Analysis of Environmental Data

Chapter 6c. Conceptual Foundations

Confidence Intervals and More

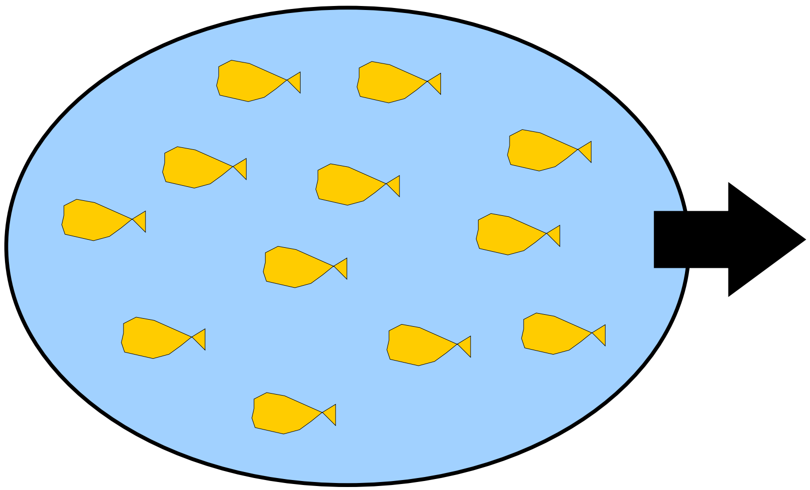
<u>Topics</u>:

1. Population distribution of random variable
2. Z-standardization of population distribution
3. Sample distribution of random variable
4. Z-standardization of sample distribution
5. Sample estimate of population parameter
6. Standard errors
7. Confidence intervals

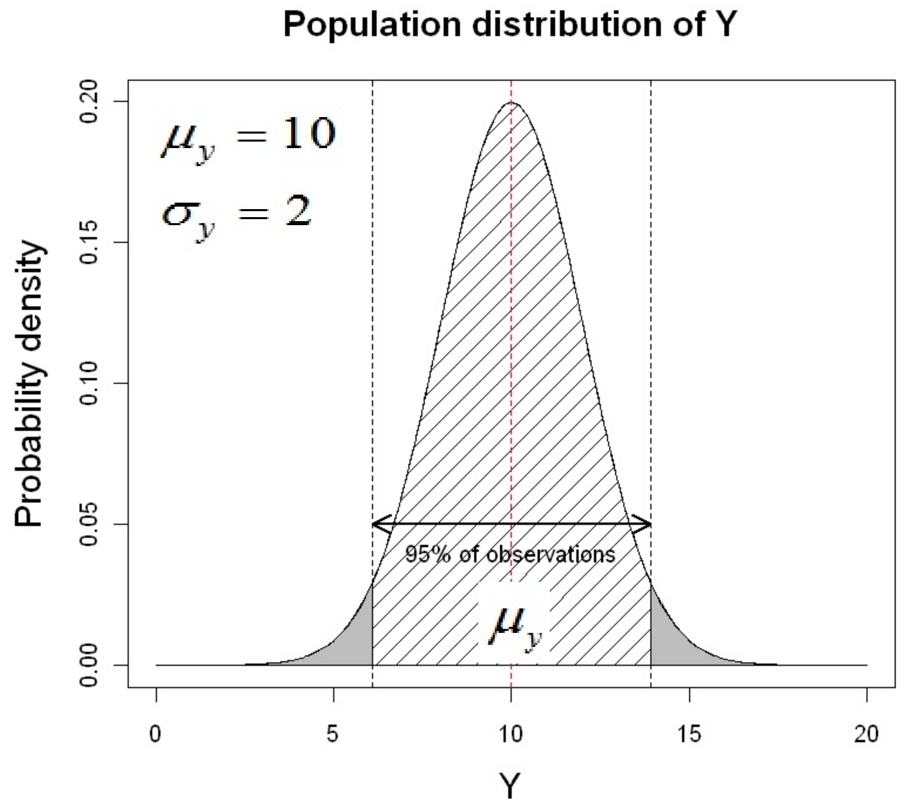
Primer on confidence intervals and more...

Population distribution of a random variable

Population of fish



Y = fish size



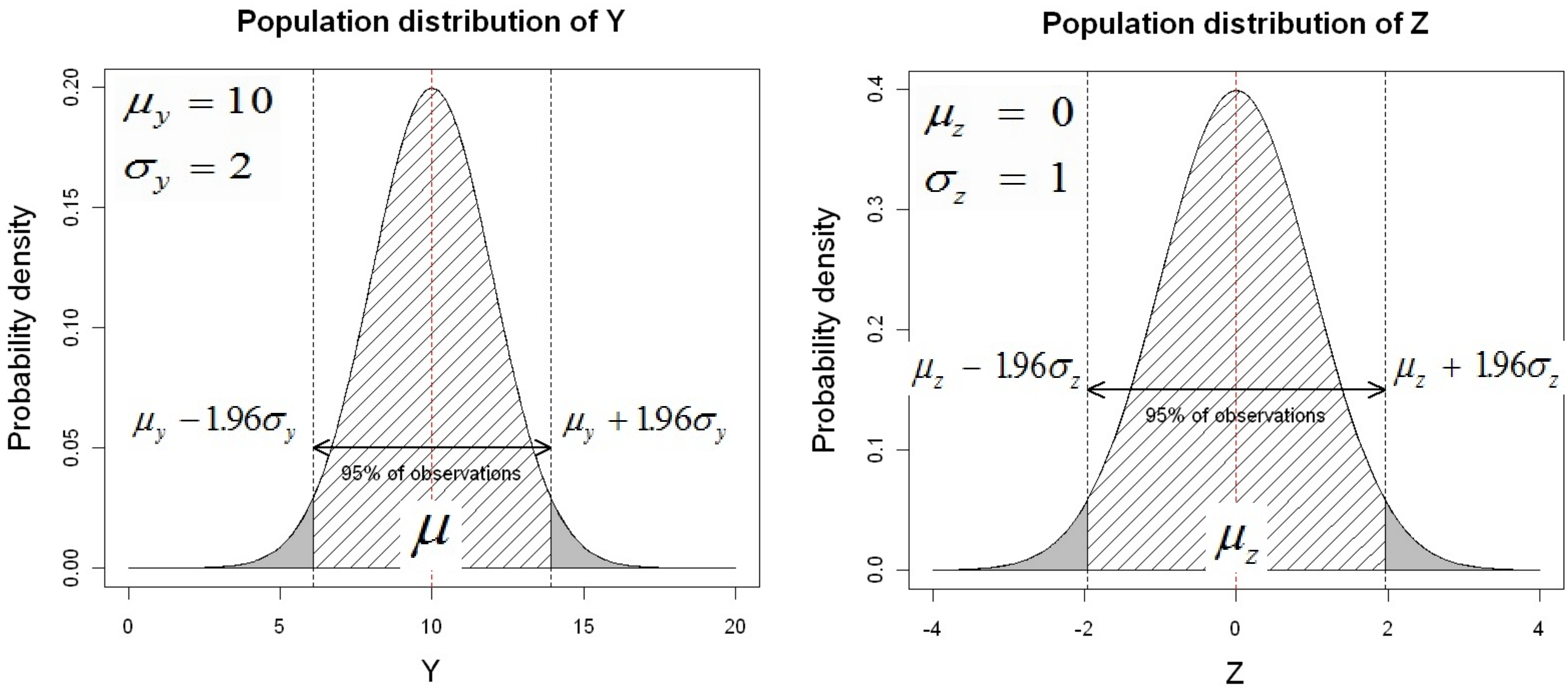
$$\Pr\left\{\mu_y - 1.96\sigma_y \leq Y \leq \mu_y + 1.96\sigma_y\right\} = 0.95$$

This is <u>not</u> a confidence interval!

Primer on confidence intervals and more…

Z standardization of population distribution

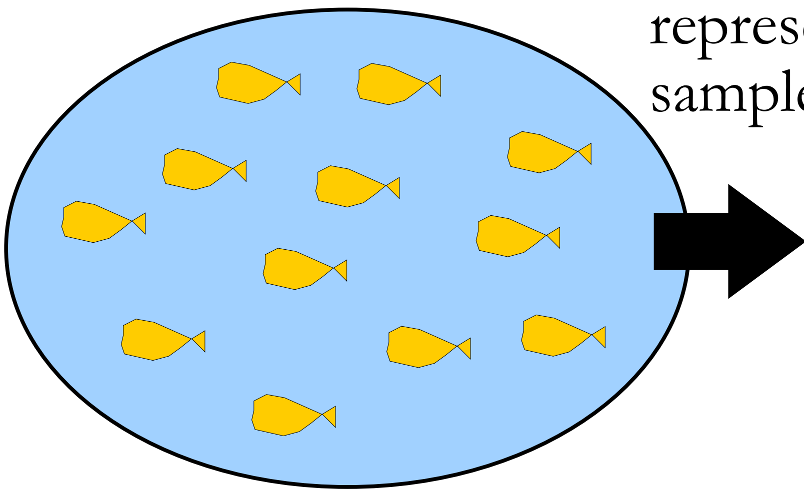
$$z = \frac{y_i - \mu_y}{\sigma_y}$$



Primer on confidence intervals and more…

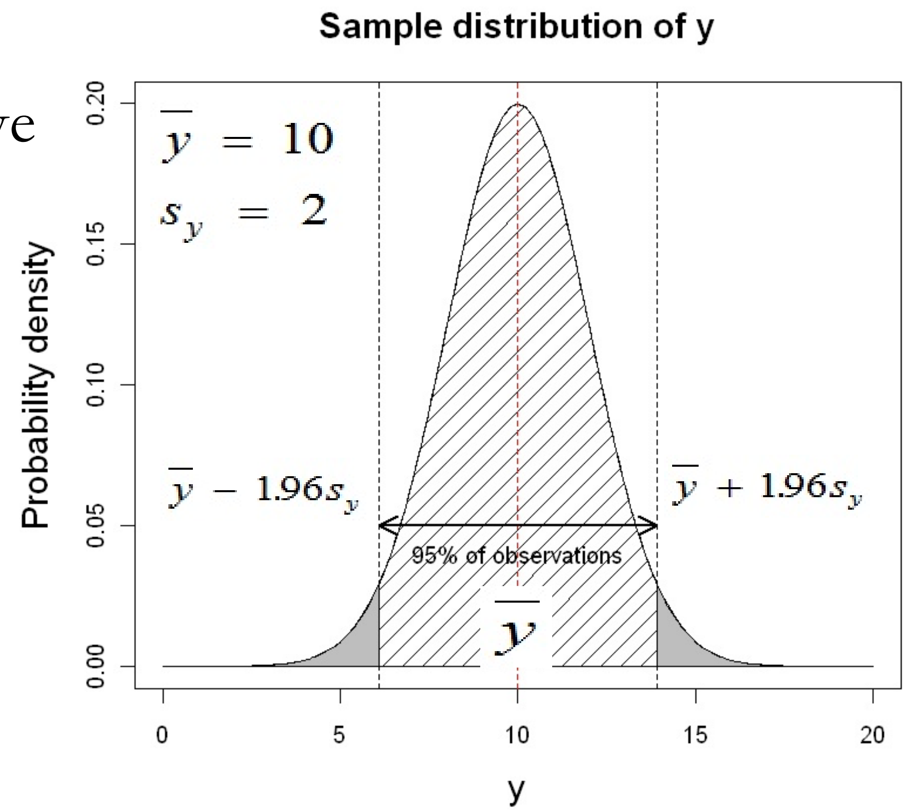
Sample distribution of a random variable

Sample of fish



A perfectly representative sample

Y = fish size



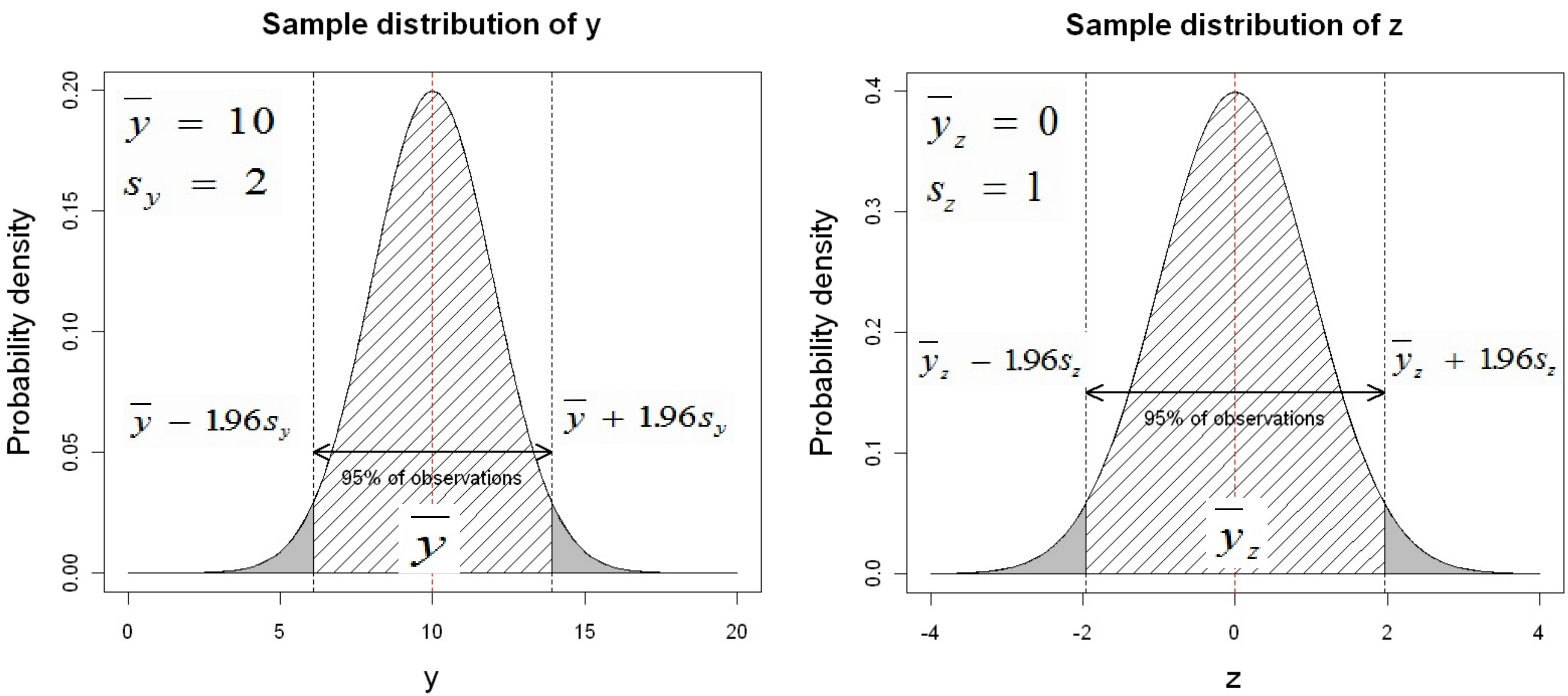
$$\Pr\left\{\overline{y} - 1.96s_y \le y \le \overline{y} + 1.96s_y\right\} = 0.95$$

This is <u>not</u> a confidence interval!

Primer on confidence intervals and more…

Z standardization of sample distribution

$$z = \frac{y_i - \bar{y}}{s_y}$$

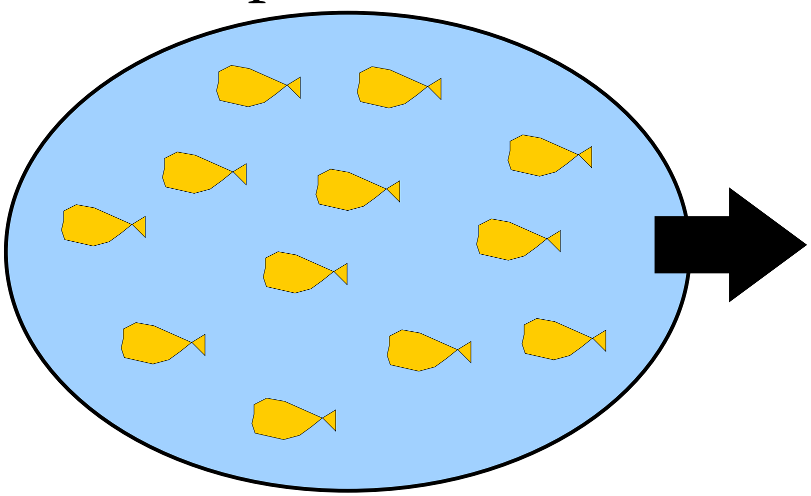


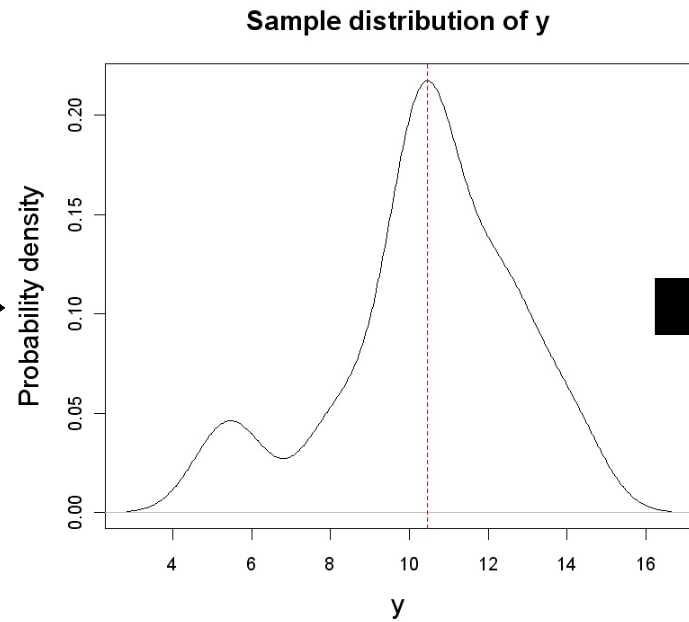
Primer on confidence intervals and more…

Sample estimate of population parameter

Our goal is to estimate the population mean, μ_y, from the sample y

Sample of fish





$$\hat{\mu}_{y1} = \overline{y}_1$$

OLS/MLE

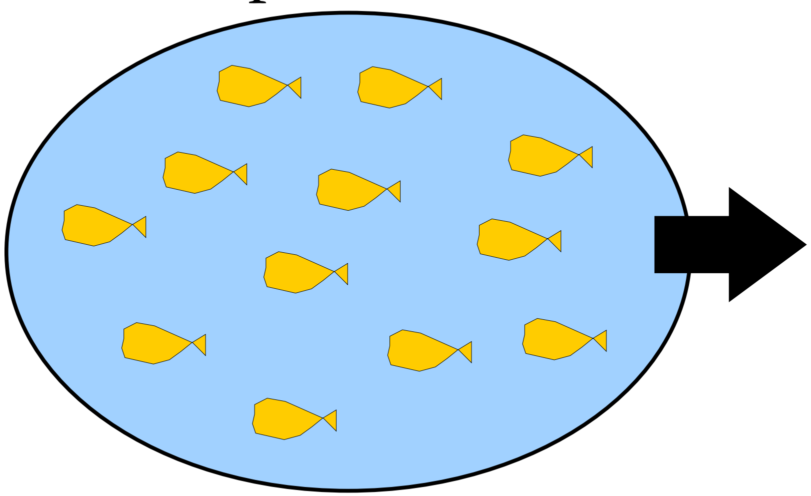
Remember the sample mean is a random variable because it is derived from a random variable

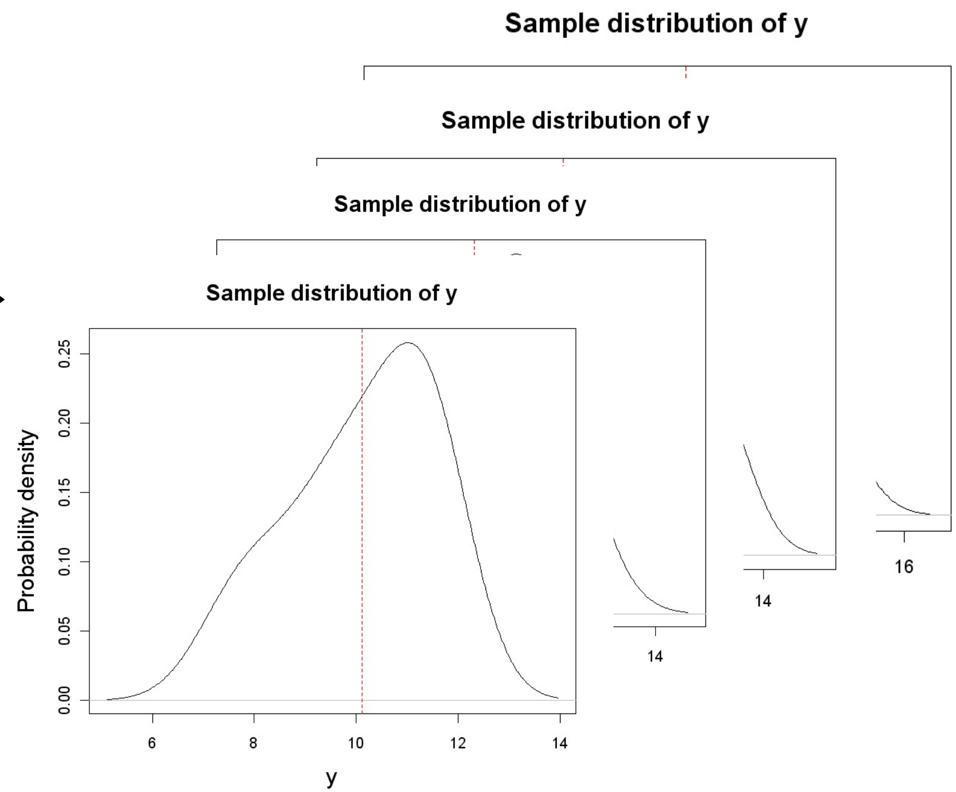
Primer on confidence intervals and more...

Many sample estimates of population parameter

What if we could collect many samples (or even every possible sample) and for each sample compute an estimate of the population parameter?

Sample of fish



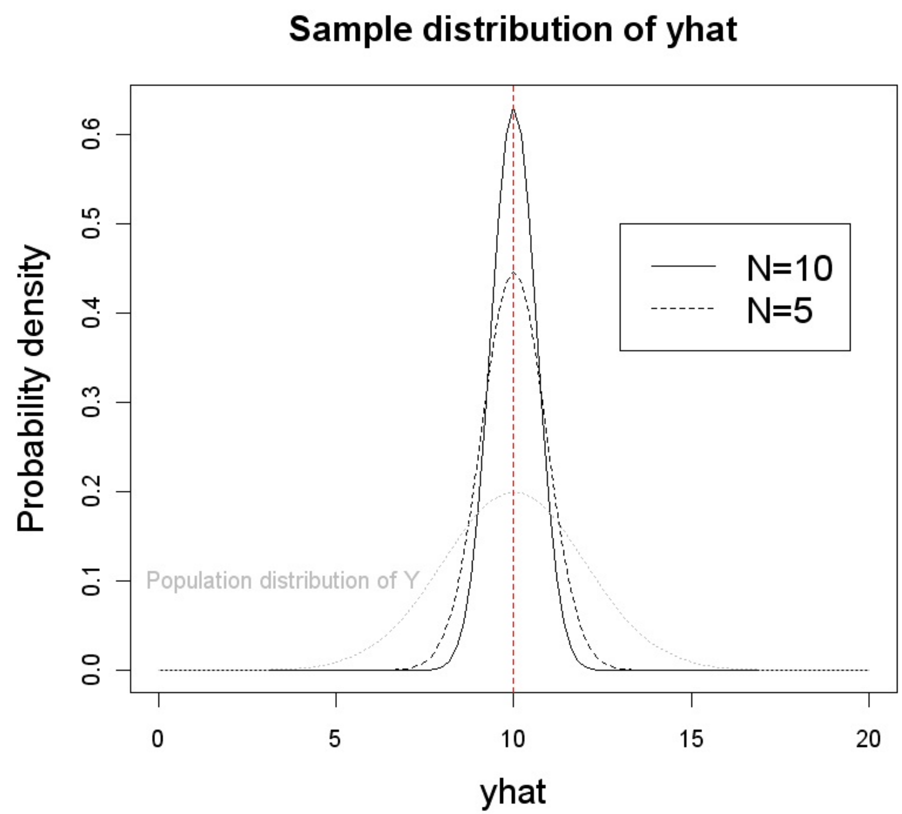


Primer on confidence intervals and more...

Standard error of sample estimates of population parameter

Standard error of the mean:

- If the distribution of the sample means is normal (and CLT says they always are), we can calculate the variance and standard deviation of the sample means – known as the *standard error*
- But with only a single sample, we have to estimate the *standard error* from our sample



Primer on confidence intervals and more…

Standard error of sample estimates of population parameter

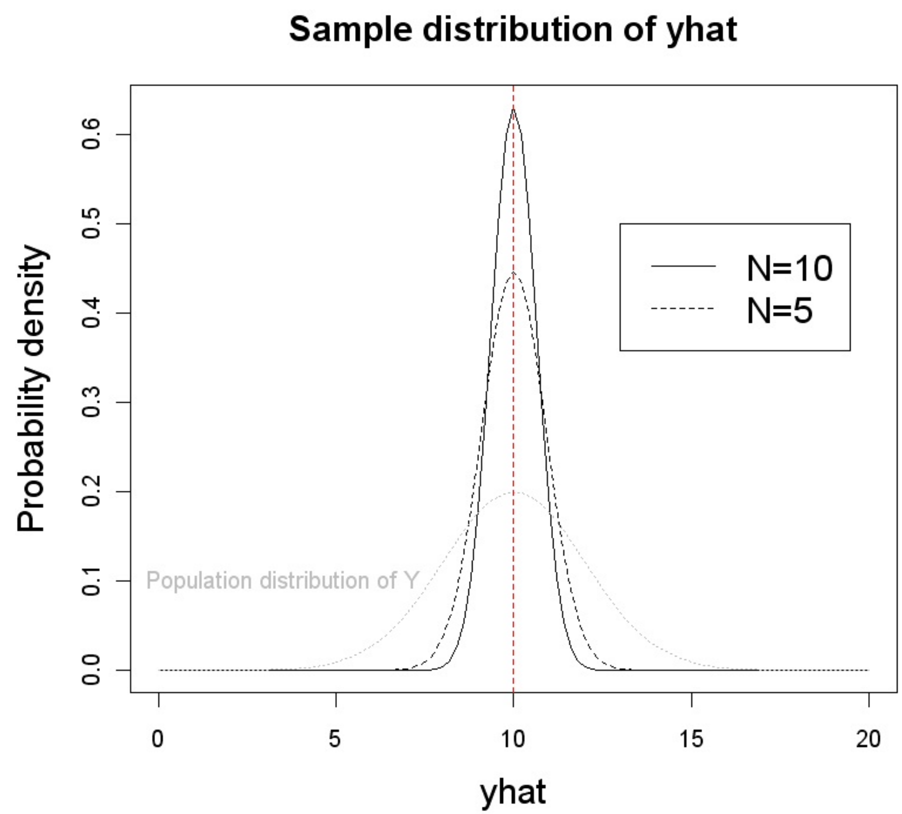
Standard error of the mean:

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}$$

σ_y = population standard deviation

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}}$$

s_y = sample standard deviation

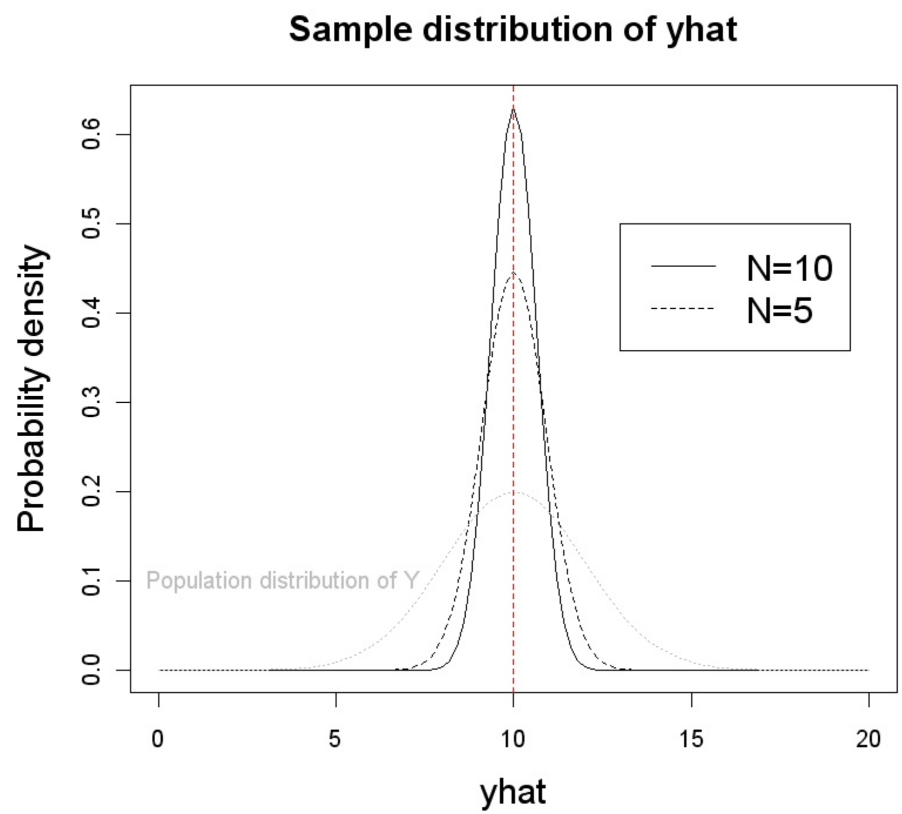


Primer on confidence intervals and more…

Standard error of sample estimates of population parameter

Standard error of the mean:

- Tells us about the variation in our sample mean (under repeated sampling)
- Tells us about the "error" in using the sample mean to estimate the population mean
- Smaller variance in the population and larger sample size decrease the error in our estimate



Primer on confidence intervals and more...

Confidence interval for the sample estimate of population parameter

Confidence interval for the mean:

- Convert the distribution of *sample means* into a standard normal distribution via the *z*-score standardization

σ_y = population standard deviation

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}} \qquad z = \frac{\bar{y} - \mu_y}{\sigma_{\bar{y}}}$$

$$\Pr\left\{\bar{y} - 1.96\sigma_{\bar{y}} \le \mu_y \le \bar{y} + 1.96\sigma_{\bar{y}}\right\} = 0.95$$

This is a confidence interval!

Primer on confidence intervals and more...

Confidence interval for the sample estimate of population parameter

Confidence interval for the mean:

- z variable (standard normal) is called a t statistic when we use the *sample estimate of the standard error* of the mean

s_y = sample standard deviation

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}}$$

$$t = \frac{\bar{y} - \mu_y}{s_{\bar{y}}}$$

$$\Pr\left\{\bar{y} - t_{0.025(n-1)} s_{\bar{y}} \leq \mu_y \leq \bar{y} + t_{0.025,n-1} s_{\bar{y}}\right\} = 0.95$$

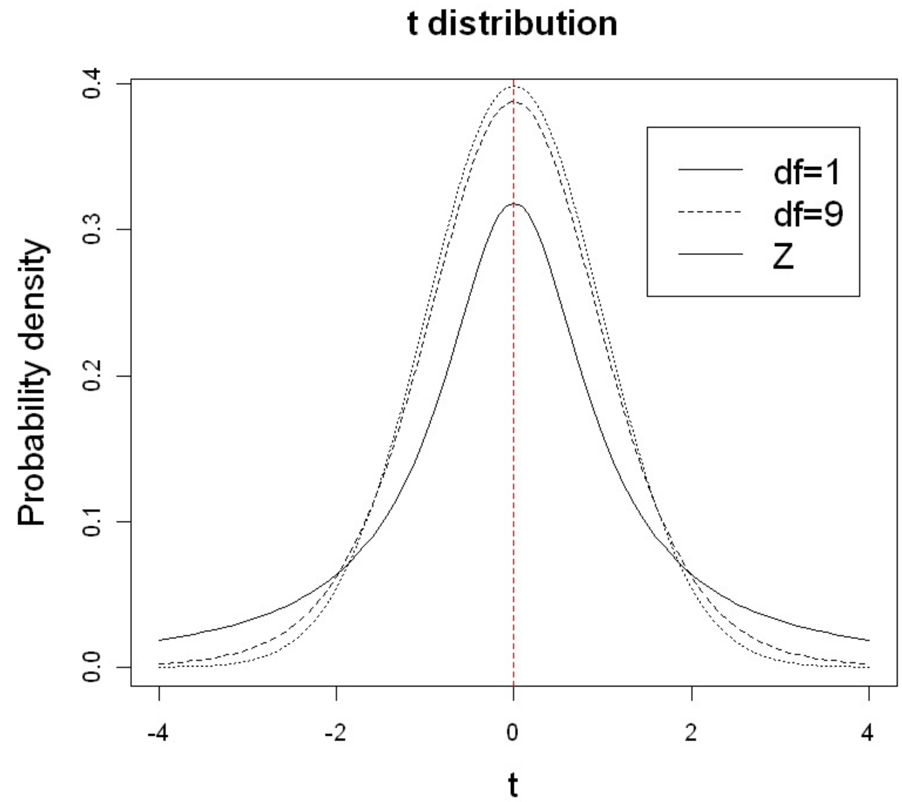
This <u>is</u> a confidence interval!

Primer on confidence intervals and more...

Confidence interval for the sample estimate of population parameter

What is a *t* statistic?

- *t* distribution is a symmetrical probability distribution centered around zero
- Similar to the normal distribution except varies with sample size (actually degrees of freedom, *n*-1); has slightly fatter tails than the normal but approaches the normal when *n* (>30) is large

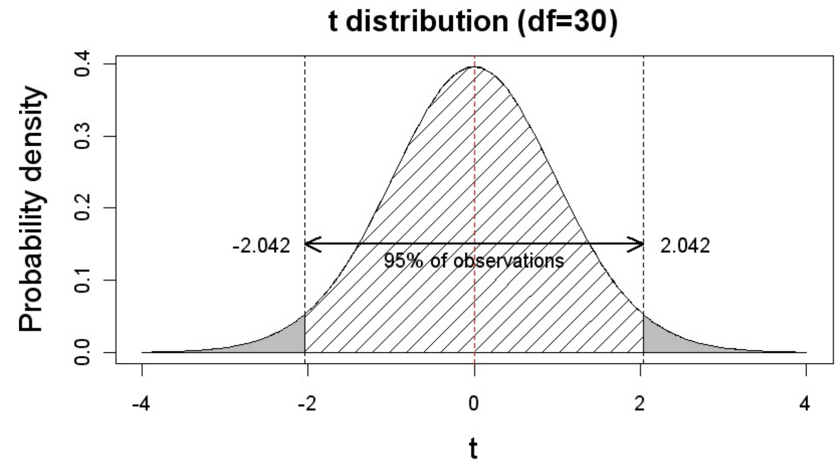


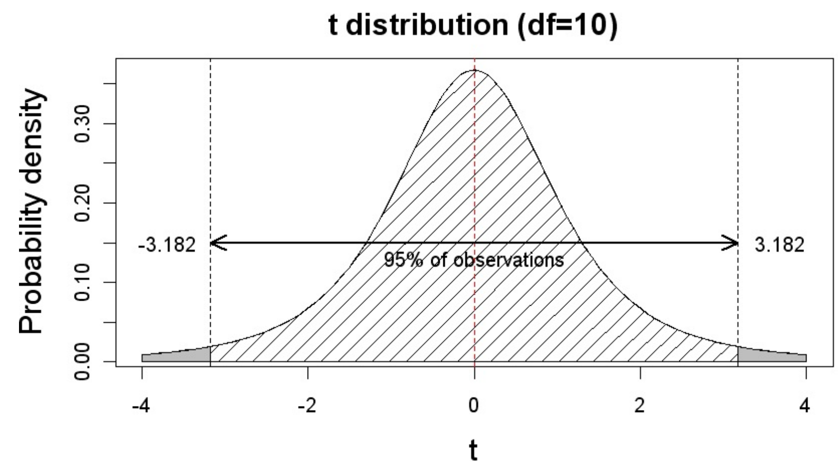
Primer on confidence intervals and more...

Confidence interval for the sample estimate of population parameter

What is a *t* statistic?

- *t* distribution is a symmetrical probability distribution centered around zero
- Similar to the normal distribution except varies with sample size (actually degrees of freedom, *n*-1); has slightly fatter tails than the normal but approaches the normal when *n* (>30) is large





Primer on confidence intervals and more...

Confidence interval for the sample estimate of population parameter

What is a *t* statistic?

- *z* distribution is the probability distribution of the *z*-standardized data <u>or</u> the *z*-standardized sample means based on <u>population</u> standard error of mean

$$z = \frac{y_i - \mu_y}{\sigma_y} \qquad z = \frac{y_i - \bar{y}}{s_y}$$

$$z = \frac{\bar{y} - \mu_y}{\boxed{\sigma_{\bar{y}}}}$$

- *t* distribution is the probability distribution for the *z*-standardized sample means based on <u>sample estimate</u> of the standard error of mean

$$t = \frac{\bar{y} - \mu_y}{\boxed{s_{\bar{y}}}}$$