Chapter 7: Sampling and Sampling Distributions

We collect data on a sample to make inferences about the sampled population.

Sampled population: population the sample is drawn from.
Frame: list of all the elements from the sampled population.

A sample mean provides an estimate of a population mean.
A sample proportion provides an estimate of a population proportion.
Estimation error can be expected with these estimates.

7.1 The Electronics Associates Sampling Problem

Population Parameters: numerical characteristics of a population.
e.g.: A firm's population of managers *mean annual salary* or *proportion who completed training*.

7.2 Selecting a Sample

7.2.1 Sampling from a Finite Population

Simple random sample: each possible sample of the same size has the same probability of being selected (table of random #s, *Microsoft Excel function* **rand**).

https://support.microsoft.com/en-us/help/828795/description-of-the-rand-function-in-excel
https://www.ablebits.com/office-addins-blog/2015/07/08/random-number-generator-excel/
http://mathworld.wolfram.com/RandomNumber.html

7.2.2 Sampling from an Infinite Population

Random sample (when a frame cannot be constructed): each element is selected independently (from the same population).

7.3 Point Estimation

Sample Statistics: numerical characteristics of a sample.
e.g. A firm's sample of managers *mean annual salary* or *proportion who completed training*.

Point estimation: using sample statistics as point estimates of population parameters.
e.g. The value $51,814, obtained for the **point estimator** $\bar{x}$, is the **point estimate** of population mean μ.

7.3.1 Practical Advice

Target population: population we want to make inferences about.

When making inferences about a population based on a sample, it is important to have a close correspondence between the sampled population and the target population.

7.4 Introduction to Sampling Distributions

Each simple random sample of size n is different than the next.
Each simple random sample of size n will yield different point estimates than the next.
This means that point estimators (such as $\bar{x}$) are random variables (see Ch.5).
A distribution of every possible point estimates of a point estimator is called a **sampling distribution**.
Like any other, a sampling distribution has its own mean & standard deviation (such as E($\bar{x}$) & $\sigma_{\bar{x}}$).

7.5 Sampling Distribution of (the sample mean) $\bar{x}$

Sampling distribution of $\bar{x}$: probability distribution of all possible values of the sample mean $\bar{x}$.

Expected value of $\bar{x}$ (*assuming μ is known*):
E($\bar{x}$) = μ
where
μ = sampled population mean

Standard deviation (or standard error) of $\bar{x}$ (*assuming σ is known*):

Finite population: $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$ (1)

Infinite population: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (2)

where
σ = (sampled) population standard deviation
N = (sampled) population size
n = sample size

NOTE: If N is large relative to n, the finite population correction factor, $\sqrt{\frac{N-n}{N-1}}$, is closed to 1.
Consequently, use (2) for finite population whenever n/N ≤ .05.

Form of the sampling distribution of $\bar{x}$: normal distribution
(Sampled) population has a normal distribution: the sampling distribution of $\bar{x}$ is normally distributed.
(Sampled) population does not have a normal distribution: the sampling distribution of $\bar{x}$ can be approximated by a normal distribution as the sample size becomes large (**Central Limit Theorem; CLT**).

NOTE: General statistical practice appeals to the CLT whenever n ≥ 30, even though CLT theoretically applies only to sampling *with* replacement, for finite populations.

Practical value of $\bar{x}$ being normally distributed:
Once we've identified E($\bar{x}$) and $\sigma_{\bar{x}}$, we can answer probability questions, like we did in Chapter 6.

Relationship between the sample size and the sampling distribution of $\bar{x}$:
Because $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, when n increases, $\sigma_{\bar{x}}$ decreases, which means the values of $\bar{x}$ have less variation and tend to be closer to the (sampled) population mean. As a result, a larger sample size provides a higher probability that the sample mean is within a specified distance of the (sampled) population mean.

7.6 Sampling Distribution of (the sample proportion) $\bar{p}$

Sample proportion $\bar{p}$:
$\bar{p}$ = x/n
where
x = the # of elements in the sample that possess the characteristic of interest (binomial random variable)
n = sample size

Sampling distribution of $\bar{p}$: probability distribution of all possible values of the sample proportion $\bar{p}$.

Expected value of $\bar{p}$ (assuming *p* is known):
$E(\bar{p}) = p$
where
p = sampled population proportion

Standard deviation (or standard error) of $\bar{p}$ (assuming *p* is known):

Finite population: $\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$ (1)

Infinite population: $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ (2)

where
p = (sampled) population proportion
N = (sampled) population size
n = sample size

NOTE: If N is large relative to n, the finite population correction factor, $\sqrt{\frac{N-n}{N-1}}$, is closed to 1.
Consequently, use (2) for finite population whenever n/N ≤ .05.

Form of the sampling distribution of $\bar{p}$: normal distribution
If $np < 5$ and $n(1-p) < 5$: sampling distribution of $\bar{p}$ is binomial.
If $np \geq 5$ and $n(1-p) \geq 5$: sampling distribution of $\bar{p}$ can be approximated by a normal distribution (Ch. 6)

NOTE: In practical applications, we find that sample sizes are almost always large enough to permit the use of a normal approximation for the sample distribution of $\bar{p}$.

Practical value of $\bar{p}$ being normally distributed:
Once we've identified $E(\bar{p})$ and $\sigma_{\bar{p}}$, we can answer probability questions, like we did in Chapter 6.

7.7 Properties of Point Estimators

Unbiased: The sample statistic $\hat{\theta}$ is an unbiased estimator of the population parameter θ if $E(\hat{\theta}) = \theta$. In other words, a point estimator is unbiased when its expected value (mean of all possible values) equals its sampled population parameter.

$\bar{x}$ = sample mean
$\bar{p}$ = sample proportion
s^2 = sample variance

Since $E(\bar{x}) = \mu$, where μ = (sampled) population mean, $\bar{x}$ is an unbiased estimator of μ.
Since $E(\bar{p}) = p$, where p = (sampled) population proportion, $\bar{p}$ is an unbiased estimator of p.
Since $E(s^2) = \sigma^2$, where σ^2 = (sampled) population variance, s^2 is an unbiased estimator of σ^2.

Proof that $E(\bar{x}) = \mu$ and $Var(\bar{x}) = \sigma^2$:

$E(\bar{x}) = E(\Sigma x_i/n)$
$= (1/n)E(\Sigma x_i)$
$= (1/n)\Sigma E(x_i)$
$= (1/n)n\mu$, using $E(x_i) = \mu$
$= \mu$

$Var(\bar{x}) = Var(\Sigma x_i/n)$
$= (1/n)^2\Sigma Var(x_i)$ since $Cov(x_i,x_j) = 0$, where $i \neq j$ (see Ch.5)
$= (1/n)^2 n Var(x_i)$
$= \sigma^2/n$, using $Var(x_i) = \sigma^2$

Proof that $E(\bar{p}) = p$:

$E(\bar{p})$ = E(x/n), where x = the # of elements in the sample that possess the characteristic of interest
$= (1/n)E(x)$
= p, using E(x) = np since x is a binomial random variable (see 7.6 and Ch.5)

Proof that $E(s^2) = \sigma^2$:

NOTE, $Var(x_i) = E(x_i^2) - [E(x_i)]^2$ (see Ch.5) ⇔ (1): $E(x_i^2) = \sigma^2 + \mu^2$ since $Var(x_i) = \sigma^2$ and $E(x_i) = \mu$.
Similarly, $Var(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$ ⇔ (2): $E(\bar{x}^2) = \sigma^2/n + \mu^2$ since $Var(\bar{x}) = \sigma^2/n$ and $E(\bar{x}) = \mu$ (see above).

$E(s^2) = E[\Sigma(x_i - \bar{x})^2/(n-1)]$, using definition of variance for a sample (See Ch. 3.)
$= (1/(n-1))E[\Sigma(x_i - \bar{x})^2]$
$= (1/(n-1))E[\Sigma(x_i^2 - 2\bar{x}x_i + \bar{x}^2)]$
$= (1/(n-1))E[\Sigma x_i^2 - 2\bar{x}\Sigma x_i + n\bar{x}^2)]$
$= (1/(n-1))E[\Sigma x_i^2 - 2n\bar{x}^2 + n\bar{x}^2)]$, using $\bar{x} = \Sigma x_i/n \Leftrightarrow \Sigma x_i = n\bar{x}$
$= (1/(n-1))E[\Sigma x_i^2 - n\bar{x}^2)]$
$= (1/(n-1))(E[\Sigma x_i^2] - E[n\bar{x}^2])$
$= (1/(n-1))(\Sigma E[x_i^2] - nE[\bar{x}^2])$
$= (1/(n-1))(\Sigma E[x_i^2] - nE[\bar{x}^2])$
$= (1/(n-1))(\Sigma(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2))$, using (1) and (2)
$= (1/(n-1))(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2)$
$= (1/(n-1))((n-1)\sigma^2)$
$= \sigma^2$

Additional Reference: jbstatistics.com Sections 4.4, 4.5 & 8.2, viewed 4/21/17.

Efficiency: Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased point estimators of the population parameter θ, with standard error of $\hat{\theta}_1$ less than the standard error of $\hat{\theta}_2$; then, $\hat{\theta}_1$ is said to be relatively more efficient than $\hat{\theta}_2$ and is the preferred point estimator. Note that, because the standard error of $\hat{\theta}_1$ is less than the standard error of $\hat{\theta}_2$, the values of $\hat{\theta}_1$ have a greater chance of being close to the population parameter θ than do values of $\hat{\theta}_2$; that is why the more efficient point estimator $\hat{\theta}_1$ is preferred to the less efficient one $\hat{\theta}_2$.

Consistency: A point estimator is consistent if the values of the point estimator tend to become closer to the population parameter as the sample size becomes larger.

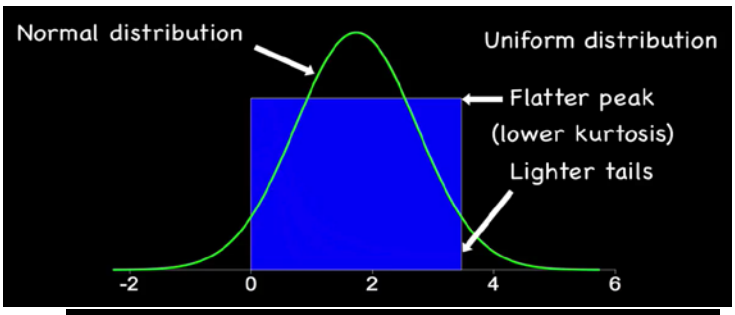
7.7b Sampling Distribution of (the sample variance) s^2 *ADDITIONAL MATERIAL*

Sampling distribution of s^2: probability distribution of all possible values of the sample variance s^2.

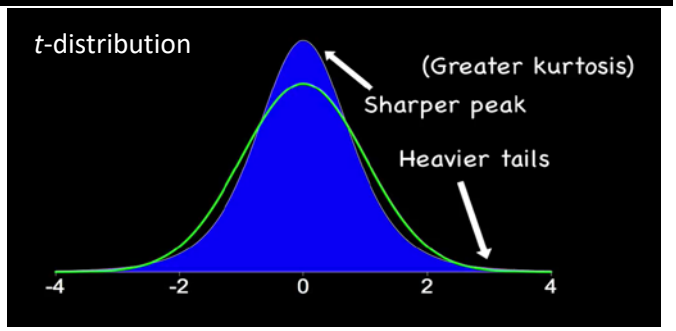
Expected value of s^2 (*assuming* σ^2 *is known*):
$E(s^2) = \sigma^2$
where
σ^2 = sampled population variance

Standard deviation (or standard error) of s^2
The variance (and standard error) of the sampling distribution of the sample variance depends on whether the distribution from which we are sampling is peaked or flat relative to the normal distribution (**kurtosis**).

For (sampled) population with lower kurtosis (flatter peak, lighter tails), we do not get extreme values, so the sample variances do not take on large values as often as it would if the sampled population was normally distributed.



For (sampled) population with higher kurtosis (sharper peak, heavier tails), we get more extreme values, so there is more variability in s^2.

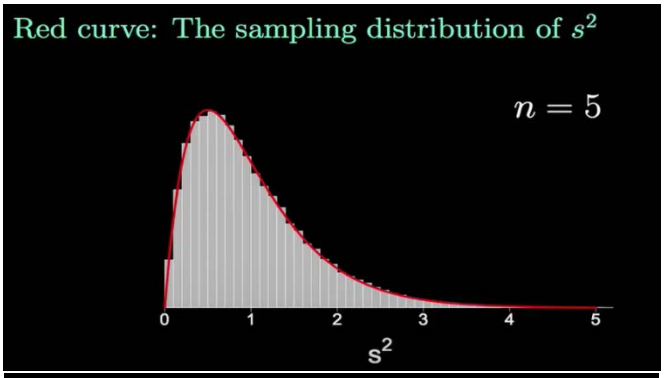


In short, the variance (and standard error) of the sampling distribution of s^2 can be much less or much greater than when we are sampling from a normally distributed population with the same value of σ^2.

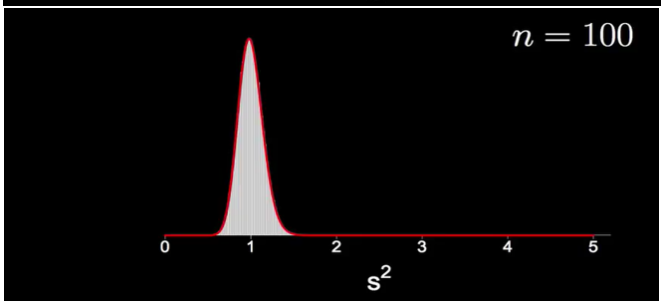
The variance (and standard error) of the sampling distribution of the sample variance also depends on the sample size: the higher the sample size, the less variation in the sampling distribution of the sample variance.

Form of the sampling distribution of s^2:
$(n-1)\,s^2/\sigma^2$ ischi-square (χ^2) distributed with (n – 1) degrees of freedom, *assuming σ^2 is known*.

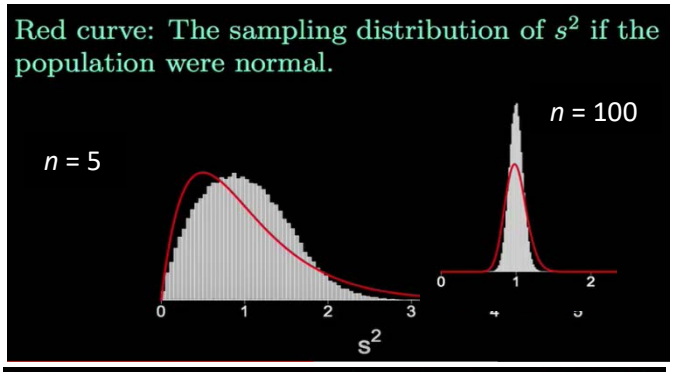
NOTE: The grey histograms are built from 100,000 simulated values of s^2 for samples of size *n* and $\sigma^2 = 1$.



(Sampled) population has a normal distribution: the sampling distribution of s^2 is Chi-Square distributed (with n - 1 degrees of freedom), more & more normally distributed as the sample size increases and approximately normal for large sample sizes.

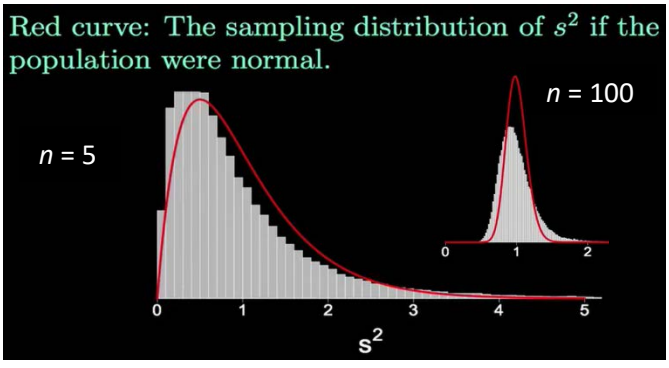


Sampling from:
A uniform distribution with $\sigma^2 = 1$



(Sampled) population does not have a normal distribution: here, less skewness in the distribution than when the sampled population is normal.

Sampling from:
A heavier-tailed distribution with $\sigma^2 = 1$



(Sampled) population does not have a normal distribution: here, more skewness in the distribution than when the sampled population is normal.

(Sampled) population does not have a normal distribution: more & more normally distributed as the sample size increases and approximately normal for large sample sample sizes (may need to be very large).

Reference: jbstatistics.com Section 10.4, viewed 4/27/17.

7.8 Other Sampling Methods

Strata: Groups that elements in the population are divided into. Each element belongs to one and only one stratum. Variance among elements in each stratum should be relatively small.

Stratified Random Sampling: a simple random sample is taken from each stratum. Results from the individual stratum samples are then combined into one estimate of the population parameter of interest.

Clusters: Groups that elements in the population are divided into. Each element belongs to one and only one cluster. Each cluster should provide a small-scale representation of the population.

Cluster Sampling: a simple random sample of the clusters is taken. All elements within each sampled cluster form the sample.

Systematic Sampling: It is an alternative to simple random sampling in some sampling situations with large populations. If a sample size of n is desired, from a population containing N elements, then one element for every N/n elements in the population is sampled: one of the first N/n elements from the population is ramdomly selected; the other sample elements are identified by starting with the first sampled element and then selecing every (N/n)th element that follows in the population list.

Probability Sampling Techniques: elements selected from the population have a known probability of being included in the sample. With probability sampling, the sampling distribution of the sample statistics can be identified and used to make probability statements about the error associated with using the sample results to make inferences about the population.

Convenience Sampling: the sample is identified primarily by convenience—easy sample selection and data collection. It is a *non*probability sampling technique.

Judgment Sampling: the person most knowledgeable on the subject of the study selects elements of the population that he or she feels are most representative of the population. It is a *non*probability sampling technique.