

Topic 8: Transformations of the data [S&T Ch. 9.16]

8. 1. The assumptions of ANOVA

The fundamental model for the completely randomized design is

$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. There are four key assumptions implicit in this model:

8. 1. 1. Additive effects

The treatment terms τ_i and error terms ε_{ij} are *added* (as opposed to, for example multiplied). A common way to say this is that "The treatment remains constant over replications and the replication effect remains constant over treatments". In terms of the model above, the treatment effect τ_i is independent of j , the index of replications, and the replication effect, which is the error term ε_{ij} , acts the same for all treatments.

If multiplicative effects are present the model would be more appropriately described by:

$$Y_{ij} = \mu\tau_i\varepsilon_{ij}$$

Tukey's test (Syllabus Topic 6.4.1) provide a tool to test if the departure from additivity due to multiplicative effects is statistically significant in a RCBD with one observation per cell (treatment x block combination).

8. 1. 2. Independence of errors

This says that the values of the ε_{ij} are statistically independent. Failure of this assumption is often caused by a failure to properly randomize the plots, e.g.

Rep	I	II	III
1	A	2 A	3 A
4	B	5 B	6 B
7	C	8 C	9 C

The self-similarity of experimental units adjacent in space or time is called **positive autocorrelation**. Regular alternation of positive and negative errors is a manifestation of **negative autocorrelation**. Independence of errors in a sequence of continuous variates may be tested using a test based on the differences between adjacent values (Biometry p394 –395). However, the physical process of randomly allocating the treatments to the experimental units ensures that the ε_{ij} will be independent. One ordinarily does not need to do more than a true randomization to satisfy this requirement. Often, however, if another assumption is violated then the independence of errors assumption will be violated, too.

8. 1. 3. Normally distributed errors

Violation of normality means that the error terms ε_{ij} are not normally distributed. This is the least influential assumption on the F test, and some people don't worry too much about it. A plot of the residuals or a check of normality (PROC UNIVARIATE Normal;) of the residuals generated by the OUTPUT statement is useful. The test for normality used by the Normal option is the Shapiro-Wilk (ST&Dp.567). Normality is rejected if the Pr<W in the SAS output is lower than 0.05. If the sample size is ≥ 2000 the Kolmogorov-Smirnov statistic (ST&Dp.564-568) is calculated.

8. 1. 4. Homogeneity of Variance

This is usually the most important assumption. The variances of the sample data may depend on the treatment mean. This is often the case when the treatments have substantial effects on the response variable, but can also be true for certain types of data as discussed next. The variances may also be unequal but have no apparent relation to the treatment means. The following example from Little and Hills show what can happen when homogeneity is violated.

Treatment	Replicate					Total	Mean	s^2
	1	2	3	4	5			
A	3	1	5	4	2	15	3	2.5
B	6	8	7	4	5	30	6	2.5
C	12	6	9	3	15	45	9	22.5
D	20	14	11	17	8	70	14	22.5

Little and Hills first analyze the data using Fisher's protected LSD

Source of variation	df	SS	MS	F
Treatments	3	330	110	8.8*
Error	16	200	12.5	

Since the F statistic is significant, we compute the LSD for α 0.05 significance level as $LSD = 4.74$. Therefore the difference between means A and B is not significant while that between C and D is. Analyzing A and B separately from C and D yields the opposite result. We do a two-sample test between A and B and between C and D:

Source of variation	df	SS	MS	F
Treatments A B	1	22.5	22.5	9*
Error	8	20.0	2.5	

Source of variation	df	SS	MS	F
Treatments C D	1	62.5	62.5	2.8 NS
Error	8	180	22.5	

The difference between A and B is significant while that between C and D is not, when analyzed separately. The reason is that the variance for treatments C and D is larger than for A and B.

The consequences of moderate heterogeneity of variances are not too serious for the overall test of significance, but single degree-of-freedom comparisons may be far from accurate

Among the different tests for homogeneity of variances, Bartlett's (S&T, p. 471) and Levene's test are widely used and are available in the Analyst application for single-way ANOVA. Bartlett's test is a chi-square test and Levene's test is an ANOVA of the square of the residuals.

If Levene's test is significant and the problem of heterogeneity of variances can not be corrected by transformation, you can perform a **Welch's variance-weighted ANOVA** to test for differences between group means. This option is available in Analyst (One-Way ANOVA -> Tests) but only for CRD (one-way ANOVA). Another alternative is the use of **non-parametric statistics**.

8. 2. Transformations

If the evidence indicates that the assumptions for an analysis of variance cannot be maintained, two courses of action are available. The first one is to carry out a different test not requiring the rejected assumptions, such as a non-parametric test or a variance-weighted ANOVA. The second one is to transform the variable to be analyzed in such a manner that the resulting transformed variables meet the assumptions of the analysis and then perform the entire analysis on the transformed variables.

When a statistical test may be made significant after transformation of a set of data, though it would not have been so without such a transformation, people may feel suspicious. What is the justification for transforming the data? There is really no scientific necessity to employ the common linear or arithmetic scale to which we are accustomed. If a relationship is multiplicative on a linear scale, it may make much more sense to think of it as an additive system on a logarithmic scale. The square root of a variable is another frequent transformation. The square root of the surface area of an organism may be a more appropriate measure of the fundamental biological variable subjected to physiological and evolutionary forces than is the area. This concept is reflected in the normal distribution of the square root of the variable as compared with the skewed distribution of areas. In many cases experience has shown that is better to express some variables in logarithms (pH values), square roots (areas) or reciprocals (microbiological titrations). As soon as the idea that the scale of measure is arbitrary is accepted it is not difficult to understand that is valid to look at the distributions of transformed variables to decide which transformation most closely satisfies the assumptions of the ANOVA.

A fortunate fact about transformations is that very often several departures from the assumptions of ANOVA are simultaneously cured by the same transformation to a new scale. Simply by making the data homoscedastic (with uniform variances), normality and additivity of treatment effects may be obtained.

Four transformations will be discussed: the logarithmic transformation, the square root transformation, the angular or arcsine transformation and the power transformation.

8. 2. 1 The log transformation (Little and Hill)

Whenever the **standard deviations** (not the variances) of samples are roughly proportional to the means, the most effective transformation is a log transformation. Another criterion for deciding on this transformation is the evidence of multiplicative rather than additive main effects (significant Tukey's test). Finally, frequency distributions skewed to the right are often made more symmetrical by transformation to logarithmic scale.

Data with negative values cannot be transformed using logarithms, because the log of zero is minus infinity. In these cases a 1 must be added to each data point before transforming. Data containing a large number of zeros would probably be handled better by some other method. Logarithms to any base can be used, but common logarithms (to the base 10) are generally the easiest. Before transforming, it is legitimate to multiply all data points by a constant, since this has no effect on the subsequent analyses. This is particularly useful if any of the data points are less than 1 to avoid negative logarithms.

The following data set (Little and Hills 1978) will be used as an example of the effects of the log transformation on the assumptions of the ANOVA. The dependent variable is weight in pounds of vitamin-treated and control animals, in a RCBD.

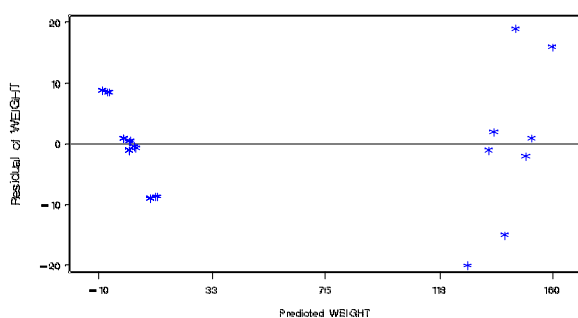
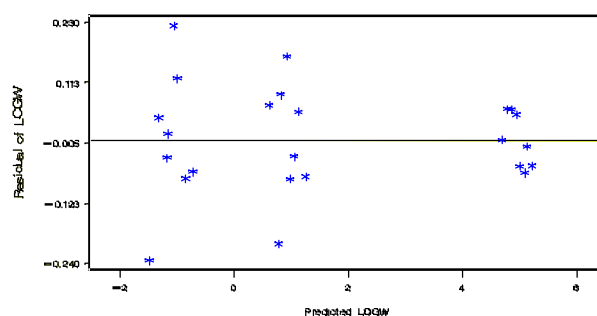
Species—Treatment	Block				Total	Mean	s _i ²
	I	II	III	IV			
Mice—control	0.18	0.30	0.28	0.44	1.2	0.3	0.0115
Mice—vitamin	0.32	0.40	0.42	0.46	1.6	0.4	0.0035
Subtotals	0.50	0.70	0.70	0.90	2.8	0.35	
Chickens—control	2.0	3.0	1.8	2.8	9.6	2.40	0.3467
Chickens—vitamin	2.5	3.3	2.5	3.3	11.6	2.90	0.2133
Subtotals	4.5	6.3	4.3	6.1	21.2	2.65	
Sheep—control	108.0	140.0	135.0	165.0	548.0	137.0	546.0
Sheep—vitamin	127.0	153.0	148.0	176.0	604.0	151.0	425.3
Subtotals	235.0	293.0	283.0	341.0	1152.0	144.0	
Totals	240.0	300.0	288.0	348.0	1176.0	49.0	

In SAS a new transformed variable is generated after the input statement

```
data XXX;
input trtmnt $ block weight @@;
logwt= log10(weight);
cards;
```

In this data, there are no zeros, but since the lowest value is 0.18, the data was multiplied by 10 before taking logs. The next table compares the analysis of the transformed and original data.

	Original data	Transformed data
Shapiro-Wilks W Statistic for residuals	W= 0.95, p= 0.32	W= 0.97, p= 0.56
Tuckey's test of nonadditivity	F= 545, p= 0.0001	F= 1.74, p= 0.21
Levene's test of homogeneity of s_i^2	F= 2.5, p= 0.07	F= 1.78, p= 0.17
	0.0115, 0.0035, 0.347, 0.213, 546.0, 425.3	0.134, 0.024, 0.062, 0.025, 0.030, 0.018
Treatment effect	F= 174.4, p= 0.0001	F= 1959.4, p= 0.0001
Species effect	F= 434.5, p= 0.0001	F= 4883.0, p= 0.0001
Vitamin effect	F= 1.14, p= 0.30	F= 16.62, p= 0.011
Species * Vitamin	F= 1.0, p= 0.39	F= 1.58, p= 0.24

Residuals vs. predicted. **Original data**Residual vs. predicted **Log10(original data)**

When this data was analyzed without any transformation no significant differences were detected between vitamin and control. This seems very strange because every animal in every replicate receiving the vitamin showed a greater weight than the corresponding control animal. It also seems strange that no significant differences were detected for the interaction between vitamin effects and species, since the apparent response to vitamins is so different in the different species.

These unexpected results suggest that there is something wrong with the ANOVA. When the assumptions of the ANOVA are checked, significant Tukey and Levene's test demonstrate that there is no additivity and that there is a significant heterogeneity of variances. A quick look at the residuals confirms these violations of the ANOVA assumptions.

The presence of multiplicative effects and a rough proportionality between standard deviations and means suggest a logarithmic transformation. The analysis using the logarithm of the original data is more satisfying than the analysis of the original data as far as positive results are concerned. The interaction between species and vitamins is still no significant, but the question is now different. With the original data the question was, "Does the amount of change in weight due to the addition of vitamins vary from species to species?" With the logarithm transformation of the original data the question is, "Does the *proportion or percent* change in weight due to vitamins vary from species to species?"

Did we get more positive with the transformed data because we were simply “playing with figures” until we got a result we liked? Or was the transformation we used justified and is the new analysis valid? Values of Tukey and Levene’s test (and also the better normality of the residuals) are significant for the original data but non-significant for the transformed data. The new analysis is valid, since the transformed data satisfied *all* the assumptions of the analysis of variance. With the original data, none of the assumptions were true.

Effect of log transformation on the mean and the variances

This example shows the effect of the \ln transformation on a multiplicative effect of a treatment than increases the control values by 50% (*1.5). See the similar variance of 0.2740 after the \ln transformation.

						Mean	Variance
Control							
Y	20	40	50	60	80	50	500
$\ln(Y)$	2.9957	3.6889	3.9120	4.0943	3.820	3.8146	0.2740
Treatment = (1.5)*control							
Y	30	60	75	90	120	75	1125
$\ln(Y)$	3.4012	4.0943	4.3175	4.4998	4.7875	4.2201	0.2740

8. 2. 2. The square root transformation (Little and Hill)

Whenever we are dealing with counts of rare events -insects on a leaf or blood cells in a hemacytometer, for example- the data tend to follow a special distribution called a *Poisson distribution*. By a rare event, we mean one that has a very low probability of occurring in any individual. For example, suppose that in a lot of lettuce seed, 0.1% of the seed was carrying mosaic disease virus. The probability that any individual seed contains mosaic is then only 1/1000, so as far as a single seed is concerned, this is a rare event. If we take 100 samples of 1000 seeds each from such a lot, we will get approximately these results:

37	“	“	“	1	“	“
18	“	“	“	2	“	“
6	“	“	“	3	“	“
2	“	“	“	4	“	“

It is obvious that this looks very little like a normal distribution. This Poisson distribution has a very interesting characteristic—the **variance** is equal to the mean. In actual practice, the variance is generally somewhat larger than the mean because other factors, in addition to sampling variation, are affecting the occurrence of the events being counted. At any rate, the variance tends to be proportional to the mean, thus violating the assumption that the variances and means are not correlated.

Another example of data of this kind is found in insect counts, such as those made from a standard number of sweeps with a net. The probability of finding an insect at a particular spot selected at random at one particular time is indeed a rare event.

Data of this kind can be made more nearly normal and at the same time the variances can be made relatively independent of the means by transforming them to square roots. Actually, it is better to use:

$$\sqrt{Y + \frac{1}{2}}$$

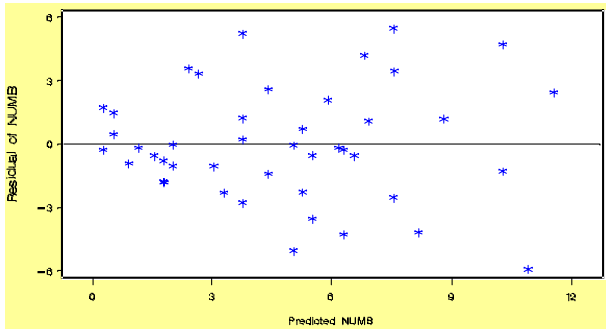
if there are counts under 10 [**SAS: SQRT(variable+0.5)**].

Example. Number of lygus per 50 sweeps. In each plot of an experiment testing 10 insecticides and a check treatment, replicated four times in a randomized complete block design.

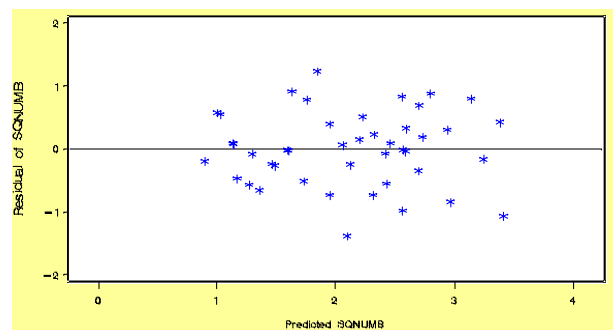
Treatment	Block				Total	Mean	s _i ²
	I	II	III	IV			
A	7	5	4	1	17	4.25	6.25
B	6	1	2	1	10	2.50	5.67
C	6	2	1	0	9	2.25	6.92
D	0	1	2	0	3	0.75	0.92
E	1	0	1	2	4	1.00	0.67
F	5	14	9	15	43	10.75	21.58
G	8	6	3	6	23	5.75	4.25
H	3	0	5	9	17	4.25	14.25
I	4	10	13	5	32	8.00	18.00
J	6	11	5	2	24	6.00	14.00
K	8	11	2	6	27	6.75	14.25

Comparison of transformed and original data

	Original data	Transformed data
Shapiro-Wilks W Statistic for residuals	W= 0.98, p= 0.73	W= 0.987, p= 0.95
Tuckey's test of nonadditivity	F= 0.63, p= 0.44	F= 0.07, p= 0.80
Levene's test of homogeneity of s ²	F= 1.61, p= 0.15	F= 0.90, p= 0.55
Treatment effect	F= 3.7, p= 0.0026	F= 4.04, p= 0.0014



Original data. Residuals vs. Predicted
Correlation mean-variance $r=0.89^{**}$ ($p=0.001$)



Transformed data. Residuals vs. Predicted
Correlation mean-variance $r=0.37^{NS}$ ($p=0.1$)

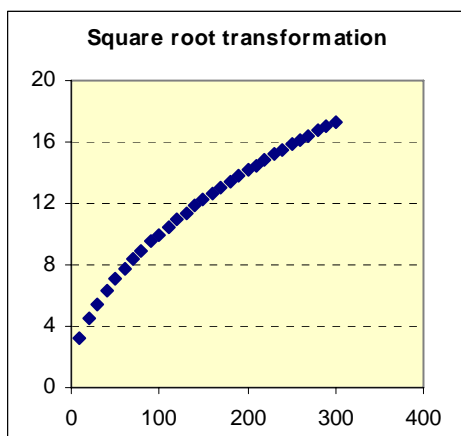
The two analyses are not very different, since they both show a highly significant treatment effect. The F value is about 10% higher after transformation. Some differences will occur in mean separation, as shown below.

The *weighted means* below are obtained by “detransforming” the means of the transformed data back to the original units. This is done by squaring the transformed means and subtracting one-half. The means obtained in this way are smaller than those obtained directly from the raw data because more weight is given to the smaller variates. This is as it should be, since in a Poisson distribution the smaller variates are measured with less sampling error than the larger ones. Actually in reporting the results of such an experiment, it is better to use these weighted means, making it clear in the report how they were obtained.

Duncan’s multiple range test on raw and transformed data, (5% level). Mean Separation

	D	E	C	B	H	A	G	J	K	I	F
ORIGINAL	0.75	1.00	2.25	2.50	4.25	4.25	5.75	6.00	6.75	8.00	10.75
	<hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/>										
SQRT	0.62	0.89	1.81	2.22	3.50	3.95	5.50	5.60	6.31	7.57	10.32
	<hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/> <hr style="border: none; border-top: 1px solid black; margin: 5px 0;"/>										

Letters connected by the same line are not significantly different. As an example, note that in the transformed data, G and D, G and E, J and D, and J and E were declared significantly different, whereas they were not in the raw data.



- **The general effect of the square root transformation is to increase the precision with which we measure the differences between small means.**
- This is highly desirable in insect control work, since we are generally not as interested in differences between two relatively ineffective treatments as we are in comparing treatments that give good control.
- Note that the values of the means of the square root values, transformed back to the original scale are smaller than the original means. This is because of the higher weight placed to the smaller values with smaller variance and higher information content.

The assumption of independence of means and variances was violated in the original data, and this was remedied by the transformation. There do not appear to be any additional serious violations of the assumptions. An examination of the error components shows no striking deviation from a random and normal distribution. Carrying out Levene's test for homogeneity of variance on the raw data gives an F of 1.61, which has a 15% probability of being exceeded by chance alone. After transformation, this F value was reduced to 0.90, which has a 55% probability of being exceeded by chance. Thus, transformation reduced the amount of heterogeneity over that in the raw data. In carrying out Tukey's test for additivity, even with the raw data, the F value for nonadditivity was less than one.

In general, we can say that data requiring the square root transformation do not violate the assumptions of the analysis of variance nearly as drastically as data requiring a log transformation. Consequently, the changes in the analysis brought about by the transformation are not nearly so spectacular.

8. 2. 3. The arcsine or angular transformation (Little and Hill)

Another kind of data that may require transformation is that based on counts expressed as percentages or proportions of the total sample. Such data generally have what is called a *binomial distribution* rather than a normal distribution. One of the characteristics of this distribution is that the variances are related to the means but in quite a different way than the types of data discussed before. Data discussed in 8.2.1 and 8.2.2 showed a positive correlation between means and variances. In binomial data, variances tend to be small at the two ends of the range of values (close to zero and 100%), but larger in the middle (around 50%).

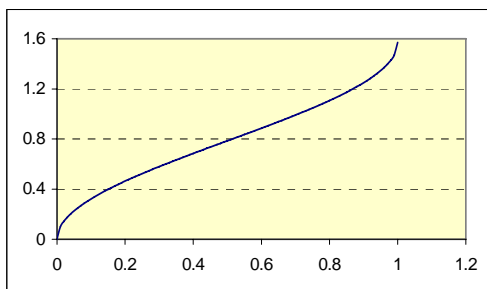
The data in Table 12.13 are from a completely randomized experiment on lettuce seed with 24 treatments, each replicated three times. Treatments are arranged in order of magnitude of their means. Note that there is a strong tendency for the variances at the extremes to be smaller than those in the middle of the range. This is typical of binomial data.

TABLE 12.13. Number of lettuce seeds germinating in samples of 50.

Treatment	Blocks			Mean	s_i^2
	1	2	3		
1	0	0	1	0.33	0.33
2	0	1	0	0.33	0.33
3	0	0	1	0.33	0.33
4	0	2	0	0.67	1.33
5	2	0	0	0.67	1.33
6	0	2	3	1.67	2.33
7	7	10	7	8.00	3.00
8	11	12	15	12.67	4.33
9	13	18	18	16.33	8.33
10	22	16	13	17.00	21.00
11	24	13	18	18.33	30.33
12	23	21	29	24.33	17.33
13	24	29	29	27.33	8.33
14	37	28	27	30.67	30.33
15	42	41	40	41.00	1.00
16	39	41	45	41.67	9.33
17	41	45	40	42.00	7.00
18	47	41	43	43.67	9.33
19	45	42	48	45.00	9.00
20	46	42	48	45.33	9.33
21	49	46	48	47.67	2.33
22	48	49	48	48.33	0.33
23	50	49	48	49.00	1.00
24	49	49	50	49.33	0.33
				Totals	178.00

Comparison of transformed and original data

	Original data	Transformed data
Levene's test of homogeneity of s^2	F= 2.43, p= 0.0048	F=0.99, p= 0.49
Treatment effect	F= 148.12, p= 0.0001	F= 100.14, p=0.0001



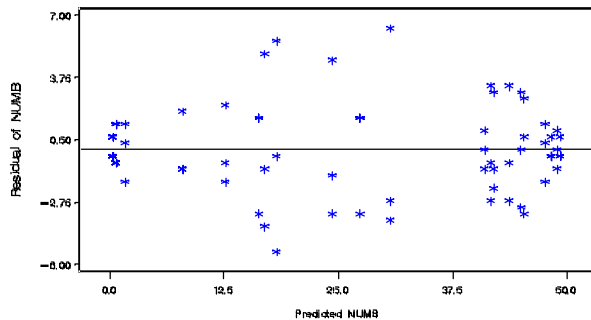
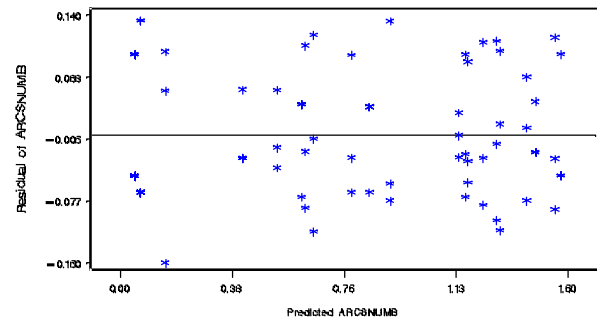
The appropriate transformation for data of this kind is called the angular or arcsine transformation. It is obtained by finding the angle whose sine is the square root of the proportion (percentage/100).

Written in mathematical shorthand, this is $\arcsin(Y)$ or $\text{sine}^{-1}(Y)$

• **In SAS: ARSIN(SQRT(proportion))].**

- ARCSINE: Returns the SINE of a number in radians in the range $-\pi/2$ to $+\pi/2$

- This transformation will spread the values at both ends of the distribution compared with the central part.
- Data should be transformed if the range of percentages is greater than 40. Otherwise, it is scarcely necessary.

Residuals vs. predicted. **Original data**Residual vs. pred. $\text{SIN}^{-1}(\text{SQRT}(\text{original} * 2/100))$

Since the data were based on samples of 50, each variate had to be multiplied by 2 to convert it to a percentage. The pattern of the residuals observable in the raw data is no longer apparent in the transferred data. An analysis of variance of the transformed data does not seem to lead us to a different conclusion than the analysis of the raw data:

The important difference is not in the overall analysis, but in mean separation. A Duncan's multiple range test shows that:

1. Five differences were declared significant before transformation but not after: 7–8, 8–11, 10–12, 11–12 and 12–14.
2. Five differences were declared significant after transformation but not before: 18–22, 19–23, 19–24, 20–23, and 20–24.

Which set of conclusions should we accept? The answer is simple: we should accept the conclusions based on the more valid analysis, in this case, the analysis of the transformed data.

We do not transform data to give us results more to our liking. We transform data so that the analysis will be *valid* and the conclusions *correct*.

8. 2. 4. The power transformation

(Hinz, P. N. and H. A. Eagles. 1976. Estimation of a transformation for the analysis of some agronomic and genetic experiments. *Crop Science* 16: 280-283.)

Field experiments are often conducted using replicated trials over a broad range of locations and environmental conditions. Often the means and the residual variances differ

markedly among environments. Furthermore there tends to be a marked positive correlation between the mean and the residual variance. In such instances, an analysis of the combined data over environments must contend with this heterogeneity in the residual variance. The transformation of the dependent variable is a well-known technique for reducing the relationship between the mean and the residual variance.

The choice of an optimal transformation from the many possible alternatives is not always obvious specially if the functional relationship between variance and mean is not known. The power transformation method provides a mean of selecting a transformation from a broad class of power transformations and employs the data themselves to estimate the exponent used in transforming the original measures.

The power transformation is to transform X to Y empirically by “a”,

$Y = X^a \quad \text{if } a \neq 0$ $Y = \log X \quad \text{if } a = 0$

Note that

$Y = \sqrt{X}$	if $a = 1/2$
$Y = \log X$	if $a = 0$
$Y = 1 / X$	if $a = -1$

Generally (1) If variances \uparrow as means \uparrow , use $a < 1$
 (2) If variances \uparrow as means \downarrow , use $a > 1$.

The exact value of the power of the transformation ‘a’ can be estimated by obtaining the slope ‘b’ of the regression of $\log s_i^2$ on $\log \bar{x}_i$, and then solving for ‘a’ = 1-b/2

Suppose there are k groups of observations and s_i^2 and \bar{x}_i are the sample variance and mean of the ith group. Let, $\log s_i^2 = \alpha + \beta \log \bar{x}_i$.

Use SAS Analyst, Excel or any other program to calculate a lineal regression between the logarithms of the observed variances on the logarithms of the means. Solve the equation and find the least squared estimate (b) of β .

Then $a = 1 - b/2$, which can be used to transform X to Y.

Examples of power transformation

Using means and variances from Table **Topic 8.2.2.**

$$\log S_i^2 = \alpha + \beta \log \bar{x}_i$$

$$\log S_i^2 = 0.1086 + 1.238 \bar{x}$$

$b = 1.238$ and the estimated power for transformation is: $a = 1 - (1.238/2) = 0.38$ that is close to 0.5, the square root transformation.

Using $Y = X^{0.38}$, Levene’s test is non significant (p=0.32)

Using means and variances from Table **Topic 8.2.1.**

$$\log S_i^2 = \alpha + \beta \log \bar{x}_i$$

$$\log S_i^2 = -1.337 + 1.855723 \bar{x}$$

$b = 1.855723$ and the estimated power for transformation is: $a = 1 - (1.855723/2) = 0.072$ that is close to 0. The logarithmic transformation should be appropriate in this case.

8.3. Reporting results from transformed data

When a transformation is applied, tests of significance are performed on the transformed data, but estimates of means are usually given in the familiar untransformed scale. Furthermore, it is better to calculate means of the transformed data before transforming back to original units. In this way we obtain correctly weighted means.

Since the transformations discussed before are non-linear, confidence limits computed in the transformed scale and change back to the original scale would be asymmetrical. Stating the standard error in the original scale would therefore be misleading. In reporting results of research with variables that require transformation, it is better to report means in the untransformed scale followed by their asymmetrical confidence limits rather than by the standard errors.

8.3.1 Transforming *ln* transformations back to the original scale

This example shows that when the mean of the logarithms is transformed back to the original scale, the resulting mean is a Geometric Mean of the original data.

- Geometric mean $G = (Y_1 * Y_2 * \dots * Y_n)^{1/n}$ and \log of $G = (1/n) \sum \log Y_i$.
- Geometric mean $G = (20 * 40 * 50 * 60 * 80)^{1/5} = 45.3586$

	Control					Mean	Variance
Y	20	40	50	60	80	50	500
$\ln(Y)$	2.9957	3.6889	3.9120	4.0943	3.820	3.8146	0.2740

- Note that when you transform back the mean of the \ln to the original scale: $e^{3.8146} = 45.3586$ and NOT the arithmetic mean = 50