5. Sampling, Point Estimation and the CLT

Readings: Rosner Ch. 6; Chihara and Hesterberg Ch. 4
R: dchisq, pchisq, qchisq, rchisq, runif, rbinom, rpois, rnorm
SAS: function RANNOR, PROBCHI, CINV

R: Lab 0 and Lab 1

Homework: Homework 2 due by 11:59 pm on September 17
Homework 3 due by 11:59 pm on September 24

Overview

A) Populations and Samples
B) Estimation – point vs. interval
C) s.e.($\bar{X}$) and the Sampling Distribution of $\bar{X}$
D) Central Limit Theorem
E) Sampling distribution of the Sample Variance

Recall what you know about the most common measure of central tendency for samples from any distribution - the sample average, $\bar{X}$.

It is an *unbiased* estimator of the true population mean μ, which means that on repeated sampling from the population the average value of $\bar{X}$ will be μ.

You already know about some theoretical distributions from which we could sample and obtain values of $\bar{X}$, e.g. binomial, Poisson, normal. In the R Labs we learned how to obtain *random* samples using, for example, random digits from various functions: runif, rbinom, rpois, rnorm.

It's important for us to know the *sampling distribution* of $\bar{X}$- the central tendency and variability of $\bar{X}$ itself, as well as the parametric form of the distribution - so that we can make inferences about the population mean μ based on $\bar{X}$, which is often of interest.

So, what might we want to know about a single value of the sample average $\bar{X}$?

- How accurate is it?
- How variable is it?
- How does the variability depend on *n*, the number of observations in the sample?
- What is the distribution of $\bar{X}$?
- What probability statements can we make about values that $\bar{X}$ could take on?

We'll need to use some theoretical results to answer these questions, since, in practice, we'll have only one value of $\bar{X}$ to work with.

A) Populations and Samples

Population: reference, target or study population is the group we wish to study. It contains all of the items we want to study. We select a sample from the study population.

Sample: a subset of items from the population.

Random Sample: random selection of some members of the population of interest such that each member is independently selected and has a *known* nonzero probability of being selected.

Simple Random Sample: each population member has the *same* (known) probability of being selected.

Probability and Statistics:

With probability we can study the behavior of samples from a known population.
With statistics we can make inferences about an unknown population from a sample.

A *statistic* (e.g. $\bar{X}$, median, s.d., s^2) is a combination of the random variables X_1, …, X_n, so it is itself a random variable. That is, the value of a statistic will vary from sample to sample, and has its own probability distribution. We call the distribution of a statistic its *sampling distribution.*

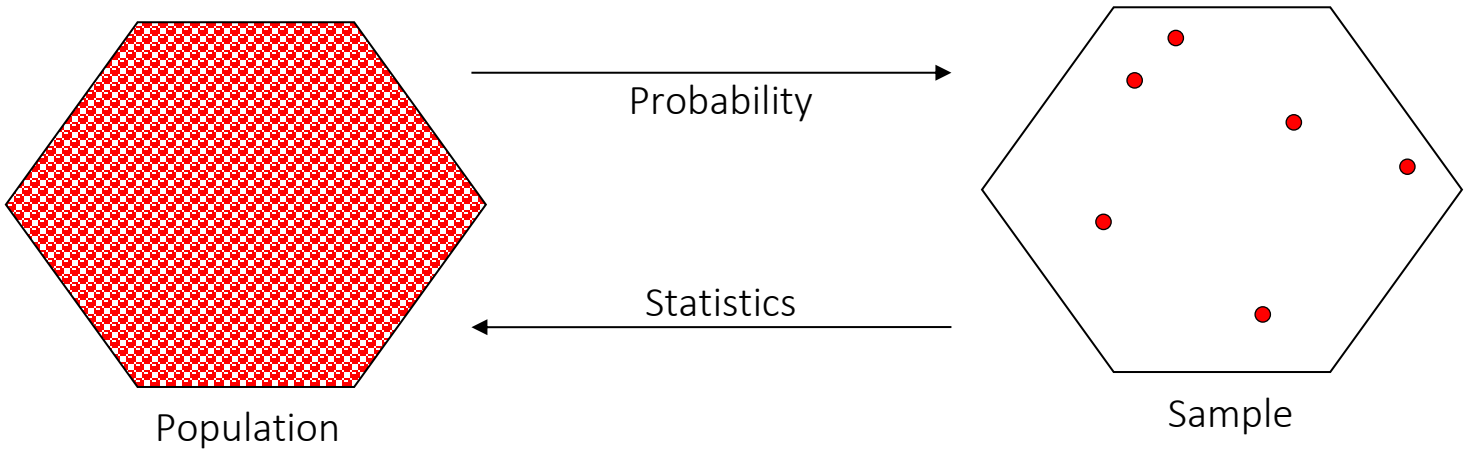


Table 6.2 Sample of birthweights (oz) obtained from 1000 consecutive deliveries at Boston City Hospital

ID Numbers	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
000–019	116	124	119	100	127	103	140	82	107	132	100	92	76	129	138	128	115	133	70	121
020–039	114	114	121	107	120	123	83	96	116	110	71	86	136	118	120	110	107	157	89	71
040–059	98	105	106	52	123	101	111	130	129	94	124	127	128	112	83	95	118	115	86	120
060–079	106	115	100	107	131	114	121	110	115	93	116	76	138	126	143	93	121	135	81	135
080–099	108	152	127	118	110	115	109	133	116	129	118	126	137	110	32	139	132	110	140	119
100–119	109	108	103	88	87	144	105	138	115	104	129	108	92	100	145	93	115	85	124	123
120–139	141	96	146	115	124	113	98	110	153	165	140	132	79	101	127	137	129	144	126	155
140–159	120	128	119	108	113	93	144	124	89	126	87	120	99	60	115	86	143	97	106	148
160–179	113	135	117	129	120	117	92	118	80	132	121	119	57	126	126	77	135	130	102	107
180–199	115	135	112	121	89	135	127	115	133	64	91	126	78	85	106	94	122	111	109	89
200–219	99	118	104	102	94	113	124	118	104	124	133	80	117	112	112	112	102	118	107	104
220–239	90	113	132	122	89	111	118	108	148	103	112	128	86	111	140	126	143	120	124	110
240–259	142	92	132	128	97	132	99	131	120	106	115	101	130	120	130	89	107	152	90	116
260–279	106	111	120	198	123	152	135	83	107	55	131	108	100	104	112	121	102	114	102	101
280–299	118	114	112	133	139	113	77	109	142	144	114	117	97	96	93	120	149	107	107	117
300–319	93	103	121	118	110	89	127	100	156	106	122	105	92	128	124	125	118	113	110	149
320–339	98	98	141	131	92	141	110	134	90	88	111	137	67	95	102	75	108	118	99	79
340–359	110	124	122	104	133	98	108	125	106	128	132	95	114	67	134	136	138	122	103	113
360–379	142	121	125	111	97	127	117	122	120	80	114	126	103	98	108	100	106	98	116	109
380–399	98	97	129	114	102	128	107	119	84	117	119	128	121	113	128	111	112	120	122	91
400–419	117	100	108	101	144	104	110	146	117	107	126	120	104	129	147	111	106	138	97	90
420–439	120	117	94	116	119	108	109	106	134	121	125	105	177	109	109	109	79	118	92	103
440–459	110	95	111	144	130	83	93	81	116	115	131	135	116	97	108	103	134	140	72	112
460–479	101	111	129	128	108	90	113	99	103	41	129	104	144	124	70	106	118	99	85	93
480–499	100	105	104	113	106	88	102	125	132	123	160	100	128	131	49	102	110	106	96	116
500–519	128	102	124	110	129	102	101	119	101	119	141	112	100	105	155	124	67	94	134	123
520–539	92	56	17	135	141	105	133	118	117	112	87	92	104	104	132	121	118	126	114	90
540–559	109	78	117	165	127	122	108	109	119	98	120	101	96	76	143	83	100	128	124	137
560–579	90	129	89	125	131	118	72	121	91	113	91	137	110	137	111	135	105	88	112	104
580–599	102	122	144	114	120	136	144	98	108	130	119	97	142	115	129	125	109	103	114	106
600–619	109	119	89	98	104	115	99	138	122	91	161	96	138	140	32	132	108	92	118	58
620–639	158	127	121	75	112	121	140	80	125	73	115	120	85	104	95	106	100	87	99	113
640–659	95	146	126	58	64	137	69	90	104	124	120	62	83	96	126	155	133	115	97	105
660–679	117	78	105	99	123	86	126	121	109	97	131	133	121	125	120	97	101	92	111	119
680–699	117	80	145	128	140	97	126	109	113	125	157	97	119	103	102	128	116	96	109	112
700–719	67	121	116	126	106	116	77	119	119	122	109	117	127	114	102	75	88	117	99	136
720–739	127	136	103	97	130	129	128	119	22	109	145	129	96	128	122	115	102	127	109	120
740–759	111	114	115	112	146	100	106	137	48	110	97	103	104	107	123	87	140	89	112	123
760–779	130	123	125	124	135	119	78	125	103	55	69	83	106	130	98	81	92	110	112	104
780–799	118	107	117	123	138	130	100	78	146	137	114	61	132	109	133	132	120	116	133	133
800–819	86	116	101	124	126	94	93	132	126	107	98	102	135	59	137	120	119	106	125	122
820–839	101	119	97	86	105	140	89	139	74	131	118	91	98	121	102	115	115	135	100	90
840–859	110	113	136	140	129	117	117	129	143	88	105	110	123	87	97	99	128	128	110	132
860–879	78	128	126	93	148	121	95	121	127	80	109	105	136	141	103	95	140	115	118	117
880–899	114	109	144	119	127	116	103	144	117	131	74	109	117	100	103	123	93	107	113	144
900–919	99	170	97	135	115	89	120	106	141	137	107	132	132	58	113	102	120	98	104	108
920–939	85	115	108	89	88	126	122	107	68	121	113	116	94	85	93	132	146	98	132	104
940–959	102	116	108	107	121	132	105	114	107	121	101	110	137	122	102	125	104	124	121	111
960–979	101	93	93	88	72	142	118	157	121	58	92	114	104	119	91	52	110	116	100	147
980–999	114	99	123	97	79	81	146	92	126	122	72	153	97	89	100	104	124	83	81	129

From Tables 6.2 and 6.3 in Rosner

For this population,

$\mu = 112$ and

$\sigma = 20.6$

Five random samples of size 10 from the population of infants whose birthweights (oz) appear in Table 6.2

	Sample				
Individual	1	2	3	4	5
1	97	177	97	101	137
2	117	198	125	114	118
3	140	107	62	79	78
4	78	99	120	120	129
5	99	104	132	115	87
6	148	121	135	117	110
7	108	148	118	106	106
8	135	133	137	86	116
9	126	126	126	110	140
10	121	115	118	119	98
$\bar{x}$	116.90	132.80	117.00	106.70	111.90
s	21.70	32.62	22.44	14.13	20.46

B) Estimation

Note: consistent with commonly used notation, X denotes a random variable while *x* denotes a specific realization of the random variable X in a sample.

For any sample we can obtain *point* estimates, $\bar{X}$, s^2, s, $\hat{p}$ for a population parameters μ, σ^2, σ, p, respectively. Because we know that these estimates will vary from sample to sample, we could also obtain interval estimates for these parameters: $\bar{X} \pm k$, $\hat{p} \pm k$, etc.

The *k* values reflect both the *inherent variability* in the random variable being measured and the *sampling variability* due to the size and nature of the sample.

As commonly interpreted: Interval estimates give a range of parameter values with which the data are consistent.

The usual estimator for the population mean μ is the sample mean: $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$. The *sampling distribution* of $\bar{X}$ is the (probability) distribution of values of $\bar{X}$ over *all* possible samples of size *n* that could have been selected from the reference population.

Desirable properties: A good statistic or point estimator for a population parameter should be unbiased and have high precision (i.e., low variance).

$\bar{X}$ is unbiased: $E[\bar{X}] = \mu$

We already know this, i.e. that the average or expected value of the estimator, $\bar{X}$, is the value of the population parameter, μ.

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

Thus, in any given sample we "expect" the sample average to be near the true population mean.

Note: other unbiased estimators of μ exist including the sample median (when the underlying distribution is symmetric, as seen in R Lab 1). However, $\bar{X}$ is also **the *minimum variance* unbiased estimator**. Of all the unbiased estimators for μ, $\bar{X}$ can be shown to be the estimator with the *smallest* variance. This makes $\bar{X}$ a very useful estimator indeed.

C) s.e. ($\bar{X}$) and the Sampling Distribution of $\bar{X}$

Standard Error of the Mean (sem): Measures the precision with which $\bar{X}$ estimates μ.

The larger the sample size, the more precise an estimator $\bar{X}$ will be:

$$V[\bar{X}] = V\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

σ^2 is the variability of the r.v. X in the population

$\frac{\sigma^2}{n}$ is the variability of the estimator $\bar{X}$

So, (on repeated sampling) the variability of the sample mean is *less than or equal to* the variability of the population, i.e. sample means are *less spread out about the true mean* than are the values of the individual population members.

Standard error of the mean (standard error, sem): the standard deviation of $\bar{X}$ over repeated (all possible) samples of size *n* from a population with underlying variance σ^2:

$$\sqrt{V[\bar{X}]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \sigma_{\bar{X}}$$

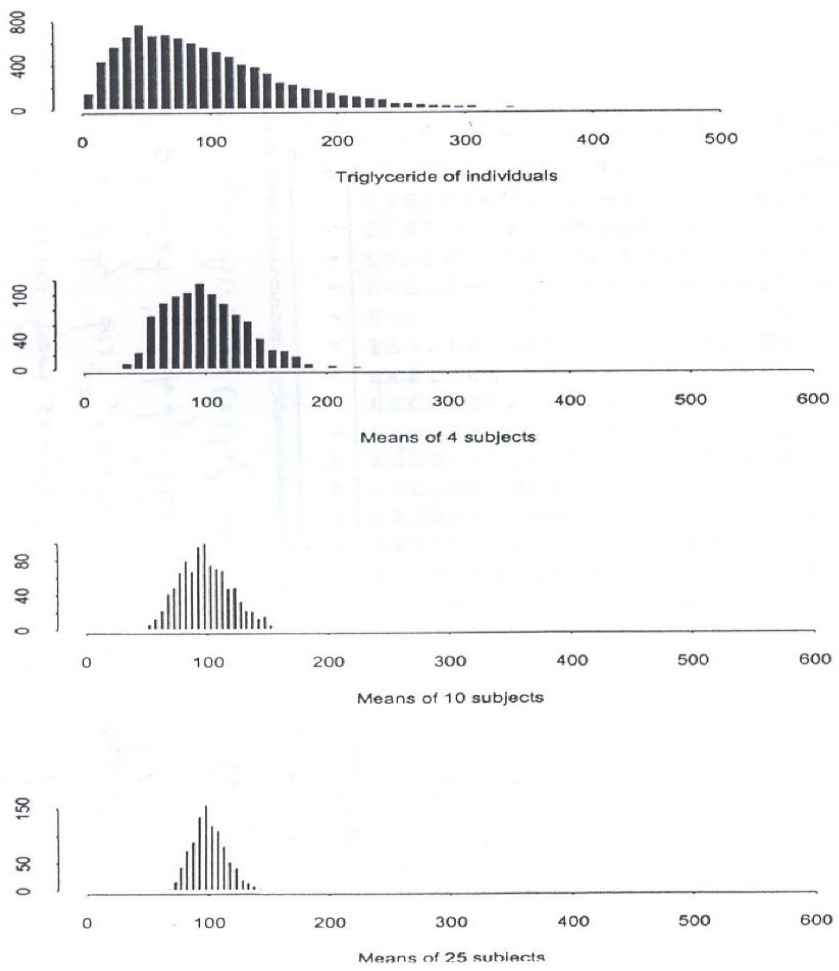
We estimate the sem by $\frac{s}{\sqrt{n}}$, where *s* is the standard deviation estimated from our sample.

We use the term *standard error* to refer to the (sampling) standard deviation of estimators of population parameters. For example, the standard deviation of (the sampling distribution of) $\hat{p}$ is the *standard error* of $\hat{p}$.

Sampling Distribution of $\bar{X}$

We can see that, as *n* increases, the variability of the sample mean decreases. In other words, as $n \to \infty$, $\sigma_{\bar{X}} \to 0$. If we had all of the data from the population in every sample we draw, then $\bar{X}$ would be equal to the true mean μ every time we sampled.

e.g. serum triglyceride data. Population has mean 100 mcg/dl and s.d. 71 mcg/dl. A histogram is shown to the right along with plots of mean values for samples of 4 subjects, 10 subjects, and 25 subjects.



Notice that as n ↑, the sampling distribution of $\bar{X}$ is narrower (smaller s.d., or *standard error*), more symmetric, and closer to being normally distributed. The variability can be predicted theoretically.

For $X_1, \ldots, X_n$ representing measurements on a random sample (i.e. $X_1, \ldots, X_n$ are independent and identically distributed, *and not necessarily normal*, random variables) from a population with mean μ and variance σ^2, then for "large" n,

$$\bar{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$, where $\dot{\sim}$ means "is approximately distributed as".

This very powerful and useful result is called the ***Central Limit Theorem*** (or CLT).

D) Central Limit Theorem

If the underlying distribution of X is normal, then, as we have seen, the sample mean (a linear combination of the X_i) will also be normally distributed: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

If, however, the underlying distribution of a random sample is *not* normal with mean μ and variance σ^2, then, for *large enough n*, $\bar{X}$ is *still* distributed approximately normal:

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

So, to reiterate, no matter what the underlying distribution is, the sampling distribution of $\bar{X}$ is approximately normal, as long as the sample size *n* is large enough.

Many of the distributions we'll encounter in data analysis won't be normal. What the CLT allows us to do is to perform statistical inference based on the *approximate* normality of the *sample mean*, despite the non-normality of the individual observations.

For many continuous distributions, the sampling distribution of $\bar{X}$ will become fairly normal for samples of size $n \geq 30$. How quickly this happens as n increases will depend on the shape of the (original) distribution of the X_i (i.e., the skewness, kurtosis, etc.).

Later we'll read a conference paper by Dr. Tim Hesterberg of Google arguing that the rate of convergence of the sampling distribution of $\bar{X}$ to a normal distribution is on "geological time scales". That means that, even though the sampling distribution *looks* normal, the *error* (i.e., variability) involved in, for example, confidence interval estimation can remain astonishingly large even for sample sizes much greater than 30.

Further comments on accuracy

Use of the CLT works well for samples sizes of ≥ 30 unless the distribution has extreme skewing or other oddities. For binomial data, a common rule of thumb is to use the CLT, *with continuity correction*, if both np ≥ 10 and n(1-p) ≥ 10 -OR- if np(1-p) ≥ 5.

For more generality, an expanded form of the CLT gives an estimate of the error in the typical use of the CLT:

$$P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = P(Z \leq z) \approx \Phi(z) + \frac{\kappa_3}{6\sqrt{n}}(1-z^2)\Phi'(z)$$

where Φ = cdf of normal distribution

κ_3 = skewness of the distribution being sampled (aka "3rd central moment")

Φ' = first derivative of cdf of normal, aka _____

Example from section 4.3.3 of Chihara and Hesterberg

- Suppose we are interested in the error of the CLT approximation for an exponential distribution at $z = 2.33$ since the probability $P(Z > z) = P(Z > 2.33) \approx 0.01$ is important in statistical practice.

- For accuracy, we want to use the CLT and have the approximation of this probability to be within 0.001 accuracy, i.e. from 0.009 to 0.011.

- What size sample size is needed to achieve this level of accuracy?

- Note, for an exponential distribution, $\kappa_3 = 2$.

- Also, $\Phi'(2.33) = \varphi(2.33) = 0.02642649$. Found using dnorm(2.33) in R:

```
little_phi <- dnorm(2.33)

little_phi

[1] 0.02642649
```

Now, let's solve for n …

$$0.001 = \frac{\kappa_3}{6\sqrt{n}}(1-z^2)\Phi'(z)$$
$$= \frac{2}{6\sqrt{n}}(1-2.33^2)(0.0264264)$$
$$= \frac{0.03901342}{\sqrt{n}}$$

Final step: $n = \left(\frac{0.03901342}{0.001}\right)^2 = 1522.047$

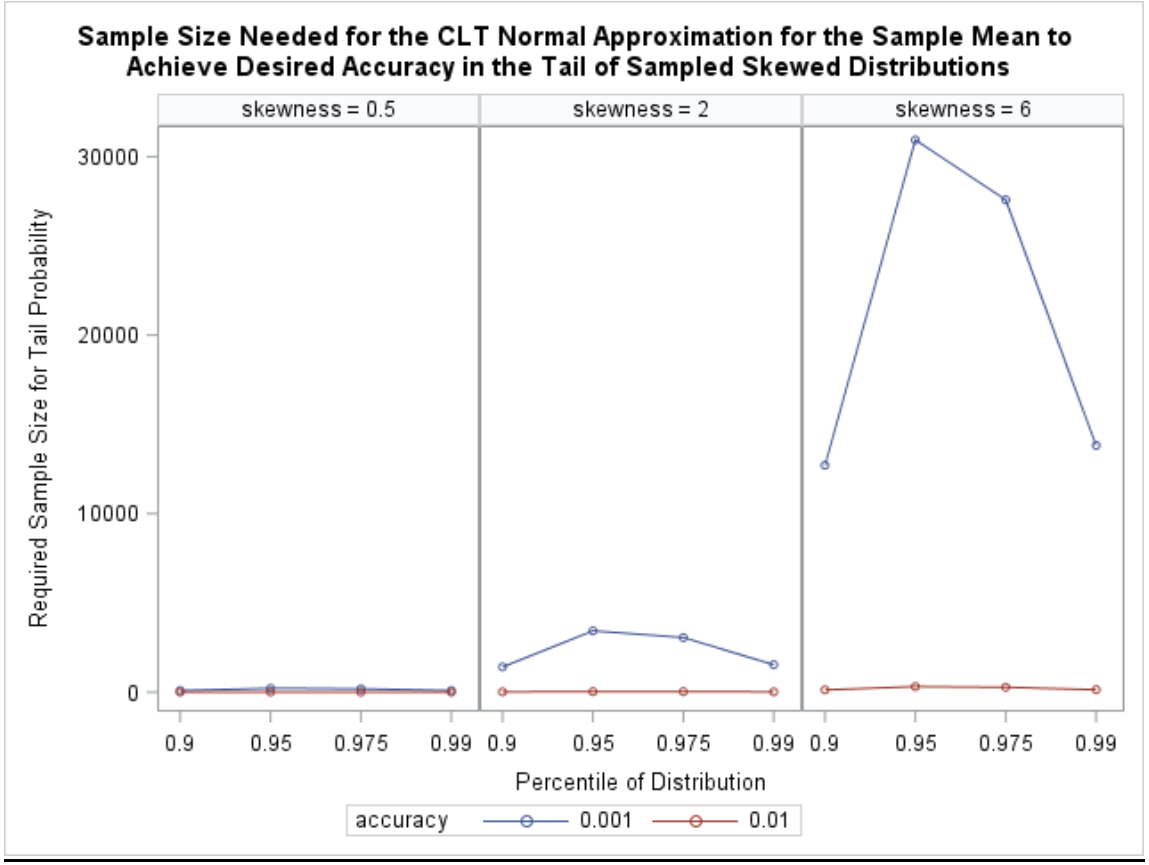
Hence, a direct application of the CLT for this situation and the desired accuracy would require n ≥ 1523!

Bottom line: depending on the level of accuracy and the distribution for the population, much larger sample sizes may be required in order to justify use of the CLT.

Our use of the CLT has assumed an infinite sample size. In cases where the population is finite, other formulations must be used, because a finite population doesn't have the possibility of $n \to \infty$!

If you are working with such a scenario, look for resources on survey sampling and statistics on finite populations.

For general values of skewness, let's take a look at the graph below to see how greater skewness affects the sample size calculation…



So, when would we want a high level of accuracy in applying the CLT?

Compelling argument from Hesterberg (2008)

Confidence interval coverage at each extreme can be off by (much) more than 10% when the underlying distribution being sampled is highly skewed and the sample size is only slightly larger than 30, i.e.

$P\left(\mu \leq \bar{X} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right)$ is not $\leq \pm 1.1 \times \frac{\alpha}{2}$ (i.e., within 10% of the target α)

and/or

$P\left(\mu \geq \bar{X} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right)$ is not $\leq \pm 1.1 \times \frac{\alpha}{2}$ (i.e., within 10% of the target α)

Bottom line: If we use the CLT to obtain normal or even t-distribution <u>confidence intervals</u> for population means (and other parameters) in order to draw conclusions when the distribution being sampled is very skewed, we can be wrong more often than we realize!

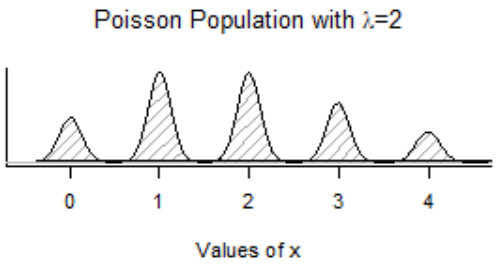
Lower bound of CI:

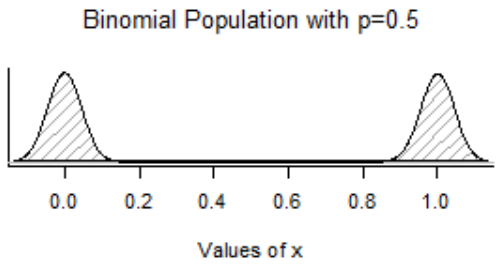
e.g. quality control: we want to be assured that *at least* a certain threshold is achieved (a minimum quantity exists)

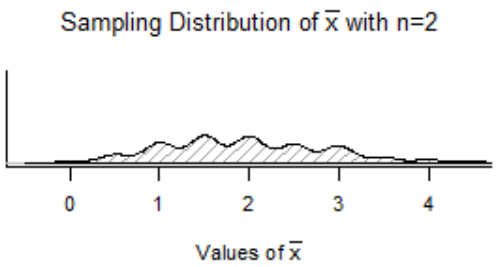
Upper bound of CI:

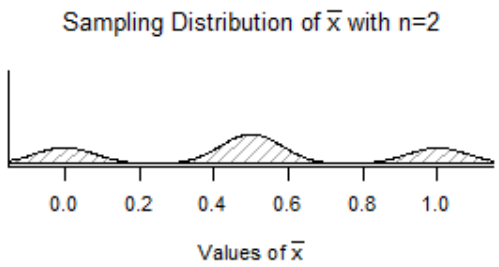
e.g. ensure that a toxic substance *does not exceed* a maximum

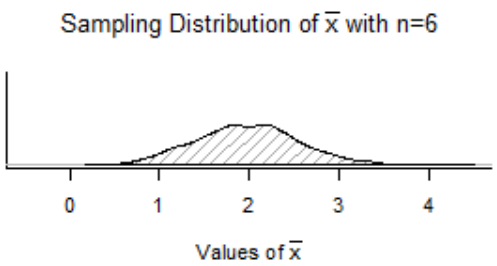
Related to this but slightly different: Under the right conditions, the normal distribution can approximate discrete distributions like the binomial and Poisson distributions (see Lecture 4).

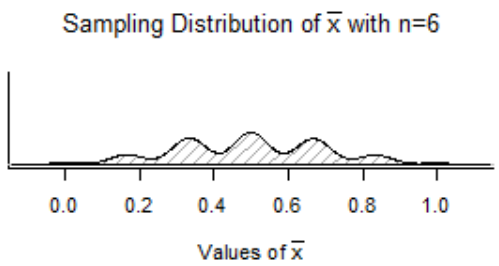


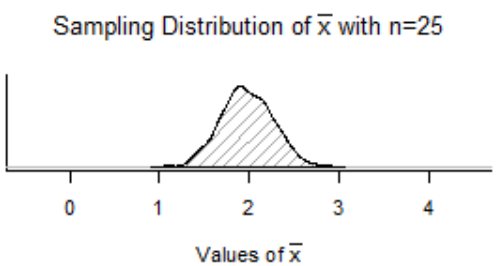


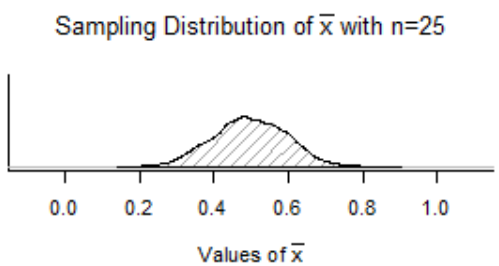












Example: The weight of 6-year old boys is normally distributed with $\mu = 40$ lbs and $\sigma^2 = 25$ lbs^2.

What is the probability that the sample mean weight of 50 boys would be between 38 and 42 pounds, $P(38 \leq \bar{X} \leq 42)$?

If $X \sim N(40, 25)$. What is the distribution of $\bar{X}$?

Compare this to $P(38 \leq X \leq 42)$.

Why, then, is the CLT important? …

E) Distribution of the Sample Variance

Thinking in terms of repeated random sampling from the population of interest, the sample variance, s^2, will also vary from sample to sample.

Recall: $V[X] = \sigma^2 = E[(X-\mu)^2]$, the average value of $(X-\mu)^2$ over all possible samples of size n. The intuitive estimator of σ^2 is $\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n}$.

However, $E\left[\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n}\right] = \frac{n-1}{n}\sigma^2$ and so $\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n}$ is a **biased** estimator of σ^2.

Instead, $E[s^2] = E\left[\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n-1}\right] = \sigma^2$.

However, $E[s] \neq \sigma$, so s is a **biased** estimator of σ.

Thus, $E\left[\frac{s^2}{n}\right] = \frac{\sigma^2}{n}$ and is **unbiased,** but $E\left[\frac{s}{\sqrt{n}}\right] \neq \frac{\sigma}{\sqrt{n}}$ and is **biased**.

Measure of how precisely s^2 estimates σ^2:

$$V[s^2] = \frac{2\sigma^4}{(n-1)} \text{ with estimate } \frac{2s^4}{(n-1)}$$

$$s.d.[s^2] = se[s^2] = \sigma^2\sqrt{\frac{2}{(n-1)}} \text{ with estimate } s^2\sqrt{\frac{2}{(n-1)}}$$

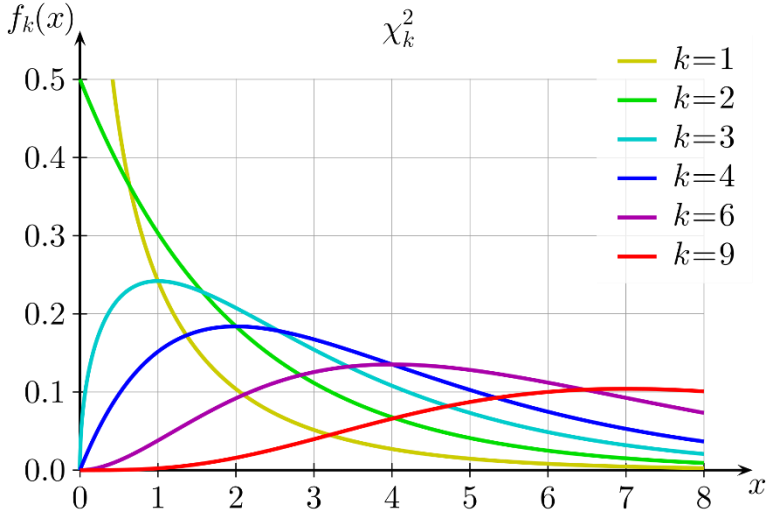
This derives from the following:

If the individual observations in a sample, $X_1, X_2, \ldots, X_n$ are *iid* $N(\mu, \sigma^2)$, then

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

χ^2 = "chi-square": $\chi^2(n-1)$ is a chi-square sampling distribution where *n-1* indicates the **degrees of freedom (df),** which are the number of independent data points that are free to vary.

The χ^2 distribution is skewed to the right with no negative values and a range from 0 to $+\infty$. As the *df* increase, the distribution becomes more symmetric, more normal shaped.



https://en.wikipedia.org/wiki/Chi-squared_distribution

The chi-square distribution also applies to a sum of independent standard normal random variables: If $X_1, X_2, \ldots, X_n \sim N(0,1)$, then $G = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$ (also denoted as $\chi^2(n)$).

If $X \sim \chi_v^2$, then E[X] = v and V(X) = 2v.

So, if $X_1, X_2, \ldots, X_n$ iid $N(\mu, \sigma^2)$ and $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$:

$$E\left[\frac{(n-1)s^2}{\sigma^2}\right] = (n-1) \quad \Rightarrow \quad E(s^2) = \sigma^2$$

$$V\left[\frac{(n-1)s^2}{\sigma^2}\right] = 2(n-1) \quad \Rightarrow \quad V(s^2) = \frac{2\sigma^4}{n-1}$$

The χ^2 distribution is the *sampling distribution* for the statistic: $\frac{(n-1)s^2}{\sigma^2}$.

This is analogous to the standard normal distribution being the sampling distribution for the statistic $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.

The percentiles of the chi-square distribution can be found in textbook tables. You can also use functions in R (or SAS, Stata, SPSS). In R, the pchisq(x, df) function gives values of the chi-square cdf, and the qchisq(prob, df) function gives quantiles of the chi-square cdf.

Example: Let X~N(0,1). What is the distribution of X^2?

If 1.96 is the 97.5 percentile of the N(0,1) distribution, what value and percentile does it map to in the χ^2 distribution?

Using R:

```
# Quantile of standard normal distribution
z <- 1.96
# Transformation result chi-square with 1 df ~ z^2
chisq <- z^2
# Print chisq value
chisq
[1] 3.8416
# cdf of chi-square distribution
pchisq(chisq, df = 1)
[1] 0.9500042
# Cumulative probability
p <- 0.95
# Quantile function for chi-square distribution
qchisq(p, df = 1)
[1] 3.841459
```

Using SAS:

```
DATA probs;
   z = 1.96;
      chisq = z**2;
      prob = probchi(chisq, 1);/*gives P(CHI < chi)
for X~Chisq(1)*/
   p = 0.95;
      chi = cinv(p, 1);  /*gives chi so that P(CHI <
chi) = p*/
RUN;

PROC PRINT; RUN;
```

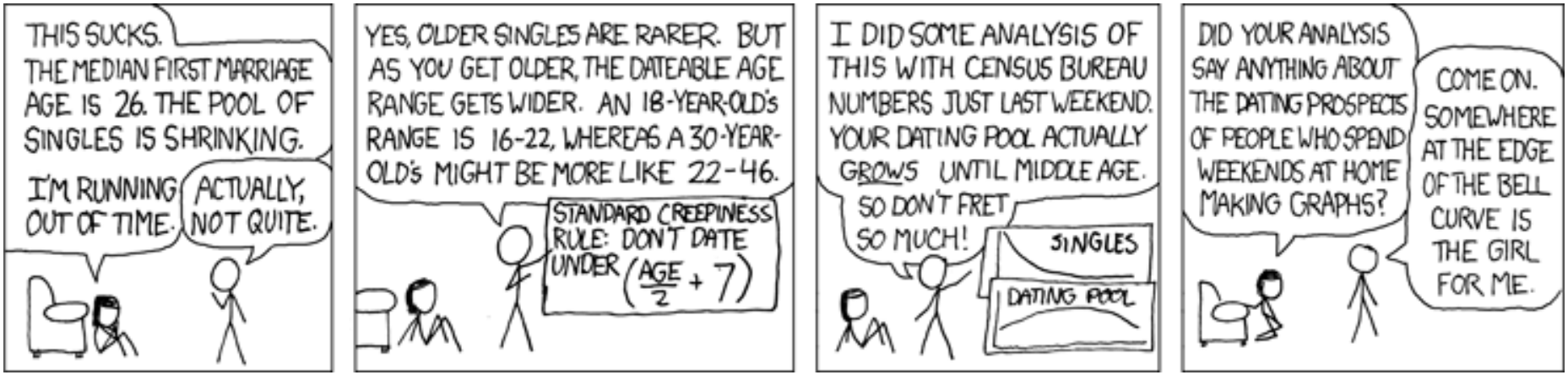
Obs	z	chisq	prob	p	chi
1	**1.96**	**3.8416**	**0.95000**	**0.95**	**3.84146**

Summary:

	Empirical	*Theoretical*	
	Data Sample	Discrete r.v. X	Continuous r.v. X
Density	Histogram or relative frequency distribution	p(x) = P (X = x)	f(x) smooth function
Properties	$\sum$ rel freq = 1	$\sum_{x} p(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$
CDF	$\frac{\#x_i \le x}{\#x_i}$	$P(X \le x) = \sum_{k \le x} p(x)$	$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$
Mean	$\overline{X}$	$\mu = E(x) = \sum_{x} xp(x)$	$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$
Variance	$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$	$\sigma^2 = V(x) = \sum_{x} (x-\mu)^2 p(x)$	$\sigma^2 = V[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$
Standard Dev	$s = \sqrt{s^2}$	$\sigma = s.d.(X) = \sqrt{V[X]}$	$\sigma = s.d.(X) = \sqrt{V[X]}$
s.e.($\overline{X}$)	$\frac{s}{\sqrt{n}}$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$
s.e.(s^2)	$s^2\sqrt{\frac{2}{(n-1)}}$	$SD[s^2] = SE[s^2] = \sigma^2\sqrt{\frac{2}{(n-1)}}$	$SD[s^2] = SE[s^2] = \sigma^2\sqrt{\frac{2}{(n-1)}}$

☺

Better living through the CLT (… *not* chemistry): "I used to be skewed, but through repeated large sampling, I'm starting to feel almost normal."



Source: xkcd.com #314