TELFER

VOTRE LIEN AVEC CE QUI COMPTE — CONNECTS YOU TO WHAT MATTERS

Sampling Distributions: a Review

ADM 2304 – Winter 2016

Probability and Statistics

- Probability Analysis: *If* a population has certain (known) characteristics, what samples and what sample statistics are likely or unlikely? (top-down “what if” analysis)
- Statistical Analysis: Given a sample statistic, what can we infer about the population? (bottom-up inference)

Estimating a Mean/Proportion

- We calculate sample statistics like the sample mean (*y-bar*) or the sample proportion (*p-hat*);
- But how “good” are these estimates of the population mean μ or of the population proportion p? *(statistical inference)*
- To answer this question, we need to know about the “Sampling Distribution of the Sample Mean/Proportion”. *(probability analysis)*

Sampling Distribution of the Mean

- Imagine taking all possible samples of size n from a population with mean μ and variance σ^2 (the population distribution can be *skewed* or *non-normal* in shape but *not extremely skewed).*

Now suppose we calculate the mean (y-bar) of each sample.

(An extremely skewed distribution would require a logarithmic transformation to become symmetric.)

Sampling Dist'n of the Mean

The Sampling Distribution of y-bar is the population of ***sample means*** derived from all possible samples of size n.

Its three important properties are:

1. The mean E(y-bar) is μ,
2. The variance is σ^2/n & its stdev (SD) is $\sigma/\sqrt{n}$,
3. The shape is approximately normal if the sample size is large (if the population is not extremely skewed, then n > 30 is sufficiently large) – this is the Central Limit Theorem.

see SDVW, 2nd ed.,10.3 and 10.4.

Example

Suppose the mean starting salary in a population of commerce graduates is \$50,000, with a standard deviation of \$15,000.

1. What proportion of incomes exceed \$71,000?
2. What is the probability that the mean income of a sample of 35 graduates exceeds \$71,000?

(What assumptions are required to answer the above questions and how reasonable are they?)

Sampling Distribution of the Sample Total (optional)

The sampling distribution of the sample total (T = n * y-bar) also has a normal shape as n gets large, with mean $n\mu$ and stdev $\sigma\sqrt{n}$.

Example:

What is the probability that the total income of 35 graduates exceeds $2,485,000?

(n.b. $2485 = 35 * $71)

Normal Approximation to the Binomial Distribution

- Suppose the population is comprised of two qualitative classes of subjects and p is the proportion who belong to the class of interest.
- If we code the data using 0s and 1s, then the population mean μ is *p* and the stdev σ is $\sqrt{(pq)}$, where q = 1-p.
  (Can you calculate this mean and standard deviation?)

The binomial random variable X (a count of the number of 1s in a sample of size n) is a *sample total*; therefore, the mean E(X) is $n\mu = np$ and the stdev is $\sigma\sqrt{n} = \sqrt{(npq)}$, and if the sample size is large enough ($np \geq 10$ and $nq \geq 10$) such that n/N < 10%, then the binomial distribution can be approximated by a normal distribution (CLT).

- *Ref. SDVW, 2nd ed., 9.11.*

Examples

- If the true value of p is .40, what is the probability that a survey of 1000 respondents finds between 390 and 410 supporting the Liberal party? *(Use normal approximation with continuity correction.)*
- If the incidence p of hemophilia is .006% in the population, what is the probability that a sample of 50000 finds 2 or fewer hemophiliacs? *(Use Poisson approximation with $\lambda = np$, since the sample size is not large enough for a normal approximation.)*

Sampling Distribution of the Sample Proportion

- The sample (binomial) proportion ***p-hat* = *X/n*** has the form *Total/n* which is a *Sample Mean* (where X is the number of 1s in the sample of size n).
- Therefore, the sample proportion ***p-hat*** has mean *E(**p-hat**)* = μ = *p* and *stdev* = $\sigma/\sqrt{n} = \sqrt{(pq/n)}$ and its distribution approaches a **normal** shape as n gets large (n$p \geq 10$ and n$q \geq 10$, provided that $n/N \leq 10\%$), *where N is the population size.*
- *Ref. SDVW, 2nd ed., 10.2*

Example

- If 40% of the electorate support the Liberals, what is the probability that between .29 and .31 of a sample of 1000 voters supports the Liberals?

Summary

Sample Statistic	Mean of Sampling Dist'n	Stdev of Sampling Dist'n
y-bar	**μ**	**$\sigma/\sqrt{n}$**
p-hat	**p**	**$\sqrt{(pq/n)}$**
Total (optional)	$n\mu$	$\sigma\sqrt{n}$
Binomial Count	np	$\sqrt{(npq)}$