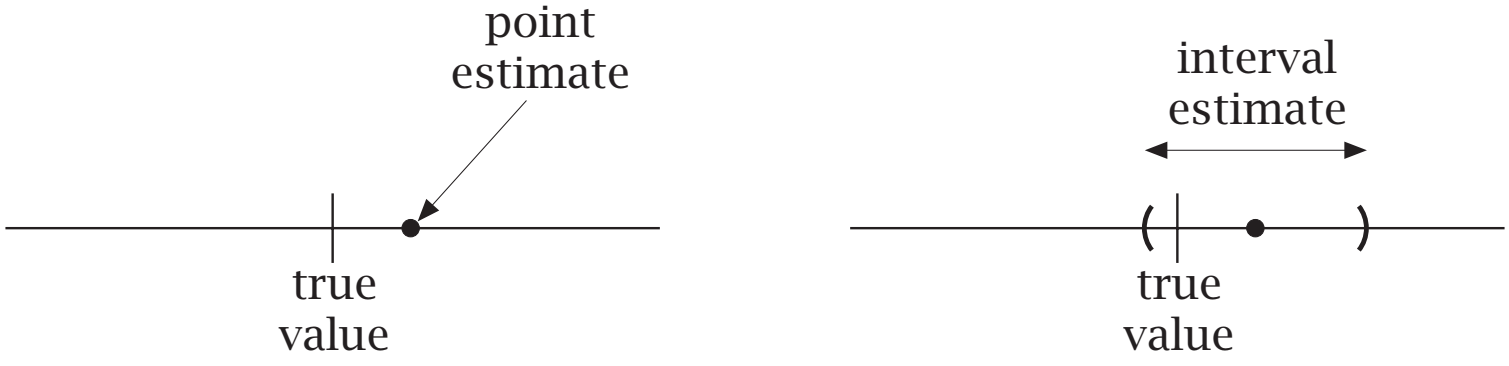
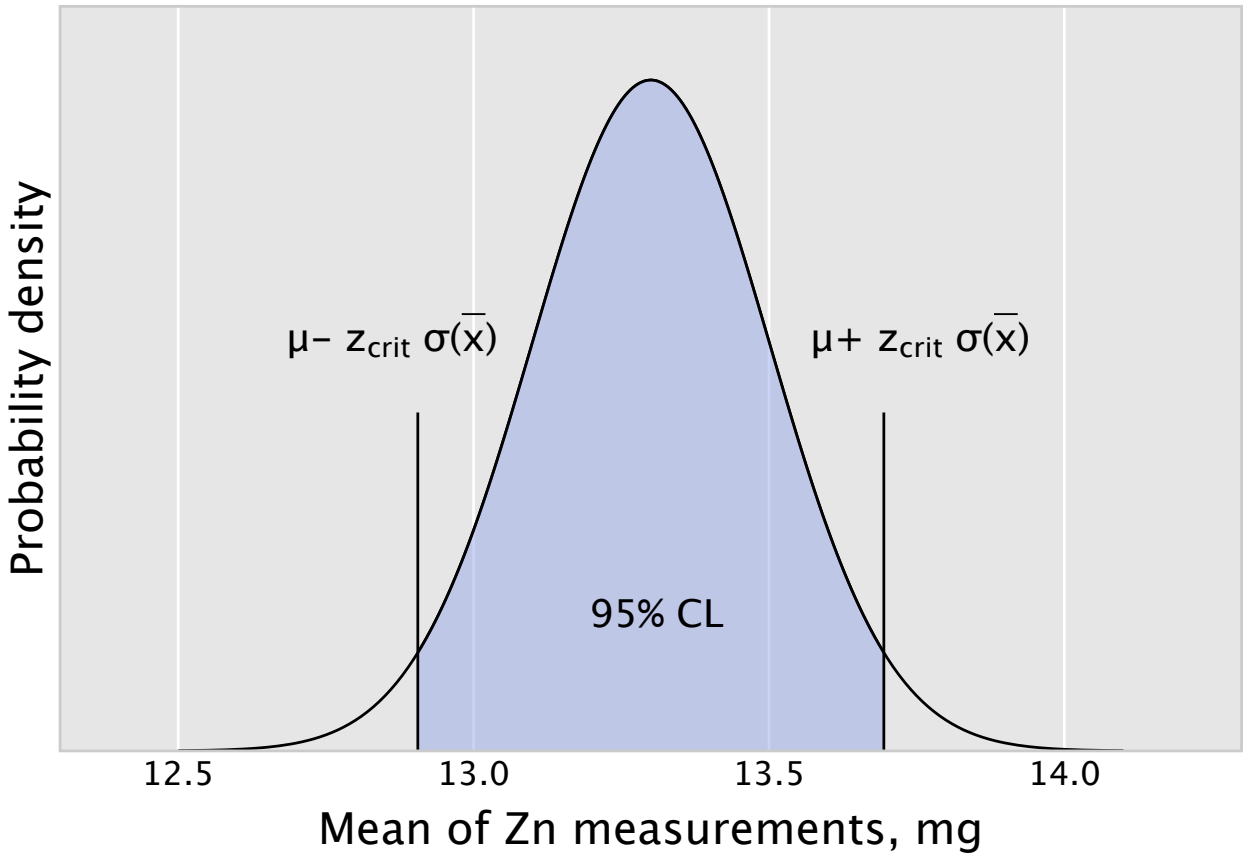
Interval Estimates

What is the difference between a point and an interval estimate?



- Point estimates gives the single, best estimate of the population parameter but gives no indication of the uncertainty of the estimate
- One can additionally provide the standard error of the estimate, but its interpretation can be a little vague, especially for non-normal sampling distributions
- A **confidence interval** is an interval that contains the population parameter with a stated level of certainty
- The *confidence level*, CL, quantifies our degree of certainty; the *significance level*, SL, is $1 - \text{CL}$

Calculate a 95% *prediction interval* for the mean of five zinc measurements taken from a normal distribution with $\mu = 13.3\,\text{mg}$ and $\sigma = 0.45\,\text{mg}$.



- Prediction intervals (PIs) are for *future measurements*.
- We want an interval that, with 95% probability will contain the mean of five future measurements
- General expression for a prediction interval for normally-distributed $\overline{x}_n$ is

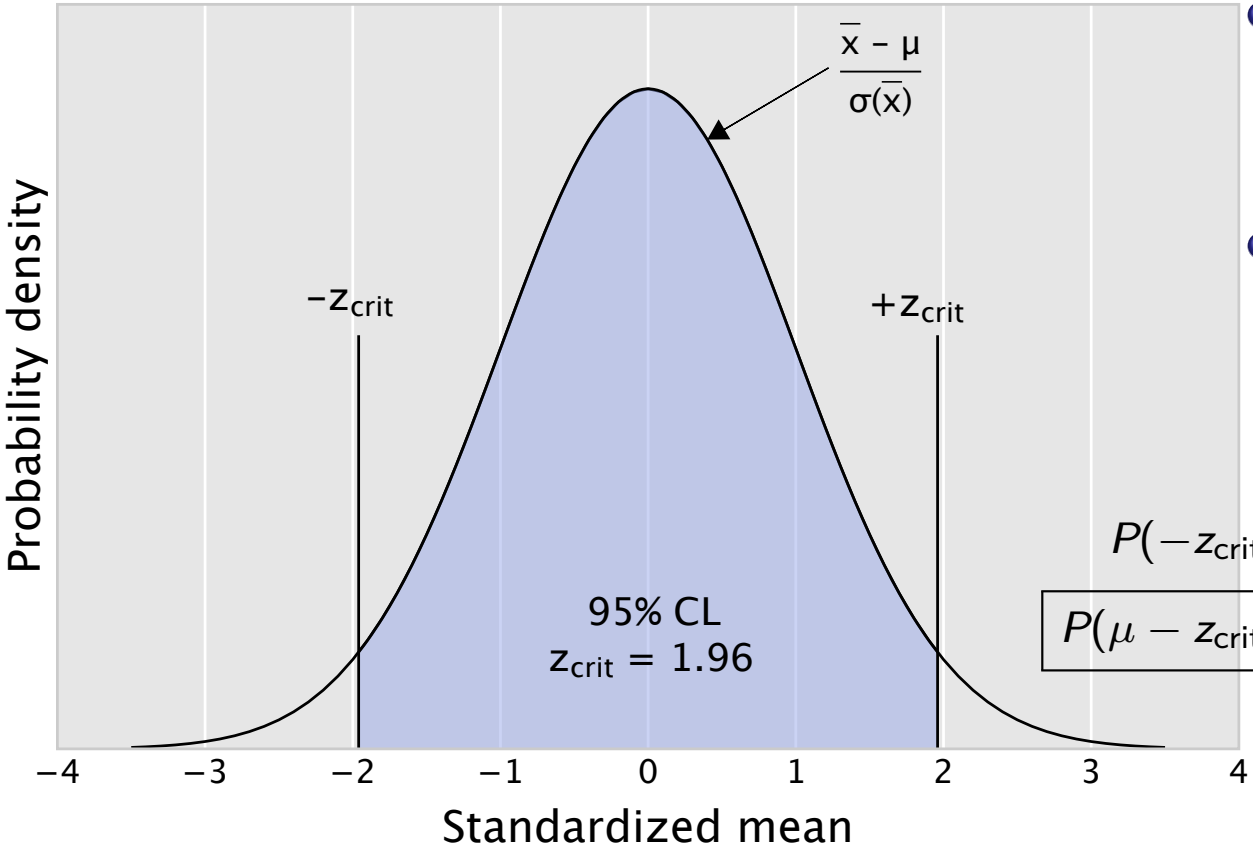
$$\mu_x \pm z_{\text{crit}}\sigma(\overline{x}_n)$$

where $\sigma(\overline{x}_n) = \sigma_x/\sqrt{n}$

- z_{crit} is the *critical value* which is determined by the sampling distribution of the mean and (shown above) and the desired confidence level
- If $\overline{x}_n$ is normally distributed then **we obtain z_{crit} from the standard normal distribution**

Critical Values

Explain why the critical values, $\pm z_{\text{crit}}$, from a standard normal distribution gives us the desired prediction interval.



- Since $\bar{x}$ is normally distributed, its standardized value follows a standard normal distribution
- Figures and eqns show connection between CL and the critical values

$$P(-z_{\text{crit}} < \frac{\bar{x} - \mu}{\sigma(\bar{x})} < z_{\text{crit}}) = CL$$

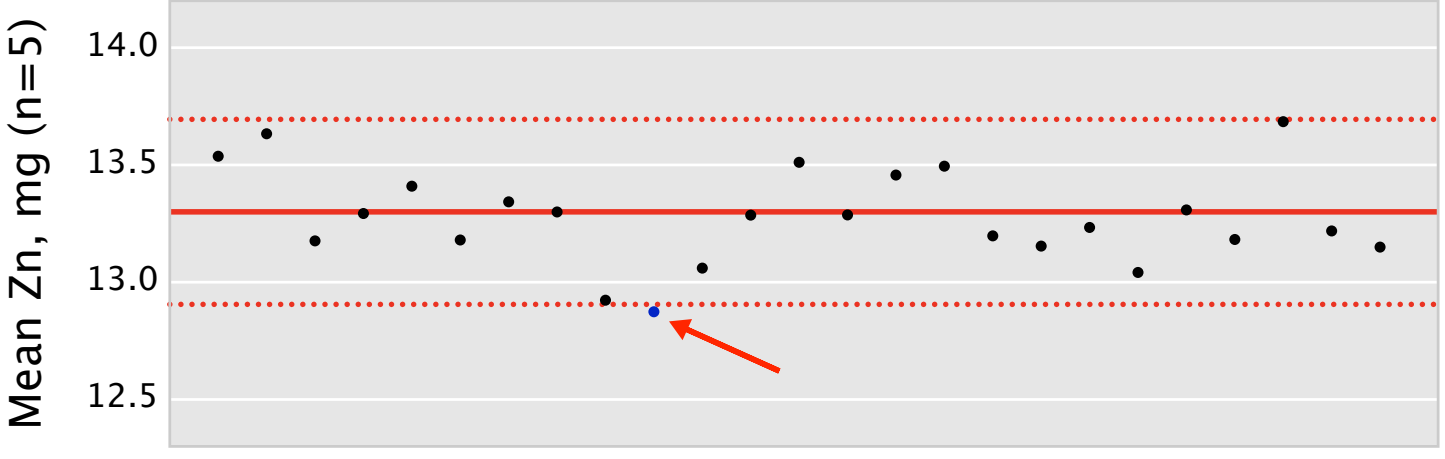
$$P(-z_{\text{crit}} \cdot \sigma(\bar{x}) < \bar{x} - \mu < z_{\text{crit}} \cdot \sigma(\bar{x})) = CL$$

$$\boxed{P(\mu - z_{\text{crit}} \cdot \sigma(\bar{x}) < \bar{x} < \mu + z_{\text{crit}} \cdot \sigma(\bar{x})) = CL}$$

The quantity $\pm z_{\text{crit}} \cdot \sigma(\bar{x})$ is sometimes called the **margin of error** (MOE). Also remember that, from the CLT, $\sigma(\bar{x}_n) = \frac{\sigma_x}{\sqrt{n}}$

Simulation Demo

Use a simulation to illustrate the prediction interval and produce a figure like the one below and to verify that $z_{\text{crit}} = 1.96$ gives a confidence level (CL) of 95%.



- The 95% PI is 13.30 ± 0.39, indicated by the dotted lines.
- First 25 (of 1000) samples are shown
- Since the CL is 95%, typically 0–3 (avg 1.25) points will be outside the PI

Prediction *vs* Confidence

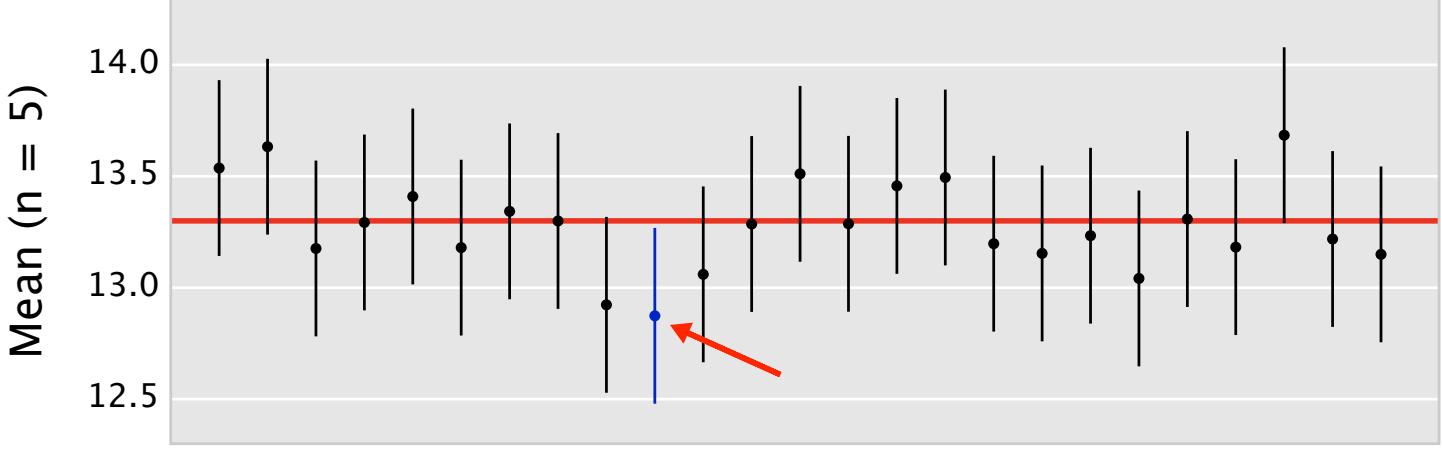
What is the connection between prediction and confidence intervals?

- While prediction intervals are for *future* observations (with stated CL), confidence intervals are meant to predict the population mean (with stated CL)

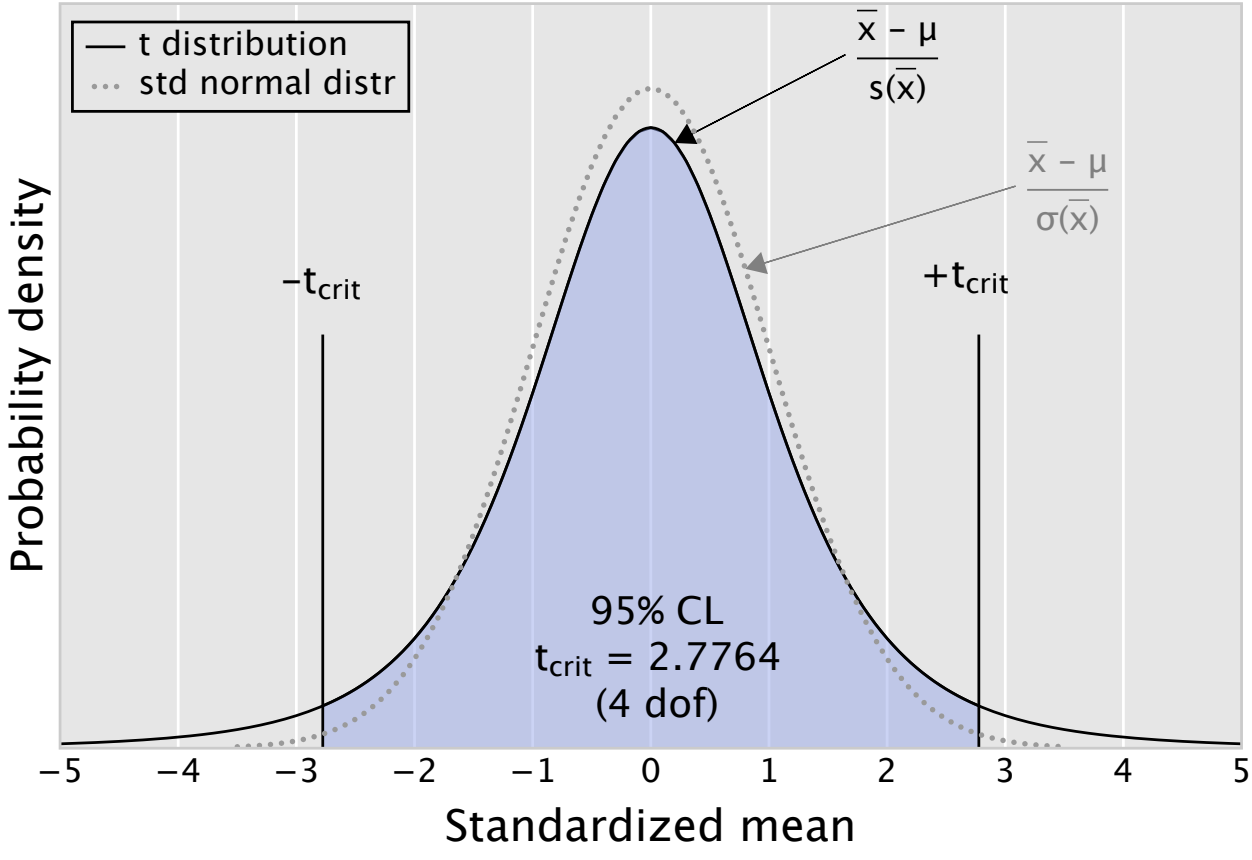
$$\bar{x} \pm z_{\text{crit}} \cdot \sigma(\bar{x}) \qquad \text{(CI for } \mu\text{, } \sigma \text{ known)}$$

  Key point: the width of both intervals at a given CL is the same (ie, it is the margin of error, MOE).

- **The figure below shows the same sample means** as the previous slide, but now the intervals are centered on each point to indicate the 95% confidence intervals.



What if we don't know the population variance? How do we construct confidence intervals for the population mean μ?



- Standardizing with $s(\bar{x}) = s_x/\sqrt{n}$ instead of $\sigma(\bar{x})$ results in a different distribution: the **Student's t-distribution**
- This is a family of distributions with a single parameter, ν, the *degrees of freedom*
- In this case, $s(\bar{x})$ will have $\nu = n - 1$ degrees of freedom
- $t_{crit} > z_{crit}$ and as $n \to \infty$, $t_{crit} \to z_{crit}$

Following the same logic as previously we can derive that

$$P(\bar{x} - t_{crit} \cdot s(\bar{x}) < \mu < \bar{x} + t_{crit} \cdot s(\bar{x})) = CL$$

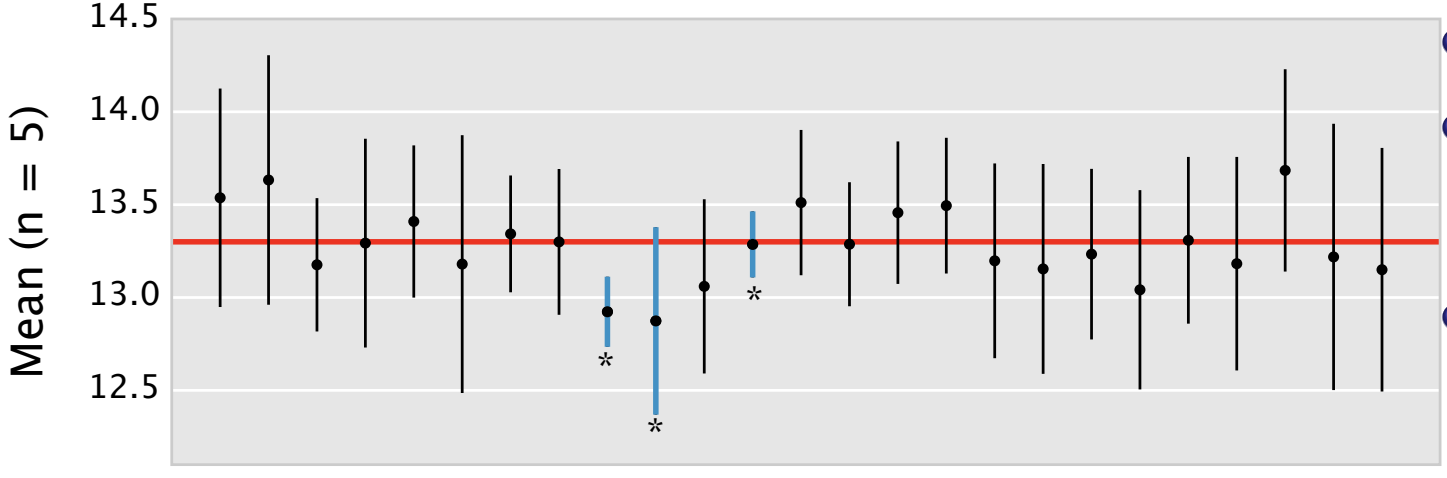
which assumes that $\bar{x}$ is normally distributed.

Confidence Intervals for μ with σ Unknown

Demonstrate that using the critical t-values with the sample standard deviation will yield a valid confidence interval.

$$\bar{x} \pm t_{\text{crit}} \cdot s(\bar{x}) \qquad \text{(CI for } \mu_x\text{)}$$

$$\bar{x} \pm 2.77645 \cdot \frac{s_x}{\sqrt{5}} \qquad \text{(this case)}$$



- CL: 95%
- MOE widths are now variable (dependent on s_x)
- Same sample means as two previous figures

- One interval above doesn't contain μ...but not the same one as before!
- Important to think of CIs **as sample statistics**, just like the sample mean or sample std deviation. They are random variables that change from sample to sample.

The Meaning of Confidence

What is the proper way to interpret a confidence interval? What, exactly, is the confidence level describing?

- Incorrect: "The confidence level is the probability that the confidence interval contains the population parameters it is estimating."

  This statement is misleading at best.

- Correct: *The confidence level is the probability that the procedure will generate an interval that contains the population parameter.*

  Another way to think if it is: if there are 100 people in a room and all obtain measurements and calculate the resulting confidence interval, on average 95 of them will have calculated intervals that contain the true value. The important principle is that ***it is the interval that is variable***.