PLSC 502: "Statistical Methods for Political Research"

Sampling Distributions
October 6, 2016

Introduction

The key thing for today is to remember the following:

Anything that is a function of random variables is, itself, a random variable.

In practical terms, this means that all the statistics we've discussed so far – means, variances, medians, etc. – all can be thought of themselves as random variables, with associated means, variances, medians, and (most important) distributions. The last of these will take up most of the class today.

Sampling Distributions, Described

The frequentist perspective on statistical inference starts with an idea: that our sample, the sample we have, is just one of many possible samples that we could have drawn from the population. A sampling distribution, then, can be thought of as the theoretical distribution of some sample statistic (a mean, a variance, etc.) in repeated sampling.

An Example

Imagine a hypothetical small town (Bitterville), where 1000 voters are registered to vote, and where 500 of them voted for the Democratic presidential candidate, and 500 for the Republican. The probability of drawing a Democratic voter from this population (using simple random sampling) is 0.5, naturally. A somewhat different question, however, is *what is the distribution of the "mean" Democratic vote percentage*?

Another way of thinking of it is this: If we sample (say) 10 voters from Bitterville, what can we know about the population as a whole? To answer this question, it is useful to know something about the properties of the statistics that we can calculate from our sample. For example, we can calculate a (sample) mean proportion of Democratic votes:

$$\bar{X} = \sum_{i=1}^{10} X_i \tag{1}$$

where X_1 is one if the voter sampled voted for the Democrat, and 0 if they voted for the Republican.

Now, notice that (1) can be rewritten as:

$$\bar{X} = \left(\frac{1}{10}\right) X_1 + \left(\frac{1}{10}\right) X_2 + ... + \left(\frac{1}{10}\right) X_{10} \tag{2}$$

or, more compactly, as

$$\bar{X} = aX_1 + aX_2 + ... + aX_{10} \tag{3}$$

where $a = 0.1$. In other words, $\bar{X}$ is a linear function of the (independent) random variables X_1, X_2, etc. Because we know from before that

$$\mathrm{E}(aX + b) = a\mathrm{E}X + b,$$

then the expectation of $\bar{X}$ is just

$$\begin{aligned}
\mathrm{E}(\bar{X}) &= \sum_{i=1}^{10} a\mathrm{E}(X_i) \\
&= \sum_{i=1}^{10} a\mu \\
&= \mu \sum_{i=1}^{10} a \\
&= \mu \sum_{i=1}^{10} \frac{1}{10} \\
&= \frac{10\mu}{10} \\
&= \mu
\end{aligned} \tag{4}$$

Similarly, the variance of $\bar{X}$ – that is, the variability in $\bar{X}$ due to variations in sampling – is

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \sum_{i=1}^{10} a^2 \mathrm{Var}(X_i) \\
&= \sum_{i=1}^{10} \left(\frac{1}{10}\right)^2 \sigma_i^2 \\
&= \left(\frac{1}{100}\right) \sum_{i=1}^{10} \sigma_i^2 \\
&= \left(\frac{1}{100}\right) 10\sigma^2 \\
&= \frac{\sigma^2}{10}
\end{aligned} \tag{5}$$

where the terms depending on the covariances of the X_is are zero because we are assuming that the observations on X are independent.

More generally, we can rewrite (4) and (5) as:

$$\mathrm{E}(\bar{X}) = \mu \tag{6}$$

and

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{N}, \tag{7}$$

respectively, where N denotes the sample size. These are the characteristics of the *sampling distributions* of the mean of a random variable X; they denote the expectation and variability – in repeated sampling – of the sample mean of a random variable.

All of this ought to make sense. Equation (6), for example, indicates that the expectation (mean) of many, many means calculated from random samples drawn from the population of X will equal the population mean itself.[1] Similarly, the variability in those $\bar{X}$s will be a function of (a) the variability in X (more variability in X will lead to samples that have higher numbers of "outliers," and so greater variability in the means) and (b) the sample size (in the limit, as $N \to \mathfrak{N}$, the sample becomes identical to the population, and the variability in $\bar{X}$ goes to zero).

As with all variances, the variance term in (7) is expressed in terms of "squared" units of X. Equation (7) also implies that

$$\sqrt{\mathrm{Var}(\bar{X})} \equiv \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}. \tag{8}$$

That is, the standard deviation of the distribution of sampling means of X is equal to the standard deviation of X divided by the square root of the sample size. (8) gives one a rough idea how much variation there will be in our estimate of the mean of X, as a function of our sample size. In particular, (8) has an important implication for selecting sample size:

One must quadruple the sample size to halve the sampling error.

For example, in our hypothetical town of Bitterville, we know that $\sigma^2 = 0.5(1 - 0.5) = 0.25$ (because X here is binary). For a sample size of 10, this means that $\sigma_{\bar{X}} = \frac{0.5}{3.162} = 0.158$. That is, with $N = 10$, the proportion of Democratic voters we would estimate based on our sample would typically be off by close to 0.16. (That's a lot.) If we wanted to reduce this to, say, 0.05 (that is, we wanted $\sigma_{\bar{X}} = 0.05$), we'd need a sample size of:

[1]Note as well that this fact *does not* depend on the sample size – *any* simple random sample from the population will have a mean that is, in expectation, μ.

$$
\begin{aligned}
0.05 &= \frac{0.50}{\sqrt{N}} \\
0.05\sqrt{N} &= 0.50 \\
\sqrt{N} &= 10 \\
N &= 100.
\end{aligned}
$$

More important, if we then wanted to cut this amount in half, to $\sigma_{\bar{X}} = 0.025$, we'd need

$$
\begin{aligned}
0.025 &= \frac{0.50}{\sqrt{N}} \\
0.025\sqrt{N} &= 0.50 \\
\sqrt{N} &= 20 \\
N &= 400.
\end{aligned}
$$

Sampling *Distributions*

It's well and good to know the mean and variance of a sampling statistic (like the mean). But, to really do inference, we need to know how such statistics are *distributed* as well as their means and variances.

The Sampling Distribution of the Mean

We made a reference a week or so ago to the *central limit theorem*; that theorem states that the sampling distribution of the mean of *any* independent random variables is asymptotically Normal. The proof requires a bit more calculus than I care to go into right now, but we can show that this is true for the special case of a normally-distributed variable X easily. Because, for $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$,

$$\sum_{i=1}^{N} X_i \sim \mathcal{N}\left(\sum_{N} \mu_i, \sum_{N} \sigma_i^2\right) \tag{9}$$

i.e., the sum of independent, random normal variates are themselves normal, then it is straightforward to show that the sampling distribution of the mean of those variables ($\bar{X}$) is

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} X_i &\sim \mathcal{N}\left[\frac{1}{N}\sum_{N} \mu_i, \left(\frac{1}{N^2}\right)\sum_{N} \sigma_i^2\right] \\
&\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right) \qquad (10)
\end{aligned}
$$

The central limit theorem allows us to extend this intuition to the mean of any i.i.d. random variables. It's hard to convey how remarkable this is – it means that, irrespective of the shape of the distribution of X itself, the sampling distribution of X's mean has a simple asymptotic form completely characterized by two easy-to-compute parameters. That will wind up being incredibly useful once we begin discussing inference.

The Sampling Distribution of the Variance

In addition to the mean, we might imagine using the sample variance as a "guess" at the population variance. We defined the sample variance as:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 \tag{11}$$

It's straightforward to demonstrate that

$$\begin{aligned}
\mathrm{E}(s^2) &= \frac{1}{N-1}\left\{\mathrm{E}\left[\sum_{i=1}^{N}(X_i-\bar{X})^2\right]\right\} \\
&= \frac{1}{N-1}\left\{\mathrm{E}\left[\sum_{i=1}^{N}(X_i-\mu)^2 - N(\bar{X}-\mu)^2\right]\right\} \\
&= \frac{1}{N-1}\left[\sum_{i=1}^{N}\mathrm{E}(X_i-\mu)^2 - N\mathrm{E}(\bar{X}-\mu)^2\right] \\
&= \frac{1}{N-1}\left(N\sigma^2 - N\frac{\sigma^2}{N}\right) \\
&= \sigma^2.
\end{aligned} \tag{12}$$

That is, the sampling variance of X has expectation equal to the population variance of X (σ^2). Moreover, as with the mean $\bar{X}$, s^2 also has a sampling distribution. It's useful to rewrite (11) as

$$\frac{(N-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(X_i - \bar{X})^2 \tag{13}$$

by multiplying both sides of (11) by $N-1$ (to get the "sum of squares") and dividing both sides by σ^2 (to "normalize" the resulting statistic to the scale of X). One can also show that the quantity in (13) is distributed as chi-square with $N-1$ degrees of freedom; it's easiest to see this in the case of $N = 2$ (a very small sample size, to be sure) normal variates X_1 and X_2. When $N = 2$, then $\bar{X} = (1/2)(X_1 + X_2)$, and we can write (11) as:

$$s^2 = \frac{(X_1 - X_2)^2}{2}.$$

From this, (13) becomes

$$\begin{aligned} \frac{(N-1)s^2}{\sigma^2} &= \frac{(X_1 - X_2)^2}{2\sigma^2} \\ &= \left(\frac{X_1 - X_2}{\sqrt{2\sigma^2}}\right)^2. \end{aligned} \tag{14}$$

Equation (14) is, in turn, nothing more than the square of a standard normal variate (because $\mathrm{E}(X_1 - X_2) = 0$ and $\mathrm{Var}(X_1 - X_2) = 2\sigma^2$), which (as we all know) is a chi-square variate with one (that is, $N-1$) degrees of freedom. More generally, it is the case that the sampling distribution of the (rescaled) variance is distributed as chi-square with $N-1$ degrees of freedom. This, in turn, means that we often use chi-square distributions for hypothesis testing involving variances; more on that below.

The Point(s)

The main point is this:

Always remember that the statistics you calculate are, themselves, random variables, with their own sampling distributions and characteristics.

Illustrations/Examples

The slides contain some examples of what I'm talking about, using some data on the U.S. Supreme Court. Specifically, those data are all fully-decided cases handed down during the Warren and Burger Courts (OT 1953-85); there are 7,161 of them, and one could think of them as the "population" of Supreme Court cases during those years ($\mathfrak{N} = 7161$). The variables in the data are:

- `us` is a string variable for the U.S citation of the case,
- `id` is just an ID variable,
- `amrev` are the number of *amicus curiae* ("friend of the court") briefs filed supporting reversal, and
- `amaff` is the number of such briefs supporting affirmance.
- `sumam` is just `amrev` + `amaff`,
- `fedpet` is 1 if the federal government was the petitioner, 0 otherwise,
- `constit` is 1 if the decision was made on constitutional grounds, 0 otherwise, and
- `sgam` is 1 if the Solicitor General (the U.S.'s attorney) filed am *amicus curiae* brief in the case, and 0 otherwise.

We'll focus mainly on the `fedpet` and `sumam` variables...

- Note that the distributions of these variables in the population are bimodal and skewed, respectively.
- For a small sample ($N = 10$), the mean of `fedpet` is based on a relatively small fraction of the population, and a small number of observations in general. That is, the mean is necessarily in $\{0, 0.1, 0.2, ..., 0.9, 1\}$ (since the variable is binary). Empirically, with 1000 such samples, we see that (a) we never get a value of $\bar{X}$ greater than 0.7, (b) the distribution, while not bimodal, is also not exactly $\mathcal{N}$ormal, and (c) there's a good bit of variation in the sample mean. On this last point, recall that the standard deviation of the sample mean for this variable is $\frac{\sigma}{\sqrt{N}} = \frac{0.0206}{3.16} = 0.0065$.
- Increasing the sample size to $N = 20$ (a) narrows the range of possible values of $\bar{X}$, and (b) makes the distribution more "normal"-looking.
- Increasing it again to $N = 100$ gives us a *very* small range of values for $\bar{X}$ and makes the distribution almost perfectly normal.
- Note that the same is true for $N = 100$ samples using the `sumam` variable, which was a count rather than binary.
- We can do the same thing for variances, showing that the rescaled variance (as discussed above) (a) approximates a chi-square, with $N-1$ d.f., and (b) becomes more chi-square-like as N gets larger.

Finally, the last slide illustrates how to conduct stratified sampling from a known population.