Sample Surveys

Statistics 110

Summer 2006

Population Parameters

Framework: Finite population with N items

- Population values: $\nu_1, \nu_2, \ldots, \nu_N$ (may be repeated values)
- Population mean:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \nu_i = \bar{\nu}$$

- Population total:

$$\tau = \sum_{i=1}^{N} \nu_i = N\mu$$

- Population variance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (\nu_i - \mu)^2$$

The quantities μ, τ, σ^2 are examples of population parameters, which are numerical summaries of the set of population values. Thus, the population standard deviation σ and

$$p = \frac{1}{N} \sum_{i=1}^{N} I\{\nu_i < 0\}$$

are also population parameters.

Population parameters of interest are often of the form

$$\frac{1}{N} \sum_{i=1}^{N} g(\nu_i) = E[g(\nu)]$$

This can be seen from the following scheme. Suppose we draw X_1 randomly from the population. Then $X_1 = \nu_i$ with probability $\frac{1}{N}$ and

$$E[X_1] = \bar{\nu} = \mu; \qquad \mathrm{Var}(X_1) = \sigma^2$$

The problem of interest is to learn about the population parameters of interest. Sample surveys are used to obtain information about a large population by examining only a small fraction of that population.

Examples of sample surveys

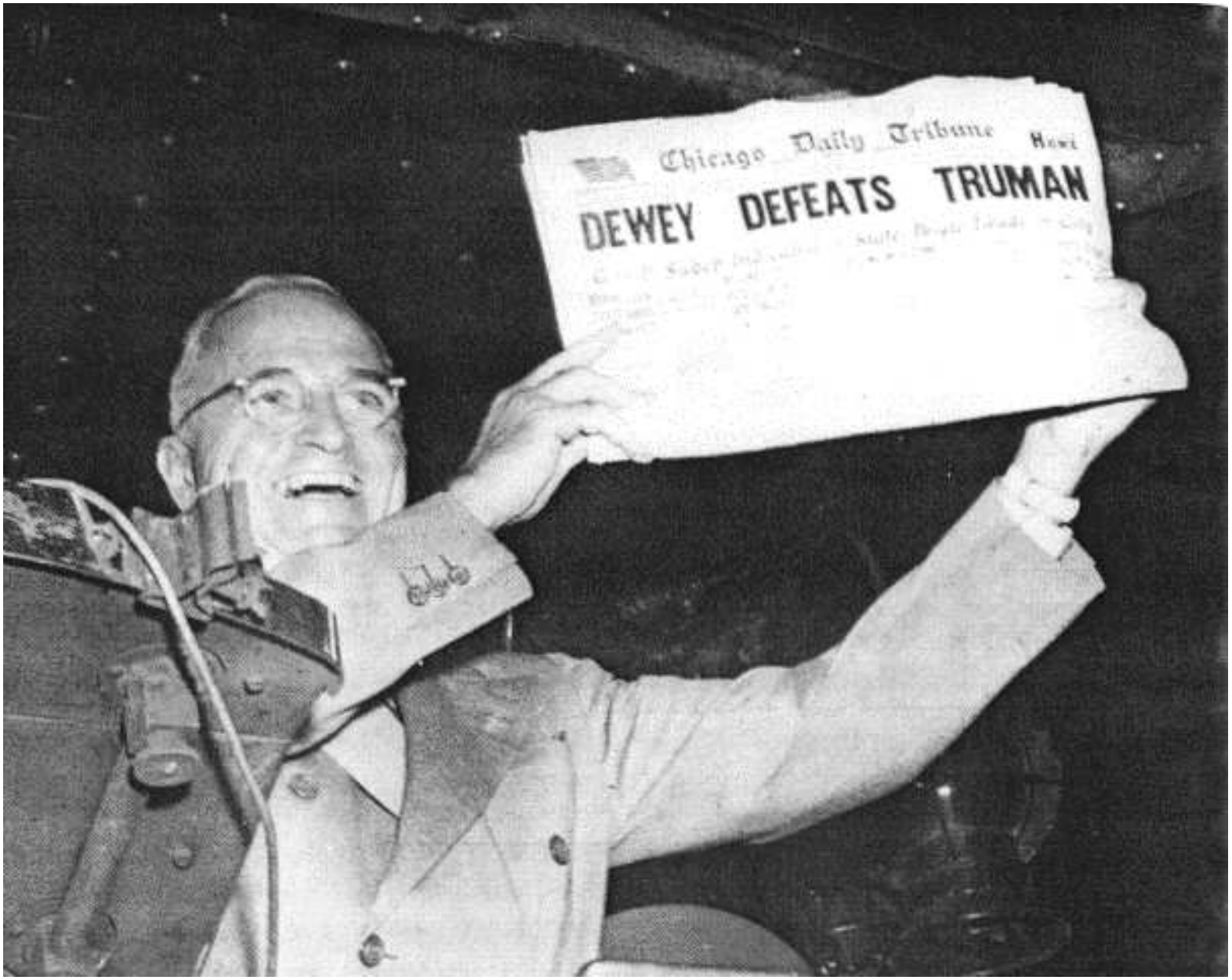
- Political polling
- Consumer preferences
- Product monitoring (quality control)
- Financial auditing

Why sample instead of doing a census (examine all population units)?

- Cheaper
- Selecting sample units at random is a guard against investigator biases
- More accurate. Better care of data quality can be taken with smaller samples
- Random sampling provides techniques for the calculation of an estimate of the sampling error.
- When designing a sample, it is usually possible to determine the sample size required to meet a desired error level.

The techniques to be discussed are probability based sampling schemes. In these schemes, each population member will have a certain probability of being sampled p_i. In fact, for these schemes the probability of any sample of size n can be determined.

Poor sampling schemes can lead to events such as the following . . .


Chicago Daily Tribune
DEWEY DEFEATS TRUMAN

This headline was partially based on polling data which suggested that Dewey would beat Truman.

	Roper	Crossly	Gallup	Election
Truman (Democratic)	38%	45%	44%	50%
Dewey (Republican)	53%	50%	50%	45%
Others	9%	5%	6%	5%

Others included Strom Thurmond (State' Rights) who won 4 states (39 electoral votes) and Henry Wallace (Progessive).

In these polls, there was a greater chance for a Republican to be sampled than a Democrat, which skewed the polls towards Dewey. Also the polls were done about a week prior to election day and there is fairly good evidence that they was a drift towards Truman in the final week of the campaign.

When designing a sampling scheme, we want it to be independent of the values we are interested in.

The schemes that will be discussed satisfy this.

Definition. *A* **Simple Random Sample (SRS)**, $\{X_1, X_2, \ldots, X_n\}$ *is equally likely to be any of the* $\binom{N}{n}$ *sample of size* n *from the* N *population values. These samples can be generated by sampling without replacement from the population.*

We will estimate these population parameters with the following estimators

- Population mean with sample average:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Population total with

$$T = \sum_{i=1}^{n} X_i = N\bar{X}$$

- Population variance with sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

We want to get an idea of how these can vary over the $\binom{N}{n}$ different samples of size n

Sampling Distributions

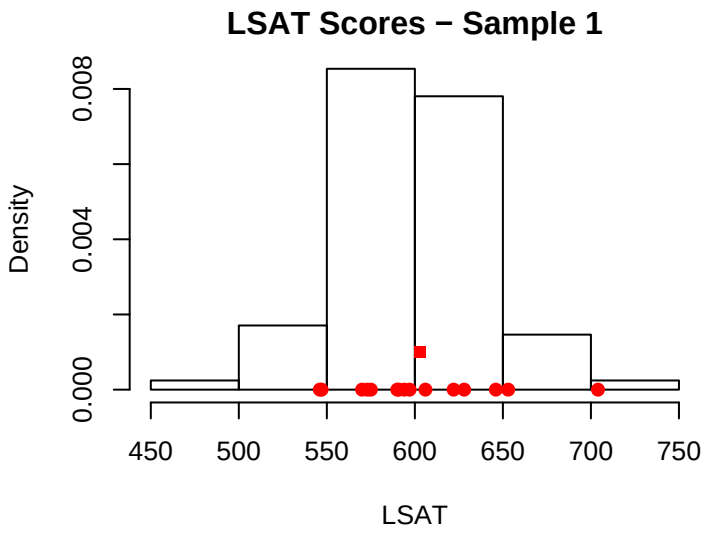
Example: Law school data

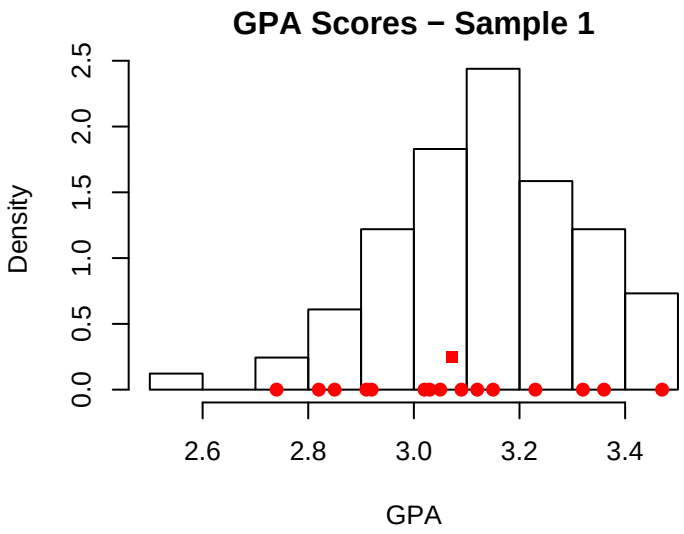
Generate 2 samples with $n = 15$ from the population of 82 schools.

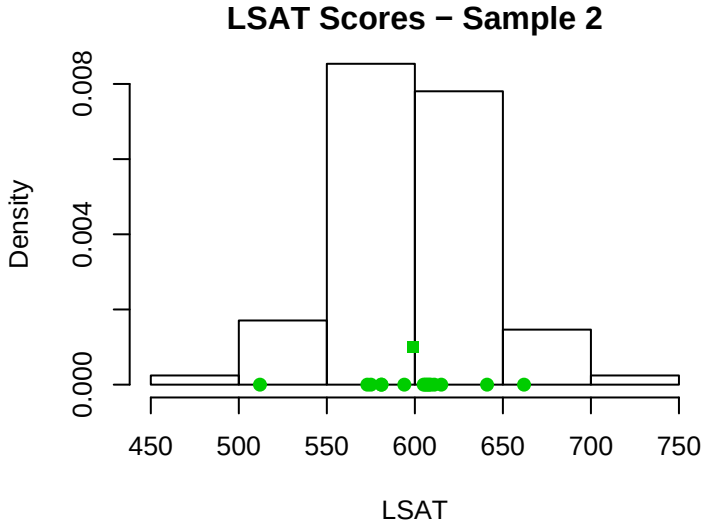
Questions of interest:

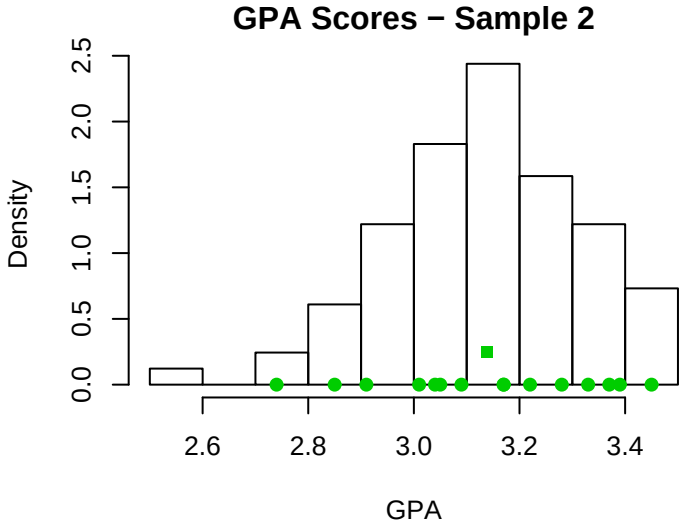
- Mean LSAT ($\mu = 597.55$)
- Proportion of LSAT < 550 ($p = 0.0976$)
- Mean GPA ($\mu = 3.135$)

Note that within each example sample, the same schools will be used for all three measures


LSAT Scores − Sample 1
Density
0.000
0.004
0.008
450
500
550
600
650
700
750
LSAT


GPA Scores − Sample 1
Density
0.0
0.5
1.0
1.5
2.0
2.5
2.6
2.8
3.0
3.2
3.4
GPA


LSAT Scores − Sample 2
Density
0.000
0.004
0.008
450
500
550
600
650
700
750
LSAT


GPA Scores − Sample 2
Density
0.0
0.5
1.0
1.5
2.0
2.5
2.6
2.8
3.0
3.2
3.4
GPA

	Sample 1 estimate	Sample 2 estimate	True value
$\bar{X}_{LSAT}$	593.93	607.27	597.55
$\hat{p}_{LSAT}$	0.133	0	0.0976
$\bar{X}_{GPA}$	3.137	3.167	3.135

So, not surprisingly, different samples give us different parameter estimates

We can ask, what the $\binom{N}{n}$ different samples samples give for each parameter estimate. Since each sample chosen is random, the estimate (say $\bar{X}, \hat{p}$, or S) is a random variables.

The probability distribution induced on a statistic by the sampling mechanism is known as its **Sampling Distribution**.

As we will see, different sampling mechanisms can lead to different sampling distributions.

Often determining the sampling distribution exactly is difficult as the number of possible samples is too large. For the law school example, the number of possible samples is

$$\binom{82}{15} = 9.97 \times 10^{15} \qquad \text{(about 10 quintillion)}$$

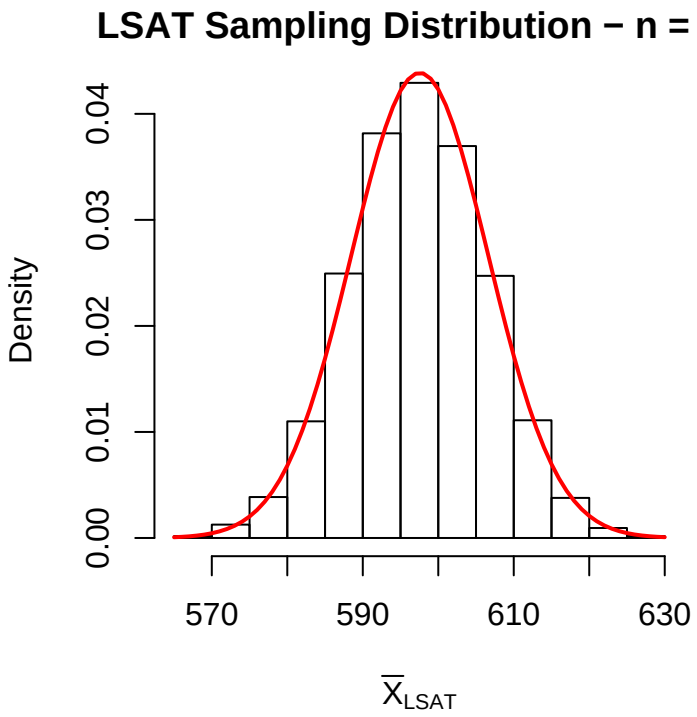
So to determine properties of the sampling distribution we will have to use approximation procedures and incomplete descriptions.

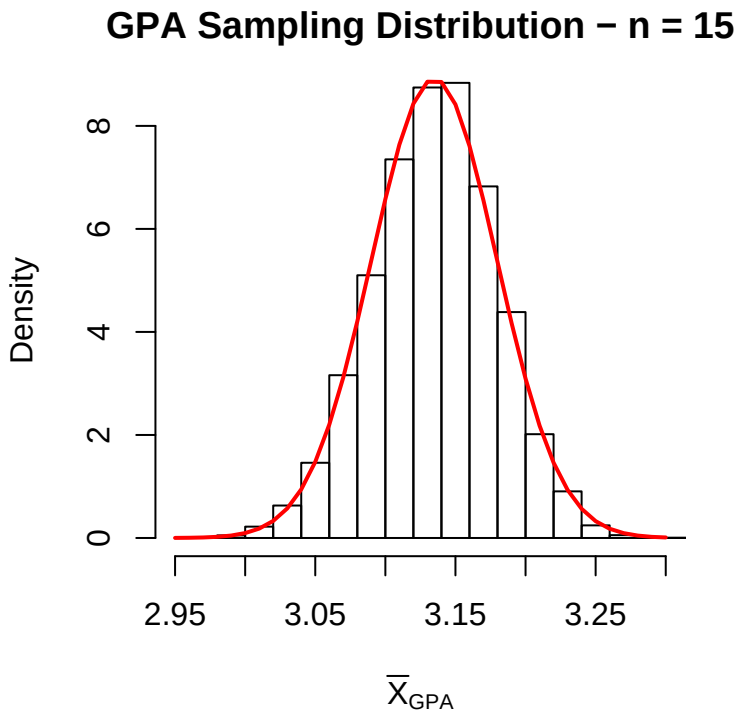
The approximations include simulation and asymptotic (convergence in distribution) approaches.

The following plot illustrates both approaches.

The histograms are based on 10,000 randomly generated simple random samples of size 15.

The curves are normal approximations based on the central limit theorem (with an adjustment for the dependency).


LSAT Sampling Distribution − n = 15
Density
0.00
0.01
0.02
0.03
0.04
570
590
610
630
$\overline{X}_{LSAT}$


GPA Sampling Distribution − n = 15
Density
0
2
4
6
8
2.95
3.05
3.15
3.25
$\overline{X}_{GPA}$

Sampling distribution of $\bar{\mathbf{X}}$

It is possible to make some general statements about the sampling distribution of $\bar{X}$. These will be justified later.

1. The histograms are centered near their population means

2. The spread in the histograms decreases as n increases.

3. The shape of the histograms is roughly symmetric, even though the population values aren't.

Theorem. *If $X_1, X_2, \ldots, X_n$ is a SRS, then*

$$E[\bar{X}] = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

Proof. Since for each i, $E[X_i] = \mu$,

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} n\mu = \mu$$

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{i\neq j}\mathrm{Cov}(X_i, X_j)\right) \\
&= \frac{1}{n^2}\left(n\mathrm{Var}(X_i) + n(n-1)\mathrm{Cov}(X_1, X_2)\right) \\
&= \frac{\sigma^2}{n} + \frac{n-1}{n}\mathrm{Cov}(X_1, X_2)
\end{aligned}$$

To get $\mathrm{Cov}(X_1, X_2)$, look at $E[X_1X_2] = E[E[X_1X_2|X_1]]$

$$\begin{aligned}
E[X_1X_2|X_1 = \nu_j] &= \nu_j E[X_2|X_1 = \nu_j] \\
&= \nu_j\left(\frac{N\bar{\nu} - \nu_j}{N-1}\right) \\
&= \frac{N}{N-1}\nu_j\bar{\nu} - \frac{\nu_j^2}{N-1} \overset{Def}{=} \theta_j
\end{aligned}$$

Thus $E[X_1X_2|X_1] = \theta_j$ with probability $\frac{1}{N}$.

$$\begin{aligned} E[X_1X_2] &= \frac{1}{N}\sum_{i=1}^{N}\theta_j = \frac{1}{N}\sum_{i=1}^{N}\frac{N}{N-1}\bar{\nu}\nu_j - \frac{1}{N}\sum_{i=1}^{N}\frac{\nu_j^2}{N-1} \\ &= \frac{N}{N-1}\bar{\nu}^2 - \frac{1}{N-1}\overline{\nu^2} \\ &= \bar{\nu}^2 - \frac{1}{N-1}(\overline{\nu^2} - \bar{\nu}^2) = \mu^2 - \frac{1}{N-1}\sigma^2 \end{aligned}$$

Therefore $\mathrm{Cov}(X_1, X_2) = E[X_1X_2] - \mu^2 = -\frac{\sigma^2}{N-1}$. Plugging this in gives,

$$\begin{aligned} \mathrm{Var}(\bar{X}) &= \frac{\sigma^2}{n} + \frac{n-1}{n}\left(-\frac{\sigma^2}{N-1}\right) \\ &= \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) \end{aligned}$$

□

Remark: If $X_1, X_2, \ldots, X_n$ are sampled with replacement, then they are independent and $E[\bar{X}] = \mu, \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$. When N is large compared to n, SRS behaves likes independent sampling. The quantity

$$1 - \frac{n-1}{N-1} = \frac{N-n}{N-1}$$

is known as the **finite population correction** (FPC). Note that it always ≤ 1, implying that ignoring the dependency in the sampling leads us to overestimating the uncertainty in $\bar{X}$ as an estimate of μ.

It is approximately equal 1 minus the fraction of the population that is sampled. Intuitively, its describing how much we gain by knowing we are sampling from a finite population without replacement. When $n \ll N$ the correction factor doesn't make much difference and is often neglected.

In the law school example, it can make as difference since

$$FPC(15) = \left(1 - \frac{14}{81}\right) = 0.82 \qquad FPC(30) = \left(1 - \frac{29}{81}\right) = 0.64$$

However for more typical examples, say sampling 1000 people living in Massachusetts (population approximately 6.4 million), the correction makes little difference

$$FPC \approx \left(1 - \frac{999}{6.4 \times 10^6}\right) = 0.9998$$

Population total

If instead we are interested in the population total (e.g. we want to know the total tax paid in Massachusetts instead of the average tax paid by Massachusetts residents), we can estimate this by

$$T = N\bar{X}$$

The first two moments of the sampling distribution are

$$E[T] = N\mu = \tau \qquad \mathrm{Var}(T) = N^2 \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

This is a rare case where the precision of an estimate depends strongly on the population size N.

Population proportion

A third common problem of interest is estimating a population proportion, such as the proportion of Massachusetts residents who pay less that $200 a year in state income tax, or the proportion of law schools with incoming average LSAT scores less than 550 (8 of 82, $p = 0.0976$)

Let X be the number of "successes" out of n draws from the population. Then X has a hypergeometric distribution with

$$E[X] = np \qquad \mathrm{Var}(X) = np(1-p)\left(1 - \frac{n-1}{N-1}\right)$$

The sample proportion, $\hat{p} = \frac{X}{n}$ has moments

$$E[\hat{p}] = p \qquad \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}\left(1 - \frac{n-1}{N-1}\right)$$

Notice that the variance looks like that of the binomial distribution times the FPC.

The usual measure of the precision of an estimator is the standard deviation of the sampling distribution, often referred to as the standard error. So for the estimators seen so far, there standard errors are

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n-1}{N-1}}$$

$$\sigma_T = N\frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n-1}{N-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{1 - \frac{n-1}{N-1}}$$

So for the law school example looking at the average LSAT scores, the standard errors are

	$n = 15$	$n = 30$
$\bar{X}$	$\frac{38.49}{\sqrt{15}}\sqrt{1 - \frac{14}{81}} = 9.04$	7.96
T	$82\frac{38.49}{\sqrt{15}}\sqrt{1 - \frac{14}{81}} = 741.16$	523.16
$\hat{p}$	$\sqrt{\frac{0.0976(1-0.0976)}{15}}\sqrt{1 - \frac{14}{81}} = 0.0697$	0.0434

Why do we care about mean and variance of an estimator, particularly with $\bar{X}$? They allow us to find regions that contain an estimate with high probability.

For example, Chebyshev gives us

$$P[-k\sigma_{\bar{X}} \leq \bar{X} - \mu \leq k\sigma_{\bar{X}}] \geq 1 - \frac{1}{k^2}$$

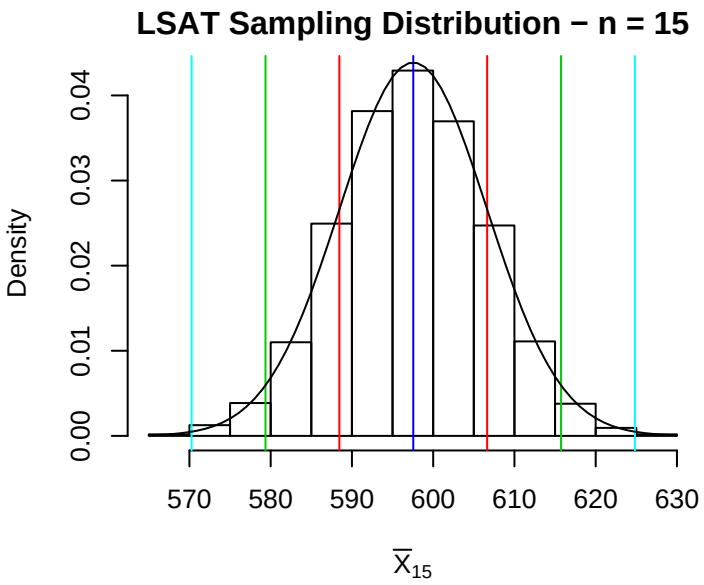
In addition, when n is large, $\bar{X}$ is approximately normal with the mean and variance given earlier. That is

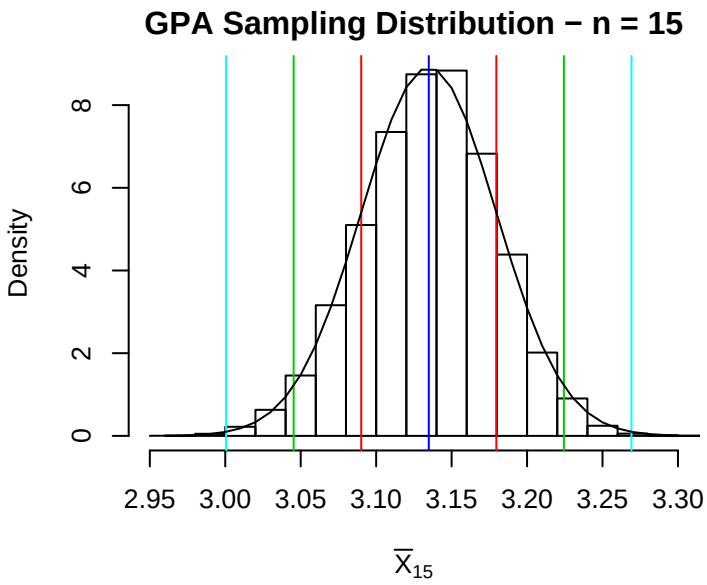
$$P\left[\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq z\right] \approx \Phi(z)$$

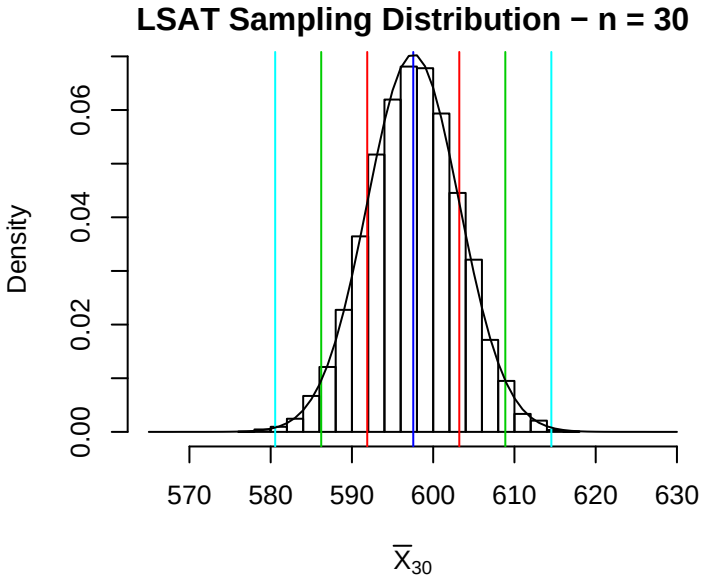
Thus we can use the normal distribution, which we have seen, often gives better approximations in probabilities than Chebyshev's inequality.

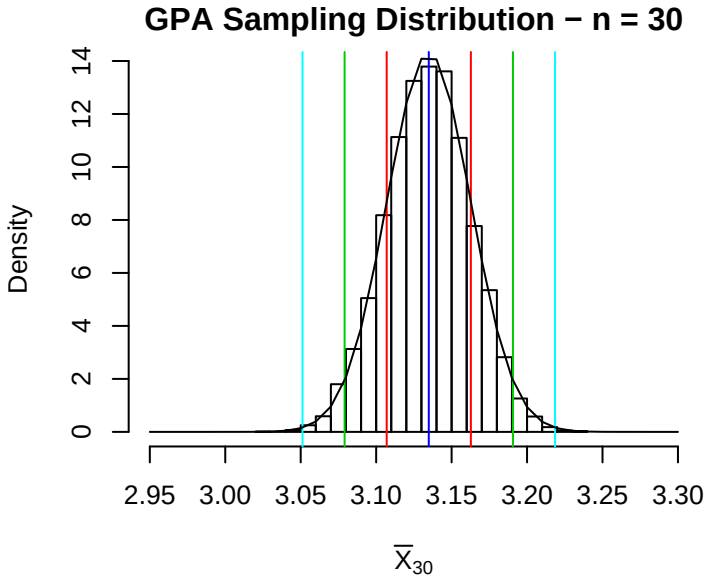
$$P[-k\sigma_{\bar{X}} \leq \bar{X}-\mu \leq k\sigma_{\bar{X}}] \approx \Phi(k)-\Phi(-k) = 1-2\Phi(-k)$$

These results also hold for T and $\hat{p}$. It is common to use the normal approximation as long n isn't to small. For many problems $n > 30$ is fine, though the choice really is problem specific.


LSAT Sampling Distribution − n = 15
Density
0.00
0.01
0.02
0.03
0.04
570
580
590
600
610
620
630
$\overline{X}_{15}$


GPA Sampling Distribution − n = 15
Density
0
2
4
6
8
2.95
3.00
3.05
3.10
3.15
3.20
3.25
3.30
$\overline{X}_{15}$


LSAT Sampling Distribution − n = 30
Density
0.00
0.02
0.04
0.06
570
580
590
600
610
620
630
$\overline{X}_{30}$


GPA Sampling Distribution − n = 30
Density
0
2
4
6
8
10
12
14
2.95
3.00
3.05
3.10
3.15
3.20
3.25
3.30
$\overline{X}_{30}$

Theorem. *If $X_1, X_2, \ldots, X_n$ is a SRS, then*

$$E[S^2] = \sigma^2 \left(1 + \frac{1}{N-1}\right) = \sigma^2 \frac{N}{N-1}$$

Proof.

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$\begin{aligned}
E[S^2] &= \frac{1}{n-1}\left(nE[X_1^2] - nE[(\bar{X})^2]\right) \\
&= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) + \mu^2\right)\right) \\
&= \frac{\sigma^2}{n-1}\left(n - \left(1 - \frac{n-1}{N-1}\right)\right) \\
&= \sigma^2\left(1 + \frac{1}{N-1}\right)
\end{aligned}$$

□

So note that with an SRS, S^2 has a small positive bias, due to the dependency of the draws. (If the draws are with replacement, it is unbiasd)

It is still preferable to

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$$

since

$$E[\hat{\sigma}^2] = \frac{n-1}{n}E[S^2] = \sigma^2\frac{n-1}{n}\frac{N}{N-1}$$

Note that since $n < N$, $\frac{n-1}{n}\frac{N}{N-1} < 1$, so $\hat{\sigma}^2$ has a negative bias.

Since $\frac{1}{N-1} \ll \frac{1}{n}$ usually, the bias of S^2 is much smaller than that of $\hat{\sigma}^2$.

If an unbiased estimate is needed, the following can be used

$$S^2\frac{N-1}{N}$$

Note that if we want to make probabilistic statement about $\bar{X}$, we need to know its variance, which depends on σ^2. By plugging the above unbiased estimate of σ^2, we get the following unbiased estimate of $\mathrm{Var}(\bar{X})$

$$S^2_{\bar{X}} = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

To get an unbiased estimate of the variance of $\hat{p}$, the following is usually used

$$S^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$$

Thus we can get the estimated standard errors

$$s_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}$$

$$s_T = N\frac{S}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}$$

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1 - \frac{n}{N}}$$

Note that since that these depend on the sample data, they are also random variables, and thus have there own sampling distribution, which we won't figure out.

	Sample 1 estimate	Sample 2 estimate	True value
$\bar{X}_{LSAT}$	593.93	607.27	597.55
$S_{\bar{X}_{LSAT}}$	12.23	7.31	8.983
$\hat{p}_{LSAT}$	0.133	0	0.0976
$S_{\hat{p}}$	0.0821	0	0.0697
$\bar{X}_{GPA}$	3.137	3.167	3.135
$S_{\bar{X}_{GPA}}$	0.0593	0.0306	0.0442

In all but one case ($P[LSAT < 550]$), the sample estimates are close to their true parameter value (error $<$ 1 SE). And even for the estimate of the sample proportion in the 2nd sample, it is not particularly surprising, since $P[X = 0] = 0.19$

For this one parameter, the standard error isn't the greatest measure of the precision of the estimate since for such a small sample size, the normal distribution approximation for $\hat{p}$ isn't very good.

Confidence Intervals

As discussed earlier

$$\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \overset{approx.}{\sim} N(0,1) \qquad \text{and} \qquad \bar{X} \overset{approx.}{\sim} N(\mu, \sigma^2_{\bar{X}})$$

so we can use the normal distribution to approximate probabilities about $\bar{X}$. For example, for the law school example when sampling 15 LSAT scores,

$$\begin{aligned} P[\bar{X} \geq 615] &= P\left[\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \geq \frac{615 - \mu}{\sigma_{\bar{X}}}\right] \\ &= P[Z \geq 1.94] \approx 1 - \Phi(1.94) = 0.0262 \end{aligned}$$

So its unlikely to see $\bar{X}$ this big or bigger when sampling 15 LSAT scores from this observation.

In addition

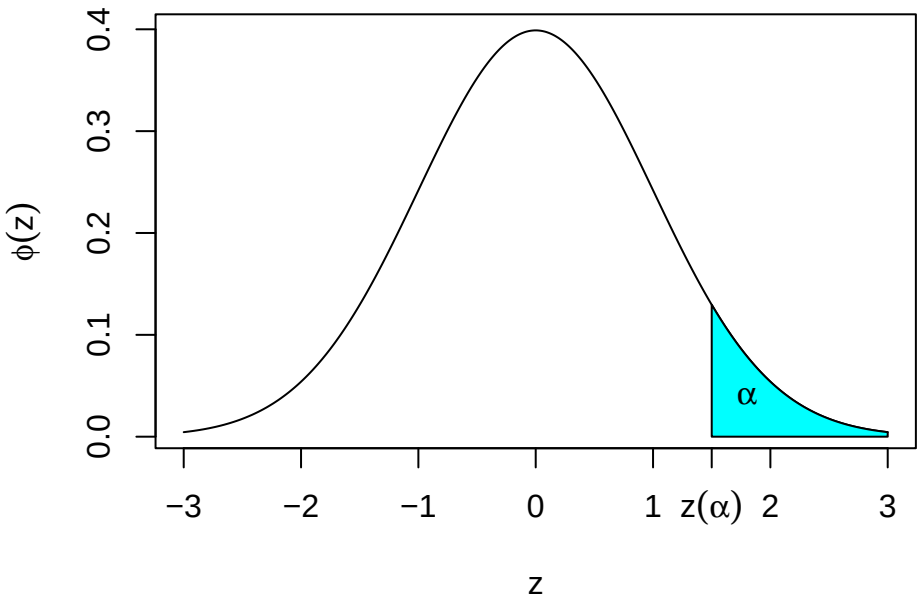
$$P[-k\sigma_{\bar{X}} \leq \bar{X} - \mu \leq k\sigma_{\bar{X}}] \approx \Phi(k) - \Phi(-k) = 1 - 2\Phi(-k)$$

which says with probability approximately $1 - 2\Phi(-k)$, $\bar{X}$ will be within k standard errors of μ. For example, 80% of the time, $\bar{X}$ will be within 1.282 standard errors, 90% of the time within 1.645 standard errors, and 95% of the time within 1.96 standard errors.

Now if we "switch" the way of looking at things, if $\bar{X}$ is within k standard error of μ, then μ must be within k standard errors of $\bar{X}$. This approach of switching gives us an approach for describing plausible values for μ based on a sample.

Lets assume that $Z \sim N(0,1)$. Let $z(\alpha)$ be the value satisfying $P[Z \geq z(\alpha)] = \alpha$.

These values can be determined by inverting the normal table in the back of the text or, for selected values of α, from the t-table (Table 4) from the ∞ row by noting that $z(\alpha) = t_{1-\alpha}$. This is where I got the number for the previous page.



Since $\bar{X}$ is approximately normal

$$P\left[-z(\alpha/2) \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq z(\alpha/2)\right] \approx 1 - \alpha$$

by rearranging the terms, this gives

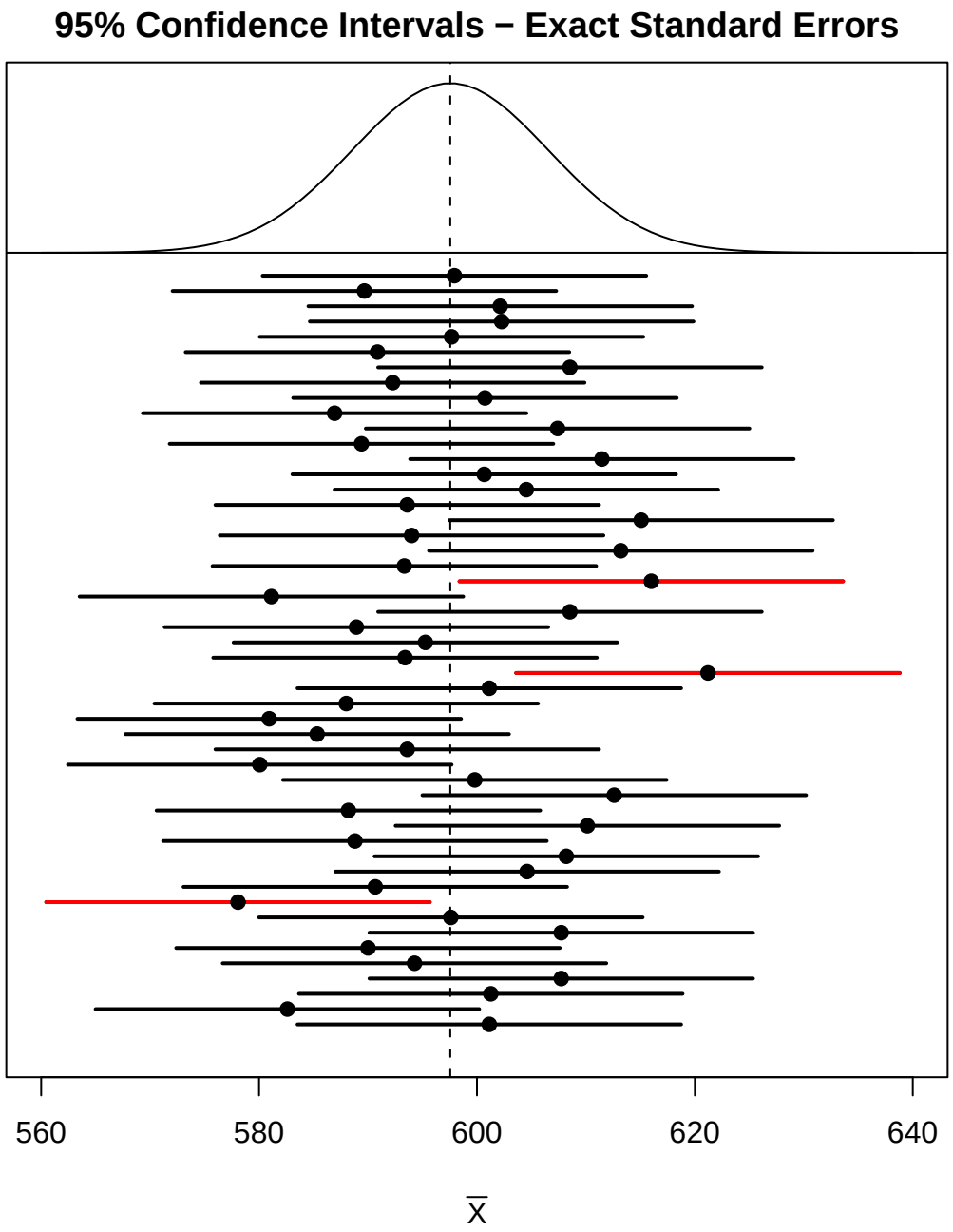
$$P\left[\bar{X} - z(\alpha/2)\sigma_{\bar{X}} \leq \mu \leq \bar{X} + z(\alpha/2)\sigma_{\bar{X}}\right] \approx 1 - \alpha$$

i.e. What the probability that μ is within $z(\alpha)$ standard errors of $\bar{X}$ (in the interval $\bar{X} \pm z(\alpha/2)\sigma_{\bar{X}}$).

So the probability that we select a sample that gives us a $\bar{X}$ so that μ is in that interval is approximately $1 - \alpha$. Note that these intervals are random as they depend on the random sample generated.

The interval $\bar{X} \pm z(\alpha/2)\sigma_{\bar{X}}$ is known as a $100(1-\alpha)\%$ confidence interval for μ. $1 - \alpha$ is known as the confidence level and is usually chosen to be large (≥ 0.9) usually with 95% being the most popular choice.

If we are working with 95% confidence intervals (CIs), we expect about 95% of them to contain the true mean and about 5% to miss the true mean. However for any particular interval, we cannot know which situation happens, unless its simulation like the following.



The vertical line is at the true mean

The dots are at the sample means for 50 different samples

The horizontal lines are the intervals for each of the 50 samples.

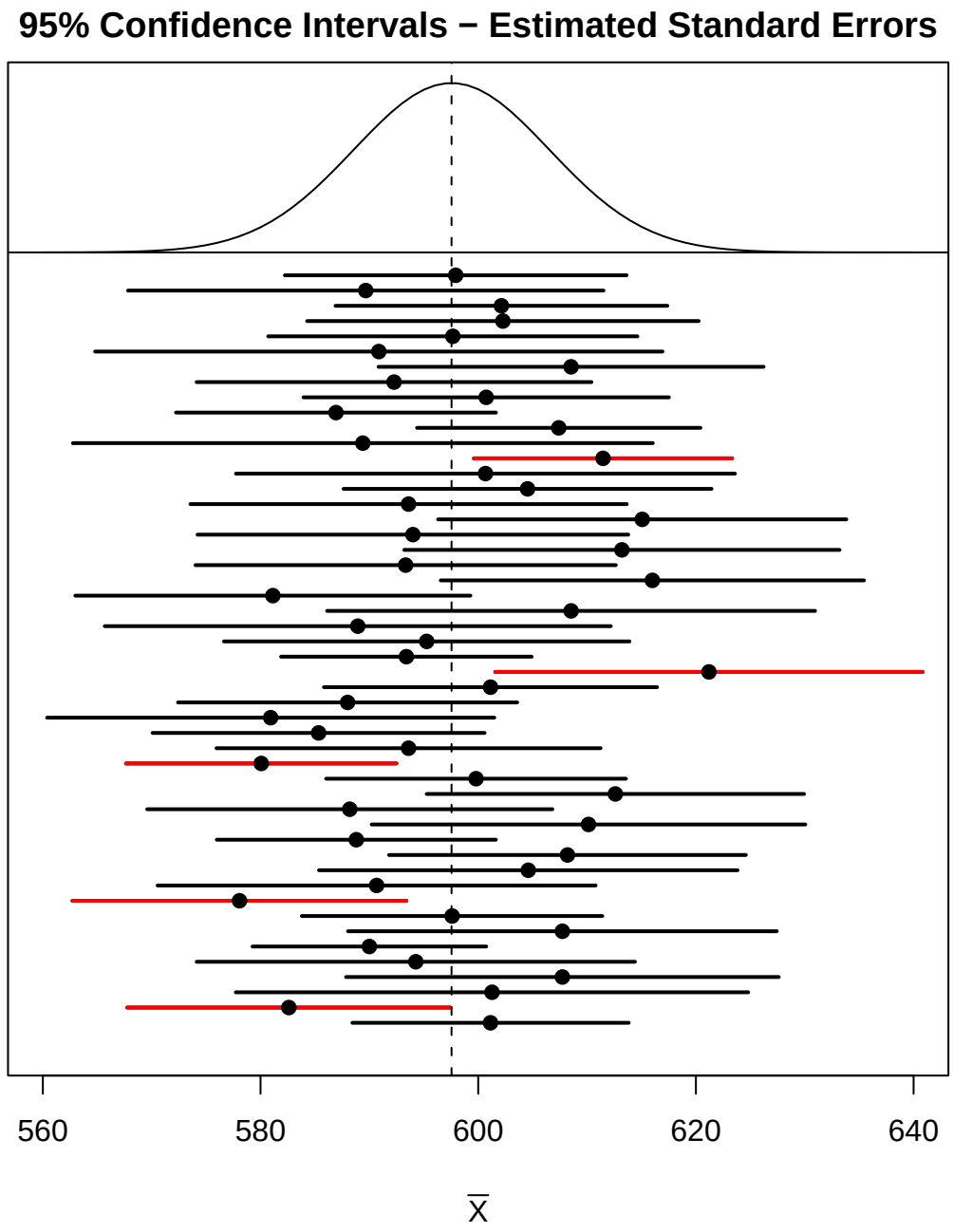
The density at the top is the normal approximation to the sampling distribution of $\bar{X}$.

Note: as mentioned before $\sigma_{\bar{X}}$ is unknown and must be estimated. So replace $\sigma_{\bar{X}}$ with $s_{\bar{X}}$ in the CI formula giving

$$\bar{X} \pm z(\alpha/2)s_{\bar{X}}$$

Also the normal approximation only holds when n is "large". A rule of thumb suggests 25 or 30 is often adequate.

Another modification that is often made is to replace the normal critical value $z(\alpha/2)$ with the t critical value $t_{1-\alpha/2}$ with $n-1$ degrees of freedom. However for $n \geq 30$, this change makes little difference.



We can also get CIs for population proportions.

The form of it matches the form of many CIs:

$$\text{Estimate} \pm \text{Critical Value} \times \text{Standard Error}$$

$$\text{Estimate} \pm \text{Margin of Error}$$

Thus a $100(1-\alpha)\%$ CI for p is

$$\hat{p} \pm z(\alpha/2)s_{\hat{p}}$$

where

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1-\frac{n}{N}}$$

Note that this interval will work better when n is large and $\hat{p}$ isn't close to 0 or 1. A modification of the rule of thumb for the normal approximation to the binomial should work here to suggest when this should be a reasonable interval. We want $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$.

Example: CIs for law school example (sample 1 only)

- 95% CI for μ_{LSAT}:

$$\bar{X} \pm z(\alpha/2)s_{\bar{X}} = 593.93 \pm 1.96 \times 12.23$$
$$= 595.93 \pm 23.97 = (571.96, 619.90)$$

- 90% CI for μ_{GPA}:

$$3.137 \pm 1.645 \times 0.0593 = 3.137 \pm 0.098 = (3.039, 3.235)$$

- 95% CI for p_{LSAT}:

$$0.133 \pm 1.96 \times 0.0821 = 0.133 \pm 0.161 = (-0.028, 0.294)$$

Here is an example where the normal approximation breaks down. An alternative procedure is needed to determine the CI in this case.

There are two main factors which determine the size of a confidence interval.

- Sample size n: As n increases, the standard error decreases
- Confidence level $1-\alpha$: As the confidence level increases $z(\alpha/2)$ increases.

Usually a narrow interval with high confidence is desired.

A problem of interest is to determine how big a sample is needed so that the margin of error $\leq$ as desired level δ. When doing this the FPC is usually ignored. What n such that

$$z(\alpha/2)\frac{\sigma}{\sqrt{n}} \leq \delta$$

To do this, you need some guess of σ. Given that this exists somewhere, then

$$n \geq \frac{z^2(\alpha/2)\sigma^2}{\delta^2}$$

In the situation of trying to estimate a population proportion, the problem is to solve

$$z(\alpha/2)\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \delta$$

which depend on p. However this margin of error is maximized when $p = 0.5$, so one can choose n satisfying

$$z(\alpha/2)\frac{1}{\sqrt{4n}} \leq \delta$$

or

$$n \geq \frac{z^2(\alpha/2)}{4\delta^2}$$

This choice of n will meet the criterion for any p.