

Transformations of Non-Normal Data

- ▶ Sometimes non-normally distributed data can be transformed to normality.
- ▶ **Important:** The transformations used should not change the relative ordering of the values but **alter the distance between successively ordered values** to change the overall shape of the distribution.
- ▶ For example, if a dataset is transformed by squaring each of the values the larger values will be pulled further apart than the smaller values.

Choosing Your Transformation Method

- ▶ There are a variety of transformations that can be used to correct for skewing to a greater or lesser extent. The correct transformation to use will depend on both the direction and extent of skew.
- ▶ **Important** - It is possible to over-correct by using too powerful a transformation and change the direction of the skew.
- ▶ For example, a small amount of downward skewing might be over-balanced by squaring the measurements and result in an upward skew distribution.

Skewness

Negative or downward or left skew

Positive or upward or right skew



Tukey's Ladder

Tukey's ladder of transformations (shown below) gives several common transformations to correct skew in each direction and illustrates the relative effectiveness of these.

TUKEYS LADDER OF TRANSFORMATIONS

$\frac{-1}{x^2}$ $\frac{-1}{x}$ $\log(x)$ \sqrt{x}
-----> correct upward skew

x^2 x^3 $\text{antilog}(x)$
correct downward skew ----->

Tukey's Ladder

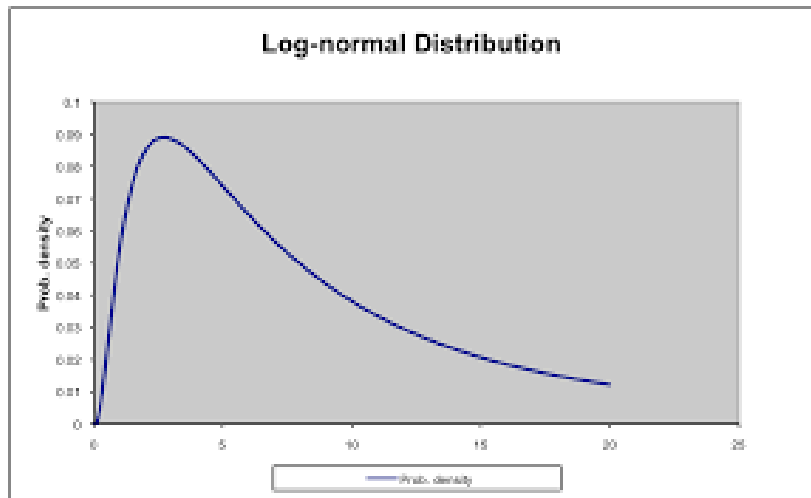
- ▶ For example, the ladder shows that squaring corrects downward skew and that cubing the data gives an even stronger correction.
- ▶ If we cube rather than square the values then the right hand (higher) values are pulled apart even more, creating a more extreme upper tail.

(Remark from last slide - Antilogging essentially is getting the exponential of a number)

Tukey's Ladder

- ▶ Upwardly skew data is not uncommon in medical applications and many measurements which display upward skewing are what is known as **lognormally distributed**.
- ▶ When data is lognormally distributed, taking logarithms (or logs) of the data values will normalise the data.

The Log-Normal distribution

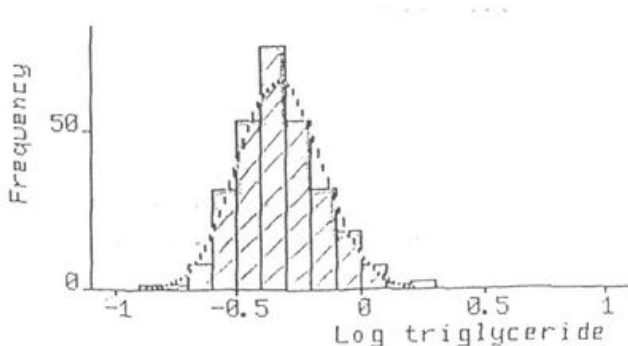


Example: Serum triglyceride values in cord-blood are lognormal :-
Before Log Transformation

For example: Serum triglyceride values in cord-blood are lognormal :-



Example: Serum triglyceride values in cord-blood are lognormal :-
After Log Transformation



Choosing Your Transformation Method

- ▶ **Important:** Choosing a suitable transformation can be a matter of trial and error.
- ▶ **Important:** Transformation may not necessarily work.
- ▶ Log-Transformation corrects upward skewness, if data is downwardly skew then logging will make the skewness worse.
- ▶ Downward skew may be corrected (to varying extents) by squaring, cubing or computing the exponential (anti-logging).

Choosing Your Transformation Method

- ▶ Often it is possible to use a transformation that has some biological basis.
- ▶ For example, taking square roots of areas or cube roots of volumes may be effective.
- ▶ Taking logarithms may not seem intuitive but this transformation is particularly useful when there are different groups to be transformed and compared.

Review

- ▶ Describe the purpose of transformations
- ▶ Describe the process of transformations
- ▶ Describe the purpose of Tukey's Ladder (referencing direction and relative strength)
- ▶ Give an example of a transformation for various types of skewed data (use Tukey's Ladder, with an example for both directions)
- ▶ Describe the limitations of transformations

Remark: Box-Cox transformations are omitted from the course, but I would still like you to be familiar with the name, and that is related to this topic.