STAT 509: Statistics for Engineers

Chapter 7: Point Estimation of Parameters and Sampling Distributions

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

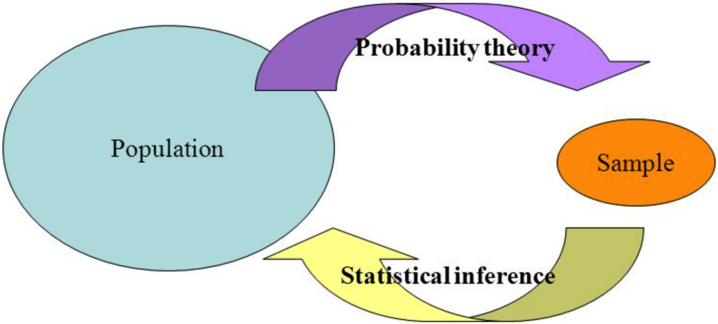
Fall 2020

Chapter 7: Point Estimation of Parameters and Sampling Distributions

Learning Objectives:

1. Explain the general concepts of estimating the parameters of a population or a probability distribution
2. Explain the important role of the normal distribution as a sampling distribution and the central limit theorem
3. Explain important properties of point estimators, including bias, variance, and mean square error

Statistical Inference



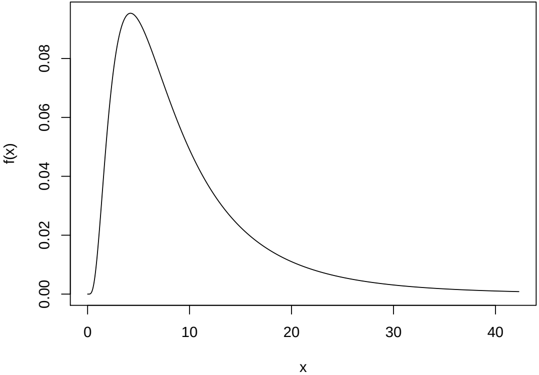
Point estimation is the first statistics inference covered in this course.

Sampling error

In statistics, sampling error is the error caused by observing a sample instead of the whole population. A direct result is that two samples of the same size very likely give you totally different observations.

Example

Suppose the population can be modeled by a lognormal distribution with $\theta = 2$ and $\omega = 0.75$.



Sampling error

We have twenty samples of size $n = 10$ from the lognormal distribution on the right.

They are all different, caused by the sampling error.

Consequently, each sample has its own sample mean.

	Sample									
Obs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	3.9950	8.2220	4.1893	15.0907	12.8233	15.2285	5.6319	7.5504	2.1503	3.1390
2	7.8452	13.8194	2.6186	4.5107	3.1392	16.3821	3.3469	1.4393	46.3631	1.8314
3	1.8858	4.0513	8.7829	7.1955	7.1819	12.0456	8.1139	6.0995	2.4787	3.7612
4	16.3041	7.5223	2.5766	18.9189	4.2923	13.4837	13.6444	8.0837	19.7610	15.7647
5	9.7061	6.7623	4.4940	11.1338	3.1460	13.7345	9.3532	2.1988	3.8142	3.6519
6	7.6146	5.3355	10.8979	3.6718	21.1501	1.6469	4.9919	13.6334	2.8456	14.5579
7	6.2978	6.7051	6.0570	8.5411	3.9089	11.0555	6.2107	7.9361	11.4422	9.7823
8	19.3613	15.6610	10.9201	5.9469	8.5416	19.7158	11.3562	3.9083	12.8958	2.2788
9	7.2275	3.7706	38.3312	6.0463	10.1081	2.2129	11.2097	3.7184	28.2844	26.0186
10	16.2093	3.4991	6.6584	4.2594	6.1328	9.2619	4.1761	5.2093	10.0632	17.9411
$\bar{x}$	9.6447	7.5348	9.5526	8.5315	8.0424	11.4767	7.8035	5.9777	14.0098	9.8727
Obs	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	7.5528	8.4998	2.5299	2.3115	6.1115	3.9102	2.3593	9.6420	5.0707	6.8075
2	4.9644	3.9780	11.0097	18.8265	3.1343	11.0269	7.3140	37.4338	5.5860	8.7372
3	16.7181	6.2696	21.9326	7.9053	2.3187	12.0887	5.1996	3.6109	3.6879	19.2486
4	8.2167	8.1599	15.5126	7.4145	6.7088	8.3312	11.9890	11.0013	5.6657	5.3550
5	9.0399	15.9189	7.9941	22.9887	8.0867	2.7181	5.7980	4.4095	12.1895	16.9185
6	4.0417	2.8099	7.1098	1.4794	14.5747	8.6157	7.8752	7.5667	32.7319	8.2588
7	4.9550	40.1865	5.1538	8.1568	4.8331	14.4199	4.3802	33.0634	11.9011	4.8917
8	7.5029	10.1408	2.6880	1.5977	7.2705	5.8623	2.0234	6.4656	12.8903	3.3929
9	8.4102	6.4106	7.6495	7.2551	3.9539	16.4997	1.8237	8.1360	7.4377	15.2643
10	7.2316	11.5961	4.4851	23.0760	10.3469	9.9330	8.6515	1.6852	3.6678	2.9765
$\bar{x}$	7.8633	11.3970	8.6065	10.1011	6.7339	9.3406	5.7415	12.3014	10.0828	9.1851

Random sample

Definition

To acknowledge the sampling error, we denote a **random sample** of size n by $X_1, \ldots, X_n$ (capital letters):

(a) the X_i's are **independent** random variables

(b) every X_i has the **identical** probability distribution.

We also call $X_1, \ldots, X_n$ are **iid** samples from the population.
The observed sample values are denoted by (lower cast) $x_1, \ldots, x_n$.

Example

In the previous example, we have a random sample of size $n = 10$: $X_1, \ldots, X_{10}$, where

(a) the X_i's are independent and

(b) having the sample lognormal distribution.

In the first observed sample, we have
$x_1 = 3.9950, \ldots, x_{10} = 16.2093$,..., in the 20th observed sample,
$x_1 = 6.8075, \ldots, x_{10} = 2.9765$.

Statistic

Definition

A **statistic** is any function of a random sample.

Example

of a random sample $X_1, \dots, X_n$, commonly used statistics are

- the **sample mean:** $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$
  (observed value denoted by $\bar{x}_n$)
- the **sample variance:** $S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$
  (observed value denoted by s_n^2)
- the **sample standard deviation:** $S_n = \sqrt{S_n^2}$
  (observed value denoted by s_n)

Sampling distribution

Definition

The probability distribution of a statistic is called a **sampling distribution**.

Recall that, a statistic is a function of a random sample which consists of n random variables. Thus the statistic itself is also a random variable. Hence, it has its own distribution, which is now called its **sampling distribution**

- Sampling distribution is different than the population distribution
- Sampling distribution depends on the sample size n

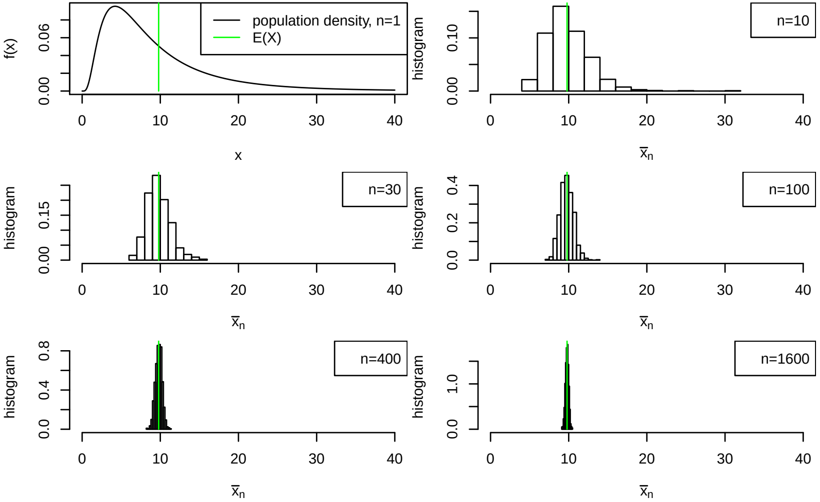
Example

The sample mean $\bar{X}_n$ is a statistic. Now we use simulation and histogram to illustrate the sampling distribution of $\bar{X}_n$:

- Suppose the population distribution is $\text{logN}(\theta = 2, \omega^2 = 0.75^2)$.
- For each $n \in \{10, 30, 100, 500, 1000\}$, we generate 1000 samples of size n.
- From each of the 1000 samples, we calculate the sample mean $\bar{x}_n$. In total, we have 1000 observed sample means, each is an average of 1000 observed values from the population distribution.
- We plot the histogram of the 1000 sample means, which approximates the sampling distribution of the statistic $\bar{X}_n$.

Example

Result: the distribution of $\bar{X}_n$ becomes more and more centered at the population mean $\mu = E(X) = 9.7889$ as n increases (i.e., $\bar{X}_n$ is a better and better estimator of μ as n increases).



Point estimator

As we just see, the sample mean $\bar{X}_n$ is a (point) estimator of the population mean $\mu = E(X)$. In general,

- let θ denote a parameter of the population distribution, which is unknown and of interest to us.
- From a random sample $X_1, \ldots, X_n$, we calculate a statistic $\widehat{\Theta}_n = h(X_1, \ldots, X_n)$ to estimate θ and call $\widehat{\Theta}_n$ a **point estimator** of θ.

Example

- $\bar{X}_n = h(X_1, \ldots, X_n) = n^{-1} \sum_{i=1} X_i$ is a point estimator of the population parameter (population mean) $\mu = E(X)$.
- $S_n^2 = h(X_1, \ldots, X_n) = (n-1)^{-1} \sum_{i=1} (X_i - \bar{X}_n)^2$ is a point estimator of the population variance $\sigma^2 = V(X)$.
- Later, we will learn how to estimate population parameters: p (population proportion), $\mu_1 - \mu_2$ (difference in means of two population), $p_1 - p_2$ (different in two population proportions).

Point estimate

A **point estimate** of some population parameter θ is a single numerical value $\widehat{\theta}_n$ of a statistic $\widehat{\Theta}_n$ (e.g., point estimator: $\bar{X}_n$ and point estimate $\bar{x}_n = 9.6447$).

Upper (random variable)	Lower (observed value)
Random variable X	an observed value x
Random sample: $X_1, \ldots, X_n$	an observed sample: $x_1, \ldots, x_n$
An estimator: $\widehat{\Theta}_n$	an estimate: $\widehat{\theta}_n$
Sample mean: $\bar{X}_n$	mean of an observed sample: $\bar{x}_n$
Sample variance: S_n^2	variance of an observed sample: $\bar{s}_n^2$

Table 1: Notation

Sampling distribution

Let $X_1, \ldots, X_n$ be a random (or equivalently, an **iid**) sample from a population with pdf $f(x)$. The **iid** means

- these X_i's are identically distributed and have the same pdf $f(x)$,
- and these X_i's are mutually independent.

Let $\widehat{\Theta}_n$ be a point estimator of θ, a parameter in $f(x)$. For example $\theta = \mu = E(X) = \int xf(x)dx$. We know that

$$\widehat{\Theta}_n = h(X_1, \ldots, X_n)$$

is also a random variable has its own distribution which is called the **sampling distribution**.

- The sampling distribution usually does not have the density function $f(x)$,
- The sample distribution changes with n while $f(x)$ does not change with n.

Central limit theorem

What is the sampling distribution of a point estimator $\widehat{\Theta}_n$?
Often it **does not** have an analytic form.

For example, suppose the population distribution is $\text{logN}(\theta, \omega^2)$. We have a random sample of size n and the sample mean is $\bar{X}_n$. What is the (sampling) distribution of $\bar{X}_n$? No analytic form! But we can approximate it when n is large.

Central limit theorem

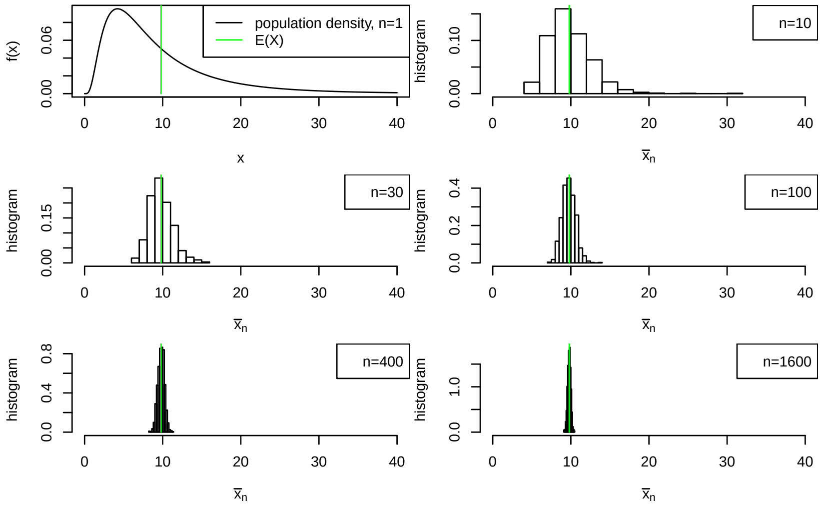
If $X_1, \dots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and variance σ^2 and if $\bar{X}_n$ is the sample mean, then

$$E(\bar{X}) = \mu, V(\bar{X}_n) = \frac{\sigma^2}{n}, \text{ and as } n \to \infty, \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \underbrace{N(0,1)}_{\text{standard normal}};$$

or $\bar{X}_n \sim \text{AN}(\mu, \sigma^2/n)$, AN stands for **A**pproximately **N**ormal.

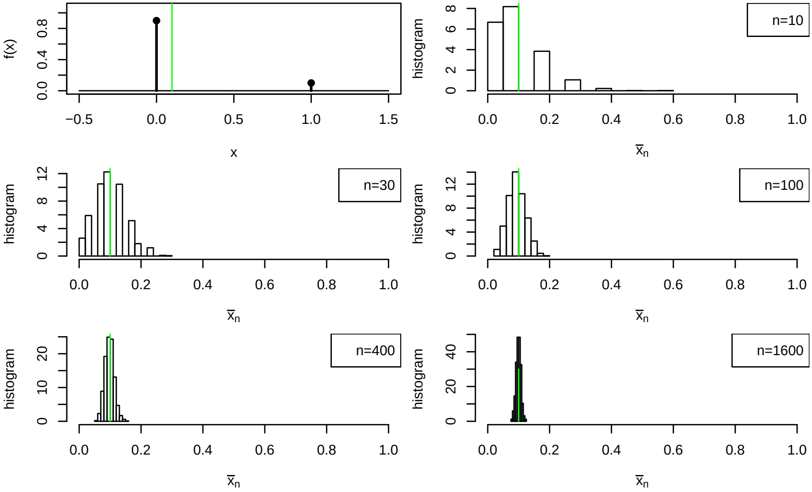
Example

We have seen these histograms of observed sample means from the population distribution $\log N(\theta = 2, \omega^2 = 0.75^2)$. We know $\mu = 9.7889, \sigma^2 = 72.3514$. Thus from CLT, $\bar{X}_n \sim AN(9.7889, 72.3514/n^2)$



Example

Suppose the population distribution is Bernoulli($p = 0.1$). Then $\mu = 0.1$ and $\sigma^2 = p(1-p) = 0.09$. From CLT, $\bar{X}_n \sim N(0.1, 0.09/n^2)$.



The use of CLT

- Do we have a random sample $X_1, \ldots, X_n$?
- Is it about $\bar{X}_n$?
- What are μ and σ^2?
- CLT: $\bar{X}_n \sim \mathrm{AN}\left(\mu, \frac{\sigma^2}{n}\right)$
- Be careful that CLT may not work well if n is small (rule of thumb: $n \geq 25$).

A special case

If the population distribution is exactly $N(\mu, \sigma^2)$, then

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

holds exactly (not approximately).

Example

Suppose a random variable $X \sim \text{Uniform}(4, 7)$. Find the distribution of the sample mean $\bar{X}_n$ of a random sample of size $n = 40$. Find/approximate $P(\bar{X}_{40} \geq 5.6)$?

Solution: This example asks for the distribution and a probability of the sample mean $\bar{X}_n$. Thus, we need to find/approximate the sampling distribution of $\bar{X}_n$. CLT tells $\bar{X}_n \sim \text{AN}(\mu, \sigma^2/n)$.

Using the information $X \sim \text{Uniform}(4, 6)$, we can find $\mu = 5.5$ and $\sigma^2 = 0.75$ by formula or StatEngine $\mathit{uniform.summary}(4, 7)$. Thus

$$\bar{X}_n \sim \text{AN}(5.5, 0.75^2/40).$$

Finally, we can use this normality to approxiamte $P(\bar{X}_{40} \geq 5.6)$. Using StatEngine

$$\mathit{normal.prob}(5.5, \mathit{sqrt}(0.75^2/40), 5.6, \mathit{Inf}) = 0.1995$$

Expectation of a point estimator

An estimator $\widehat{\Theta}_n$ should be close in some sense to the true value of the unknown parameter θ.

Bias

The point estimator $\widehat{\Theta}_n$ is an **unbiased estimator** for the parameter θ if

$$E(\widehat{\Theta}_n) = \theta.$$

If the estimator is not unbiased (or biased), then the difference

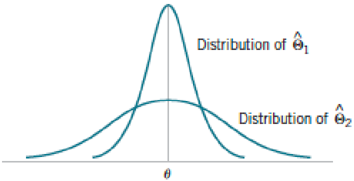
$$E(\widehat{\Theta}_n) - \theta$$

is called the **bias** of the estimator $\widehat{\Theta}_n$ and denoted by $\text{bias}(\widehat{\Theta}_n)$

For example, we know from CLT, $E(\bar{X}_n) = \mu$ always holds. Thus the sample mean $\bar{X}_n$ is always an unbiased estimator of the population mean parameter μ.

Variance of a point estimator

Suppose that $\widehat{\Theta}_{n1}$ and $\widehat{\Theta}_{n2}$ are unbiased estimators of θ. This indicates that the distribution of each estimator is centered at the true value of zero. However, the variance of these distributions may be different.



Though both estimators are unbiased, because $\widehat{\Theta}_{n1}$ has a smaller variance than $\widehat{\Theta}_{n2}$, the estimator $\widehat{\Theta}_{n1}$ is more likely to produce an estimate close to the true value of θ.

Standard error of a point estimator

The standard error of an estimator $\widehat{\Theta}_n$ is its standard deviation given by

$$\sigma_{\widehat{\Theta}_n} = \sqrt{V(\widehat{\Theta}_n)}.$$

If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\widehat{\Theta}_n}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\widehat{\Theta}_n}$ or $\mathrm{SE}(\widehat{\Theta}_n)$.

For example, Suppose that we are sampling from a normal distribution with mean μ and variance σ^2, the standard error of $\bar{X}_n$ is

$$\sigma_{\bar{X}_n} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

which involves the unknown parameter σ. We can estimate σ by the sample standard deviation S_n and substitute it into the preceding equation. The esitamed standard error of $\bar{X}_n$ would be

$$\mathrm{SE}(\bar{X}_n) = \frac{S_n}{\sqrt{n}}.$$

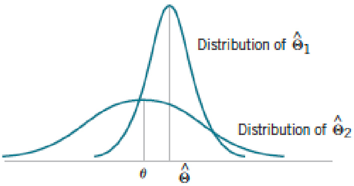
Mean squared error of a point estimator

Sometimes it is necessary to use a biased estimator. In such cases, the mean squared error of the estimator can be important.

MSE

The mean squared error of an estimator $\widehat{\Theta}_n$ of the parameter θ is defined as

$$\mathrm{MSE}(\widehat{\Theta}_n) = E(\widehat{\Theta}_n - \theta)^2 = \mathrm{bias}^2(\widehat{\Theta}_n) + V(\Theta_n).$$



Though $\widehat{\Theta}_{n1}$ is a biased estimator while $\widehat{\Theta}_{n2}$ is unbiased, the mse of $\widehat{\Theta}_{n1}$ is smaller than the one of $\widehat{\Theta}_{n2}$, thus $\widehat{\Theta}_{n1}$ is more likely to produce an estimate close to the true value of θ.

Mean squared error of a point estimator

The mean squared error is an important criterion for comparing two estimators $\widehat{\Theta}_{n2}$ and $\widehat{\Theta}_{n1}$. Define the **relative efficiency** of $\widehat{\Theta}_{n2}$ to $\widehat{\Theta}_{n1}$ by

$$\mathrm{RE}(\widehat{\Theta}_{n2} \text{ to } \widehat{\Theta}_{n1}) = \frac{\mathrm{MSE}(\widehat{\Theta}_{n1})}{\mathrm{MSE}(\widehat{\Theta}_{n2})}.$$

If $\mathrm{RE}(\widehat{\Theta}_{n2} \text{ to } \widehat{\Theta}_{n1}) < 1$, then $\widehat{\Theta}_{n1}$ is a more efficient estimator of θ than $\widehat{\Theta}_{n2}$ in the sense that it has a smaller mean squared error.

For example, we estimate the population mean μ by $\bar{X}_n$. Let $\widehat{\mu}_{n1} = \bar{X}_{n_1}$ and $\widehat{\mu}_{n2} = \bar{X}_{n_2}$, where $n_1 < n_2$. Then

$$\mathrm{RE}(\widehat{\mu}_{n2} \text{ to } \widehat{\mu}_{n1}) = \frac{\mathrm{MSE}(\widehat{\mu}_{n1})}{\mathrm{MSE}(\widehat{\mu}_{n2})} = \frac{\sigma^2/n_1}{\sigma^2/n_2} = \frac{n_2}{n_1} > 1.$$

Thus $\widehat{\mu}_{n2} = \bar{X}_{n_2}$ is a more efficient estimator of μ than $\widehat{\mu}_{n1} = \bar{X}_{n_1}$, which makes sense because $n_2 > n_1$; i.e., more samples bring more information and thus lead to more accurate estimation.

Different types of point estimation

Data on the oxide thickness of semiconductor wafers are as follows: 425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416.

```
#Define your data
ot=c(425, 431, 416, 419, 421, 436, 418, 410, 431, 433,
     423, 426, 410, 435, 436, 428, 411, 426, 409, 437,
     422, 428, 413, 416)
#Summarize it
data.summary(ot)
```

```
Summary:
           [,1] [,2]     [,3]     [,4]   [,5] [,6]  [,7] [,8]   [,9] [,10]
statistics min  mean     variance std    max  range Q1   Median Q3   IRQ
result     409  423.3333 82.4928  9.0826 437  28    416  424    431  15
```

Different types of point estimation

```
Summary:
           [,1] [,2]     [,3]     [,4]   [,5] [,6]  [,7] [,8]   [,9] [,10]
statistics min  mean     variance std    max  range Q1   Median Q3   IRQ
result     409  423.3333 82.4928  9.0826 437  28    416  424    431  15
```

(a) Calculate a point estimate of the mean oxide thickness for all wafers in the population.

Solution: We use the sample mean: $\widehat{\mu} = \bar{x}_n = 423.3333$.

(b) Calculate a point estimate of the standard deviation of oxide thickness for all wafers in the population.

Solution: We use the sample standard deviation: $\widehat{\sigma} = 9.0826$.

(c) Calculate the standard error of the point estimate from part (a).

Solution: $\mathrm{SE}(\bar{X}_n) = \frac{S_n}{\sqrt{n}}$. Thus we have it as $\frac{9.0826}{\sqrt{24}} = 1.854$.

```
Summary:
           [,1] [,2]     [,3]     [,4]   [,5] [,6]  [,7] [,8]   [,9] [,10]
statistics min  mean     variance std    max  range Q1   Median Q3   IRQ
result     409  423.3333 82.4928  9.0826 437  28    416  424    431  15
```

(d) Calculate a point estimate of the median oxide thickness for all wafers in the population.

Solution: We use the sample median: 424.

(e) Calculate a point estimate of the proportion of wafers in the population that have oxide thickness of more than 430 angstroms.

Solution: We use the sample proportion: 29.17%, which is the proportion of wafers in the sample that have oxide thickness of more than 430 angstroms.

```
mean(ot>430)
```