11.1 An ANOVA variance-stabilizing transformation stabilizes variances in the following approximate way. Let Y have mean θ and variance $v(\theta)$.

(a) Use arguments as in Section 10.1.3 to show that a one-term Taylor series approximation of the variance of $g(y)$ is given by $\mathrm{Var}\,(g(Y)) = [\frac{d}{d\theta}g(\theta)]^2 v(\theta)$.

(b) Show that the approximate variance of $g^*(Y)$ is independent of θ, where $g^*(y) = \int [1/\sqrt{v(y)}]dy$.

11.2 Verify that the following transformations are approximately variance-stabilizing in the sense of Exercise 11.1.

(a) $Y \sim$ Poisson, $g^*(y) = \sqrt{y}$

(b) $Y \sim$ binomial$(n, p), g^*(y) = \sin^{-1}(\sqrt{y/n})$

(c) Y has variance $v(\theta) = K\theta^2$ for some constant $K, g^*(y) = \log(y)$.

(Conditions for the existence of variance-stabilizing transformations go back at least to Curtiss 1943, with refinements given by Bar-Lev and Enis 1988, 1990.)

11.3 The Box–Cox family of power transformations (Box and Cox 1964) is defined by

$$g_\lambda^*(y) = \begin{cases} (y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0, \end{cases}$$

where λ is a free parameter.

(a) Show that, for each $y, g_\lambda^*(y)$ is continuous in λ. In particular, show that

$$\lim_{\lambda \to 0} (y^\lambda - 1)/\lambda = \log y.$$

(b) Find the function $v(\theta)$, the approximate variance of Y, that $g_\lambda^*(y)$ stabilizes. (Note that $v(\theta)$ will most likely also depend on λ.)

Analysis of transformed data in general and the Box–Cox power transformation in particular has been the topic of some controversy in the statistical literature. See Bickel and Doksum (1981), Box and Cox (1982), and Hinkley and Runger (1984).

11.5 Suppose that random variables Y_{ij} are observed according to the overparameterized oneway ANOVA model in (11.2.2). Show that, without some restriction on the parameters, this model is not identifiable by exhibiting two distinct collections of parameters that lead to exactly the same distribution of the Y_{ij}s.

11.8 Show that under the oneway ANOVA assumptions, for any set of constants $\mathbf{a} = (a_1, \ldots, a_k)$, the quantity $\sum a_i \bar{Y}_{i\cdot}$ is normally distributed with mean $\sum a_i \theta_i$ and variance $\sigma^2 \sum a_i^2/n_i$. (See Corollary 4.6.10.)

11.9 Using an argument similar to that which led to the t test in (11.2.7), show how to construct a t test for

(a) $H_0\colon \sum a_i\theta_i = \delta$ versus $H_1: \sum a_i\theta_i \neq \delta$.
(b) $H_0\colon \sum a_i\theta_i \leq \delta$ versus $H_1: \sum a_i\theta_i > \delta$, where δ is a specified constant.

11.11 For any sets of constants $\mathbf{a} = (a_1, \ldots, a_k)$ and $\mathbf{b} = (b_1, \ldots, b_k)$, show that under the oneway ANOVA assumptions,

$$\text{Cov}\left(\sum a_i \bar{Y}_{i\cdot}, \sum b_i \bar{Y}_{i\cdot}\right) = \sigma^2 \sum \frac{a_i b_i}{n_i}.$$

Hence, in the oneway ANOVA, contrasts are uncorrelated (orthogonal) if $\sum a_i b_i/n_i = 0$.

11.13 Under the oneway ANOVA assumptions, show that the likelihood ratio test of $H_0: \theta_1 = \theta_2 = \cdots = \theta_k$ is given by the F test of (11.2.14).

11.15 (a) Show that for the t and F distributions, for any ν, α, and k,

$$t_{\nu,\alpha/2} \leq \sqrt{(k-1)F_{k-1,\nu,\alpha}}.$$

(Recall the relationship between the t and the F. This inequality is a consequence of the fact that the distributions $kF_{k,\nu}$ are stochastically increasing in k for fixed ν but is actually a weaker statement. See Exercise 5.19.)

(b) Explain how the above inequality shows that the simultaneous Scheffé intervals are always wider than the single-contrast intervals.

(c) Show that it also follows from the above inequality that Scheffé tests are less powerful than t tests.

11.19 Let $X_i \sim \text{gamma}(\lambda_i, 1)$ independently for $i = 1, \ldots, n$. Define $Y_i = X_{i+1}/\left(\sum_{j=1}^{i} X_j\right)$, $i = 1, \ldots, n-1$, and $Y_n = \sum_{i=1}^{n} X_i$.

(a) Find the joint and marginal distributions of $Y_i, i = 1, \ldots, n$.

(b) Connect your results to any distributions that are commonly employed in the ANOVA.