? Confidence Intervals ?

86408

10/09/2019

What is a confidence interval?

-The 95% confidence interval *defines a range of values* that you can be *95% certain* contains the population mean.

Confidence levels correspond to the percentage of area under the normal distribution curve.

- A *0.90 confidence interval* would include *90% of the curve's area*, and the probability of observing a value outside of the confidence interval would be 0.10.
- A *0.95 confidence interval* would include *95% of the curve's area*, and the probabilty of an observation outside of the confidence interval would be 0.05.

How do we use confidence intervals?

- For practical purposes, we study "samples" and extrapolate results to the population from where the sample is drawn.
- The purpose of statistical inference is to estimate population parameters using observed sample data without having to actually study the whole population.
- A **95% confidence interval (95% CI)** is a population parameter expressed as a range that contains the true population parameter 95% of the times. 2.5% of the values will be below and 2.5% of the times above this range.
- The confidence interval helps gauge how adequately a sample represents the population.
- A confidence interval indicates the range that's likely to contain the true population parameter, so the confidence interval focuses on the population.

Creating a Vector

```
(randomVector = round(rnorm(5, mean=10, sd=1), 2))
```

```
## [1] 11.85 11.49 11.55  8.51  8.95
```

Generating a Sample

```
(vectorSample = sample(randomVector))
```

```
## [1]  8.95 11.85 11.49 11.55  8.51
```

```
(sampleMean = mean(vectorSample))
```

```
## [1] 10.47
```

```
(sampleSD = sd(vectorSample))
```

```
## [1] 1.601811
```

$$\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right)$$

Figure 1: Equation with Known Variance

$$\left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right)$$

Figure 2: Equation

How to Calculate a Confidence Interval

When the population variance is *known*, use Figure 1.

When the population variance is *unknown*, use Figure 2.

Calculating Confidence Intervals with Unknown Variance

```r
#sesm: standard error of sample mean
#ucl: upper confidence level
#lcl: lower confidence level
#ci: confidence interval

confidenceLevel = function(sample, t, s){
  sesm = s/sqrt(length(sample))
  ucl = mean(sample)-qt((1-t/2)/2, 4)*sesm
  lcl = mean(sample)+qt((1-t/2)/2, 4)*sesm
  ci = c(lcl, ucl)
  return(ci)
}
#Confidence Level 0.95
(confidenceLevel95_vectorSample = confidenceLevel(vectorSample, 0.95, sd(randomVector)))
```

```
## [1]  9.971736 10.968264
```

```r
#Confidence Level 0.90
(confidenceLevel90_vectorSample = confidenceLevel(vectorSample, 0.9, sd(randomVector)))
```

```
## [1] 10.00298 10.93702
```

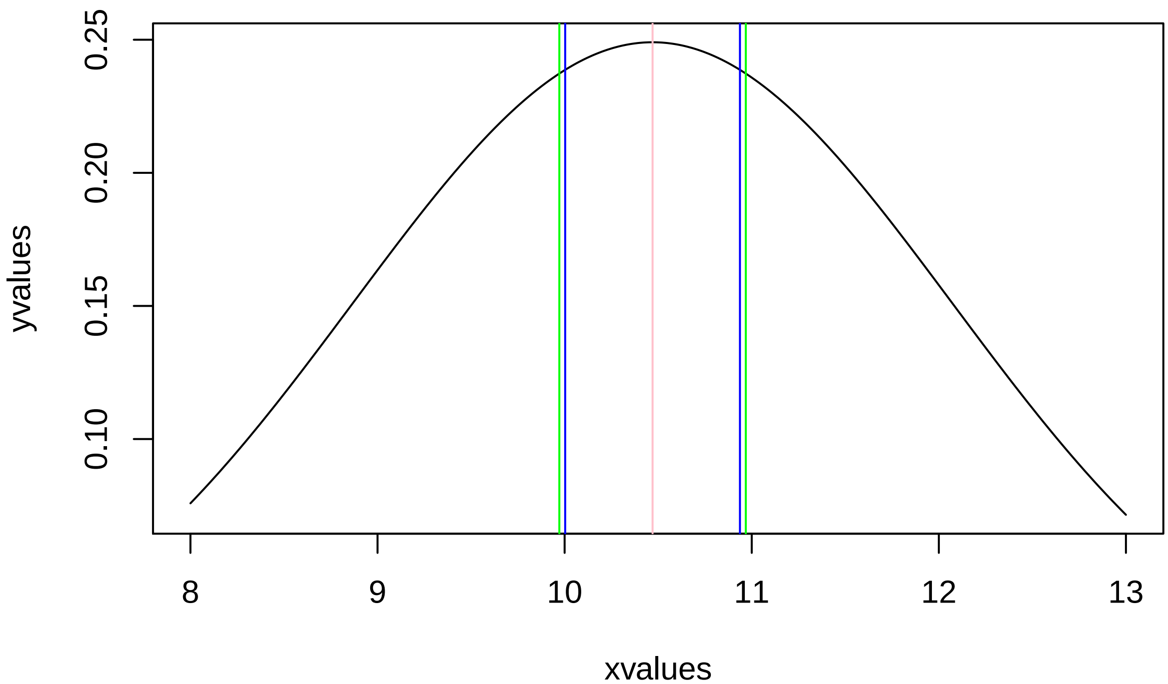
Probability Distribution with Confidence Intervals

Normal Distribution with 90% and 95% Confidence Intervals

90% Confidence Interval represented in blue

95% Confidence Interval represented in green

Sample Mean represented in pink

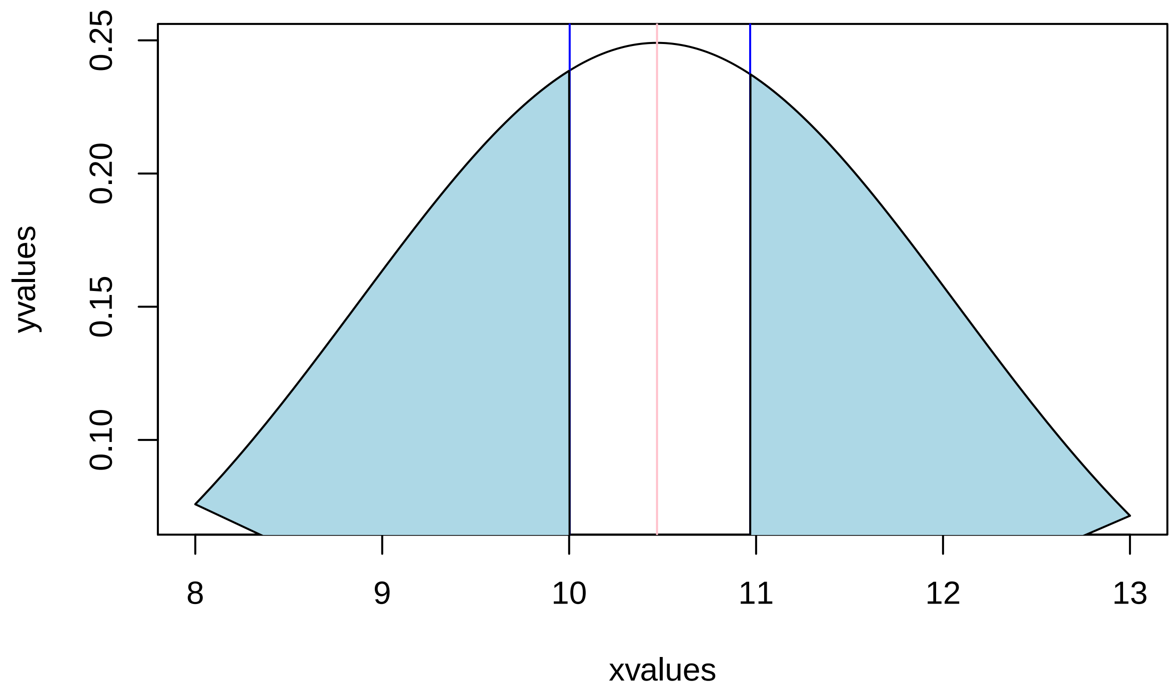


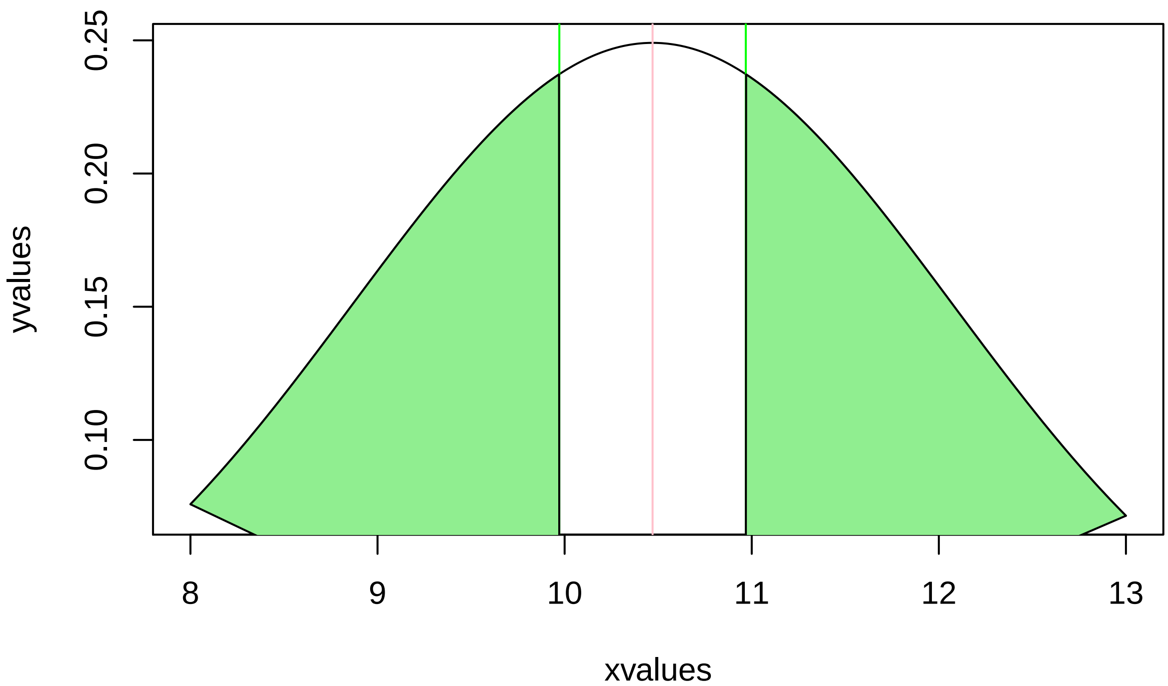
Normal Distribution with Shaded 90% and 95% Confidence Intervals

90% Confidence Interval represented in blue

95% Confidence Interval represented in green

Sample Mean represented in pink


0.10
0.15
0.20
0.25
yvalues
8
9
10
11
12
13
xvalues

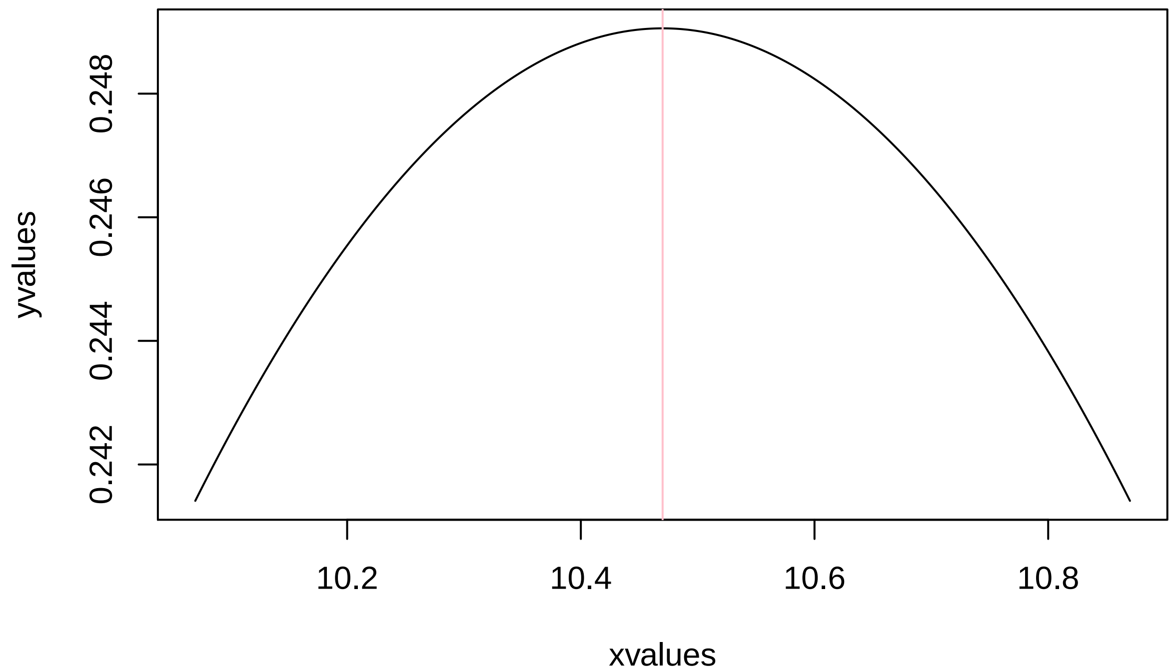


Comparative Normal Distribution with Shaded 90% and 95% Confidence Intervals

90% Confidence Interval represented in blue

95% Confidence Interval represented in green

Sample Mean represented in pink


0.242
0.244
0.246
0.248
yvalues
10.2
10.4
10.6
10.8
xvalues