TRANSFORMATIONS TO OBTAIN EQUAL VARIANCE (Section 5.6.2)

General method for finding variance-stabilizing transformations: If Y has mean μ and variance σ^2, and if U = f(Y), then by the first order Taylor approximation around μ,

$$U \approx f(\mu) + (Y - \mu) f'(\mu),$$

so
$$\begin{aligned} Var(U) &\approx Var[f(\mu) + (Y - \mu) f'(\mu)] \\ &= [f'(\mu)]^2 Var(Y - \mu) \\ &= [f'(\mu)]^2 \sigma^2. \end{aligned}$$

If we have an ANOVA situation in which the group variances σ_i^2 are a function of the group means μ_i, say

$$\sigma_i^2 = g(\mu_i),$$

then if we choose the function f so that

$$f'(y) = [g(y)]^{-1/2},$$

and take $U_i = f(Y_i)$ (where Y_i denotes the response variable), we will have

$$\begin{aligned} Var(U_i) &\approx [f'(\mu_i)]^2 \sigma_i^2. \\ &= [g(\mu)]^{-1} g(\mu) = 1. \end{aligned}$$

Thus such a transformation (or any scalar multiple of it) should give a transformed variable U with approximately equal variance.

Example: If $\sigma_i^2 \approx k(\mu_i)^q$ for some constant k and exponent q, then $g(y) = ky^q$, so we want $f'(y) \propto y^{-\frac{q}{2}}$, giving

$$f(y) \propto \begin{cases} y^{1-\frac{q}{2}}, \text{if } q \neq 2 \\ \ln(y), \text{if } q = 2 \end{cases}.$$

(If some of the y's are zero or negative, then we will need to add a suitable constant to y before taking a negative power or log.)

Often a suitable value of q can be determined empirically, using the following idea:
If $\sigma_i^2 \approx k(\mu_i)^q$, then $\ln(\sigma_i^2) \approx \ln(k) + q\ln(\mu_i)$, so

- If a plot of $\ln(\sigma_i^2)$ vs $\ln(\mu_i)$ is close to a straight line, then a power transformation is a suitable choice.
- In this event, q can be estimated as the slope of a line approximately fitting this plot.

Cautions when transforming data:

- Other model assumptions (especially normality) need to be checked before running the analysis, since the transformation might mess up other assumptions.
- Significance levels and confidence levels using transformed data will only be approximate, if the model has been changed *based on the data.*
- Interpretations need to be made in terms of the transformed units, or transformed back to the original units with care not to misinterpret.

Example: Battery data, with response "battery life" (rather than life per dollar).

Transformations based on theoretical considerations: Sometimes theoretical considerations point to a particular relationship between mean and variance, suggesting a particular transformation. Examples:

Type of Distribution	**Mean/Variance relationship**	**Type of Transformation**	**Comments**
Poisson	Variance = mean (so q = 1)	Square root (1-q/2 = 1/2)	1. Likely to occur with count data for rare events -- e.g., counts of accidents, flaws, or contaminating particles. 2. Simulations suggest that for sample size 15, the transformation does not substantially alter the probability of false rejection.
Binomial	Mean = mp, variance = mp(1-p)	$\arcsin\left(\sqrt{\frac{y}{m}}\right)$	1. Likely to occur with count data such as number of seeds in a fixed number that germinate, number of culture plates that grow visible bacteria colonies. 2. Simulations suggest that for m = 10, transformation does not change probability of false rejection.
Exponential	Variance = mean^2 (q = 2)	Log(y) (1 - q/2 = 0)	1. Likely to occur with certain kinds of reaction times, waiting times, and financial data. 2. Simulations suggest that with small sample sizes when differences in group means are large, transformation increases power, but in other cases can decrease power.

More on Transformations

Suppose we originally have response variable Y_i for the i^{th} treatment group. Our original intent was to test

H_0: $\mu_1 = \mu_2 = \ldots = \mu_t$

against

H_a: At least two of the μ_i's differ.

If we transform by f, we now have transformed response $U_i = f(Y_i)$ for the i^{th} group. We will assume that f is monotone -- that is, it either preserves order or reverses order; in either case, f is invertible -- that is, we know variable Y_i if we know U_i.

Letting $\mu_i^* = E(U_i)$we test

H_0*: $\mu_1^* = \mu_2^* = \ldots = \mu_t^*$

against

H_a*: At least two of the μ_i^*'s differ.

If the one-way ANOVA model is correct for the transformed variables, then in particular each U_i is normal with variance $(\sigma^*)^2$. Thus, if the model is correct for the transformed variables, H_0 says that all U_i's have the same distribution, so it would follow that all Y_i's have the same distribution, and hence the same mean. In other words, if the model is correct for the transformed variables, then H_0* implies H_0. Thus: if we do not reject H_0* then it is reasonable to say that the data are consistent with H_0.

Still assuming the model is correct, is the converse of the above conclusion true? Equivalently (assuming the model is correct), if we know H_0* is false, can we conclude that H_0 is false? I have seen this asserted, and it seems to be true for the transformations I have checked, but I haven't found a proof for it. It certainly is *not* the case that the mean of Y_i transforms to the mean of U_i.

However, we *can* assert that the *median* of Y_i transforms to the *median* of U_i, which, if U_i is normal, is the same as the mean of U_i. Thus if we reject H_0*, we have evidence *against* the hypothesis

H_0': The medians of the Y_i's are all the same

in favor of

H_a': At least two of the medians of the Y_i's differ.

Confidence intervals with transformed variables

- We can "backtransform" a confidence interval for the mean of a transformed variable to a CI for the *median* of the original variable. Usually the resulting confidence interval for the median is *not* symmetric about the median.
- We can form confidence intervals for differences of means (or other contrasts) with the transformed data, but the interpretation needs to be made in terms of the transformed variables. This is not always feasible.

If we need confidence intervals for differences of means or other contrasts for the original response variable, we need to work with the original data. A variety of methods of analysis have been developed. These include:

- Satterthwaite's approximation, discussed in Section 5.6.3 of the text.
- General linear models (There are entire books and courses devoted to this topic.)
- Weighted ANOVA can be used when the ratio of the variances in the different groups is known -- for example, when responses in the i^{th} group are the average of n_i measurements, but the variance of individual measurements is the same for all groups.
- The Welch procedure for contrasts. This is a generalization of the "unpooled t-test' for comparing two means.
- The Brown-Forsythe modified F-test.

Note: Another possible problem with transforming data is that transformations can produce values for the response that don't make sense in the original context. For example, if we transform by square roots, then assuming the square root has a normal distribution isn't entirely accurate because the normal distribution could have negative values. Depending on the situation, this might or might not be a concern. If, for example, the transformed variable has mean 20 and standard deviation 1, there is likely to be no problem. However, if the interest in the original question is in rare events in the negative direction, then this "negative tail" scenario could make the analysis totally unhelpful.

Box-Cox Transformations

This is a computerized method of finding possible transformations in the power family (including logs) to attempt to equalize variance and achieve normality. It is not implemented in Minitab (although there are macros available for Box-Cox there). We will not use this method in this course.