STA304

Neil Montgomery
2016-01-14

simple random sampling

Recall the definition, and one practical method

"all samples of size n have the same probability of being selected"

So there are $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ different possible samples, each with probability $\binom{N}{n}^{-1}$ of being selected.

Theorem: Selecting elements one at a time without replacement results in a simple random sample. (*Last week this was quietly assumed.*)

Proof: The sample $\{e_1, e_2, \ldots, e_n\}$ could be selected in any order, e.g. $e_4 e_2 e_n e_{22} \cdots e_{42}$. There are $n!$ such orders. Any particular order will be selected with probability $\frac{1}{N}\frac{1}{N-1}\cdots\frac{1}{N-(n-1)}$. So:

$$P(\{e_1, e_2, \ldots, e_n\}) = n!\frac{1}{N}\frac{1}{N-1}\cdots\frac{1}{N-(n-1)} = \binom{N}{n}^{-1}$$

When you encounter the population sequentially, N known

1. Let n_k be the number of elements selected *so far*
2. Select the first element with probability $\frac{n}{N}$.
3. Select the k^{th} element with probability $\frac{n-n_k}{N-k+1}$

This is a simple random sample.

When you encounter the population sequentially, N not known

1. Select the first n elements to be in your "temporary" sample
2. Then when you encounter the k^{th} element out of $n+1,\ n+2,\ \dots$, do one of the following:
    - with probability $1-\frac{n}{k}$ leave the temporary sample as is.
    - with probability $\frac{n}{k}$ replace an element in the temporary sample (chosen at random) with the k^{th} element.

This is also a simple random sample (from the "population" of those elements you happened to encounter.)

recall properties of a simple random sample (without replacement)

- The sample $\{y_1, y_2, \ldots, y_n\}$ is a list of random variables.
- Each of the y_i have the *same* distribution.
  - Name: *discrete uniform distribution over* $\{y_1, y_2, \ldots, y_N\}$.
  - $E(y_i) = \sum_{j=1}^{N} y_j \frac{1}{N} = \mu$
  - $V(y_i) = \sum_{j=1}^{N} (y_j - \mu)^2 \frac{1}{N} = \sigma^2$
- But they are *not* independent random variables. In particular (for $i \neq j$):

$$Cov(y_i, y_j) = -\frac{1}{N-1}\sigma^2$$

the details—"identically distributed"

I stated "Each of the y_i have the *same* distribution" also said what the distribution actually was.

This is a "theorem"—the proof would be based on symmetry, from the fact that the sample $\{y_1, y_2, \ldots, y_n\}$ could have arrived in any order with equal probability

the details—"mean and variance"

The following "theorems" then actually follow immediately from the previous slide:

$E(y_i) = \sum_{j=1}^{N} y_j \frac{1}{N} = \mu$

$V(y_i) = \sum_{j=1}^{N} (y_j - \mu)^2 \frac{1}{N} = \sigma^2$

These results look suspiciously like these *definititions* from last week, but are completely different in character:

- **population mean**: $\mu = \frac{1}{N} \sum_{i=1}^{N} y_i$
- **population variance**: $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \mu)^2$

the details—"$Cov(y_i, y_j)$"

But y_i and y_j are obviously not independent.

For example, $P(y_1 = y) = \frac{1}{N}$. But $P(y_1 = y \mid y_2 = y') = \frac{1}{N-1}$ for $y \neq y'$.

The derivation of $Cov(y_i, y_j) = -\frac{1}{N-1}\sigma^2$ is tricky, dull, and contained the textbook appendix.

For those who are interested: The derivation starts with the always true fact that $Cov(y_i, y_j) = E\left(y_i y_j\right) - E(y_i)E\left(y_j\right)$. The next line uses the fact (in this case) that the joint distribution of y_i and y_j is also a discrete uniform distribution on the $N(N-1)$ pairs of elements that can be chosen from the population. Then it's a matter of calculation tricks.

how to estimate the population parameters

So we have our unknown τ and μ (and σ^2) and we have a sample $\{y_1, y_2, \ldots, y_n\}$ some of whose properties we have established.

How shall we use the sample to estimate the unknown paramaters? We'll go with obvious choices for μ and τ which are:

$$\hat{\mu} = \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad \text{and} \qquad \hat{\tau} = N\hat{\mu} = N\overline{y} = \frac{N}{n} \sum_{i=1}^{n} y_i$$

We'll get to σ^2.

Note that this estimator was not derived from any principle of estimation. It is handed to you from on high. A little more on this later.

properties of $\overline{y}$

$\overline{y}$ is a function of random variables, so it is also a random variable (so it has a distribution with a mean and a variance etc.)

The *distribution* $\overline{y}$ is also discrete uniform, but this isn't so helpful this time to find its mean and variance.

Instead we can use properties of $E(\cdot)$ and $V(\cdot)$ (reviewed last week) to obtain first:

$$E\left(\overline{y}\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n} n\mu = \mu$$

So we say $\overline{y}$ is *unbiased* for μ. (In fact: $\overline{y}$ is the *only* unbiased estimator under SRS.)

properties of $\overline{y}$

The variance of $\overline{y}$ is a little more involved due to the lack of independence:

$$
\begin{aligned}
V(\overline{y}) &= V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) \\
&= \frac{1}{n^2} V\left(\sum_{i=1}^{n} y_i\right) \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n} V(y_i) + \sum_{i \neq j} Cov(y_i, y_j)\right] \\
&= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)\frac{-\sigma^2}{N-1}\right]
\end{aligned}
$$

properties of $\overline{y}$

$$
\begin{aligned}
V(\overline{y}) &= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)\frac{-\sigma^2}{N-1}\right] \\
&= \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) \\
&= \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)
\end{aligned}
$$