Analysis of Variance [Chapter 5]

Testing between two samples is readily done with the two-sample t-test. In this situation we compare two groups (also referred to as classes, categories, treatments or indicator variables) and use hypothesis testing procedures that will allow us to decide if the two samples are, statistically speaking, significantly different (with α probability of error) or not.

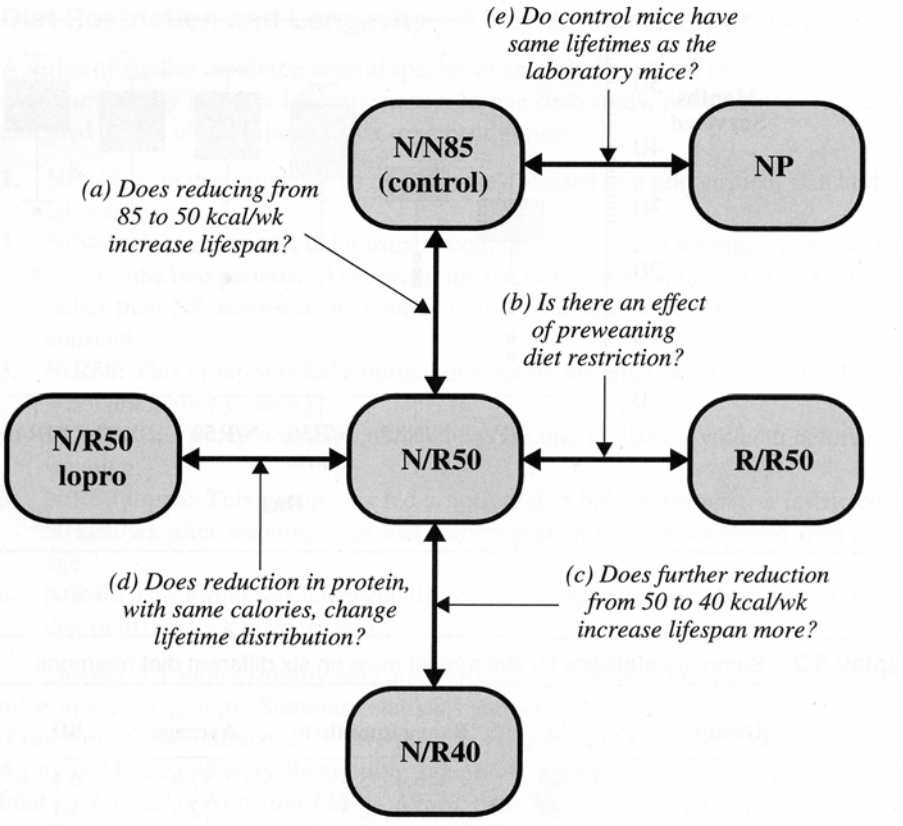
See example: Mouse Diet Experiment 01(mousefeed01.sas). Case study 5.1 from your text book.

We want to examine differences among the following 6 treatments

N/N85 fed normally before weaning and 85 kcal/wk after
N/R40 fed normally before weaning an d 40 kcal/wk after
N/R50 fed normally before weaning and 50 kcal/wk after
NP standard diet to satiation
R/R50 fed a reduced diet of 50 kcal/wk before and after weaning
lopro fed normally before weaning and 50 kcal/wk after and dietary protein decreasing with age

The text describes 5 distinct test of interest. These 5 tests are: N/N85 vrs NP (test e), N/N85 vrs N/R50 (test a), N/R50 vrs R/R50 (test b), N/R50 vrs lopro (test d) and N/R40 vrs N/R50 (test c). To do this in SAS I created 5 datasets, each with the appropriate 2 groups for one of the tests and I ran PROC TTEST to do the tests.

Display 5.3 Structure of planned comparisons among groups in the diet restriction study



The results of those 5 tests are as follows:

Test	Hypothesis $\mu_1 = \mu_2$	Means	P > |t| (unpooled)	P > |t| (pooled)
A	N/N85 = N/R50	32.691, 42.297	**<0.0001**	<0.0001
B	N/R50 = R/R50	42.297, 42.886	0.6474	**0.6532**
C	N/R40 = N/R50	45.117, 42.297	0.0275	**0.0294**
D	N/R50 = lopro	42.297, 39.686	0.0489	**0.0516**
E	N/N85 = NP	32.691, 27.402	<0.0001	**<0.0001**

According to the book, there is convincing evidence that restricting diet increases lifespan. This is based on Analysis of Variance (P>F < 0.0001). However, the individual tests are also pretty convincing.

All diets with restrictions (R) have longer life expectancy than normal (N) diets (see test A). The highest calorie restriction (85) lived longer than the unlimited calorie diet (test e). When restricted (R), the lower the level of calories the longer the life expectancy (test C, calories 40 versus 50). Finally, even on a restricted diet (R50) life expectancy increased if dietary protein decreasing with age (test D). The only test that was not significant, or near significance, was test B, suggesting that limiting calories before weaning was not beneficial.

One of the tests is ambiguous (D). We will be better able to make a decision when we have discussed ANOVA.

Why do we need Analysis of Variance (ANOVA)?

What happens if we have more than two groups? If we want to compare 3 groups, we could compare group 1 to group 2, 2 to 3 and 1 to 3. This is 3 tests, each with an α chance of error. The probability of error is not strictly additive, but Bonferroni showed that the probability of error will be no more than the sum of the individual tests. Since 0.05 + 0.05 + 0.05 = 0.15, the probability of error for the three tests would be a maximum of 0.15 (15%).

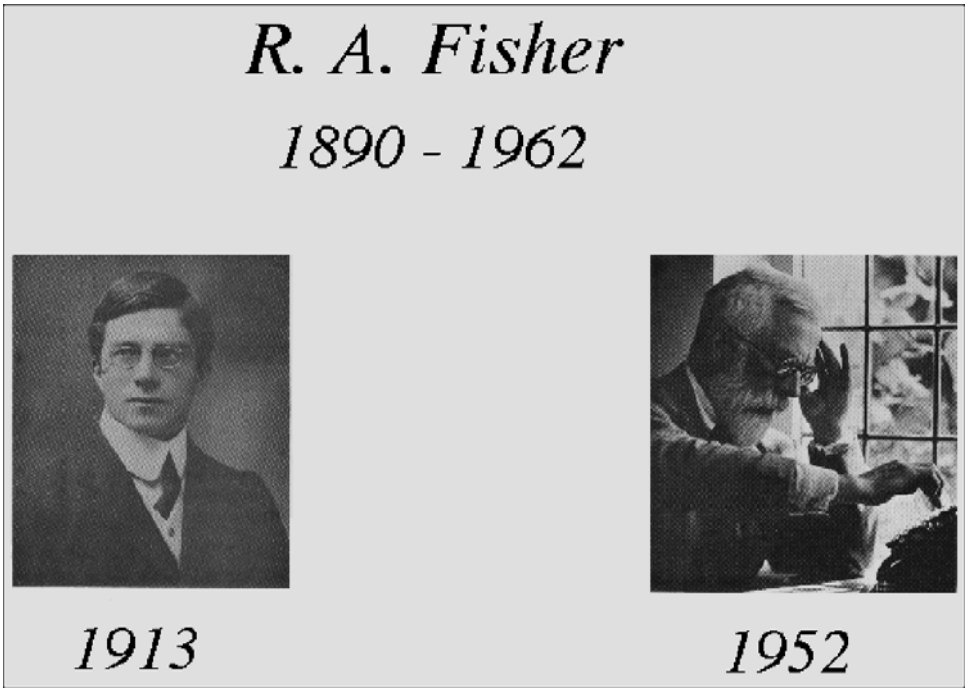
For the example above the tests are among 5 means (e.g. 1 versus 2, 1 v 3, 1 v 4, 1 v 5, 2 v 3, 2 v 4, 2 v 5, 3 v 4, 3 v 5 and 4 v 5). This is a total of 10 tests. The general expression for the total number of tests needed to test among t groups is t(t-1)/2. So, 3 groups require 3*2/2=3 tests, 4 groups require 4*3/2=6 tests, 5 groups require 5*4/2=10 tests, etc.

According to Bonferroni, testing among 4 groups (α=0.05) would have an upper bound of 6*0.05=0.30 and among 5 groups the probability of error is up to 10*0.5=0.50. At 20 groups we reach 1.00 or 100 percent, and it becomes obvious that Bonferroni is overly conservative. Consider 100 tests of 1000 tests.

But it makes one important point. More tests results a greater probability of error and this is not reflected in our selection of the α probability. So we need a test that will test multiple groups and only have a probability of error of α for all groups.

Carlo Emilio Bonferroni, 28 Jan 1892 (Bergamo, Italy) 18 Aug 1960 in Florence, Italy.

Enter R. A. Fisher!



What advantages do we get from combining the data into a single test? First the single decision making test is an advantage in and of itself because we have a test with a single α value. Second, by combining our groups into a single group we will potentially have more observations from which to estimate our variance and more degrees of freedom in doing tests.

What is Analysis of Variance? You will recall that we previously looked at the t-test and saw that for two or more means we could potentially pool the variances. This provides a single, superior estimate of the variance and a variance with more degrees of freedom for testing hypotheses. This

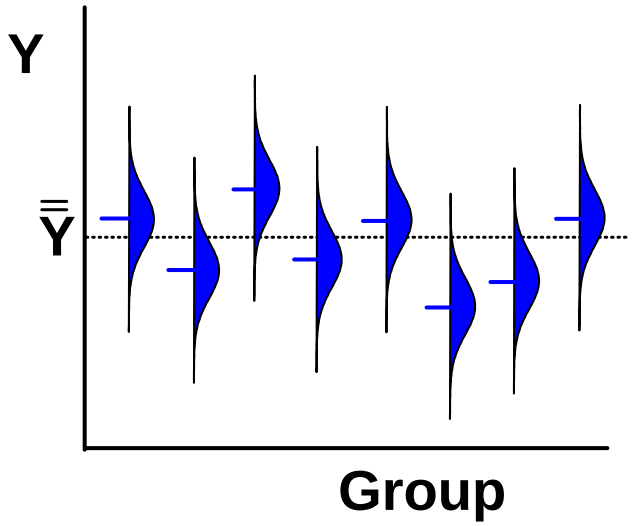
can be extended to more than the two means for the t-test. A pooled variances is just a weighted mean where

For two groups, $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$

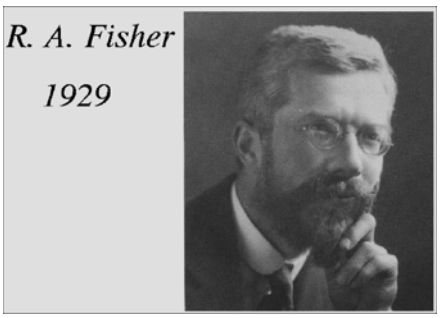
For more than two groups, $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + \ldots + (n_t-1)s_t^2}{(n_1-1)+(n_2-1)+(n_3-1)+\ldots+(n_t-1)}$

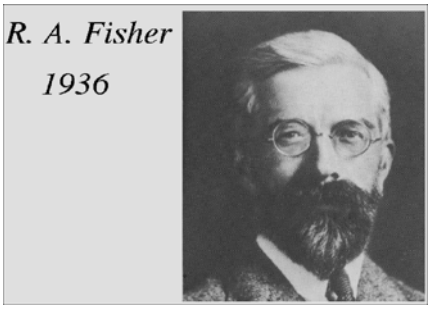
Recall that the standard error (√ variance of the mean) was estimated as $S_{\bar{Y}} = S/\sqrt{n}$. It seems odd that we cannot estimate variances from a single observation, but we can estimate the variance of the mean or a sample from a single sample. We can because we know the relationship between the variance of the observations and the variances of a mean of n observations, $S_{\bar{Y}}^2 = S^2/n$.

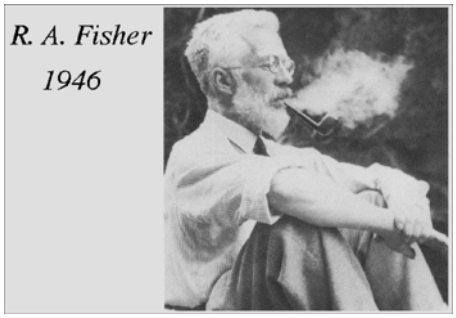
But, could we estimate the variance of the means from several samples with different means? What if we had several means ($\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_3$, $\bar{Y}_4$, ..., $\bar{Y}_t$) and we calculated the variance of these means? Would this give us the same "variance of the means"? Yes, IF the samples are all drawn from the same population and have the same mean, the estimate of $s_{\bar{Y}}$ should be the same whether it is estimated as $S_{\bar{Y}}^2 = S^2/n$ or as $S_{\bar{Y}}^2 = \frac{\sum_{i=1}^{t}(\bar{Y}_t - \bar{\bar{Y}})^2}{(t-1)}$.



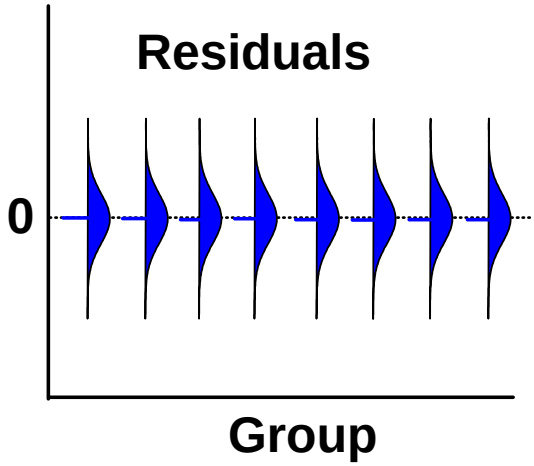
This being the case, Fisher reasoned, <u>if the null hypothesis is true</u> (H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_\tau$), then the two estimates should be the same. This is what is required for a test of hypothesis, that our known distribution of the test statistic be true if the null hypothesis is true. If the means are not equal then the estimate will not be $s_{\bar{Y}}$, it will be some larger value (since the means are different). The hypotheses are then H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$) versus the alternative (some μ_i is different). For our treatments in the mouse diet study we would state our null hypothesis as H_0: $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E$ versus the alternative (some diet is different).







Finally, we already have a good estimate of variance (S_p^2). Now we have a second, independent estimate of variance from the means, since $S_{\bar{Y}} = S/\sqrt{n}$ then $S^2 = (S_{\bar{Y}}\sqrt{n})^2$. If the null hypothesis is true these should estimate the same variance. To test for equality of two variances we use an F test.



So, Analysis of Variance consists of two estimates of variances. One is obtained by pooling the variances from within each group (within group variance) and the other is calculated between the groups as the variance of the means (between or among group variance). The calculations are not particularly complicated.

Within group variance: $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2 + \ldots + (n_t-1)S_t^2}{(n_1-1)+(n_2-1)+(n_3-1)+\ldots+(n_t-1)}$,

and when balanced so $n_i = n_j = n$, $S_p^2 = \dfrac{(n-1)S_1^2 + (n-1)S_2^2 + (n-1)S_3^2 + \ldots + (n-1)S_t^2}{t(n-1)}$

Between group variance: $S_{\bar{Y}}^2 = \dfrac{\sum_{i=1}^{t}(\bar{Y}_t - \bar{\bar{Y}})^2}{(t-1)}$

These calculations are usually expressed in an Analysis of Variance (ANOVA) table. This is the format produced by SAS. The table is expressed in terms of a "Sum of Squares" which is the calculation of variance before dividing by the degrees of freedom.

Within group sum of squares: $SS_{Within} = (n_1-1)S_1^2 + (n_2-1)S_2^2 + \ldots + (n_t-1)S_t^2 = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$

Between group sum of squares: $SS_{Between} = \sum_{i=1}^{t}(\bar{Y}_i - \bar{\bar{Y}})^2$

Analysis of Variance source table

Source	d.f.	Sum of Squares	Mean Square	F test
Between	**t–1**	$SS_{Between}$	$SS_{Between}$ / (t–1)	$MS_{Between}$ / MS_{Within}
Within	**t(n–1)**	SS_{Within}	SS_{Within} / t(n–1)	
Total	**tn–1**	**SSTotal**		

For the Mouse Life expectancy problem in SAS see example

PROC GLM gives a traditional ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12733.94181	2546.78836	57.10	<.0001
Error	343	15297.41532	44.59888		
Corrected Total	348	28031.35713			

PROC MIXED separates "fixed" effects (most treatments) from random effects (all errors and some treatments)

Covariance Parameter Estimates

Cov Parm	Estimate
Residual	44.5989

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
diet	5	343	57.10	<.0001

Note that the standard deviation (MSE or Residual variance) is calculated with a pooled estimate across all of the groups. The original total number of observations in the mousefeed dataset was 349. The pooled variance will have 349 minus one d.f. for each mean fitted, one for each of the 6 groups. The resulting d.f. for the pooled error is 349 – 6 = 343.

Expected mean squares

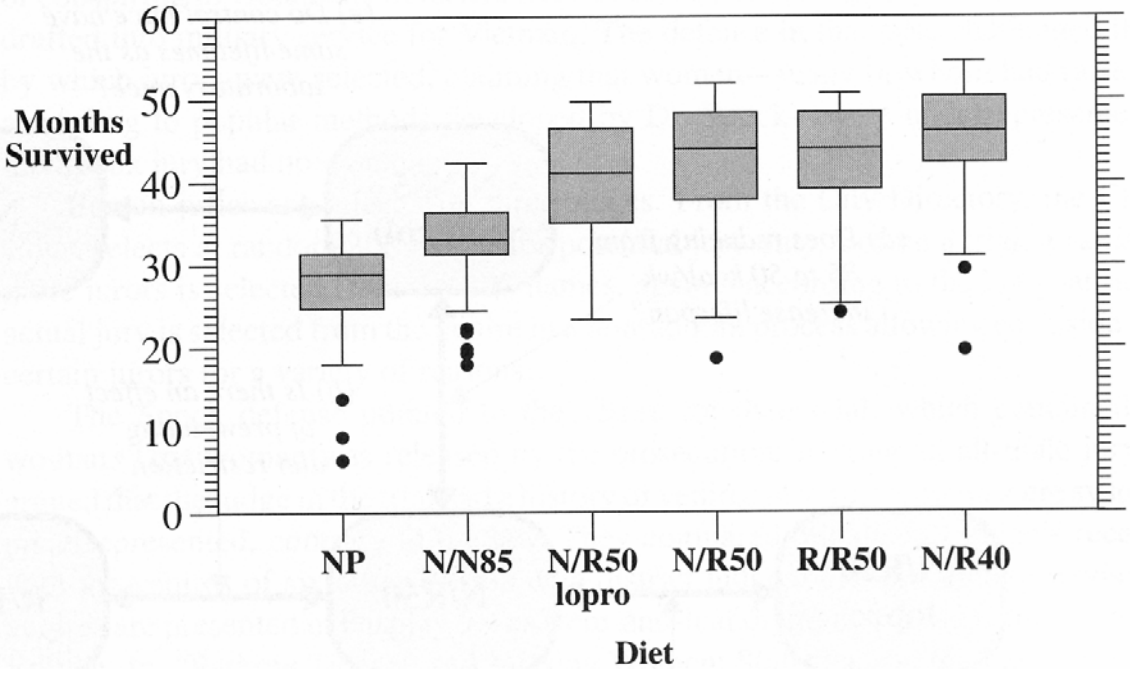
When we calculate S_p^2 we are estimating σ^2, the random variation.

When we estimate $S_{\bar{Y}}^2$ we are also estimating σ^2/n which when multiplied by n will provide a second estimate of σ^2 if the null hypothesis is true. If the null hypothesis is not true there is some additional variation in this estimate, σ_τ^2 variation due to the treatments. This is then $\sigma^2/n + \sigma_\tau^2$. When multiplied by n we have $\sigma^2 + n\sigma_\tau^2$.

The F test is then $F = \frac{\sigma^2 + n\sigma_\tau^2}{\sigma^2}$. The null hypothesis can be stated as $H_0: \sigma^2 = 0$. For fixed effects the sum of squares does not estimate a variance, so the F test is $F = \frac{\sigma^2 + \sum_{i=1}^{n} \tau_i^2 / n-1}{\sigma^2}$.

We can plot the results (see SAS) and get the following. The plot below is from your book.

Display 5.1 Lifetimes of female mice fed on six different diet regimens



ANOVA tells us there are differences among the mean survival for the treatments. The logical next step is to determine which diets are different from which other diets. This is in Chapter 6, so we will com back to this example.

Assumptions and Robustness

The assumptions are :

1) Normality – normality is assumed, though the analysis is "robust" and tends to perform well if the distribution is symmetric. Strongly skewed distributions are problematic.

2) Independence – this assumption is achieved in part by randomization. In some cases a lack of independence can be addressed by some statistical applications.

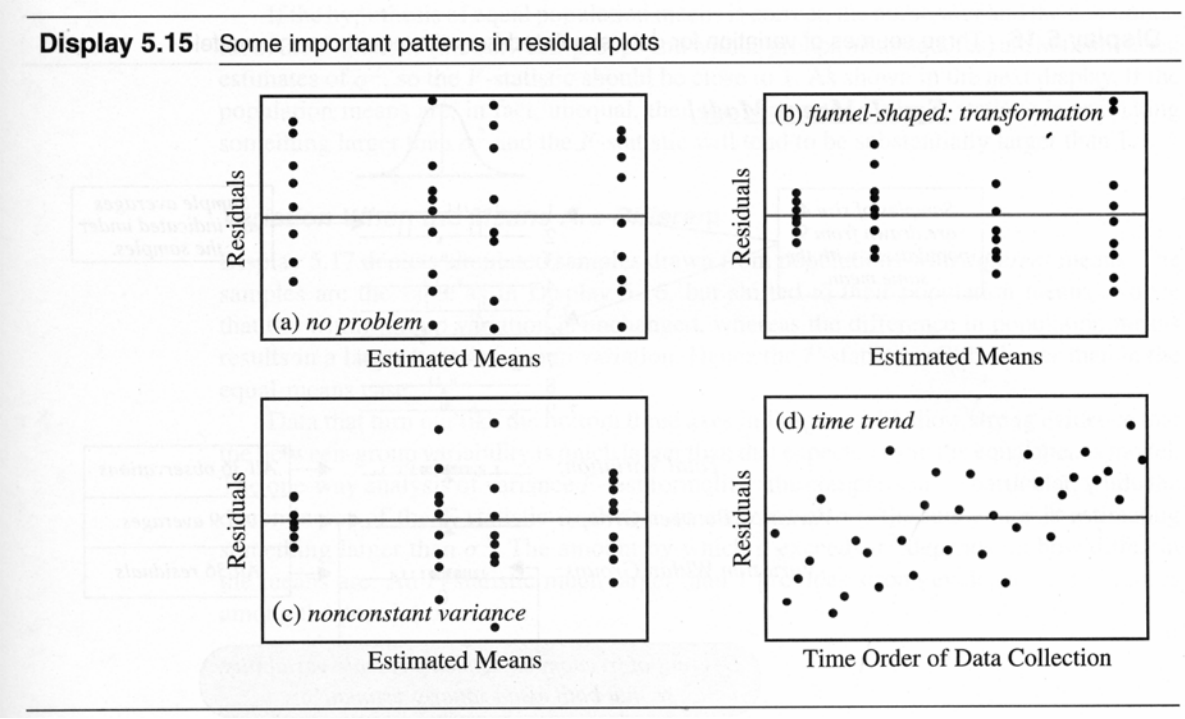
3) Homogeneity of variance – since the variances from the different treatments are pooled into a single variance this assumption is needed. Some newer analyses allow the fitting of different variances to the treatments.

4) Your text book mentions the need to avoid severe outliers. This is true, but I do not consider it a separate assumption. I think of it as a disturbance in the assumption of a normal distribution.

The term "robust" refers to the fact that the tests of hypothesis tend to perform well even if the assumptions are violated to some degree, as long as the violation is not severe.

Checking the assumptions with SAS

There are several applications that will help to check if the assumptions have been met. The first is a plot of the residuals on the means.



From these we look for the following:

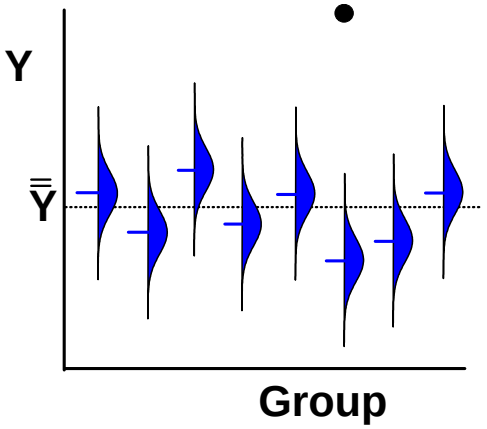
1) If there is no problem the residuals should appear to be scattered at random about zero. Residual plots often have a reference line drawn at zero.

2) One potential problem is non-homogeneous variance. If this is present we may see increasing (or decreasing) scatter among the points at larger values of the means.

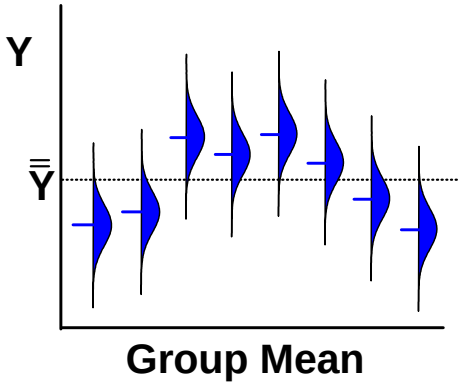
3) Another indicator of non-homogeneous variance occurs when we have large differences in the scatter of the points at different values of the mean. It is not necessary that the variance only increase or decrease with the mean.

4) If the data is truly independent the plot of the data in the order taken, or in chronological order, should also appear to be random scatter. Certain problems, such as learning processes and other factors that may occur over time, can cause a pattern (increasing or decreasing) to occur over time. Randomization or inclusion of additional variables may help solve this type of problem,

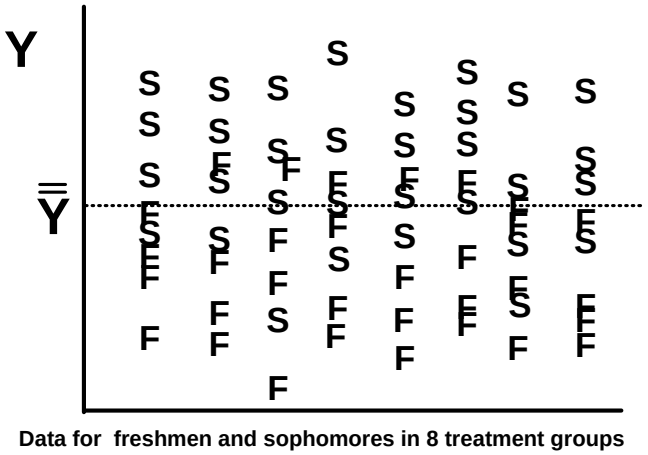
Other problems that can be detected with residual plots include the following:



1) Outliers



2) Curvature (more applicable to regression)



3) Need for additional variables