Tsvstat

wozgonon

1 Introduction

The *tsvtstat* tool can be used to generate a table of summary statistics given a table or stream of tab separated numeric observations.

2 Statistics

The formulas used to generate statistics are compatible with those used by common spreadsheets.

2.1 Mean

$$\begin{aligned} \bar{x} &= \sum \frac{x_i}{n} \\ \implies \sum x_i &= n\bar{x} \qquad (1) \end{aligned}$$

2.2 Sample Variance and Standard Deviation

$$\begin{aligned}
s^2 &= \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2 && \text{Sample Variance} \\
&= \frac{1}{n-1}\sum_1^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\
&= \frac{1}{n-1}(\sum_1^n x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2) \\
&= \frac{1}{n-1}(\sum_1^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2) && \text{substitute (1)} \\
&= \frac{1}{n-1}\sum_1^n x_i^2 - n\bar{x}^2) \\
s &= \sqrt{s^2} && \text{Sample standard deviation}
\end{aligned}$$

2.3 Skewness

$$
\begin{aligned}
skew &= \frac{n}{(n-1)(n-2)} \sum \frac{(x, -\bar{x})^3}{s^3} \\
&= \frac{n}{(n-1)(n-2)} \frac{\sum x_i^3 - 3\bar{x} \sum x_i^2 + 2n\bar{x}^3}{s^3} && \text{subst. (3)}
\end{aligned}
$$

$$
\begin{aligned}
\sum_1^n (x_i - \bar{x})^3 &= \sum_1^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2)(x_i - \bar{x}) \\
&= \sum_1^n (x_i^3 - 2x_i^2\bar{x} + x_i\bar{x}^2 - \bar{x}x_i^2 + 2x_i\bar{x}^2 - \bar{x}^3) \\
&= \sum_1^n (x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3) && (2) \\
&= \sum_1^n x_i^3 - 3\bar{x} \sum x_i^2 + 3\bar{x}^2 \sum x_i - n\bar{x}^3 \\
&= \sum_1^n x_i^3 - 3\bar{x} \sum x_i^2 + 3n\bar{x}^3 - n\bar{x}^3 && \text{by (1)} \\
&= \sum_1^n x_i^3 - 3\bar{x} \sum x_i^2 + 2n\bar{x}^3 && (3)
\end{aligned}
$$

2.4 Kurtosis

$$
\begin{aligned}
kurt &= \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_1^n \frac{(x, -\bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \\
&= \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_1^n x_i^4 - 4\bar{x}\sum_1^n x_i^3 + 6\bar{x}^2 \sum_1^n x_i^2 - 3n\bar{x}^4}{s^4} && \text{substitute (4)} \\
&- \frac{3(n-1)^2}{(n-2)(n-3)}
\end{aligned}
$$

$$
\begin{aligned}
\text{excess kurtosis} &= kurt - 3 \\
\sum_1^n (x_i - \bar{x})^4 &= \sum_1^n (x_i - \bar{x})^3 (x_i - \bar{x}) \\
&= \sum_1^n (x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3)(x_i - \bar{x}) && \text{substitute (2)} \\
&= \sum_1^n x_i(x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3) - \bar{x}(x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3) \\
&= \sum_1^n (x_i^4 - 3x_i^3\bar{x} + 3x_i^2\bar{x}^2 - x_i\bar{x}^3) - (\bar{x}x_i^3 - 3x_i^2\bar{x}^2 + 3x_i\bar{x}^3 - \bar{x}^4) \\
&= \sum_1^n (x_i^4 - 4x_i^3\bar{x} + 6x_i^2\bar{x}^2 - 4x_i\bar{x}^3 + \bar{x}^4) \\
&= \sum_1^n x_i^4 - 4\bar{x}\sum_1^n x_i^3 + 6\bar{x}^2\sum_1^n x_i^2 - 4\bar{x}^3\sum_1^n x_i + n\bar{x}^4) \\
&= \sum_1^n x_i^4 - 4\bar{x}\sum_1^n x_i^3 + 6\bar{x}^2\sum_1^n x_i^2 - 4n\bar{x}^4 + n\bar{x}^4 \\
&= \sum_1^n x_i^4 - 4\bar{x}\sum_1^n x_i^3 + 6\bar{x}^2\sum_1^n x_i^2 - 3n\bar{x}^4 && (4)
\end{aligned}
$$