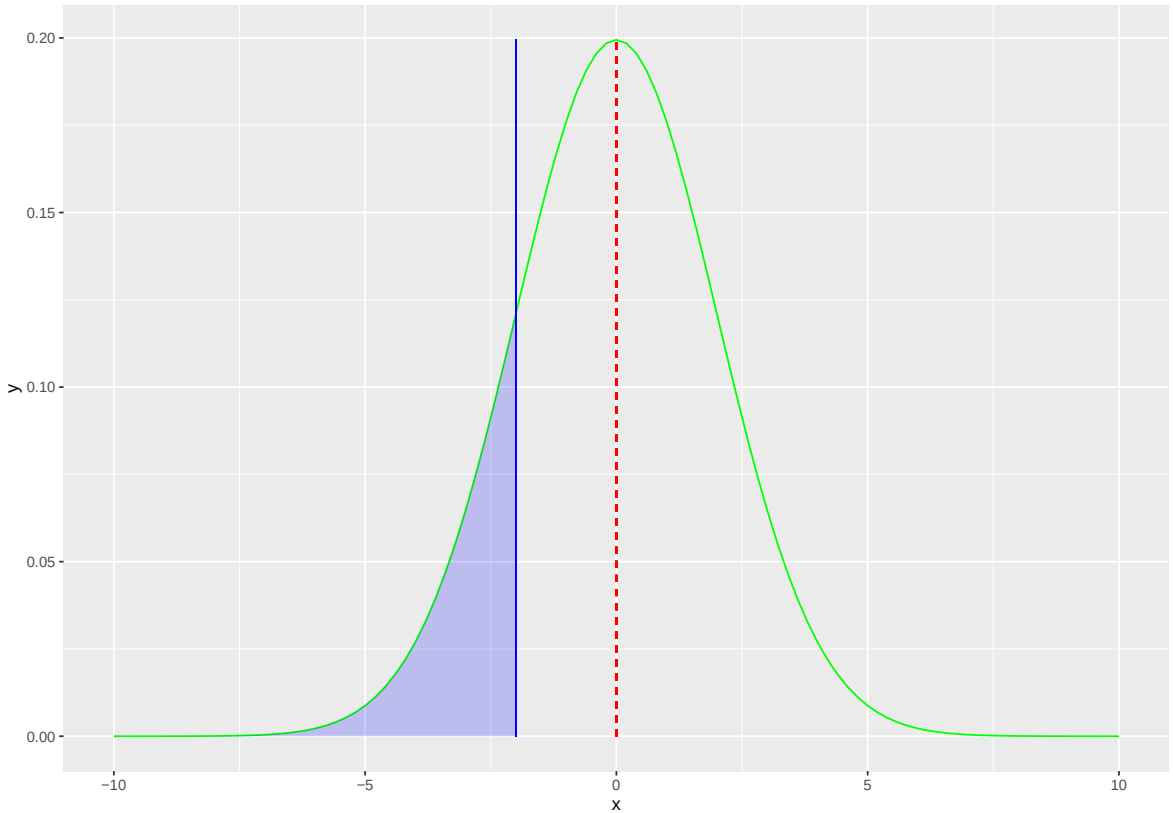
How many Standard Deviations are you away from the mean?

$$Z\text{-}score = \frac{x-\mu}{\sigma} \quad Z\text{-}score = \frac{x-\bar{x}}{s}$$



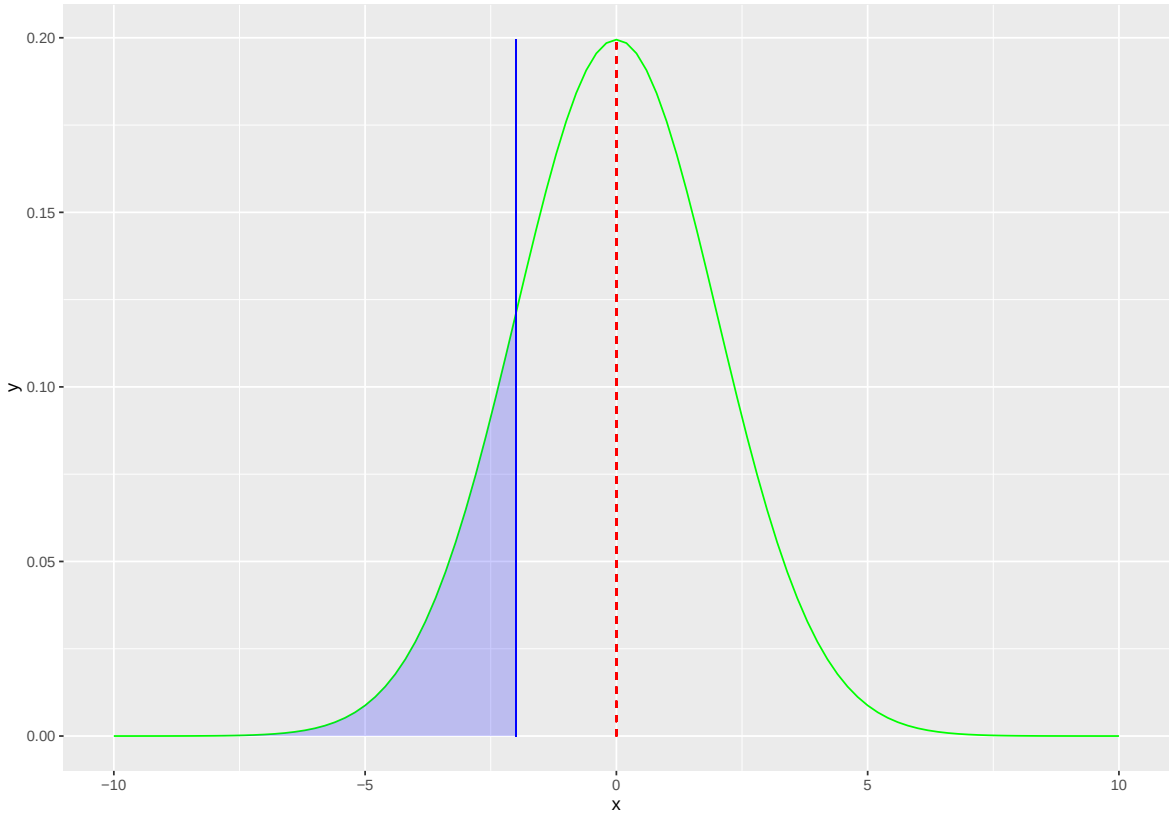
Normal Distribution Z-score

Using Z-score we calculate (or read-out) the probability.



Normal Distribution Z-score

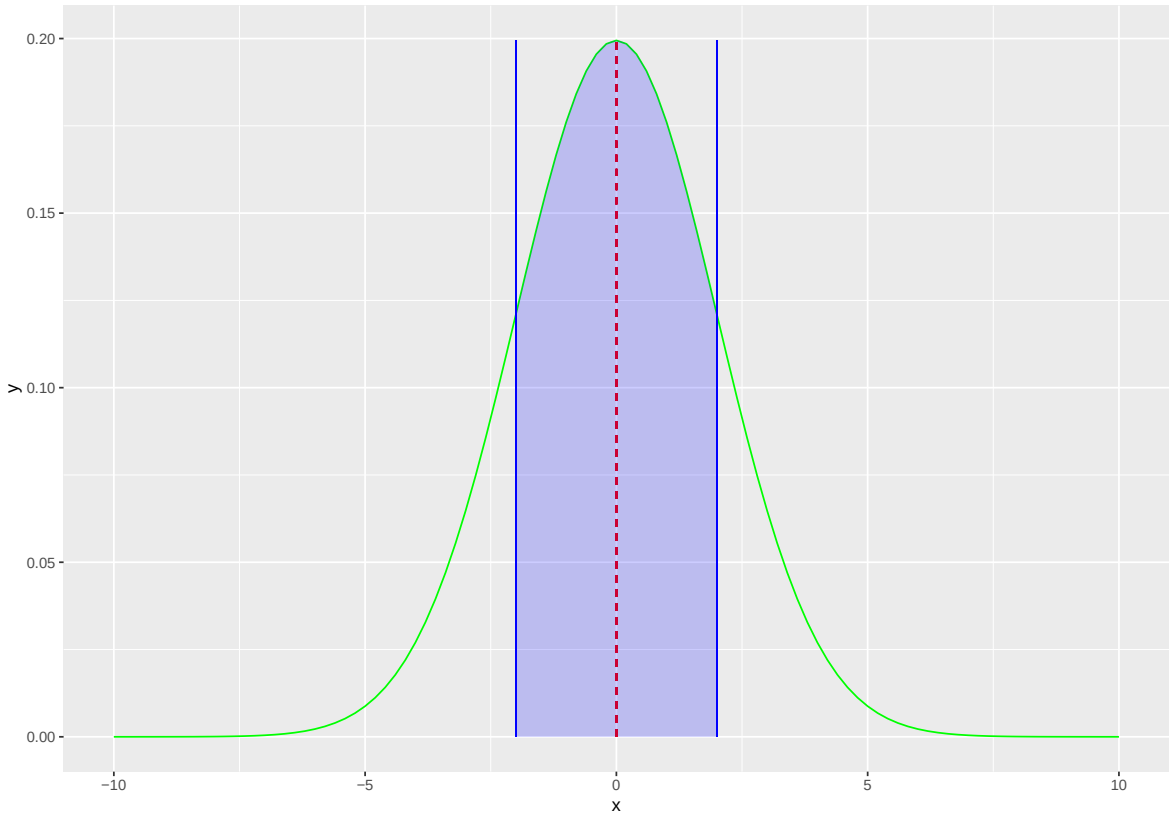
Probability 0.1586553. Read out using pnorm() function or Z-tables.



Normal Distribution Probability intervals

$$1 - 2 * pnorm(-2, mean = 0, sd = 2)$$

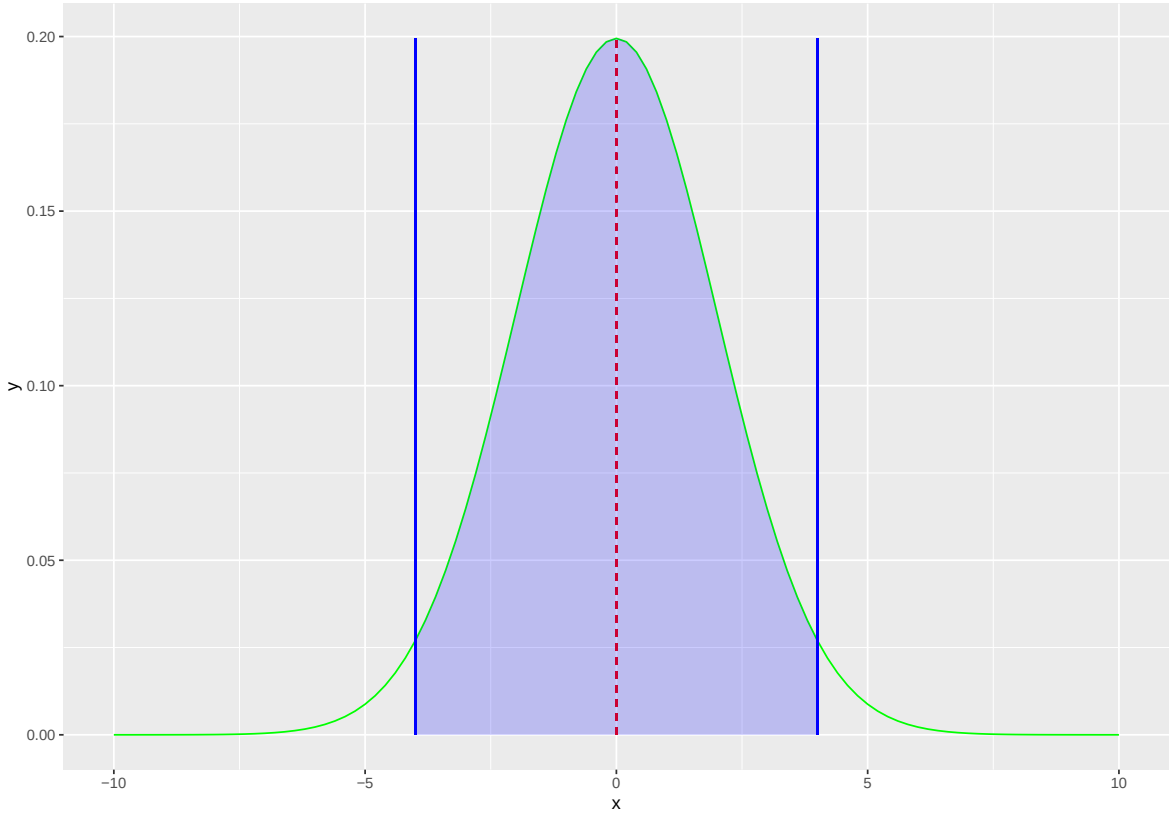
0.6826895



Normal Distribution Probability intervals

$$1 - 2 * pnorm(-4, mean = 0, sd = 2)$$

0.9544997

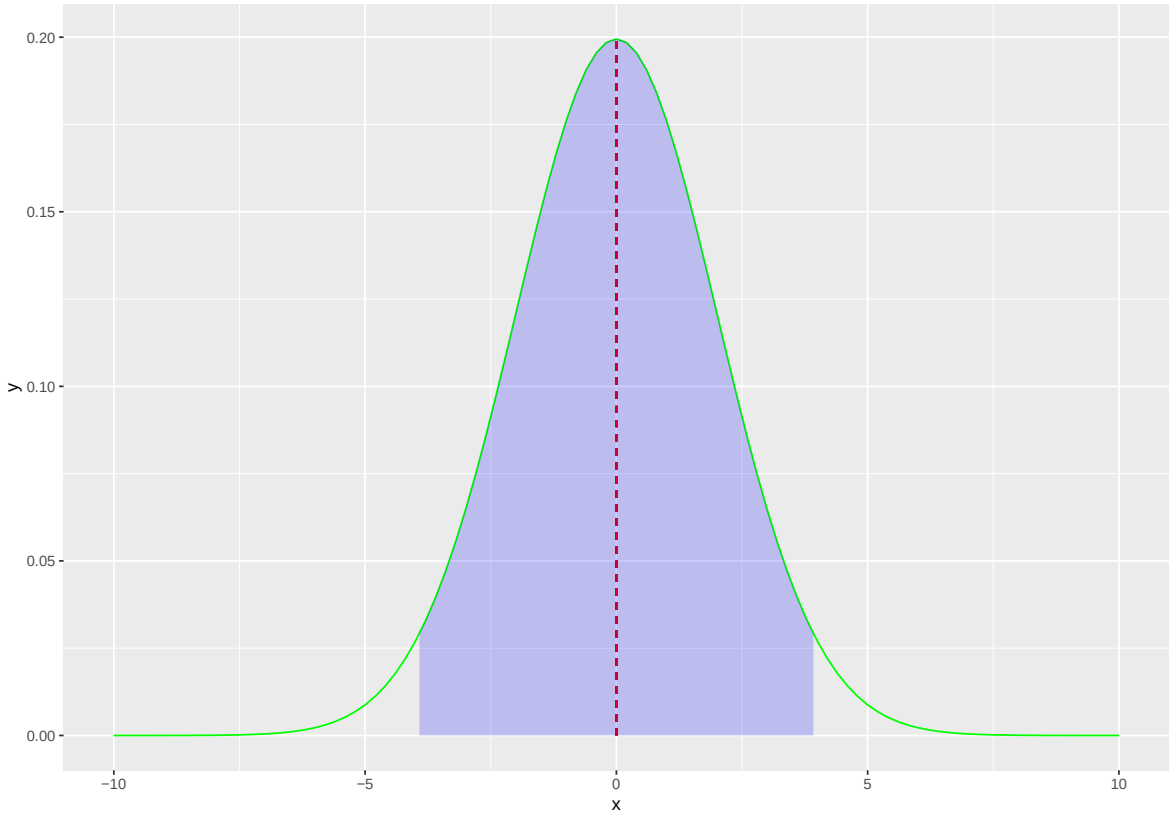


Normal Distribution Probability intervals

95% probability

$qnorm(0.025, mean = 0, sd = 1) = 1.959964$

$qnorm(0.975, mean = 0, sd = 1) = 1.959964$

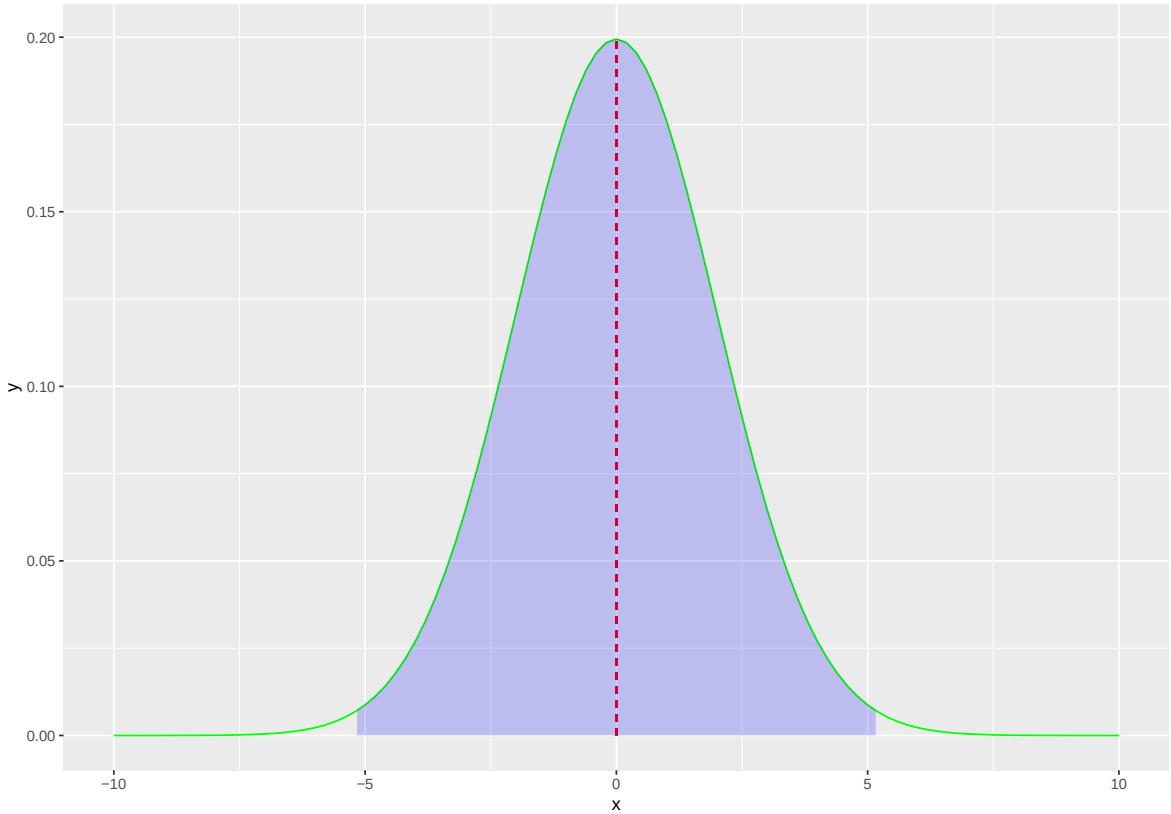


Normal Distribution Probability intervals

99% probability

$qnorm(0.005, mean = 0, sd = 1) = -2.575829$

$qnorm(0.995, mean = 0, sd = 1) = 2.575829$



Ewoks

Meet the Ewoks from the planet Endor.

Ewoks

Census on planet Endor calculated whole population to be

3 023 011 adult Ewoks is the productive age.

Out of those we have 1 500 624 females and 1 522 387 males.

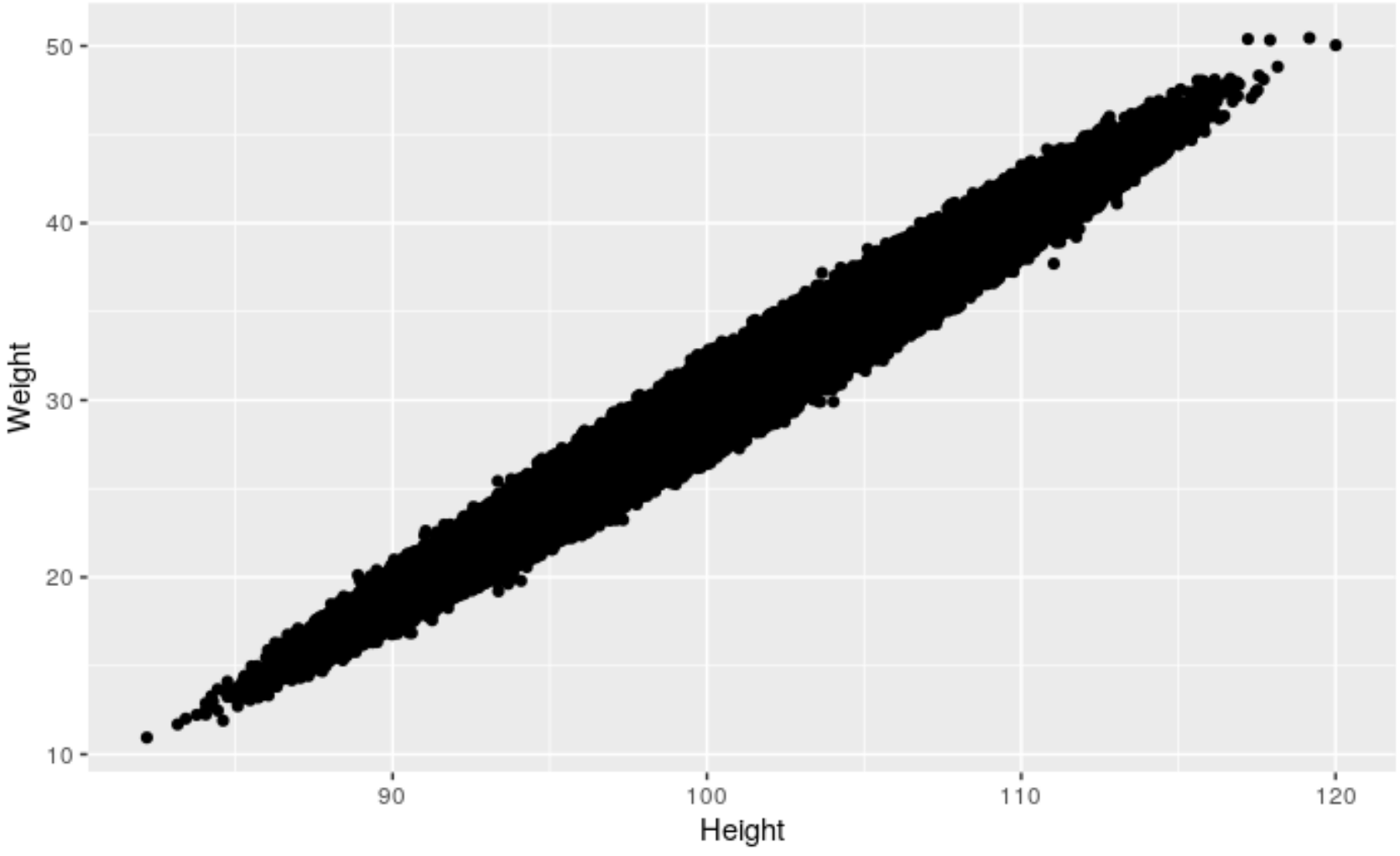
Ewoks

Id	Height	Weight	Gender
1	99.09217	28.48514	F
2	104.83346	35.57101	M
3	96.85318	27.59170	M
4	93.52584	22.05461	F
5	103.16479	32.46236	M

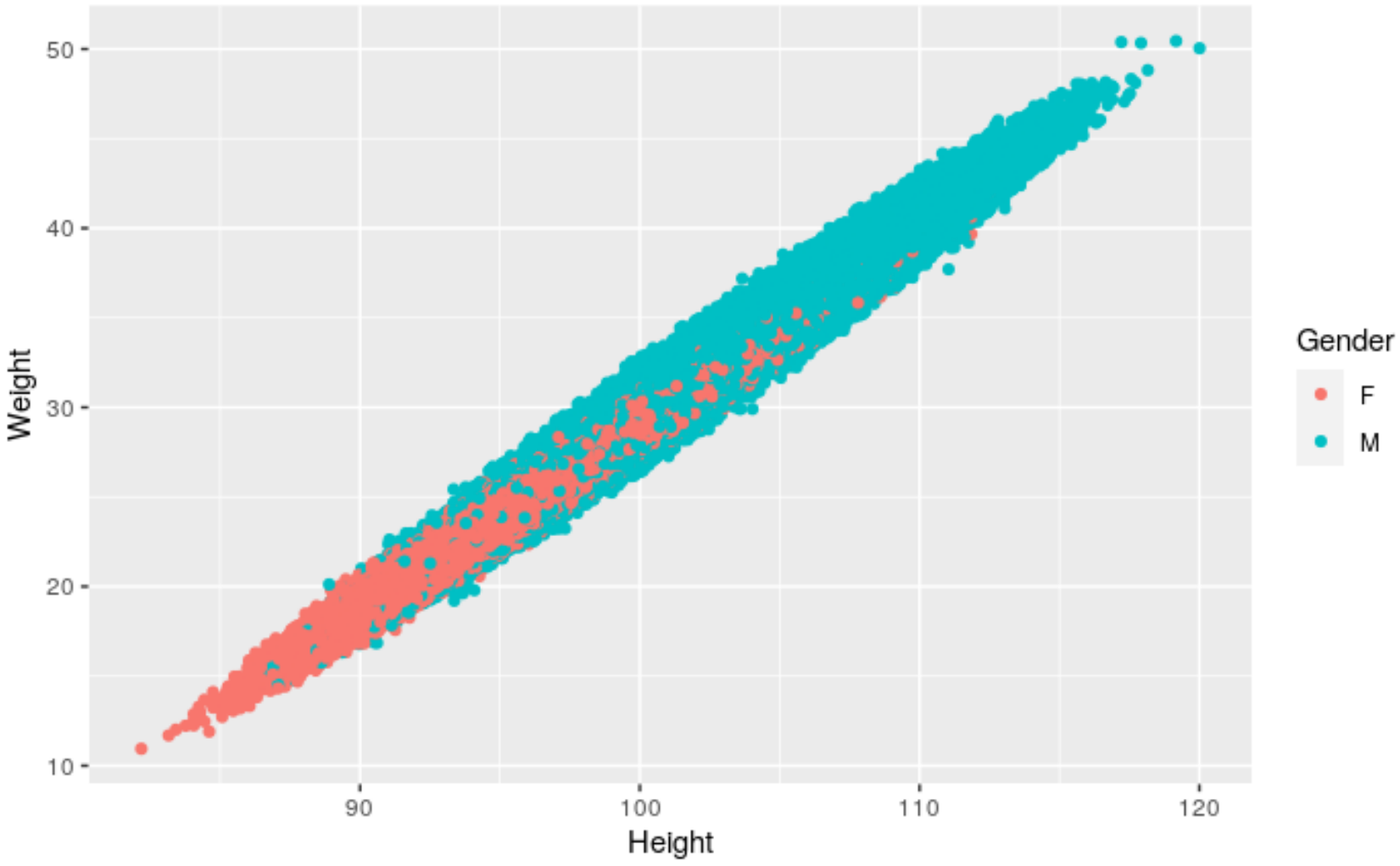
Ewoks

Gender	Mean Height	Mean Weight
F	98.0758	27.15296
M	102.2552	31.97537

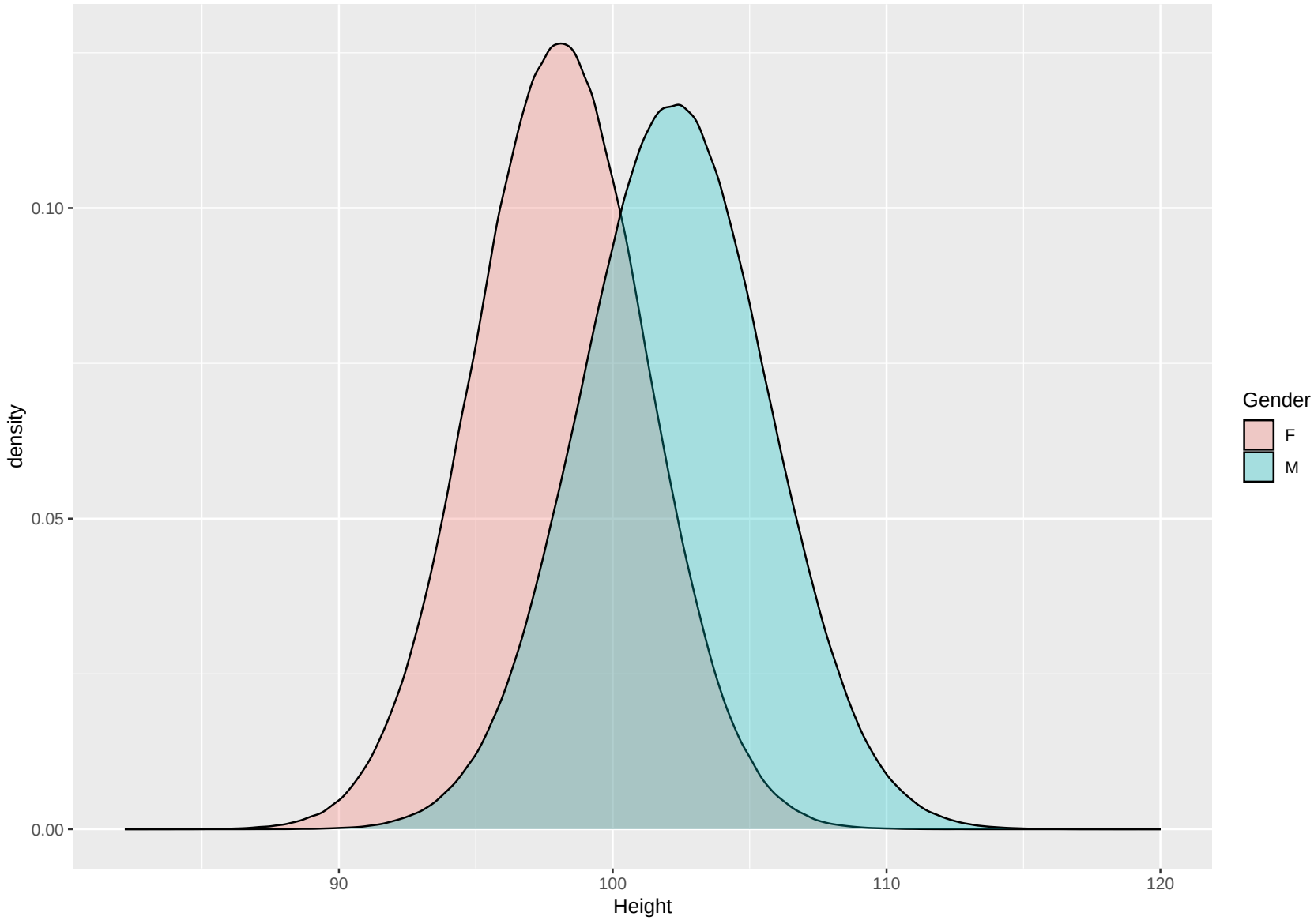
Ewoks



Ewoks

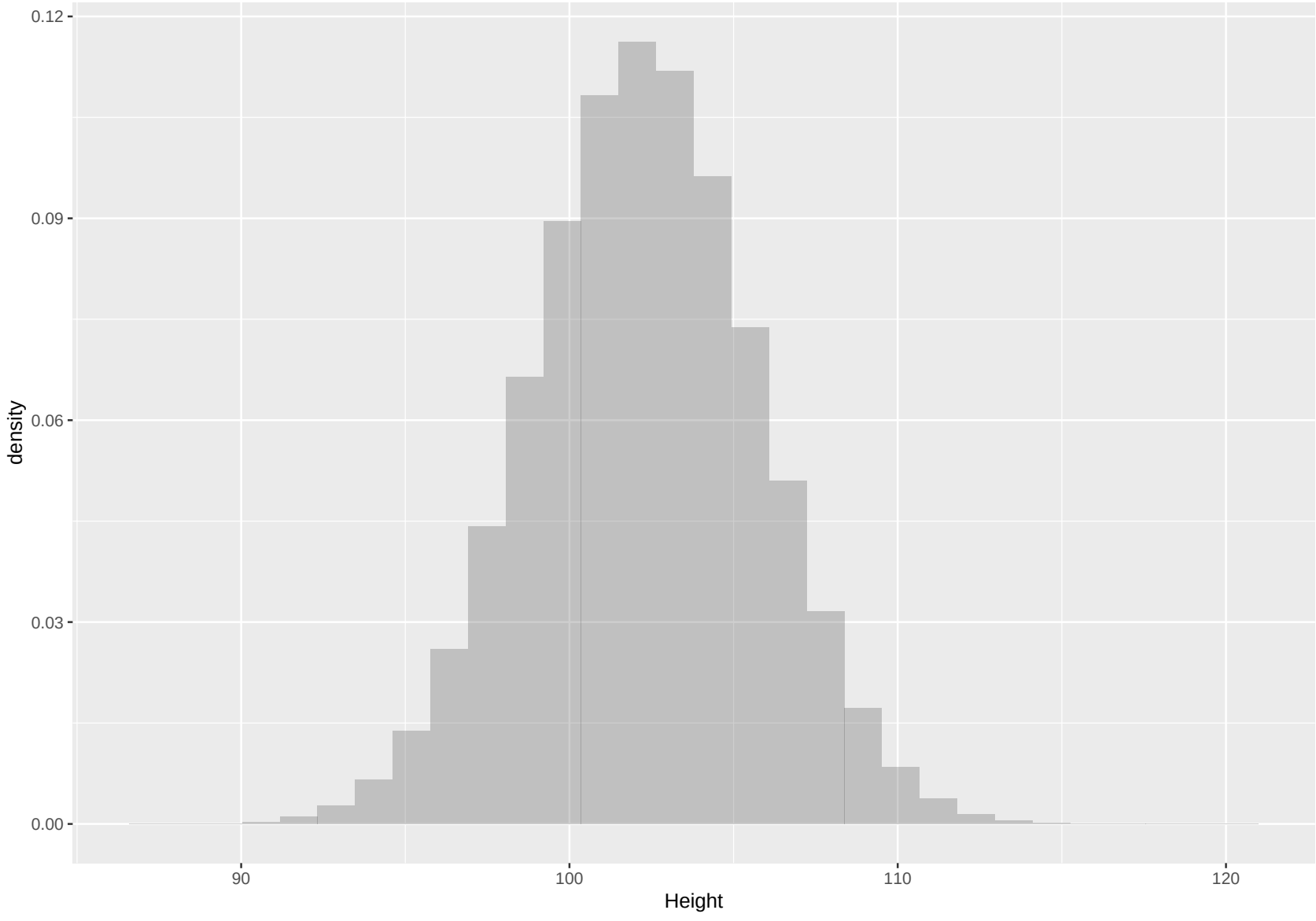


Ewoks



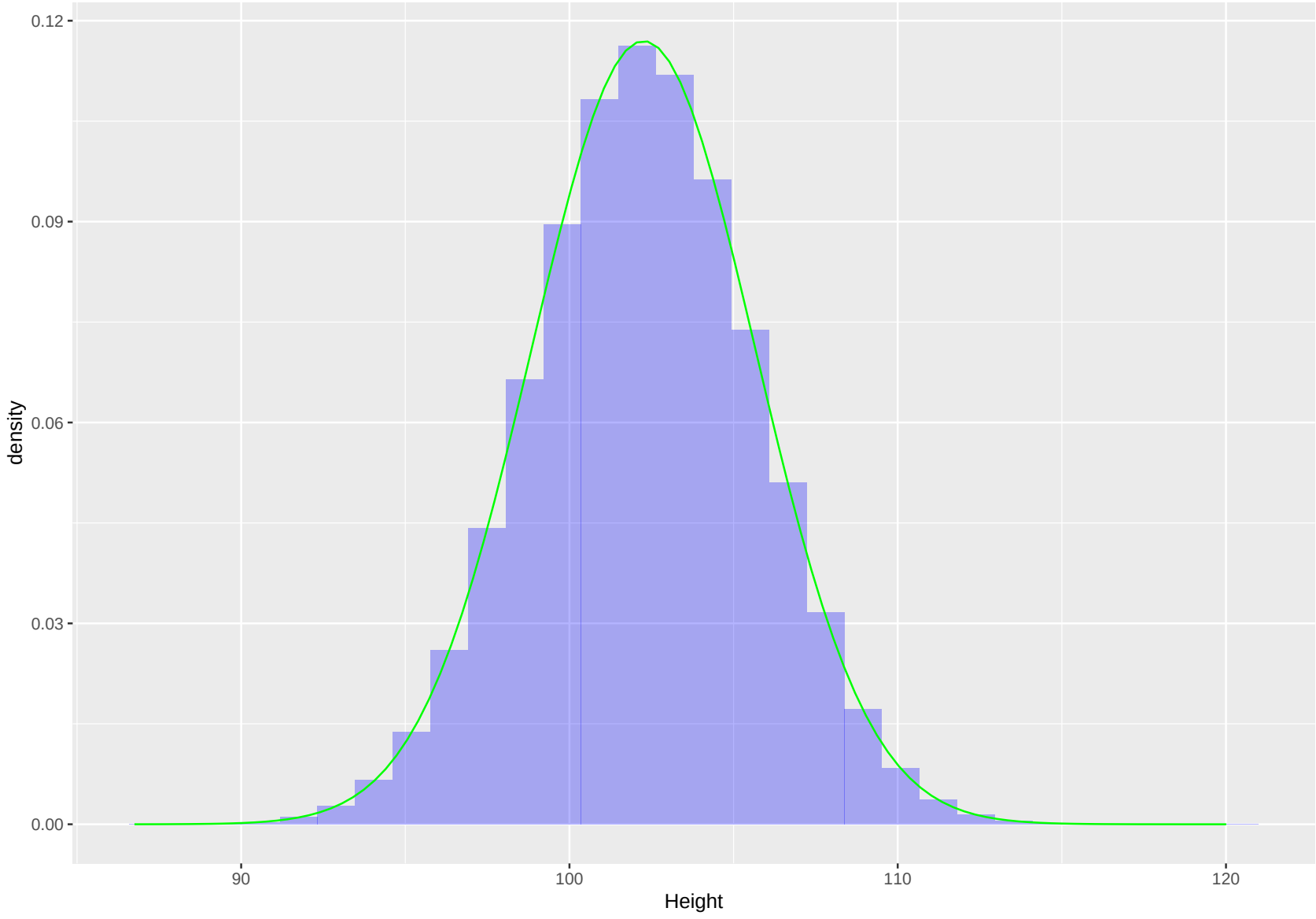
Ewoks

Do Ewok males height have Normal distribution?



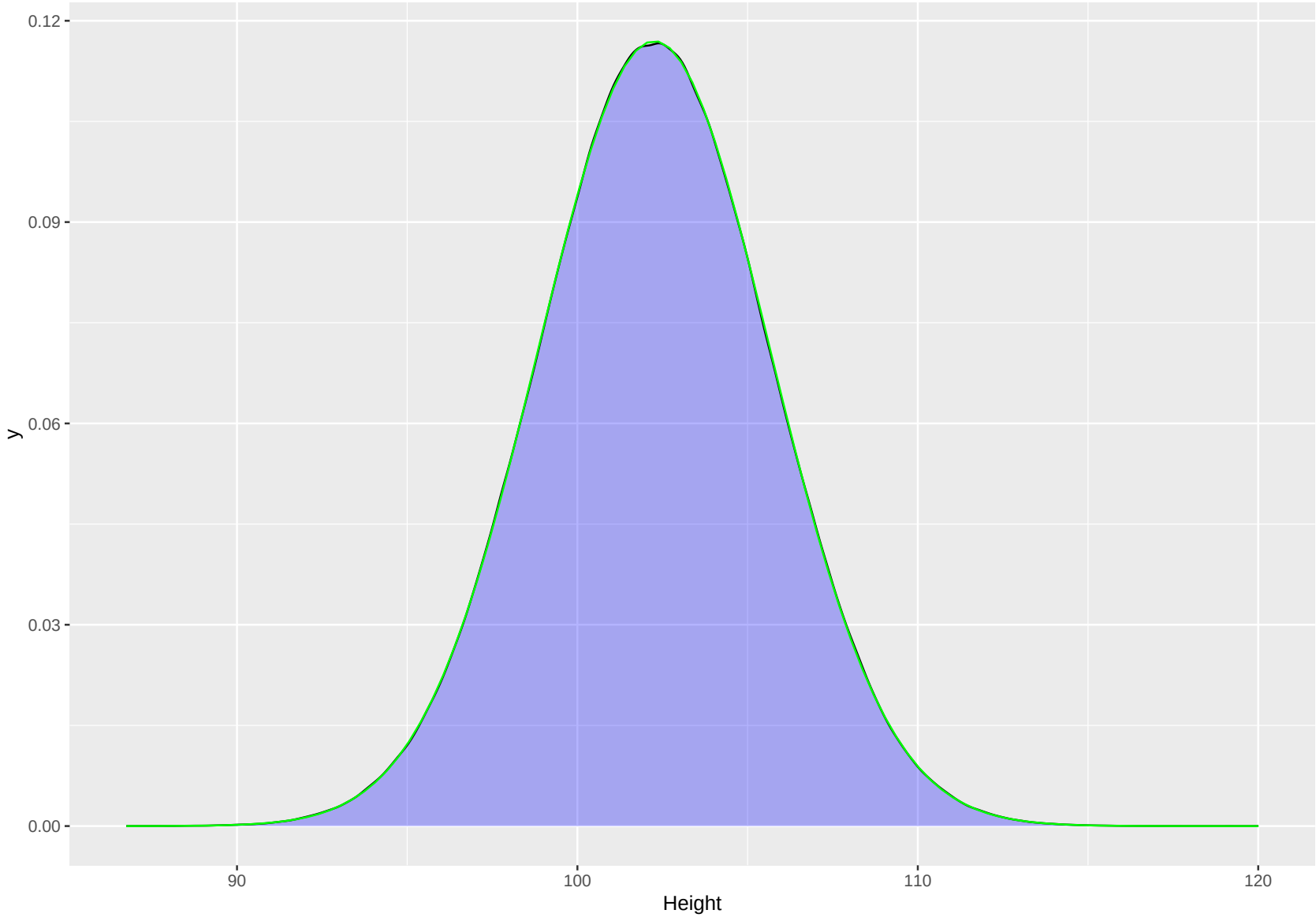
Ewoks

Compare with theoretical curve



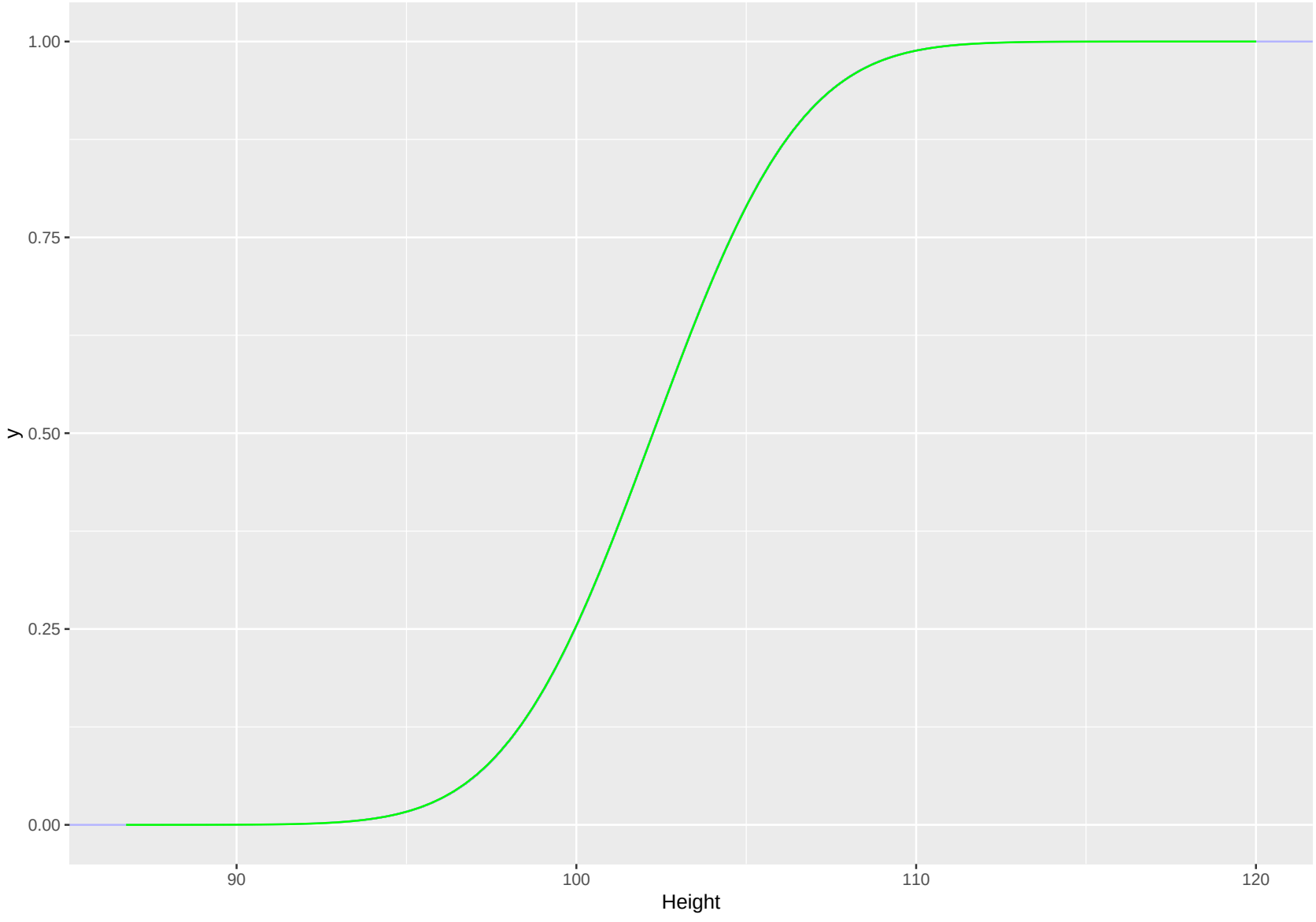
Ewoks

Compare with theoretical curve



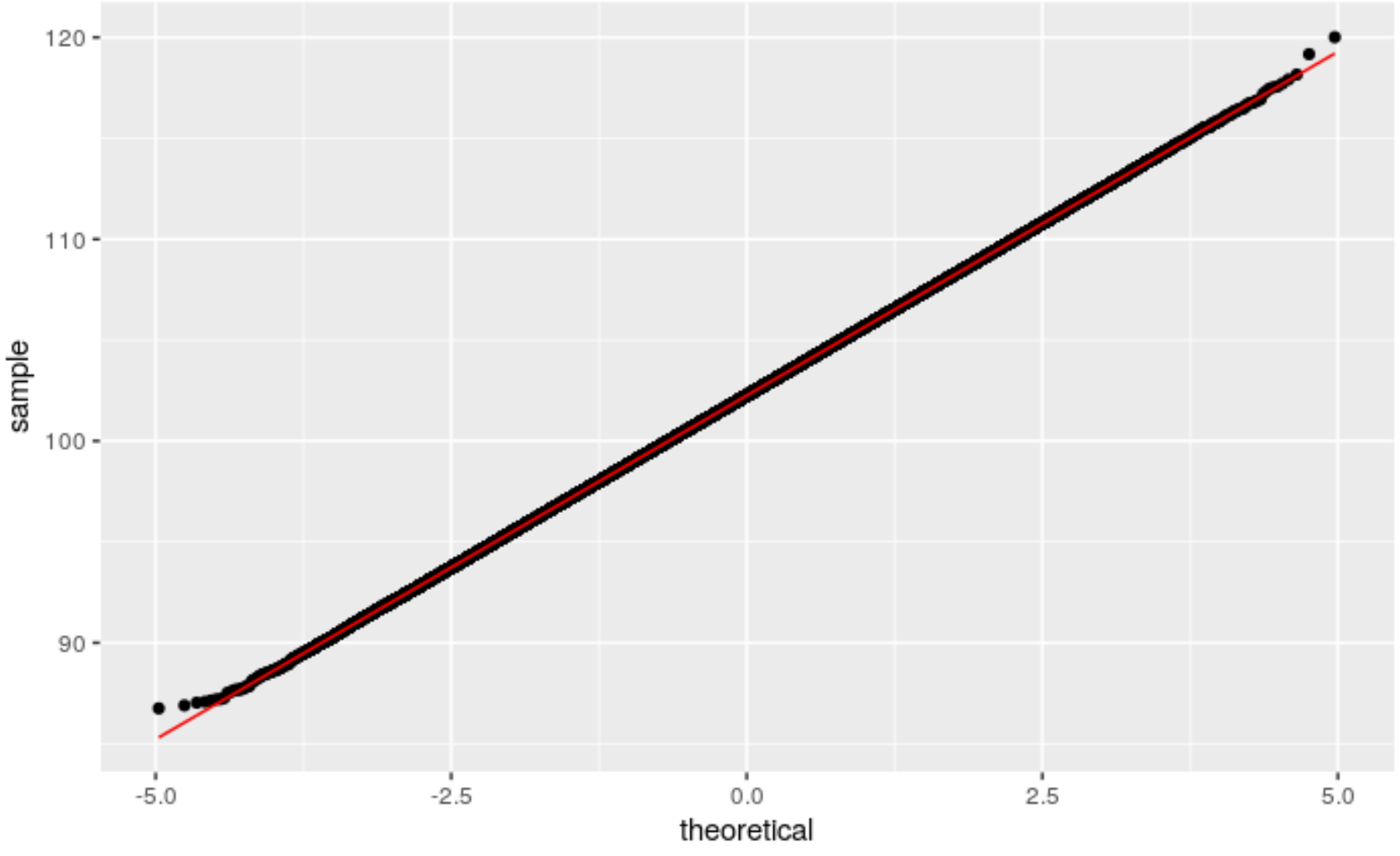
Ewoks

Compare with theoretical curve - stat_ecdf



Ewoks

Compare with theoretical curve - stat_ecdf



Ewoks

Let's check Z-score of two Ewoks

Wicket is 108.5 cm tall and weights 39.78 kg.

Paploo is 99 cm tall and weights 28.07 kg.

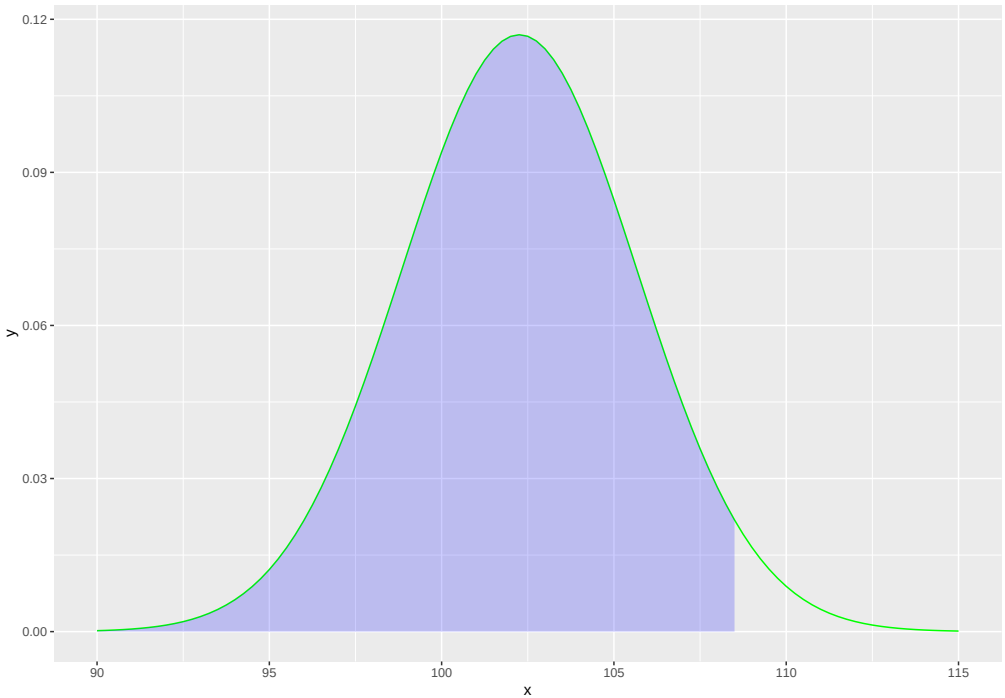
Ewoks

Z-score for Wickets height is:

$(108.5 - m_m)/sd_m = 1.831096$

And thus, percentage of shorter Ewoks:

$pnorm(1.831096) = 0.9664569$



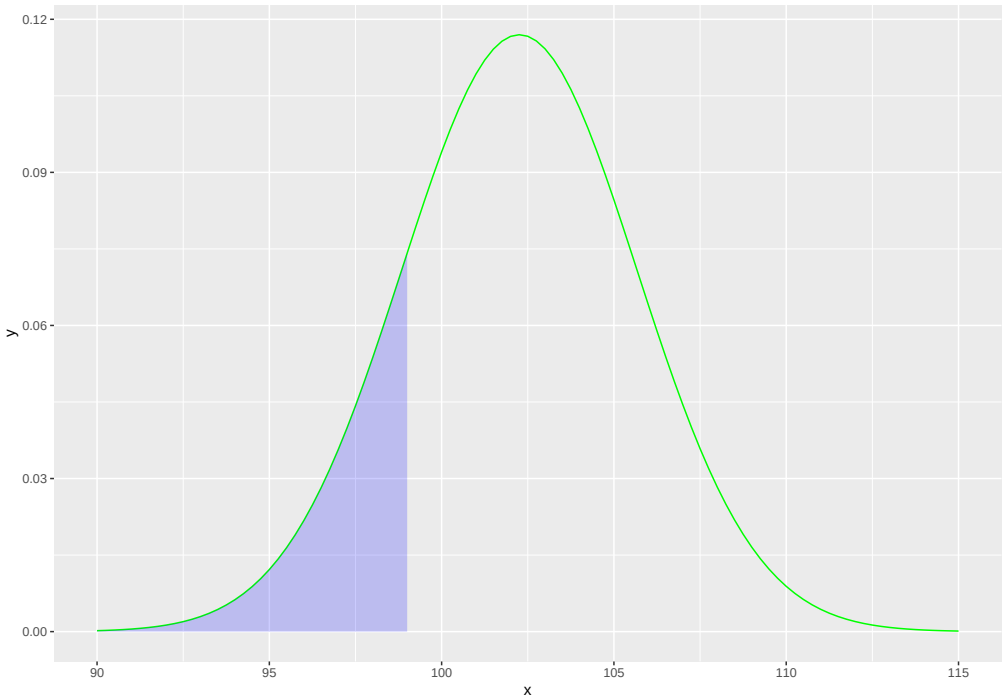
Ewoks

Z-score for Paploo height is:

$(99 - m_m)/sd_m = -0.9544674$

And thus, percentage of shorter Ewoks:

$pnorm(-0.9544674) = 0.1699235$

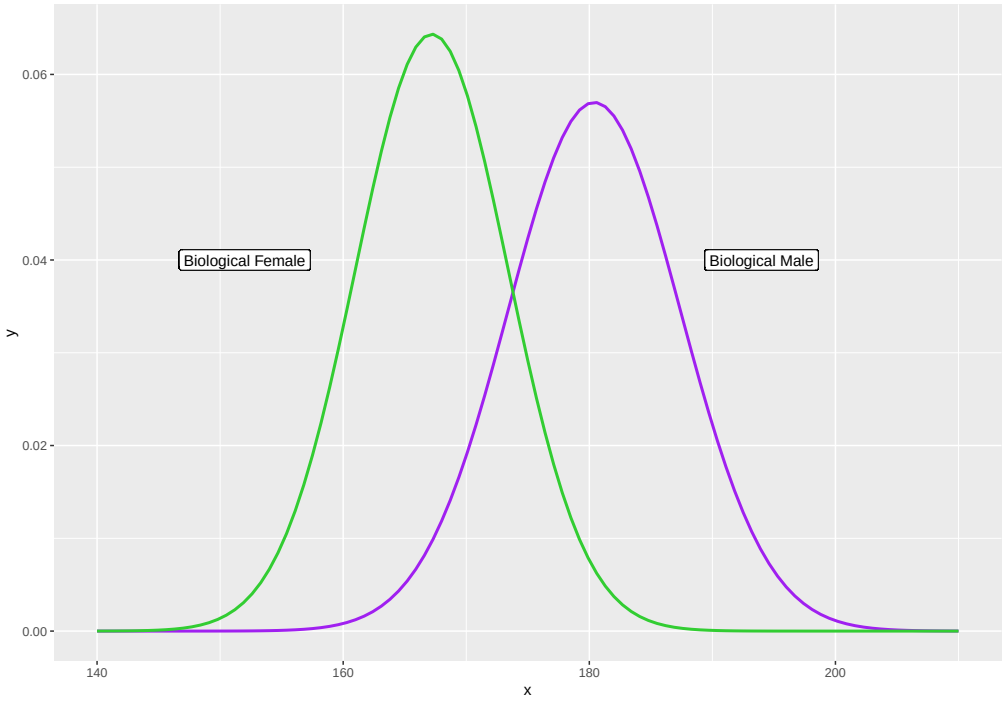


Average height in Denmark

Biological Female - mean 167.2 cm with standard deviation of 6.2 cm

Biological Male - mean 180.4 cm with standard deviation of 7.0 cm

Average height in Denmark

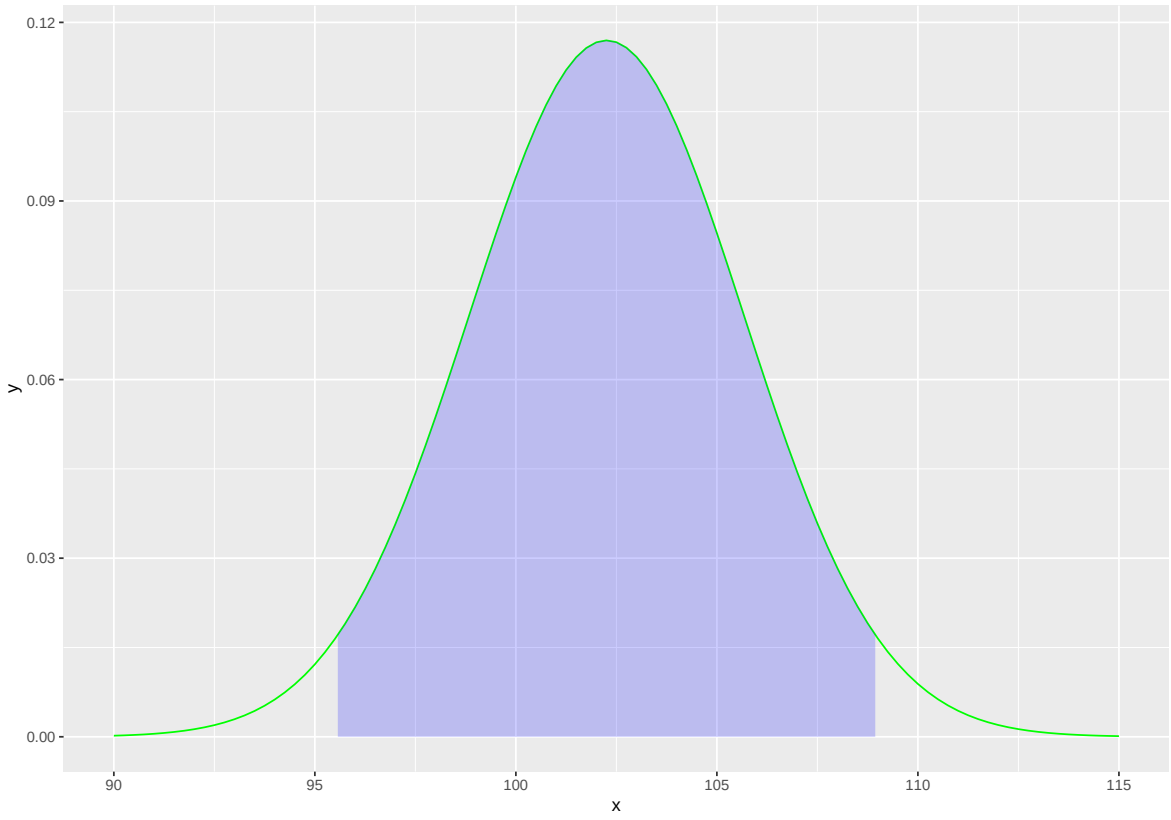


Ewoks

95% interval for male Heights

```
c(m_m - 1.96*sd_m, m_m + 1.96*sd_m)
```

```
## [1]  95.57069 108.93962
```

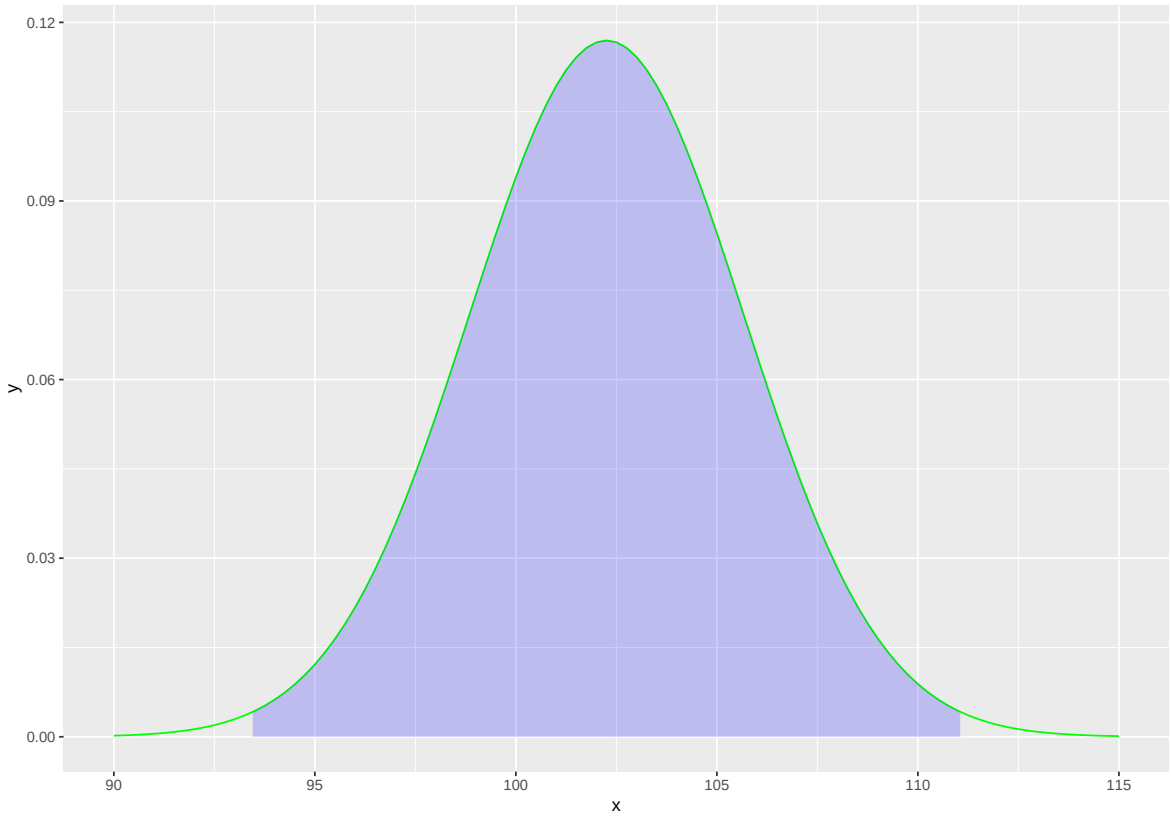


Ewoks

99% interval for male Heights

```
c(m_m - 2.58*sd_m, m_m + 2.58*sd_m)
```

```
## [1]  93.45622 111.05409
```



Population vs Sample

What do we usually deal with? Population or Sample?

Population vs Sample

What do we want to have information about? Population or sample?

Inference

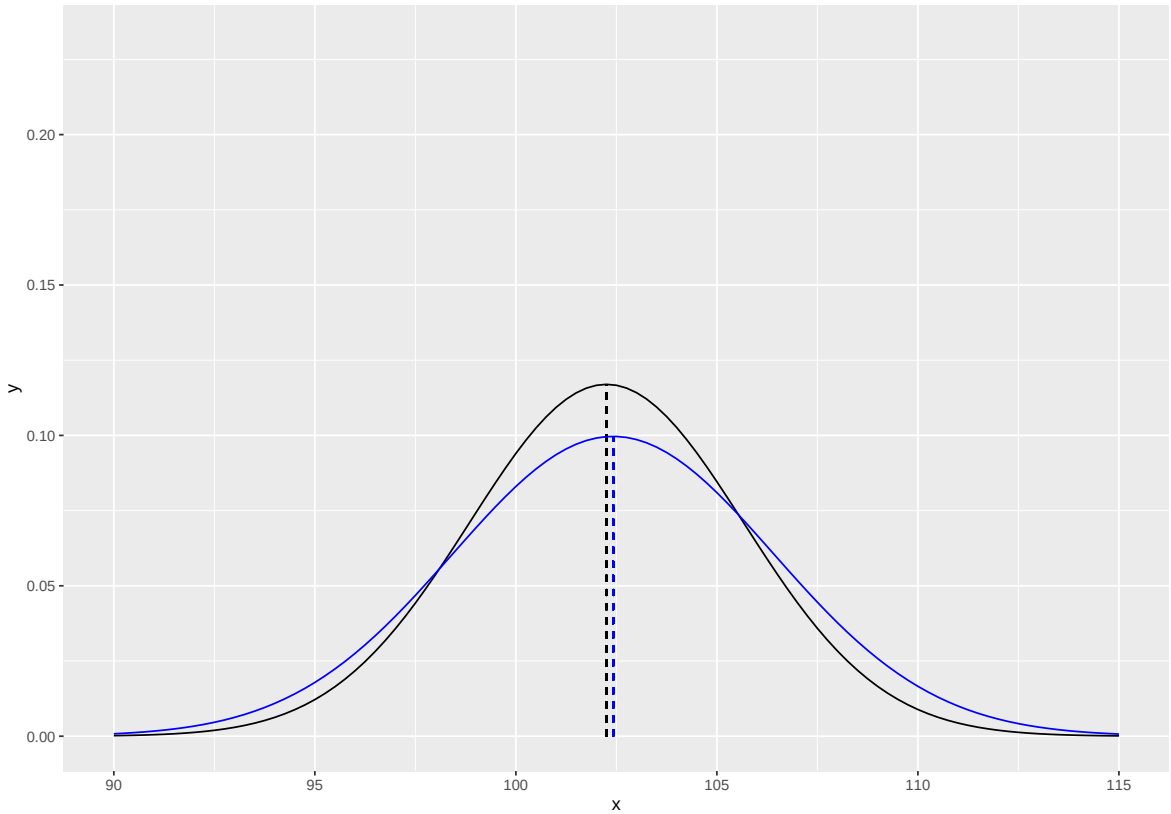
We want to infere about Population using a sample.

Inference - Point estimates

point estimate	$\longrightarrow$	*point statistic* (*population parameter*)
		$\bar{x} \longrightarrow \mu$
		$s \longrightarrow \sigma$
		$\hat{p} \longrightarrow p$

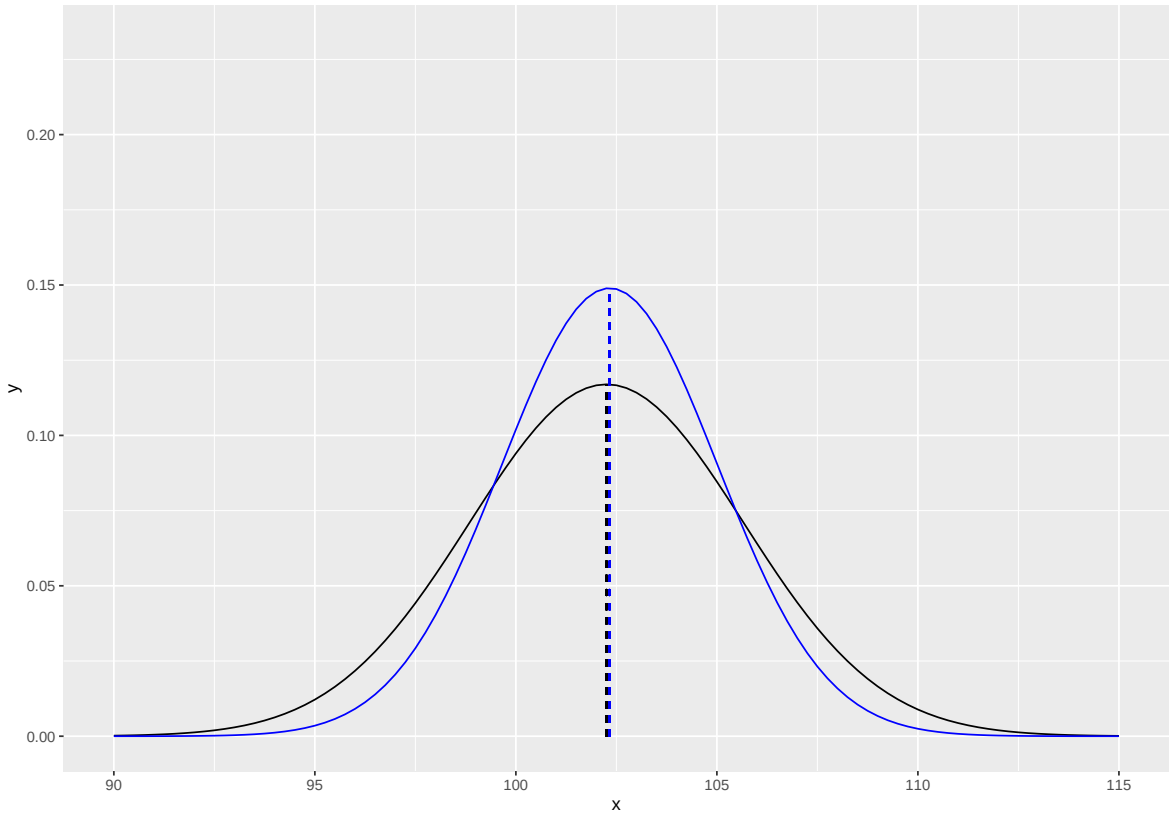
Inference - Point estimates

How accurate are our point estimates?



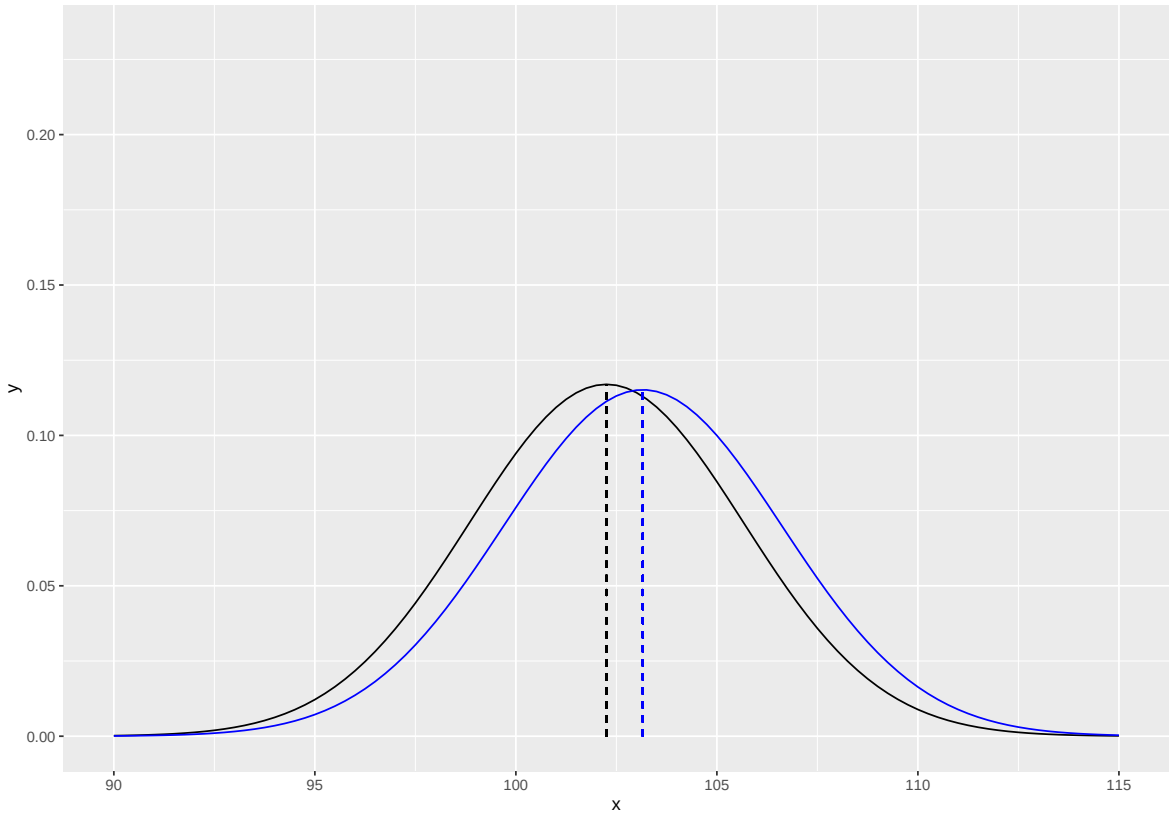
Inference - Point estimates

How accurate are our point estimates?



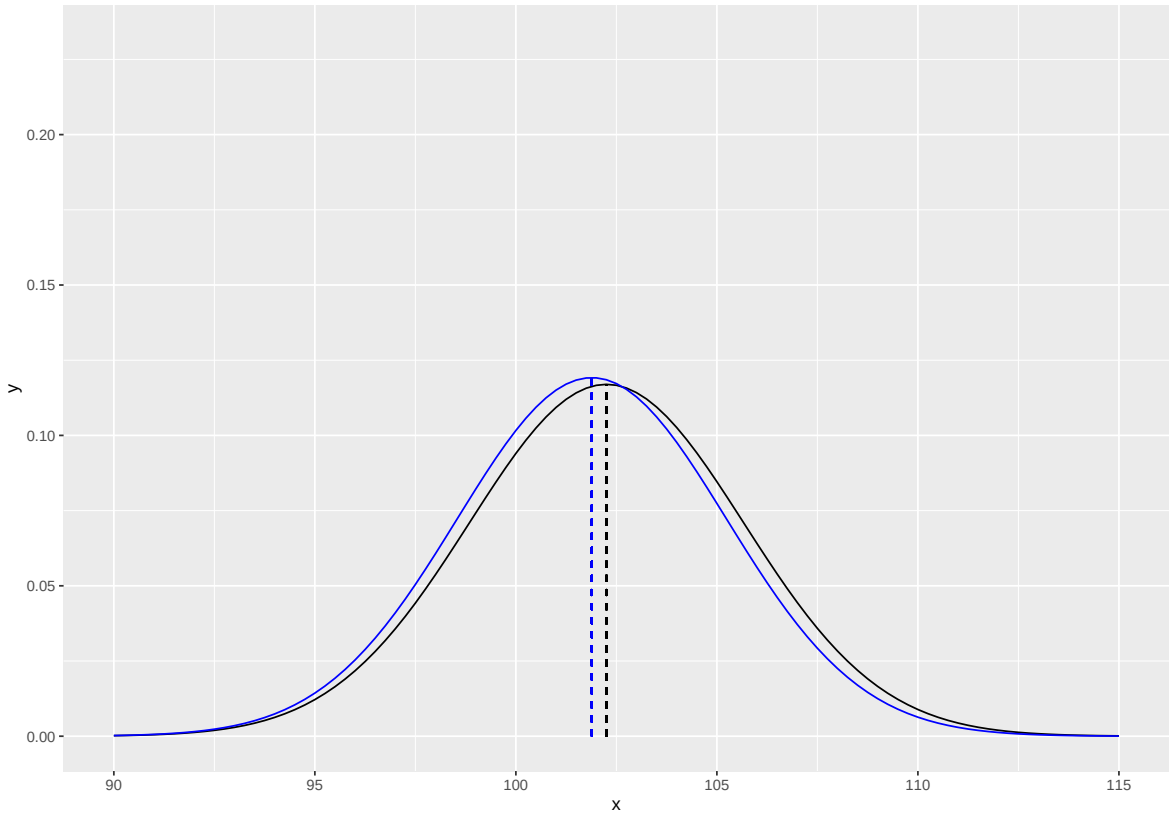
Inference - Point estimates

How accurate are our point estimates?



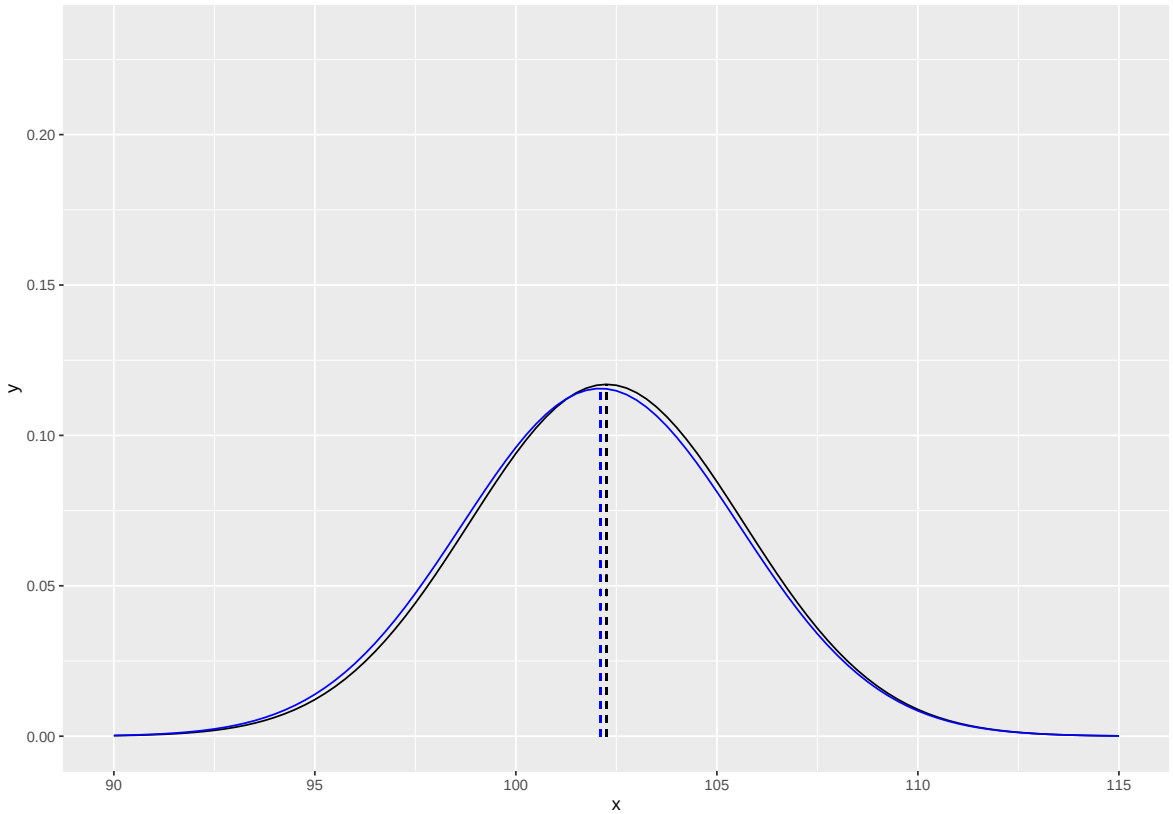
Inference - Point estimates

How accurate are our point estimates?



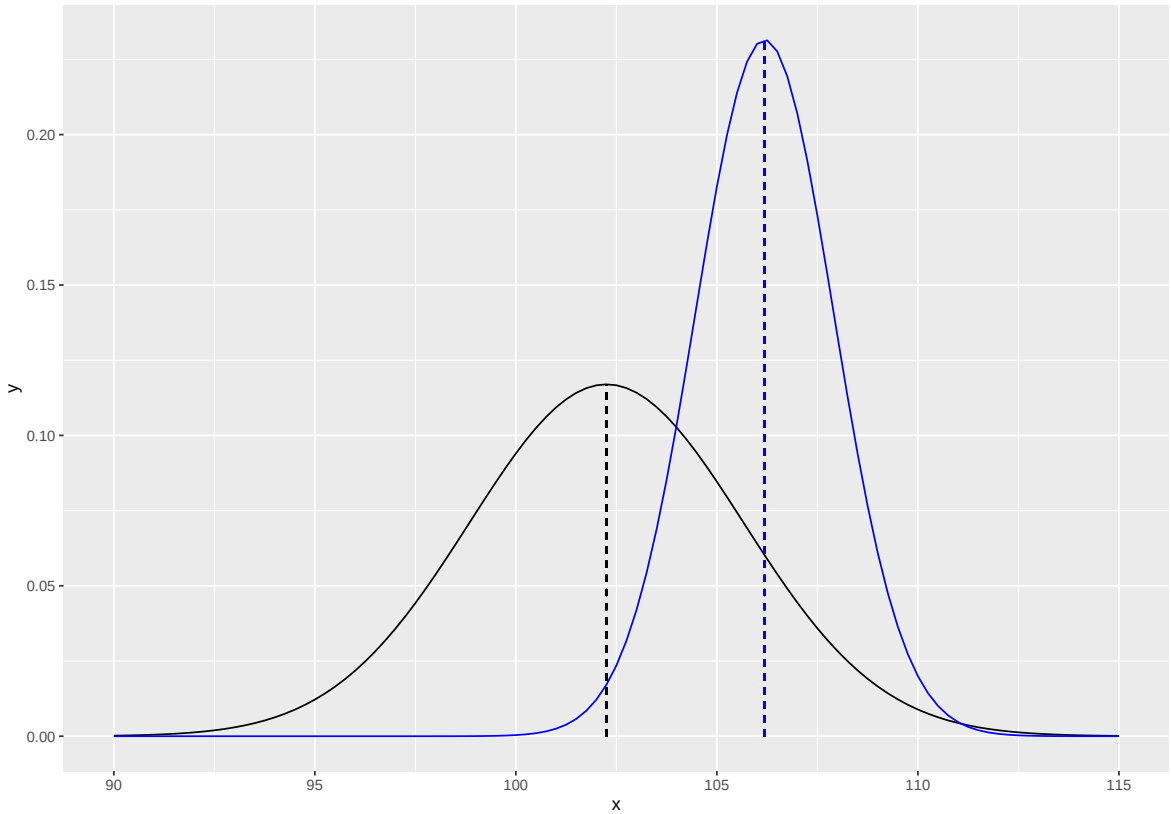
Inference - Point estimates

How accurate are our point estimates?



Inference - Point estimates

How accurate are our point estimates?



Inference

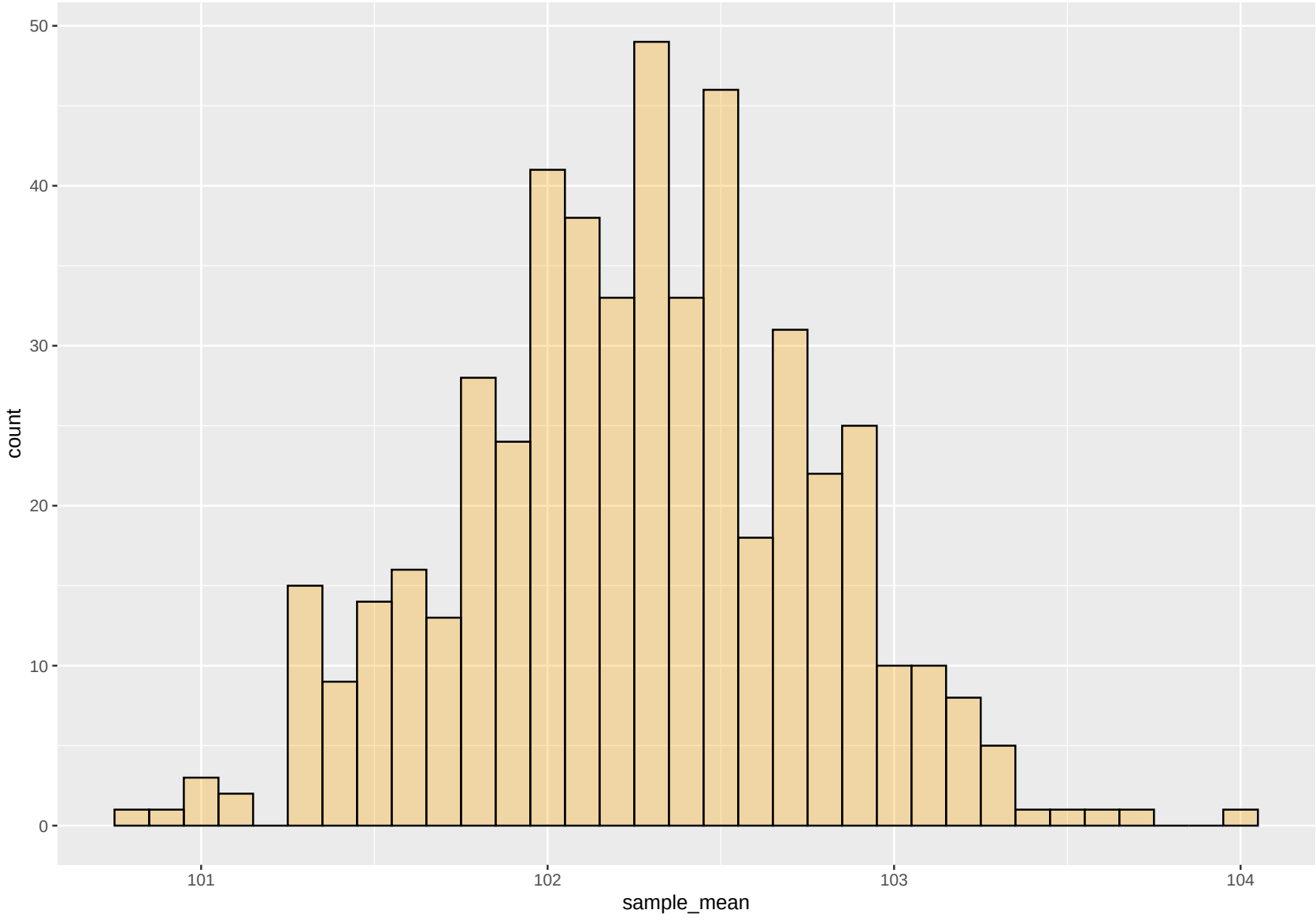
What would happen if we took same size sample (randomly) multiple times from the population?

Sampling distribution - point estimate distribution. (Not to be confused with a distribution of a sample)

Inference

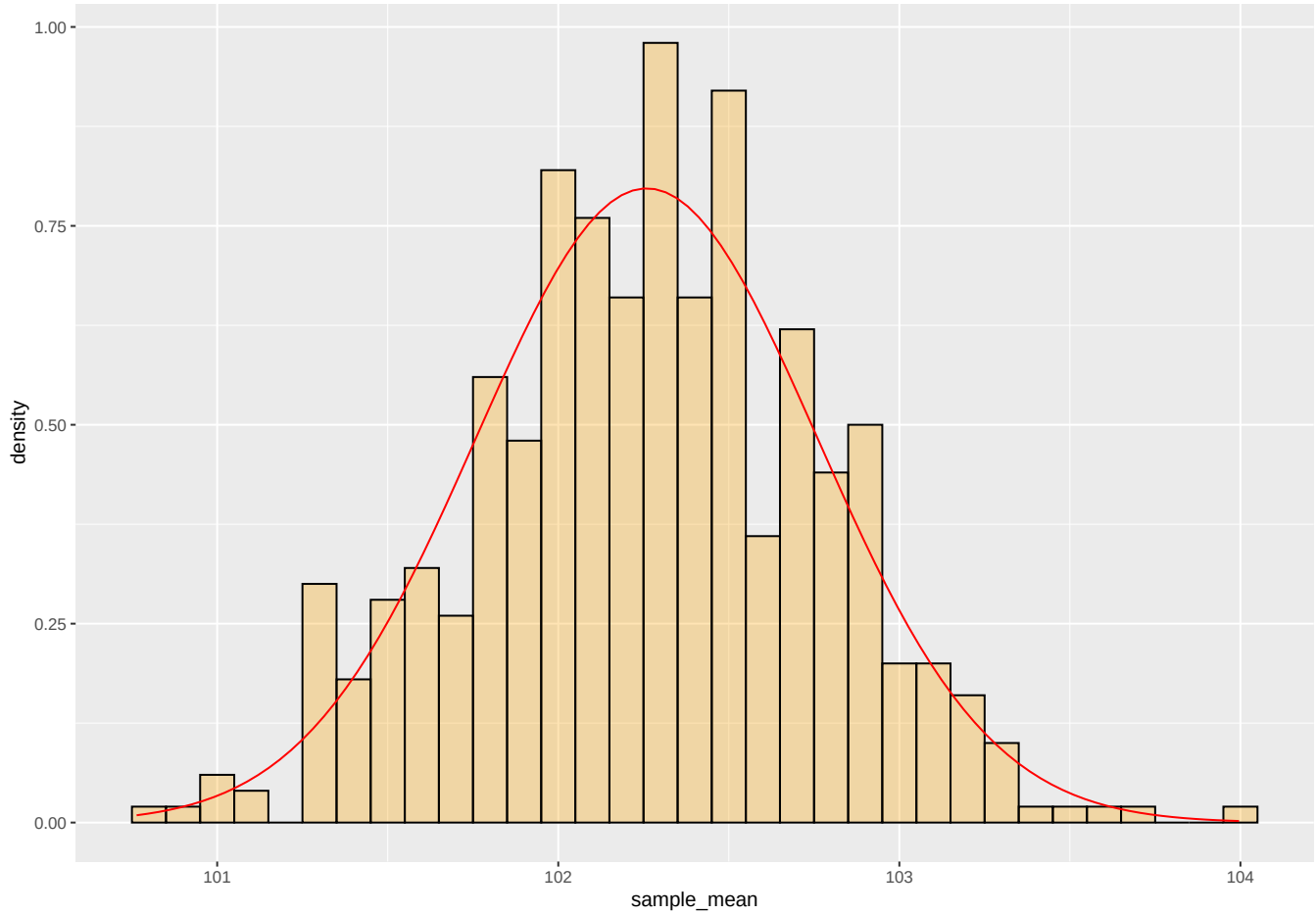
Video.

Inference



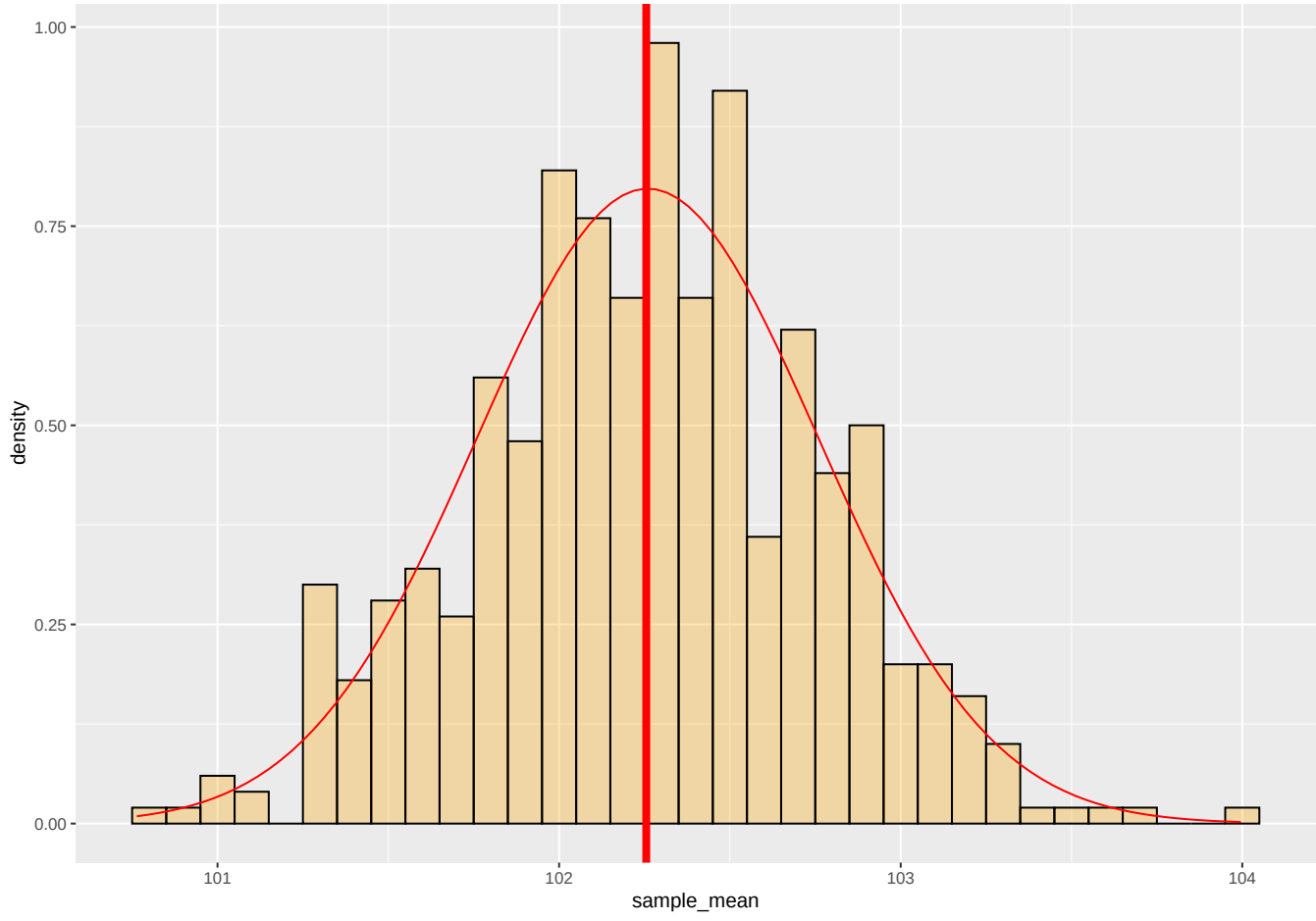
Inference

Normal distribution



Inference

Centered around Population Point Statistic



Inference

What about Standard Deviation?

Sampling distibutions standard deviation is called \STANDARD ERROR.

For point estimate MEAN its formula is:

$$SE = \frac{\sigma}{\sqrt{n}}$$

Inference

Confidence intervals.

$$\textit{point estimate} \pm 1.96 \cdot \textit{SE}$$

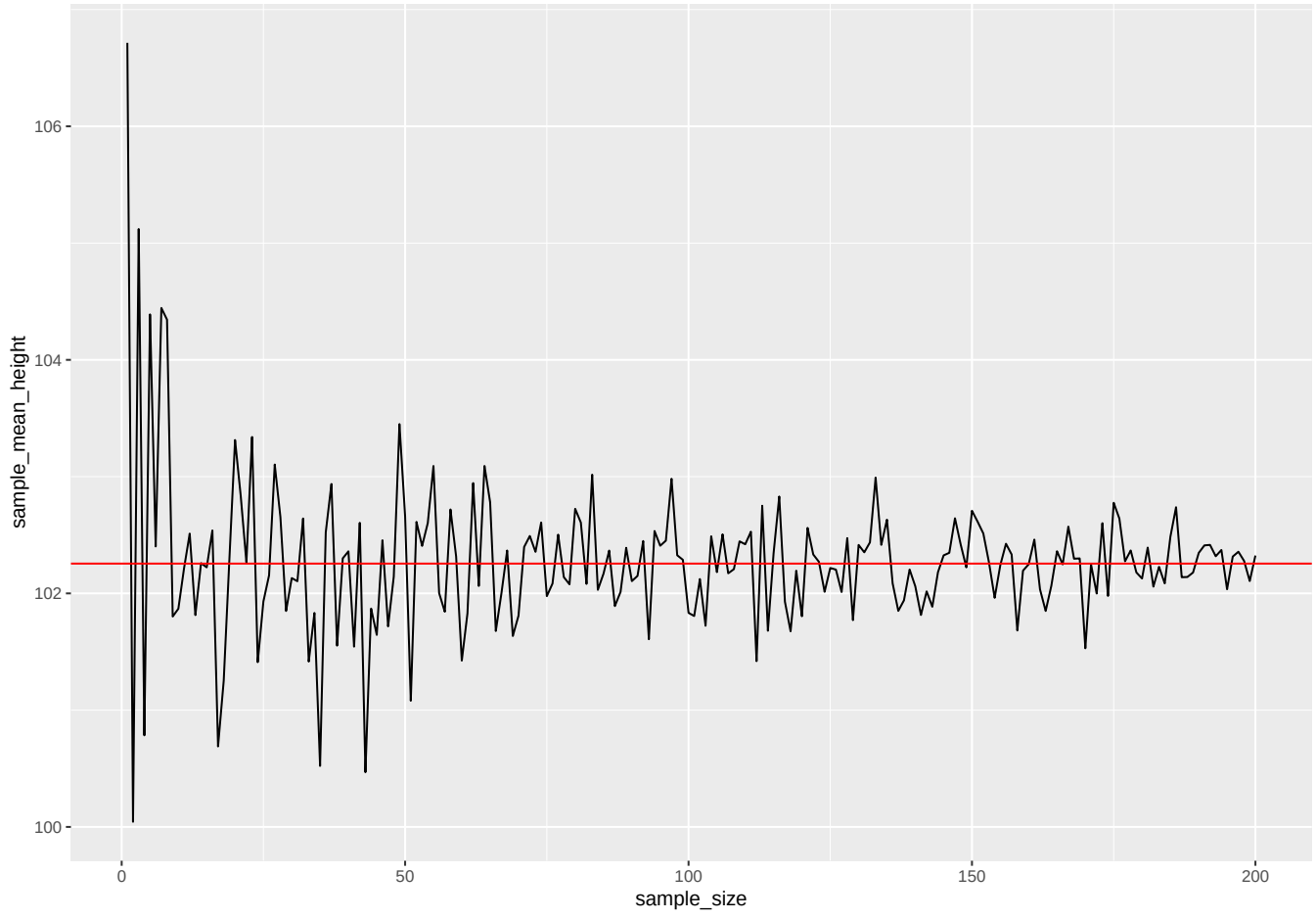
Inference

Central Limit Theorem.

- Samples are independent,
- Sample size is bigger or equal to 30,
- Population distribution is not strongly skewed.

Inference

How influential is sample size?



Inference

How influential is sample size?