

Percentile

**Bootstrap CIs**

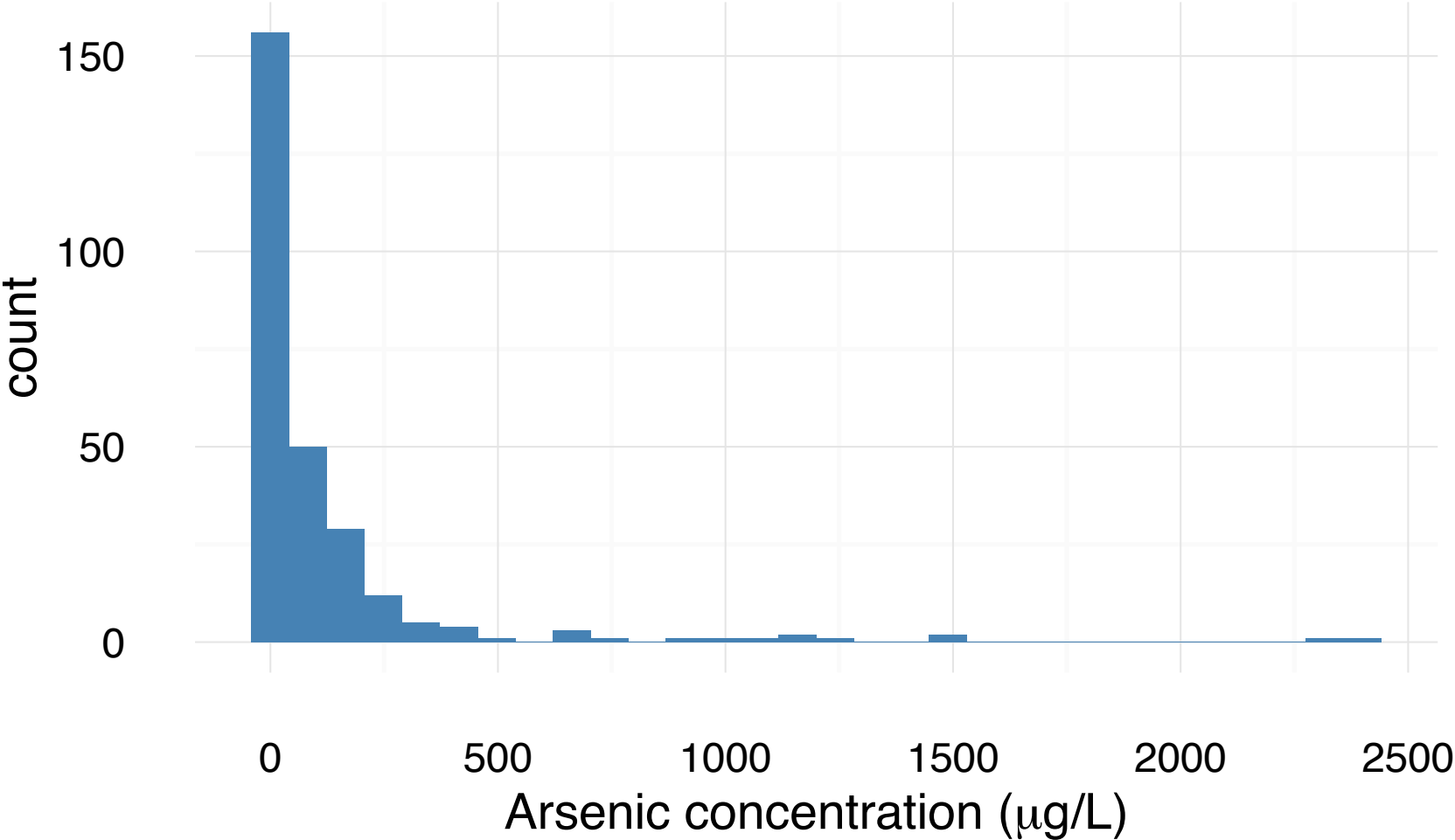
# Arsenic in groundwater

**Occurs naturally in the groundwater in Bangladesh**

**Groundwater is often used for drinking water in rural areas**

**Arsenic poisoning is a serious public health issue**

# The data



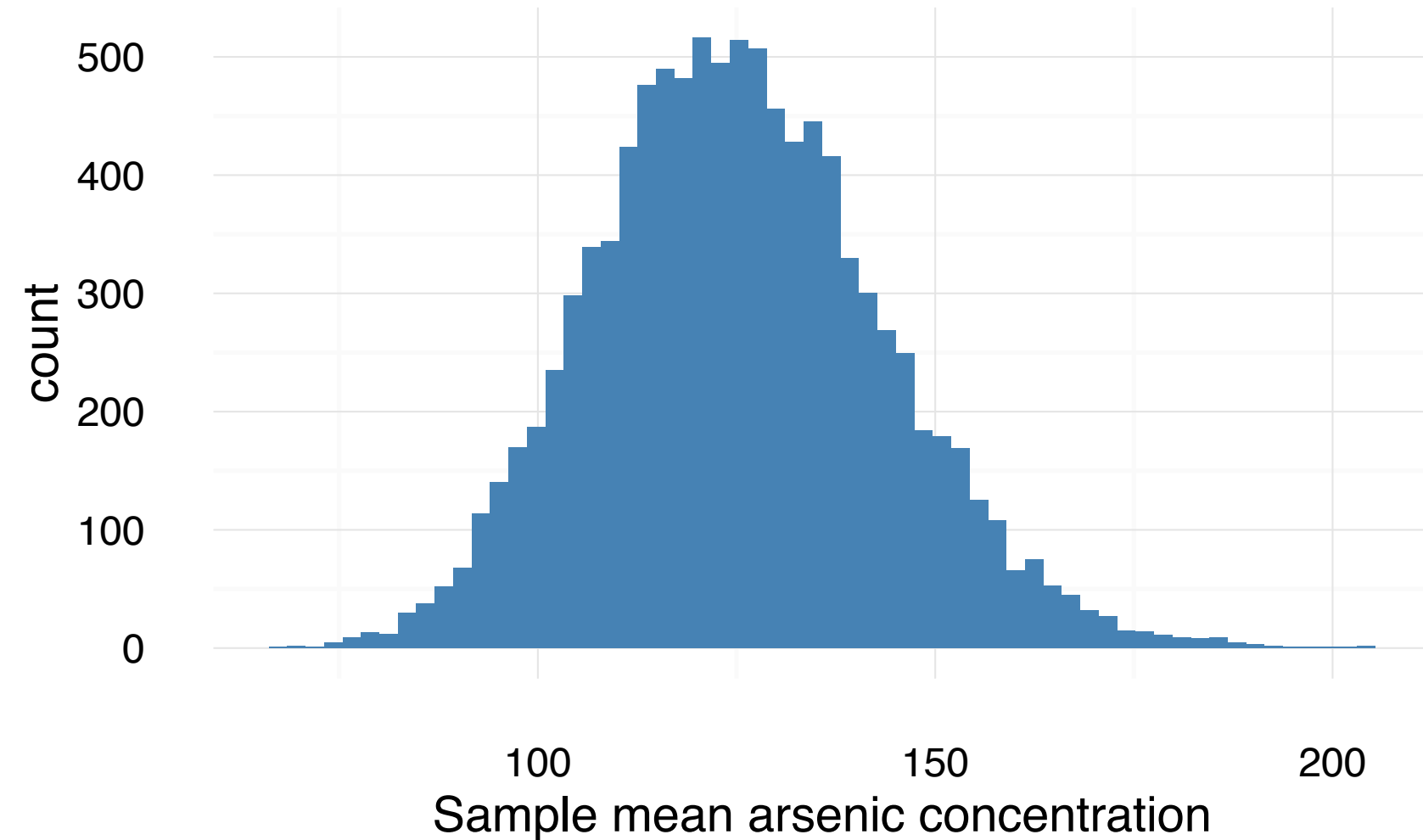
min	Q1	median	Q3	max	mean	sd	n
0.5	6	22	109	2400	125.3199	297.9755	271

# Bootstrapping arsenic concentrations

**Observed sample mean :  
125.3199**

**Mean of bootstrap distribution:  
125.2212**

**Standard error of bootstrap  
distribution: 18.24569**



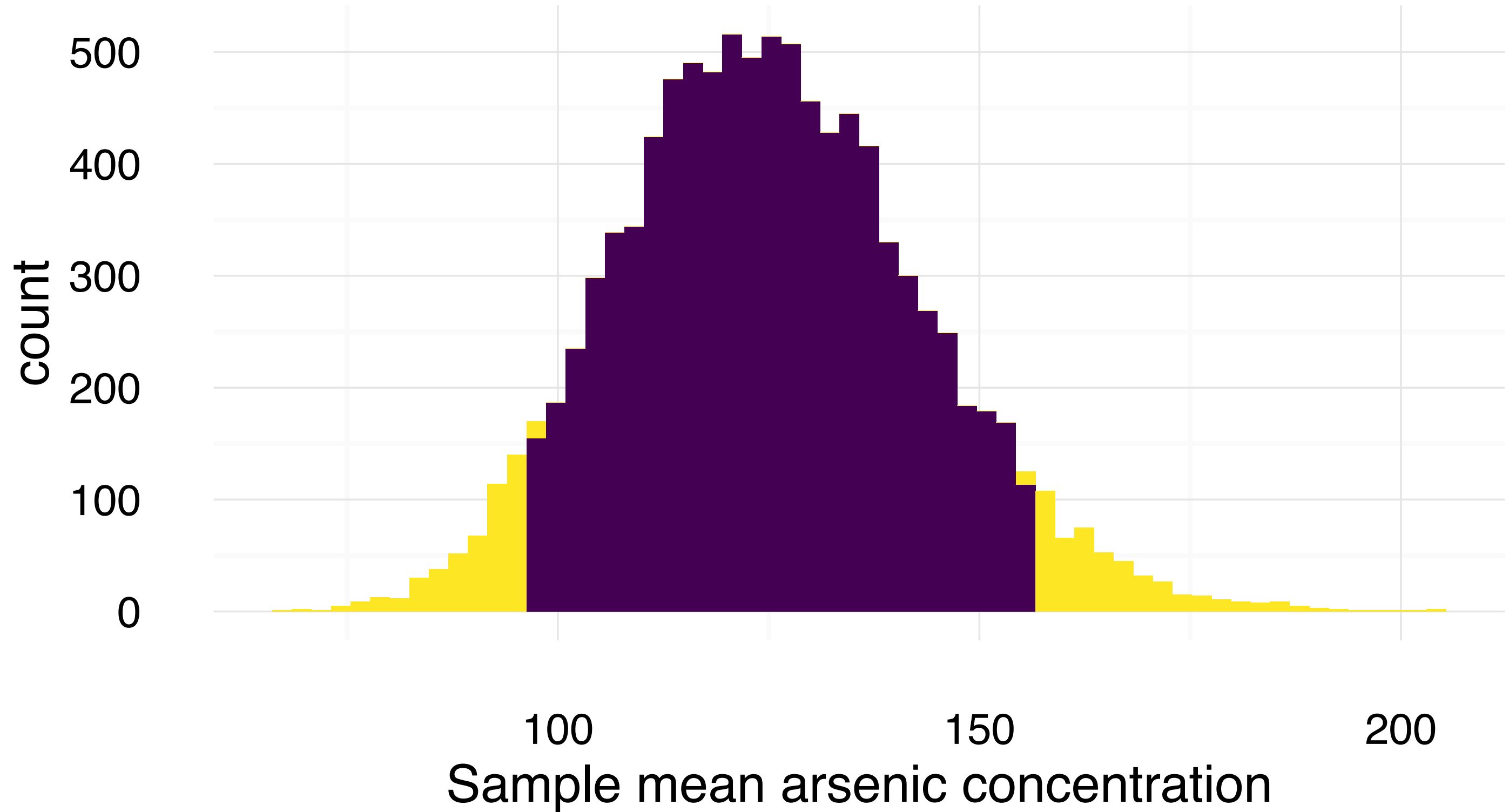
## Percentile method for bootstrap CIs

**To form  $P \cdot 100\%$  bootstrap confidence intervals use the following endpoints:**

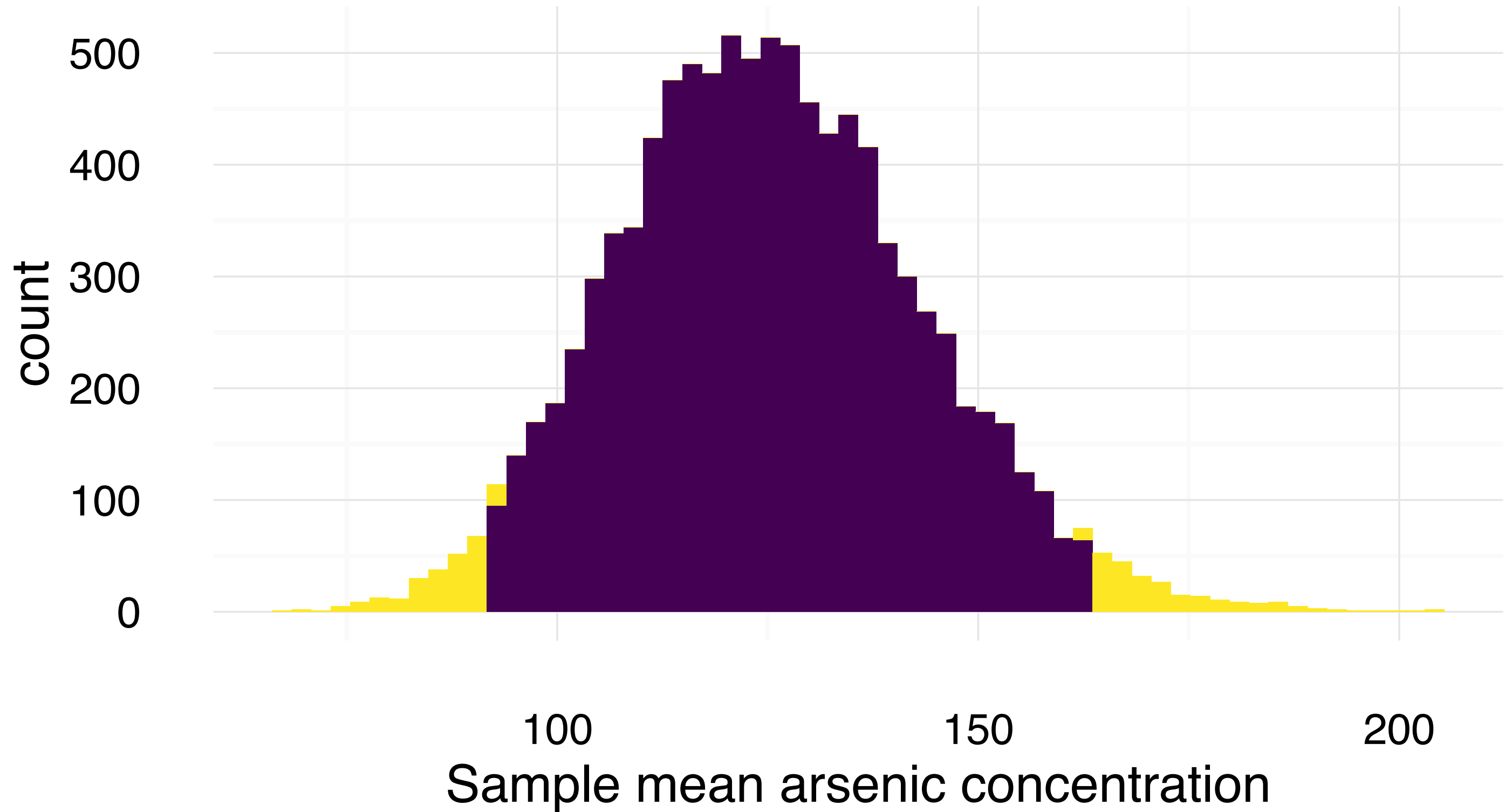
**Lower endpoint:  $(1 - P)/2$  percentile**

**Upper endpoint:  $1 - (1 - P)/2$  percentile**

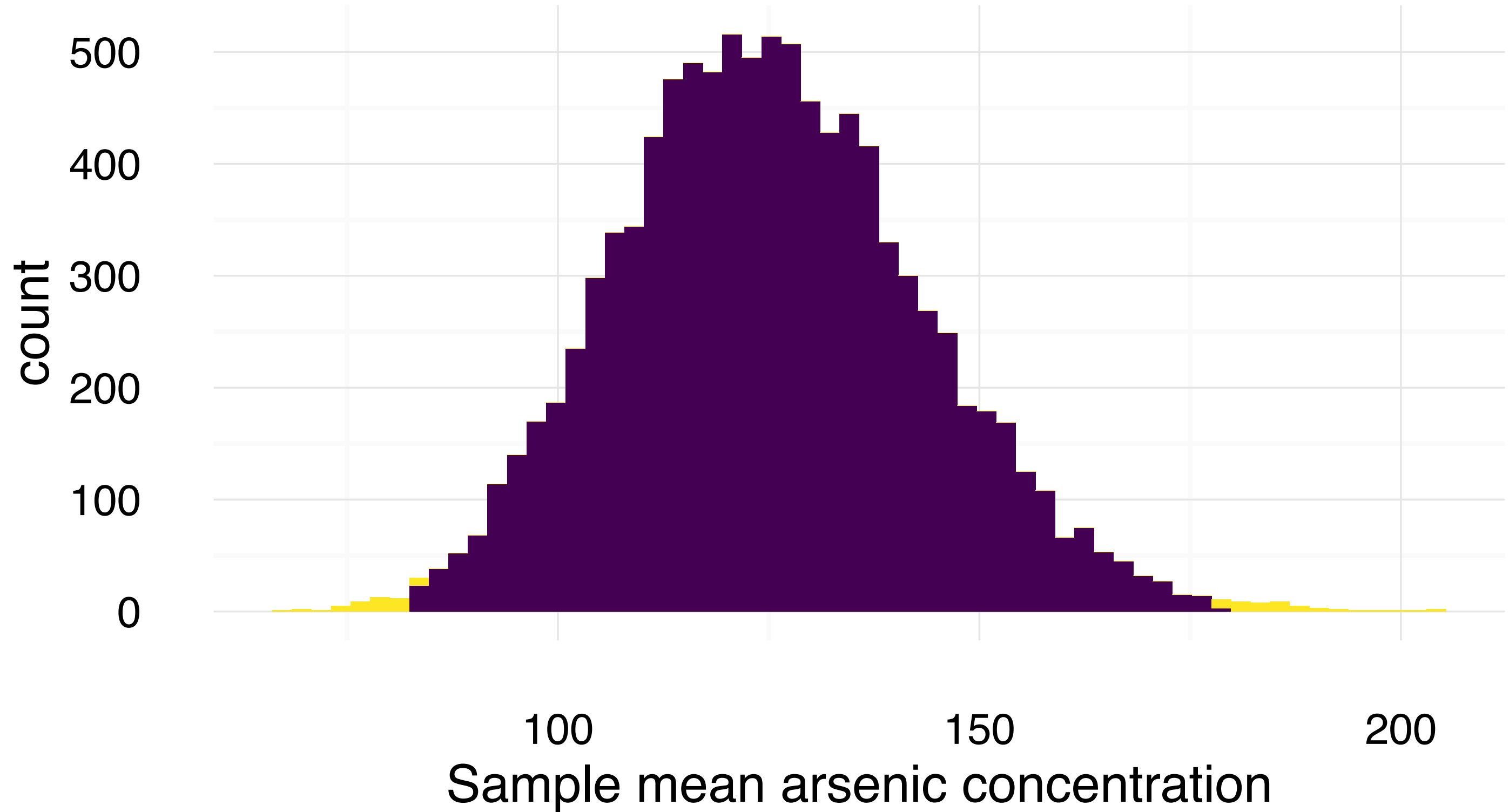
# 90% bootstrap CI



# 95% bootstrap CI



# 99% bootstrap CI





99% CI for  $\mu$ : 82.78694 to 179.2097  $\mu\text{g/L}$

**A. We are 99% sure that all arsenic levels will be between 82.79 and 179.21  $\mu\text{g/L}$ .**

B. 99% of arsenic levels will be between 82.79 and 179.21  $\mu\text{g/L}$ .

**C. We are 99% sure that the average arsenic levels for this sample will be between 82.79 and 179.21  $\mu\text{g/L}$ .**

99% CI for  $\mu$ : 82.78694 to 179.2097  $\mu\text{g/L}$

**D. We are 99% sure that the mean arsenic concentration in the groundwater is between 82.79 and 179.21  $\mu\text{g/L}$ .**

E. We are 99% sure that the confidence interval for the mean arsenic concentration in the groundwater will be between 82.79 and 179.21  $\mu\text{g/L}$ .

**F. There is a 99% chance that the mean arsenic concentration is between 82.79 and 179.21  $\mu\text{g/L}$ .**

# Cautions

**Percentile intervals should not be used for small samples**

**Rule of thumb: bootstrap distribution should appear to be smooth**

**Applies to the bootstrap in general**

# Estimation

**vs**

# Testing

# Philosophical differences

**Hypothesis testing: goal is to make a specific decision about a parameter**

**Confidence intervals: goal is to estimate a parameter**

# Connections

**Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree**

# Connections

99% CI for  $\mu$ : 82.78694 to 179.2097  $\mu\text{g/L}$

**Consider  $H_0 = 50 \mu\text{g/L}$  vs.  $H_0 \neq 50$**

**Consider  $H_0 = 50 \mu\text{g/L}$  vs.  $H_0 > 50$**