Pooled Variance

Recall If $X_1, X_2, \cdots X_n$ R.S.

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

Sample variance

$$= \frac{\text{Sum squares (SS)}}{\text{degree of freedom (df)}}$$

$$S^2 = \frac{SS}{n-1}$$

$$\boxed{SS = (n-1)\, S^2}$$

Pooled Variance (Combined variance
Overall variance)

is a method for estimating variance of several different populations when the mean of each population may be different but the variance of each pop. is the same

let $X_1, X_2, \dots X_{n_1}$ R.S. from Population 1

$X'_1, X'_2, \dots X'_{n_2}$ R.S. from " 2

assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ but σ^2 is unknown

$$S_1^2 = \frac{SS_1}{n_1 - 1} \quad , \quad S_2^2 = \frac{SS_2}{n_2 - 1}$$

If we Combine two samples

$$S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1 + n_2 - 2}$$

unbiased estimator for σ^2

why we cannot take the average

e.g. $$Sp^2 = \frac{S_1^2 + S_2^2}{2}$$

we have to take the sample size into account

Ex

Sample 1

$n_1 = 5$

$SS_1 = 100$

Sample 2

$n_2 = 5$

$SS_2 = 48$

$$S_1^2 = \frac{SS_1}{df_1} = \frac{100}{4} = 25 \qquad S_2^2 = \frac{SS_2}{df_2} = \frac{48}{4} = 12$$

$$Sp^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{100 + 48}{4 + 4} = 18.5$$

Sp^2 value between the two sample variances

here Sp^2 is value average between S_1^2, S_2^2 why ??

3

because they have the same sample size

<u>Ex2</u> $n_1 = 11$ $n_2 = 5$

$SS_1 = 100$ $SS_2 = 48$

$S_1^2 = \frac{100}{10} = 10$, $S_2^2 = \frac{48}{4} = 12$

$$Sp^2 = \frac{100 + 48}{10 + 4} = 10.57$$

Closer to 10 than 12 why?

10 has the bigger sample size

<u>So</u> the Pooled variance will be closer to the variance with the larger sample size