**AW2a Inferring the population skewness value from the sample**

The above rules of thumb are an acceptable interpretation when we have data for the whole population. But when we have a sample, the sample skewness doesn’t necessarily apply to the whole population. In that case the question is: can you conclude anything about the population skewness from the sample skewness? We advise that it is not possible to fully appreciate this section without reading Chapter 6. Accordingly, we recommend that you return to it after you have completed Chapter 6.

**Fisher - Pearson skewness coefficient** as defined by equation (1), where s=standard deviation.

$Sample skewness= \frac{n}{\left(n-1\right)\left(n-2\right)} \frac{\sum\_{i=1}^{n}\left(x\_{i}-\overbar{x}\right)^{3}}{s^{3}}$ (1)

Your data set is just one sample drawn from a population. Maybe, from ordinary sample variability, your sample is skewed even though the population is symmetric. But if the sample is skewed too much for random chance to be the explanation, then you can conclude that there is skewness in the population. But what do we mean by “too much for random chance to be the explanation”? To answer that, we need to divide the sample skewness by the standard error of skewness to get the test statistic, which measures how many standard errors separate the sample skewness from zero as illustrated by equation (2).

$Z= \frac{Sample skewness}{Standard error of skewness}$ (2)

Where

$Standard error= \sqrt{\frac{6 n \left(n-1\right)}{\left(n-2\right)\left(n+1\right)\left(n+3\right)}}$ (3)

If the sample size, n, is large then equation (3) can be approximated by $\sqrt{^{6}/\_{n}}$ .

Don’t worry about this for now – it will become clearer when you have worked through Chapters 5, 6, 7 and 8. What we are undertaking here is to conduct a hypothesis test where the stated null and alternative hypotheses are:

**Hypothesis test**

Null hypothesis: Population skewness = 0 (distribution symmetric)

Alternative hypothesis: Population skewness ≠ 0 (distribution asymmetric)

The alternative hypothesis suggests a two-tail test and if we test at 95% then the value of the normal z statistic equals ± 1.96 which is approximately ± 2.

**Interpretation:**

* If Z < - 2, the population is very likely skewed negatively.
* If Z lies between – 2 and + 2, you can’t reach any conclusion about the skewness of the population. It might be symmetric or skewed in either direction.
* If Z > + 2, the population is very likely skewed positively.

Don’t mix up the meanings of this test statistic and the amount of skewness. The amount of skewness tells you how highly skewed your sample is: the bigger the number, the bigger the skew. The test statistic tells you whether the whole population is probably skewed, but not by how much: the bigger the number, the higher the probability.

**Example**

Consider the student results obtained in a quantitative methods examination as presented in Table W3.1).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 73 | 73 | 73 | 73 | 73 | 73 | 73 | 73 | 73 | 73 |
| 78 | 78 | 78 | 78 | 78 | 78 | 78 | 78 | 78 | 78 |
| 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 |
| 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 | 75 |

Table 1

Figures 1 illustrates the Excel solution for the first 10 data values with Figure 2 the calculation for the summary statistics (skewness).



Figure 1



Figure 2

Given Z = 1.145 lies between - 2 and + 2, then you cannot reach any conclusion about the skewness of the population based upon the sample data; it might be symmetric, or it might be skewed in either direction.

Don’t mix up the meanings of this test statistic and the amount of skewness. The amount of skewness tells you how highly skewed your sample is: the bigger the number, the bigger the skew. The test statistic tells you whether the whole population is probably skewed, but not by how much: the bigger the number, the higher the probability.