Degrees of Freedom

The degrees of freedom is the number of values in a calculation that are free to vary.

Example: $s = ?$ $\qquad x_1, x_2$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

know $\bar{x} = 5$ and $x_1 = 4$

x_2 is not free to vary

$$\bar{x} = \frac{x_1 + x_2}{2} \qquad 5 = \frac{4 + x_2}{2} \Rightarrow 10 = 4 + x_2 \Rightarrow$$

$\Rightarrow x_2 = 6$ we only have one degree of freedom

Ex: $s = ?$ $\qquad x_1, x_2, x_3$

$\bar{x} = 8$ $\qquad 1, 8, ?$

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3} \qquad 8 = \frac{1 + 8 + x_3}{3} \Rightarrow$$

$$24 = 9 + x_3 \Rightarrow x_3 = 24 - 9 = 15$$

$$df = 2$$

In general, the number of degrees of freedom in a calculation is equal to the total number of observations minus the number of estimates needed in the intermediate steps

In this case $n = 3$ $\qquad df = n - 1 = 2$

www.professorserna.com 786-290-9100

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

Example CI for $\mu_1 - \mu_2$

n_1 n_2 t-distribution

"Pooled Variance" $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

In the calculation of S_1^2 we have n_1-1 df
" " " S_2^2 " " n_2-1 df

so " " S_p^2 " $(n_1-1)+(n_2-1)$ df
that is n_1+n_2-2

Example: In One-Way ANOVA

T_1	T_2	. . .	T_K	n
$n_1 \left\{ \begin{array}{c} x \\ \vdots \\ x \end{array} \right.$	$n_2 \left\{ \begin{array}{c} x \\ \vdots \\ x \end{array} \right.$		$n_K \left\{ \begin{array}{c} x \\ \vdots \\ x \end{array} \right.$	$\bar{\bar{x}}$
$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_K$	

SS total compares each of the total n observations to the overall mean $\bar{\bar{x}}$ (which must be calculated)

$SS_{tot} = (x_1 - \bar{\bar{x}})^2 + (x_2 - \bar{\bar{x}})^2 + \cdots + (x_n - \bar{\bar{x}})^2$

so we have $n-1$ df

SSTreat compares each of the k treatment means to the overall mean (which must be calculated)

$$SSTreat = (\bar{x}_1 - \bar{\bar{x}})^2 + (\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + (\bar{x}_k - \bar{\bar{x}})^2$$

so, we have $k-1$ df

SSE

$$\begin{array}{cccccccc}
 & T_1 & & T_2 & \cdots & & T_k & \\
 & x & & x & & & x & \\
n_1 \Big\{ & \vdots & n_2 \Big\{ & \vdots & & n_k \Big\{ & \vdots & n \\
 & x & & x & & & x & \\
 & \bar{x}_1 & & \bar{x}_2 & & & \bar{x}_k &
\end{array}$$

$$df = (n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)$$

$$= \underbrace{n_1 + n_2 + \cdots + n_k}_{} - k = n - k$$

So, to calculate the number of degrees of freedom in any new situation try to figure out the number of observations involved <u>minus</u> the number of estimates needed in intermediate steps.