**Purpose:** This function is used to estimate the mean CPUE and the variance around the CPUE. A jackknife estimator is calculated for each year’s CPUE data using the method of Smith (1980 – CJFAS). The ‘se’ and ‘sd’ options have been the source of some confusion over the years. First note that these are both **variances** NOT “standardized”. See details below.

**Version Control:** Surely multiple versions around but should be very similar.

**Librarys**:

**Function Arguments Summary**

1. **data**: CPUE to calculate, CPUE by month/bank/nafo, Default is "Month". Options are
   * “month": Calculate CPUE by month, area or nafo region
   * "obs": Caluculate CPUE for Observer trips only
   * "both": Calculate both the Observer trips and CPUE by month.
2. **err**: The options here are “se” or “sd”. The error we are calculating, this is in reality the variance and not the “standard” anything. The 'sd' is simply the **variance** of the Jacknife estimator, while 'se' which is the unbiased **variance** estimator which is outlined in Smith 1980. Default is 'se' Remember , so these are equivalent just different ways of presenting results. The sd method is surprisingly about 20% slower (if a big dataset expect this to take approximately 10 minutes, smaller can converge very quickly)

**Section 1**

This function is essentially a for loop which repeatedly recalculates the mean of our data with a new value excluded each time we move through the loop. There has been some confusion about the *err = ‘se’* vs *err= ‘sd’* options. In brief it is:

1. Both options actually calculate a variance not the standardized measure.
2. The ‘sd’ option calculates the raw variance of the jackknife estimator.
   1. We would rarely (if ever) really want use this or it’s standard deviation in a plot, the standard error is what we would plot
   2. To convert to the standard deviation
   3. To convert to the standard error
3. The ‘se’ option calculates the unbiased variance estimator from Smith 1980. In maths this is

but remember the raw variance of a data set is just the sum of squares divided by n-1

notice the only difference between (1) and (2) is the n in the denominator, so the Smith 1980 variance estimator in simple terms is just the variance (the exact same thing as is calculated in point 2 above)

if we take the square root of *var(SE)* we get….

So in short we can easily move between (1) and (3) the sd and se, it is rather important that you know which you are calculating and what you are using it for. Using the “se” option leads more naturally to finding the actual ‘se’ (just take square root) and this is generally what you would use in a graphic to show how precise we think the estimate of the mean is.

**Function Index**

data.frame

for

function

if

mean

na.omit

print

rep

return

sum

tapply

with

**Reference:**

Smith, S.J. 1980 Comparison of Two Methods of Estimating the Variance of the Estimate of Catch Per Unit Effort. CJFAS 37: 2346:2351.