1 Fixing problems with assumptions

For problems with either normality or constant variance, transformations are often used to attempt to satisfy model assumptions. For constant variance there are also other options to account for unequal variance in the data analysis. For problems with independence such as time or spatial dependence refer to texts such as Diggle (1991) or Cressie (1990), respectively.

1.1 Transformations

Some transformations are suggested by statistical theory, such as the arcsin-square root transformation of proportions shown in Table 6.3. In many instances, however, transformations are suggested empirically by the data. A low-tech way to select a power transformation y^λ of the data is to apply a few transformations such as the square-root or log ($\lambda < 1$) for positively skewed data or the square or cube ($\lambda > 1$) for negatively skewed data, and simply select the one that yields the best diagnostic plots (normal plot and residual-by-predicted plot). A better but more computationally demanding way is to select an optimal Box-Cox transformation. When an optimal transformation power λ is estimated, we generally use in practice a value close to it that is well-known, such as using the square-root transformation when $\widehat{\lambda} = .4$. We can obtain confidence intervals for λ to assist us in deciding both whether a transformation is needed and also which convenient power can be used, such as .5 in the example where $\widehat{\lambda} = .4$. The power family transformations are only applicable for positive-valued data. For data sets with negative values we can add a constant to make all values positive, but then the chosen transformation may depend on the value of the added constant.

1.2 Other methods to address non-constant variance

Weighted least-squares methods can be used when observations have differing (but approximately known) variances. When group variances differ, then for simple situations Welch's t or the Brown-Forsythe modified F test may help.

2 Effects of Incorrect Assumptions

Note in Tables 6.5, 6.6, and 6.7 illustrations of how incorrect assumptions affect the error rates of ANOVA and paired comparisons. In particular when group samples sizes are unequal and variances differ, Table 6.6 shows how the Type I error rate is as high as .28 in some instances. These findings reinforce the desire to insure equal sample sizes in experimental groups.

3 References

Cressie, N. 1991. *Statistics for Spatial Data.* New York: John Wiley & Sons.

Diggle, P. J. 1990. *Time Series: A Biostatistical Introduction.* Oxford: Clarendon Press.