Pooled sample estimate

Population and sample

A population is an entire set of events and its attributes. A sample is a small set of events taken from a population.

When the properties of a population are unknown we estimate them from the values drawn in the sample. Statistics derived from the sample then become "estimates" of the population values.

We need separate symbols to denote the population parameters, the sample statistics and the estimate derived from the sample.

Context	How Found	Mean	Variance
Sample	Statistics derived from a sample	$\overline{x}$	s^2
Population	Objective parameters of the population	μ	σ^2
Population	Estimates of the population parameters derived from sample statistics	$\hat{\mu}$	$\hat{\sigma}^2$

Pooling samples

Suppose we take two random samples from a population with unknown mean, μ, and unknown variance σ^2, then estimates for the joint mean and joint variance of the two samples are

$$\hat{\mu} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

where $\hat{\mu}$ is the unbiased estimator for the population mean $\hat{\mu}$; and

$$\hat{\sigma}^2 = \frac{n_1\overline{S}_1^{\,2} + n_2\overline{S}_2^{\,2}}{n_1 + n_2 - 2}$$

where, $\hat{\sigma}^2$ is an unbiased estimator for the population variance σ^2 and $S_1^{\,2}$ and $S_2^{\,2}$ are the sample variances.

Note, here, $S_1^{\,2}$ and $S_2^{\,2}$ are the sample variances, that is the *biased* estimators of the population variance; however

$$\hat{\sigma}^2 = \frac{n_1\overline{S}_1^{\,2} + n_2\overline{S}_2^{\,2}}{n_1 + n_2 - 2}$$

is an *unbiased* estimator of the population variance because of the subtraction of 2 from the divisor of the fraction (from the bottom of the fraction). So be begin by calculating the sample variances in the normal way, and then substitute into the formula for the pooled sample estimate, as will be illustrated below.

<u>Example</u>

Samples of size 30 and 40 are taken from a population with unknown mean, μ, and unknown variance σ^2. Using the data from the two samples, find the unbiased estimates of the mean and variance of the population.

Sample 1

x_1	14	15	16	17	18	19
f	2	4	9	3	7	5

Sample 2

x_1	14	15	16	17	18	19
f	4	9	11	6	7	3

Solution

Firstly, we calculate the individual sample statistics.

$$\overline{x}_1 = \frac{\sum fx}{\sum f} = \frac{504}{30} = 16.8$$

$$S_1^{\,2} = \frac{\sum fx^2}{\sum f} - \overline{x}_1^{\,2} = \frac{8536}{30} - (16.8)^2 = 2.293$$

$$\overline{x}_2 = \frac{\sum fx}{\sum f} = \frac{652}{40} = 16.3$$

$$S_2^{\,2} = \frac{\sum fx^2}{\sum f} - \overline{x}_2^{\,2} = \frac{10710}{40} - (16.3)^2 = 2.06$$

Then,

$$\hat{\mu} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2} = \frac{(30 \times 16.8) + (40 \times 16.3)}{30 + 40} = 16.51(2.D.P.)$$

$$\hat{\sigma}^2 = \frac{n_1\overline{S}_1^{\,2} + n_2\overline{S}_2^{\,2}}{n_1 + n_2 - 2} = \frac{(30\times 2.293)+(40\times 2.06)}{30+40-2} = 2.223$$

The pooled unbiased estimator of the population mean is 16.51 (2 *D.P.*) and the pooled unbiased estimator of the population variance is 2.22 (2 *D.P.*)

Pooled estimator of the population proportion

A proportion is a ratio. Given a population divided into "successes" and "failures" the proportion of successes is the ratio of the number of successes to the total number of items in the population. The proportion may be sampled. The sample proportion is the ratio of the number of successes in the sample to the total number of trials. Given a single sample, we take the sample proportion, $\hat{\rho}$, as an estimate for the population proportion, ρ.

For example, suppose that in a survey of 1200 housewives from a town in Kent 497 said they were in favour of a new by-pass. Then the sample proportion would be

$$p = \frac{497}{1200} = 0.414 \quad \text{(3 S.F.)}$$

which would also be an estimate of the population proportion.

When two sample proportions have been obtained, it makes sense to estimate the population proportion using both of them. In this case

$$\hat{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1 + n_2}$$

gives an unbiased estimator of the population proportion, where $\overline{p}_1$ and $\overline{p}_2$ are the sample proportions.

Example

Two traffic surveys were conducted on the same section of a road. Cars either travel right through the road (through traffic) or start or end somewhere along the road (residential traffic). Data for the two surveys were as follows

Sample 1

n_1	through traffic	residential traffic
80	22	58

Sample 1

n_1	through traffic	residential traffic
65	19	46

Find the unbiased pooled sample estimate for the population proportion of traffic on the road that is through traffic.

Solution

We first calculate the proportion of traffic in both samples that was through traffic.

$$p_1 = \frac{\text{through traffic}}{\text{total traffic}} = \frac{22}{80} = 0.275$$

$$p_2 = \frac{\text{through traffic}}{\text{total traffic}} = \frac{19}{65} = 0.292...$$

Then the unbiased, pooled estimate of the population proportion of through traffic is

$$\hat{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1 + n_2} = \frac{(80 \times 0.275) + (65 \times 0.292..)}{80 + 65} = \frac{41}{145} = 0.283$$

Of course, if you have the original data, you may as well calculate this pooled estimate by simply adding the total number of through cars and dividing by the total number of cars.

$$\hat{p} = \frac{\text{total through traffic}}{\text{total traffic}} = \frac{22 + 19}{80 + 65} = \frac{41}{145}$$