Comparison of "2-sample-t" and "pooled-variance 2-sample-t"

Degrees of Freedom

"2-sample-t" uses "inverse weighted average"

$$\frac{1}{\nu} = \frac{s_{\bar{x}}^2}{s_{\bar{x}}^2 + s_{\bar{y}}^2} \frac{1}{n-1} + \frac{s_{\bar{y}}^2}{s_{\bar{x}}^2 + s_{\bar{y}}^2} \frac{1}{m-1}$$

$$\nu = \frac{s_{\bar{x}}^2 + s_{\bar{y}}^2}{s_{\bar{x}}^2 \frac{1}{n-1} + s_{\bar{y}}^2 \frac{1}{m-1}}$$

summed variance

"pooled-variance 2-sample-t" uses sum

$$\nu = (n-1) + (m-1)$$
$$= n + m - 2$$
$$= \#\binom{\text{total}}{\text{samples}} - 2$$

Note: This formula appears again in ANOVA.

ν is between $(n-1)$ and $(m-1)$

→ very close to smaller of $(n-1)$ & $(m-1)$

→ increasing $s_{\bar{x}}^2$ or $s_{\bar{y}}^2$ pushes slightly in that direction

Degrees of freedom is more than twice the "2-sample-t" degrees of freedom!

Note: $s_{\bar{x}}^2 = s_x^2/n$ (sample variance of $\bar{X}$ and $\bar{Y}$)

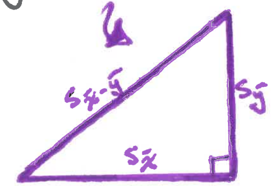
$s_{\bar{y}}^2 = s_y^2/m$

Standard Error

"2-sample-t" variance adds:

$$s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$$

So standard error is "Pythagorean Sum"



$$s_{\bar{x}-\bar{y}} = \sqrt{s^2_{\bar{x}} + s^2_{\bar{y}}}$$

$$= \sqrt{s^2_x/n + s^2_y/m}$$

standard errors of statistic

$\boxed{\text{Use } s^2_{pool} \text{ for } s^2_x \text{ \& } s^2_y}$

"pooled-variance 2-sample-t" averages:

$$s^2_{pool} = \frac{n-1}{n+m-2}\, s^2_x + \frac{m-1}{n+m-2}\, s^2_y$$

So $s^2_{\overline{pool}} = s^2_{pool} \left(\frac{1}{n} + \frac{1}{m} \right)$

"pooled variance of mean"

$$= \frac{(n-1)s^2_x + (m-1)s^2_y}{n+m-2} \cdot \left(\frac{1}{n} + \frac{1}{m} \right)$$

$$s_{\overline{pool}} = \sqrt{\frac{(n-1)s^2_x + (m-1)s^2_y}{n+m-2} \cdot \left(\frac{1}{n} + \frac{1}{m} \right)}$$

Pythagorean Sum: $\sqrt{s^2_{pool}/n + s^2_{pool}/m}$

→ If $s_x \approx s_y$ then

$$s_{\bar{x}-\bar{y}} \approx \sqrt{s^2_x \left(\frac{1}{n} + \frac{1}{m} \right)}$$

$$\approx s_x \sqrt{1/n + 1/m}$$

approximately same standard error if $s^2_x \approx s^2_y$!!

→ If $s_x \approx s_y$ then

$$s_{\overline{pool}} \approx \sqrt{\frac{(n-1)s^2_x + (m-1)s^2_x}{n+m-1} \cdot \left(\frac{1}{n} + \frac{1}{m} \right)}$$

$$\approx \sqrt{s^2_x \left(\frac{1}{n} + \frac{1}{m} \right)}$$

$$\approx s_x \sqrt{1/n + 1/m}$$

Summary:

If $s_x^2 \approx s_y^2$ (when pooled variance t-test is reasonable)

then pooled variance t-test will have

- approximately same standard error.
- more than twice the degrees of freedom.

This gives lower p-values because big degrees of freedom

$\Downarrow$

lower tails of t-distribution

$\Downarrow$

lower p-value for same score