

# Chapter 11

## Analysis of Variance (One-Way)

We now develop a statistical procedure for comparing the means of two or more groups, known as analysis of variance or ANOVA. These groups might be the result of an experiment in which organisms are exposed to different treatments. Alternately, the groups might be different species or different age classes of the same species, populations in different locations, or different genetic families. The test works by comparing the variance among the group means to the variance of the observations within each group – if the variance among group means is large (implying differences in their means) relative to the variance within groups, the test is significant. This chapter will examine tests for one-way ANOVA, in which a single factor like a treatment affects the observations. More complex designs are possible in which several factors may influence the observations and may also interact (see Chapter 14 and 19).

What do the data look like for a one-way ANOVA design? Suppose we are interested in trapping bark beetles (Coleoptera: Curculionidae: Scolytinae) using different chemical baits, which could involve the beetle's sex pheromones or odors of the trees they colonize. Suppose that three different baits (A, B, and C), with  $a = 3$  denoting the number of treatments. The baits are deployed on traps in the forest, with  $n = 5$  replicate traps for each bait type. A typical experimental design would establish 15 traps in the forest, and then randomly assign a bait to each trap. After a period of time, the traps would be checked and the number of insects caught in each trap recorded (Table 11.1). Because the data are counts, it would not be normally

distributed but more likely have a Poisson or negative binomial distribution (see Chapter 5). However, it is often possible to **transform** count data to have a distribution closer to the normal by taking the square root or log of the counts (see Chapter 15). The third column in Table 11.1 shows the count data after applying a log transformation. The notation  $Y_{ij}$  is often used to refer to the observations in ANOVA designs. The  $i$  subscript refers to the group or treatment, while  $j$  is the observation within the treatment. For example,  $Y_{13}$  refers to the third observation in the first treatment, which is 2.41.

Another one-way ANOVA design for bark beetles might simply look at variability in their density across sites. Suppose there is a large collection of study sites, and we randomly select five sites for trapping. Five traps are deployed at each of the five sites and the number of beetles caught per trap is recorded. Data for a study of this type are listed below, also with a log transformation (Table 11.2). There appears to be substantial variability in beetle abundance across sites, with Site 4 having very high beetle catches, while Site 5 has low ones.

The data sets presented in this section represent **balanced designs**, because there are the same number of replicates for each treatment or group. An **unbalanced design** would have an unequal number of replicates, possibly very unequal. We will present tests and theory for balanced designs in this chapter, because this greatly simplifies the formulas and equations. However, these results can be readily extended to unbalanced designs, and unbalanced designs require no changes in the SAS programs presented.

Table 11.1: Example 1 - Bark beetles captured in a trapping experiment comparing the attraction to different baits. There were three baits (A, B, and C) and five replicate traps per bait treatment. Also shown are the log-transformed counts ( $Y_{ij}$ ) and subscript values ( $i, j$ ), and some preliminary one-way ANOVA calculations.

Treatment	Count	$Y_{ij} =$ $\log_{10}(\text{Count})$	$i$	$j$	$\bar{Y}_i$	$(Y_{ij} - \bar{Y}_i)^2$	$\sum(Y_{ij} - \bar{Y}_i)^2$
A	373	2.57	1	1		0.0441	
A	126	2.10	1	2		0.0676	
A	255	2.41	1	3	2.3600	0.0025	0.2110
A	138	2.14	1	4		0.0484	
A	379	2.58	1	5		0.0484	
B	25	1.40	2	1		0.0999	
B	64	1.81	2	2		0.0088	
B	62	1.79	2	3	1.7160	0.0055	0.1325
B	71	1.85	2	4		0.0180	
B	54	1.73	2	5		0.0002	
C	449	2.65	3	1		0.1832	
C	249	2.40	3	2		0.0317	
C	69	1.84	3	3	2.2220	0.1459	0.4581
C	199	2.30	3	4		0.0061	
C	84	1.92	3	5		0.0912	

Table 11.2: Example 2 - Bark beetles captured in a trapping study comparing their abundance at five randomly chosen study sites. There were five replicate traps per site. Also shown are the log-transformed counts ( $Y_{ij}$ ) and subscript values ( $i, j$ ), and some preliminary one-way ANOVA calculations.

Site	Count	$Y_{ij} =$ $\log_{10}(\text{Count})$	$i$	$j$	$\bar{Y}_i$	$(Y_{ij} - \bar{Y}_i)^2$	$\sum(Y_{ij} - \bar{Y}_i)^2$
1	137	2.14	1	1		0.0164	
1	101	2.00	1	2		0.0001	
1	113	2.05	1	3	2.0120	0.0014	0.1598
1	48	1.68	1	4		0.1102	
1	155	2.19	1	5		0.0317	
2	156	2.19	2	1		0.0784	
2	165	2.22	2	2		0.0625	
2	652	2.81	2	3	2.4700	0.1156	0.4730
2	179	2.25	2	4		0.0484	
2	757	2.88	2	5		0.1681	
3	278	2.44	3	1		0.0376	
3	197	2.29	3	2		0.0019	
3	95	1.98	3	3	2.2460	0.0708	0.3419
3	395	2.60	3	4		0.1253	
3	83	1.92	3	5		0.1063	
4	2540	3.40	4	1		0.4956	
4	613	2.79	4	2		0.0088	
4	200	2.30	4	3	2.6960	0.1568	0.7600
4	251	2.40	4	4		0.0876	
4	390	2.59	4	5		0.0112	
5	18	1.26	5	1		0.0044	
5	16	1.20	5	2		0.0000	
5	11	1.04	5	3	1.1940	0.0237	0.0459
5	21	1.32	5	4		0.0159	
5	14	1.15	5	5		0.0019	

## 11.1 ANOVA models

We now examine the statistical models that are used in one-way ANOVA. There are two models for one-way ANOVA, known as fixed or random effects models, but sometimes called Model I and II. This classification is based on how the groups in the design are defined or generated. We begin by defining fixed and random effects, then present the statistical models and hypotheses for each type.

### 11.1.1 Fixed and random effects

For groups generated by different treatments in an experiment, or purposely chosen groups of organisms such as different species, sexes, or ages, the groups are classified as **fixed effects**. They are called fixed effects because these groups are the only ones of interest to the investigator, and the only ones on which a statistical inference can be made (Littell et al. 1996, McCulloch and Searle 2001). They are also incorporated in statistical models as fixed parameters. Groups that are generated by a process of random sampling are classified as a **random effects** (Littell et al. 1996, McCulloch and Searle 2001). For example, suppose we want to examine the fish populations in a large number of lakes, and are interested in how body length varies across lakes. If we randomly sample the lakes to be examined, from a large collection of lakes, then lake would be classified as a random effect. In many genetic experiments, families are chosen at random from a larger collection of families, making family a random effect. Random effects are incorporated in statistical models as random variables, typically with a normal distribution.

These definitions suggest a simple test for fixed vs. random effects – if the groups are a random sample from a large collection you have a random effect, otherwise a fixed effect. Although it is usually possible to declare an effect as either fixed or random, in practice it is sometimes difficult to decide. For example, suppose that a particular organism occurs at only a small number of locations. If we randomly select a subset of these locations to sample, seemingly implying a random effect, the overall number of locations is still finite. In this scenario, location may be better classified as a fixed effect.

### 11.1.2 Fixed effects model

Suppose that we want to model the observations in the bark beetle trapping experiment, Example 1, where different baits are used. Recall that the symbol  $Y_{ij}$  stand for the  $j$ th observation in the  $i$ th treatment group, where  $i = 1, 2, 3$  and  $j = 1, 2, 3, 4, 5$ . For example,  $Y_{11} = 2.57$  and  $Y_{12} = 2.10$ , while  $Y_{32} = 2.40$  (see Table 11.1). One commonly used model for such a design is

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (11.1)$$

where  $\mu$  is a parameter setting the grand mean (the overall mean) of the observations,  $\alpha_i$  is the deviation from the grand mean caused by the  $i$ th treatment (McCulloch and Searle 2001), and  $\epsilon_{ij} \sim N(0, \sigma^2)$ . It is usually assumed that  $\sum \alpha_i = 0$ , i.e., the treatment effect terms sum to zero. The  $\epsilon_{ij}$  term represents random departures from the mean value for the  $i$ th treatment, due to natural variability among the observations. The  $\epsilon_{ij}$  values are also assumed to be independent (Chapter 4). In practice, these parameters would be unknown but could be estimated from the data. The same model can be used to describe the observations for experiments with any number of treatments (any  $a$  value) as well as replicates per treatments (any  $n$ ), as well as experiments where the number of observations is unequal among treatments.

It follows that for the  $i$ th treatment,  $E[Y_{ij}] = \mu + \alpha_i$  and  $Var[Y_{ij}] = \sigma^2$ , using the rules for expected values and variances. Thus, for the  $i$ th treatment we have  $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$ . We can illustrate how the different parameters work in this model with a diagram that plots the distribution for each group. Suppose that we want to model an experiment similar to the bark beetle trapping one, with  $a = 3$  treatments. Suppose that  $\mu = 2.1$ ,  $\alpha_1 = 0.25$ ,  $\alpha_2 = -0.40$ , and  $\alpha_3 = 0.15$ , with  $\sigma^2 = 0.1$ . Fig. 11.1 shows the distribution of the observations in each treatment group. Note that the means for treatment 1 and 3 are shifted upward from the grand mean by their positive values of  $\alpha_i$ , while the mean for treatment 2 is shifted downward by its negative value. The distribution for each treatment has the same variance, namely  $\sigma^2 = 0.1$ .

The usual objective in ANOVA is to test whether the means of the different groups are significantly different, implying there is treatment or group effect. In terms of the fixed effects model, this amounts to testing whether the  $\alpha_i$  values are significantly different from zero, because it is these parameters that produce shifts in the group means from the grand mean. More formally, we are interested in testing the null hypothesis  $H_0 : \text{all } \alpha_i = 0$ .

Under  $H_0$ , all the groups have the same mean  $\mu$  because the  $\alpha_i$  terms are zero (Fig. 11.2). The alternative hypothesis would be  $H_1$  : some  $\alpha_i \neq 0$ , i.e., there is some treatment effect on some (perhaps all) groups (Fig. 11.1). We will discuss how this null hypothesis is actually tested later in the chapter.

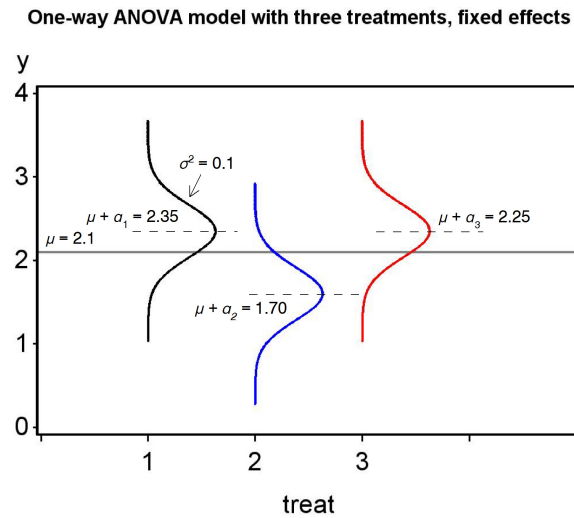


Figure 11.1: Fixed effects model for one-way ANOVA, under  $H_1$  : some  $\alpha_i \neq 0$ .

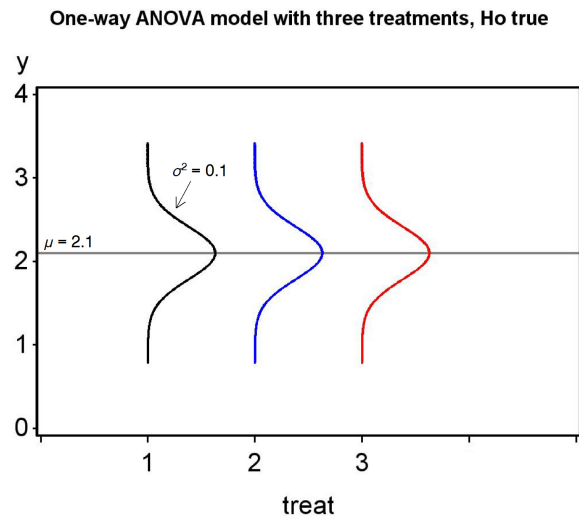


Figure 11.2: Fixed effects model for one-way ANOVA, under  $H_0$  : all  $\alpha_i = 0$ .



### 11.1.3 Random effects model

Suppose that we now want to model the variability in bark beetle abundance across different sites, such as in the Example 2 study. Let  $Y_{ij}$  stand for the  $j$ th observation at the  $i$ th sampled site, with  $i = 1, 2, 3, 4, 5$  and  $j = 1, 2, 3, 4, 5$ . We have  $Y_{11} = 4.92$ ,  $Y_{12} = 4.62$ , and so forth (see Table 11.2). A common statistical model for this design is

$$Y_{ij} = \mu + A_i + \epsilon_{ij} \quad (11.2)$$

where  $\mu$  is again a parameter setting the grand mean (the overall mean) of the observations, with  $A_i$  a random deviation from the grand mean due to the  $i$ th site (McCulloch and Searle 2001), and  $\epsilon_{ij} \sim N(0, \sigma^2)$ . It is assumed that  $A_i$  is normally distributed with mean zero and variance  $\sigma_A^2$ , or  $A_i \sim N(0, \sigma_A^2)$ . Note that in the random effects model the group effect is indeed a random variable, one whose variance is unknown but can be estimated from the data. The variances  $\sigma_A^2$  and  $\sigma^2$  are collectively called the **variance components** of the model.

For the  $i$ th group sampled, it can be shown that  $E[Y_{ij}] = \mu + A_i$  and  $Var[Y_{ij}] = \sigma^2$ , using the rules for expected values and variances. Thus, for the  $i$ th treatment we have  $Y_{ij} \sim N(\mu + A_i, \sigma^2)$ . Because the  $A_i$  values are themselves random quantities, however, the expected value is itself a random quantity and would differ for each group sampled. We again illustrate how the model works using a diagram showing the distribution for each group. Suppose that we want to model a study similar to the second bark beetle one (Table 11.2), with  $a = 5$  sites randomly selected from a larger collection of sites. Suppose that  $\mu = 2.1$  and  $\sigma^2 = 0.1$ . The first time we did this study, we might see a pattern like Fig. 11.3. If we redid the study and randomly selected another five sites, we would get a different pattern (Fig. 11.4). This illustrates that this model is not static like the fixed effects one, but instead would vary with the sites actually sampled. In the random effects model, we are usually interested in testing whether the variance of  $A_i$  is zero vs. greater than zero, or  $H_0 : \sigma_A^2 = 0$  vs.  $H_1 : \sigma_A^2 > 0$ . Under  $H_0 : \sigma_A^2 = 0$ , all the  $A_i$  values must be zero (to give  $\sigma_A^2 = 0$ ), and so all the groups have the same mean  $\mu$ . A plot of the model under  $H_0$  would therefore be similar to Fig. 11.2. This null hypothesis is tested in the same way as the one for the fixed effects model (see below).

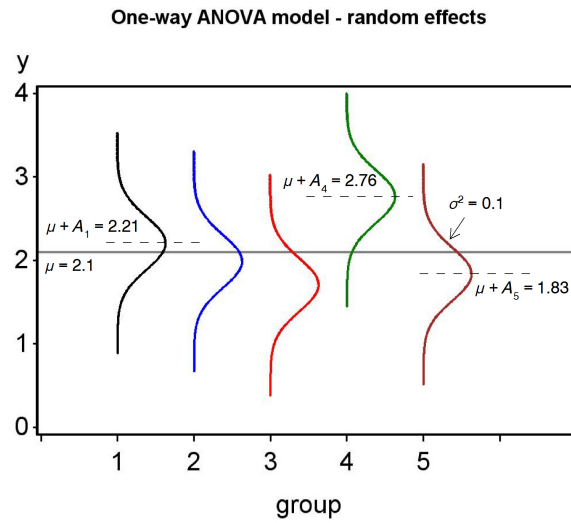


Figure 11.3: Random effects model for one-way ANOVA, for the first time sites are sampled.

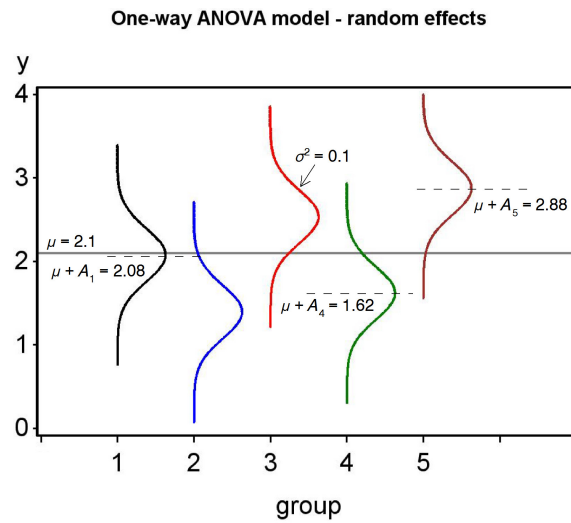


Figure 11.4: Random effects model for one-way ANOVA, for the second time sites are sampled.

## 11.2 Hypothesis testing for ANOVA

We now develop a statistical test for the null hypotheses in both fixed and random effects models, either  $H_0 : \text{all } \alpha_i = 0$  or  $H_0 : \sigma_A^2 = 0$ . We will first present the test and explain how it works in terms of different estimates of the variance, then later show it is another example of a likelihood ratio test.

### 11.2.1 Sums of squares and mean squares

Suppose the data are described by a fixed effects model, for which the hypotheses are  $H_0 : \text{all } \alpha_i = 0$  vs.  $H_1 : \text{some } \alpha_i \neq 0$ . It is clear that if  $H_1$  is true, then the observations for the different groups will be shifted from the grand mean, as shown in Fig. 11.1, and in particular  $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$  for each group. For a random effects model, we have  $H_0 : \sigma_A^2 = 0$  vs.  $H_1 : \sigma_A^2 > 0$ . If  $H_1$  is true, we would also expect the observations for the different groups to be shifted away from the grand mean (Fig. 11.3), and in particular  $Y_{ij} \sim N(\mu + A_i, \sigma^2)$ . How can we estimate this shift in actual data? How large must this shift be to be judged statistically significant?

We begin by calculating the means for each group using the data. These are labeled as  $\bar{Y}_i$  and are called group means. The ‘.’ subscript implies the mean was calculated using all the observations in that group ( $j = 1, 2, \dots, n$ ). We then calculate the mean of the group means, called the grand mean and labeled as  $\bar{\bar{Y}}$ . If the  $i$ th group is shifted from the grand mean, we can measure this shift using the quantity  $\bar{Y}_i - \bar{\bar{Y}}$ . In fact, this quantity estimates  $\alpha_i$  for the  $i$ th group, and so is a direct measure of any group effect (see section on maximum likelihood estimation below). If these quantities are small then this suggests  $H_0$  might be true, whereas if they are large this provides evidence for  $H_1$ . We can obtain a single measure of these shifts by squaring and summing them across all groups, to obtain a quantity called the sum of squares among groups or  $SS_{among}$ , because it measures variation in the observations among groups:

$$SS_{among} = n \sum_{i=1}^a (\bar{Y}_i - \bar{\bar{Y}})^2. \quad (11.3)$$

Note the sample size  $n$  in this expression, which we will justify below. To make this quantity more concrete, we will calculate  $SS_{among}$  for Example 1, the bark beetle trapping experiment. We first calculate the sample mean for

each group for the log-transformed data, as shown in Table 11.1. Then, the grand mean is estimated using the mean of these means, or

$$\bar{\bar{Y}} = \frac{\sum_{i=1}^a \bar{Y}_i}{a} = \frac{2.3600 + 1.7160 + 2.2220}{3} = \frac{6.2980}{3} = 2.0993. \quad (11.4)$$

We then have

$$SS_{among} = n \sum_{j=1}^a (\bar{Y}_i - \bar{\bar{Y}})^2 \quad (11.5)$$

$$= 5 [(2.3600 - 2.0993)^2 + (1.7160 - 2.0993)^2 + (2.2220 - 2.0993)^2] \quad (11.6)$$

$$= 5 [0.0680 + 0.1469 + 0.0151] \quad (11.7)$$

$$= 1.1500 \quad (11.8)$$

$SS_{among}$  has  $a - 1$  degrees of freedom, where  $a$  is the number of groups. There are  $a - 1$  degrees of freedom because there are  $a$  terms of the form  $\bar{Y}_i - \bar{\bar{Y}}$  in the sum of squares, but these sum to zero so there are really only  $a - 1$  independent terms (similar to the  $n - 1$  degrees of freedom for the sample variance  $s^2$ ). The next step is to convert  $SS_{among}$  to a sample variance, dividing it by  $a - 1$ . This quantity is called the mean square among groups:

$$MS_{among} = \frac{SS_{among}}{a - 1} = \frac{n \sum_{j=1}^a (\bar{Y}_i - \bar{\bar{Y}})^2}{a - 1}. \quad (11.9)$$

For the bark beetle experiment, we find that

$$MS_{among} = \frac{SS_{among}}{a - 1} = \frac{1.1500}{3 - 1} = 0.5750. \quad (11.10)$$

So, what variance does  $MS_{among}$  estimate? If  $H_0$  is true and there are no group effects, we would expect  $\bar{Y}_i$  to have a variance of  $\sigma^2/n$ , because it is a sample mean composed of  $n$  observations in the  $i$ th group (which have a variance of  $\sigma^2$ ).  $MS_{among}$  estimates this variance multiplied by  $n$ , because of the  $n$  term in numerator, and so actually estimates  $n\sigma^2/n = \sigma^2$ . On the other hand, if  $H_1$  is true then there are group effects, and we would expect the group means to be shifted away from the grand mean. This should increase the size of  $MS_{among}$ . **Thus,  $MS_{among}$  estimates  $\sigma^2$  if  $H_0$  is true but becomes larger if  $H_1$  is true.**

Expected values provide another way to understand how group effects influence  $MS_{among}$ . For the fixed effects model, we have

$$E[MS_{among}] = \frac{n \sum \alpha_i^2}{a - 1} + \sigma^2, \quad (11.11)$$

while

$$E[MS_{among}] = n\sigma_A^2 + \sigma^2. \quad (11.12)$$

for the random effects model (Winer et al. 1991). **Note how  $E[MS_{among}]$  increases as the group effects increase, above a baseline level set by  $\sigma^2$ .** For the fixed effects model, larger groups effects would involve larger values of  $\alpha_i$ , while for the random effects model a larger value of  $\sigma_A^2$ .

We next develop an estimate of the variance  $\sigma^2$  that is free of any group effects. This variance estimate is based on a quantity called the sum of squares within groups or  $SS_{within}$ , because it measures variation of the observations within each group. It is defined by the formula

$$SS_{within} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2 \quad (11.13)$$

$$= \sum_{j=1}^n (Y_{1j} - \bar{Y}_{1.})^2 + \dots + \sum_{j=1}^n (Y_{aj} - \bar{Y}_{a.})^2. \quad (11.14)$$

It has  $a(n - 1)$  degrees of freedom, because there are  $a$  sum of squares terms each with  $n - 1$  degrees of freedom. We can obtain an estimate of  $\sigma^2$  by dividing this sum of squares by its degrees of freedom, to obtain the mean square within groups:

$$MS_{within} = \frac{SS_{within}}{a(n - 1)} = \frac{\sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2}{a(n - 1)}. \quad (11.15)$$

This quantity estimates  $\sigma^2$  because it simply averages estimates of  $\sigma^2$  for

each group. With some rearrangement, we can write  $MS_{within}$  as

$$MS_{within} = \frac{\sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2}{a(n-1)} \quad (11.16)$$

$$= \frac{\sum (Y_{1j} - \bar{Y}_{1.})^2 + \dots + \sum (Y_{aj} - \bar{Y}_{a.})^2}{a(n-1)} \quad (11.17)$$

$$= \frac{\frac{\sum (Y_{1j} - \bar{Y}_{1.})^2}{n-1} + \dots + \frac{\sum (Y_{aj} - \bar{Y}_{a.})^2}{n-1}}{a} \quad (11.18)$$

$$= \frac{s_1^2 + \dots + s_a^2}{a}. \quad (11.19)$$

Each term in the numerator of this expression is the sample variance  $s^2$  for each group, which is then averaged across all groups to yield an overall or **pooled** estimate of  $\sigma^2$ . The word ‘pooled’ in statistics often indicates a combined estimate of a variance. It can also be shown that  $E[MS_{within}] = \sigma^2$ , regardless of any group effects.

We now calculate  $MS_{within}$  for the bark beetle experiment. We first need to calculate the quantity  $(Y_{ij} - \bar{Y}_{i.})^2$  for the observations in each group and then sum these for each group (see Table 11.1). Summing these quantities in turn across all groups, we obtain

$$SS_{within} = 0.2110 + 0.1325 + 0.4581 = 0.8016. \quad (11.20)$$

$$(11.21)$$

We then have

$$MS_{within} = \frac{SS_{within}}{a(n-1)} = \frac{0.8016}{3(5-1)} = 0.0668. \quad (11.22)$$

$$(11.23)$$

There is one more sum of squares that can be calculated in one-way ANOVA, the total sum of squares. It is defined as

$$SS_{total} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y})^2. \quad (11.24)$$

It measures the variability of the observations around the grand mean of the data ( $\bar{Y}$ ) and has  $an - 1$  degrees of freedom. Applying this formula to the Example 1 data set, we obtain  $SS_{total} = 1.9516$  after much calculation.

An interesting feature of the sum of squares is that they add to the total sum of squares, as do the degrees of freedom. In particular, we have

$$SS_{among} + SS_{within} = SS_{total} \quad (11.25)$$

and

$$(a - 1) + a(n - 1) = an - 1. \quad (11.26)$$

Thus, the sum of squares and degrees of freedom can be neatly partitioned into components corresponding to among group and within group variation. We will illustrate this relationship further in the section below on ANOVA tables.

### 11.2.2 $F$ statistic and distribution

We next describe the statistic used to test  $H_0 : \text{all } \alpha_i = 0$  for the fixed effect model, and  $H_0 : \sigma_A^2 = 0$  for the random effects one. It is simply the ratio of  $MS_{among}$  and  $MS_{within}$ , or

$$F_s = \frac{MS_{among}}{MS_{within}}. \quad (11.27)$$

If  $H_0$  is true for either model, both  $MS_{among}$  and  $MS_{within}$  estimate  $\sigma^2$  and we would expect their ratio,  $F_s$ , to be small and on the order of one. However, if  $H_0$  is false and  $H_1$  is true, we would expect  $MS_{among}$  to become larger and  $F_s$  to increase. We would therefore reject  $H_0$  for large values of  $F_s$ .

To complete our testing procedure and find  $P$  values, we need to know the distribution of  $F_s$  under  $H_0$ . It turns out this statistic has an  $F$  distribution under  $H_0$ , whose shape and location is governed by two parameters, the degrees of freedom for  $MS_{among}$  and  $MS_{within}$ . These are called the numerator and denominator degrees of freedom, which we abbreviate as  $df_1$  and  $df_2$ . In particular, for one-way ANOVA we have  $df_1 = a - 1$  and  $df_2 = a(n - 1)$ . Figure 11.5 shows the  $F$  distribution for three different sets of parameter values. Note that distribution can have a maximum at  $y = 0$  for small values of  $df_1$ , while larger values of  $df_2$  decrease the probability in the right tail of the distribution.

Table F gives the quantiles of the  $F$  distribution for different values of the degrees of freedom and the cumulative probability  $p$ . Statistical tests that make use of the  $F$  distribution are typically called  $F$  tests.

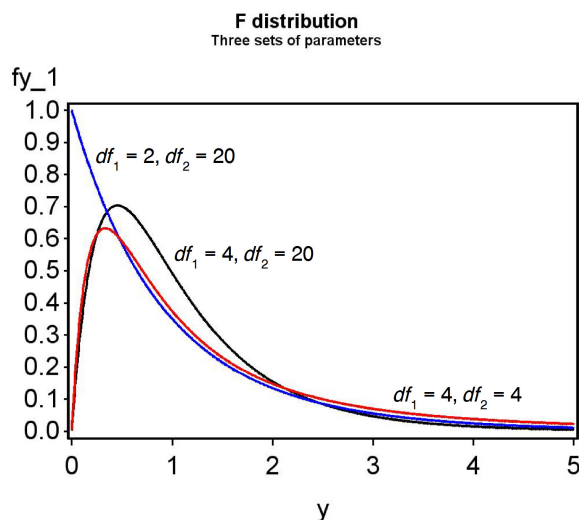


Figure 11.5: The  $F$  distribution for three different sets of parameter values

Calculating the test statistic  $F_s$  for the bark beetle experiment, we have

$$F_s = \frac{MS_{among}}{MS_{within}} = \frac{0.5750}{0.0668} = 8.6078, \quad (11.28)$$

$$(11.29)$$

with  $df_1 = a - 1 = 3 - 1 = 2$  and  $df_2 = a(n - 1) = 3(5 - 1) = 12$ .

As with previous tests, we seek acceptance and rejection regions for a particular value of  $\alpha$ , the Type I error rate. In particular, we seek a quantity  $c_{\alpha, df_1, df_2}$  such that

$$P[0 < F_s < c_{\alpha, df_1, df_2}] = 1 - \alpha. \quad (11.30)$$

The region is of this form because the test is designed to reject  $H_0$  for large values of  $F_s$ , and accept it for small ones. To find  $c_{\alpha, df_1, df_2}$ , we look in Table F for the column corresponding to  $1 - p = \alpha$ , for the appropriate degrees of freedom. The acceptance region would therefore be  $(0, c_{\alpha, df_1, df_2})$ , and we would reject  $H_0$  if  $F_s$  lies outside this region.

For  $\alpha = 0.05$ ,  $df_1 = 2$ , and  $df_2 = 12$ , we see from Table F that  $c_{0.05, 2, 12} = 3.885$ . Our acceptance region is therefore  $(0, 3.885)$ , and we reject  $H_0$  at the  $\alpha = 0.05$  level if  $F_s \geq 3.885$  (Fig. 11.6). We see this is the case because  $F_s = 8.6078 > 3.885$ . We can continue this process for increasingly smaller



$\alpha$  and eventually find that for  $\alpha = 0.005$  we can still reject  $H_0$ , but not for  $\alpha = 0.001$ . We therefore have  $P < 0.005$  for this test, because  $\alpha = 0.005$  is the smallest value of  $\alpha$  for which we can reject  $H_0$  (see Chapter 10). An  $F$  test in ANOVA would often be reported as follows: ‘There was a highly significant difference among the different baits in the number of bark beetles trapped ( $F_{2,12} = 8.6078, P < 0.005$ ).’ Note that the degrees of freedom are given as subscripts.

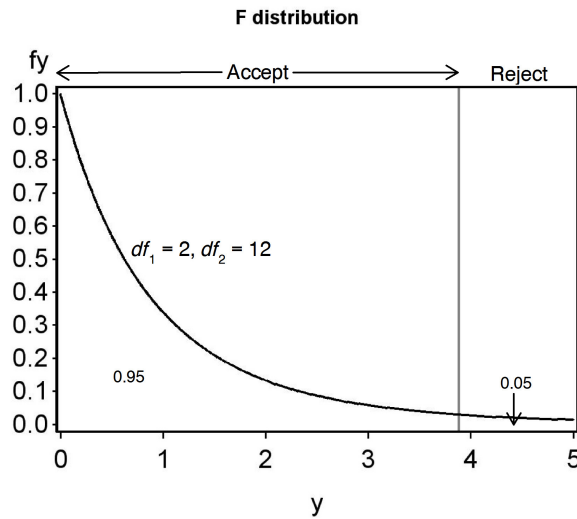


Figure 11.6: Acceptance and rejection regions for  $\alpha = 0.05$

### 11.2.3 ANOVA tables

We can organize the different sum of squares and mean squares into an ANOVA table. It lists the different sources of variation in the data (among, within, and total), their degrees of freedom, sums of squares and mean squares, and then the  $F$  statistic and its  $P$  value. Table 11.3 shows the general layout of such a table for one-way ANOVA designs, while Table 11.4 gives the results for the Example 1 analysis. Note the additive relationship for the degrees of freedom and sum of squares.

Table 11.3: General ANOVA table for one-way designs with  $a$  groups and  $n$  observations per group, showing formulas for different mean squares and the  $F$  test.

Source	$df$	Sum of squares	Mean square	$F_s$
Among	$a - 1$	$SS_{among} = n \sum_{i=1}^a (\bar{Y}_{i\cdot} - \bar{Y})^2$	$MS_{among} = SS_{among}/(a - 1)$	$MS_{among}/MS_{within}$
Within	$a(n - 1)$	$SS_{within} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i\cdot})^2$	$MS_{within} = SS_{within}/a(n - 1)$	
Total	$an - 1$	$SS_{total} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y})^2$		

Table 11.4: ANOVA table for the Example 1 data set, including a  $P$  value for the test.

Source	$df$	Sum of squares	Mean square	$F_s$	$P$
Among	2	1.1500	0.5750	8.6078	< 0.005
Within	12	0.8016	0.0668		
Total	14	1.9516			

### 11.2.4 One-way ANOVA for Example 1 - SAS demo

The same calculations for the bark beetle experiment can be carried out in SAS using `proc glm` (SAS Institute Inc. 2014a). This procedure is primarily intended for fixed effects ANOVA models, with `proc mixed` the best choice for random effects models. However, the  $F$  test would be the same in either procedure.

The SAS program for one-way ANOVA is a bit more complicated than previous programs, so we will examine it a section at a time. The first step is to read in the observations using a data step, with one variable denoting the treatment (`treat`) and a second the number of beetles captured (`count`). As discussed earlier, it is common to log-transform count data, and so we generate a variable `y` that is the log 10 (*log* base 10) of `count`. The data step is followed by a `print` statement to print the data set. See section below.

```
* bark_beetle_experiment.sas;
options pageno=1 linesize=80;
goptions reset=all;
title "One-way ANOVA for bark beetle trapping experiment";
data bark_beetle;
    input treat $ count;
    * Apply transformations here;
    y = log10(count);
    datalines;
A   373
A   126
A   255

etc.

C   199
C    84
;
run;
* Print data set;
proc print data=bark_beetle;
run;
```

We next plot the data using the SAS procedure `gplot` (SAS Institute Inc. 2014b). The basic idea is to plot, for each treatment group, the individual data points along with their mean ( $\bar{Y}$ )  $\pm$  one standard error ( $s/\sqrt{n}$ ). The `plot` statement tells `gplot` to plot the variable `y` on the  $y$ -axis and `treat` on the  $x$ -axis of the plot. The appearance of the points is controlled by the

`symbol1` statement, which among other things specifies that the points be plotted along with their means  $\pm$  one standard error, with the means joined by a line, using the option `i=std1mjt`. Other options in the `symbol` statement control the type and size of the points, and line width. The `vaxis=axis1` and `haxis=axis1` options control the visual appearance of the  $x$ - and  $y$ -axes. See below.

```
* Plot means, standard errors, and observations;
proc gplot data=bark_beetle;
    plot y*treat=1 / vaxis=axis1 haxis=axis1;
    symbol1 i=std1mjt v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;
```

The next section of the program conducts the one-way ANOVA and  $F$  test using `proc glm`. The `class` statement tells SAS that the variable `treat` is the one that defines different groups in the ANOVA (see listing below). The `model` statement basically tells SAS the form of the ANOVA model. Recall that the model for fixed effects one-way ANOVA is given by the equation

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}. \quad (11.31)$$

If we equate  $Y_{ij}$  with `y`, and  $\alpha_i$  with `treat`, we see there are similarities between the fixed effects model and the SAS `model` statement. In fact, SAS assumes you want a grand mean  $\mu$  unless otherwise specified, as well as the error term  $\epsilon_{ij}$ . As we examine more complex ANOVA models in later chapters, we will see there is nearly a one-to-one correspondence between these models and the corresponding SAS `model` statement.

```
* One-way ANOVA with all fixed effects;
proc glm data=bark_beetle;
    class treat;
    model y = treat;
    * Calculate means for each group;
    means treat;
    output out=resids p=pred r=resid;
run;
```

The `means` statement causes `glm` to calculate means for each `treat` group. The other statements generate graphs that are used to examine some of the assumptions of ANOVA – we will defer their discussion to later chapters.

The complete SAS program and output are listed below. The output shows the same  $F$  test in three different locations within the `proc glm` output.

The first is in a format resembling an ANOVA table, and then two other times corresponding to Type I and III sums of squares. These are different ways of calculating the sums of squares and tests, with Type III sums of squares more generally useful for ANOVA designs. For one-way ANOVA the results are the same, and we see that there was a highly significant difference among groups ( $F_{2,12} = 8.60, P = 0.0048$ ). Inspection of the graph and means suggests that treatment A caught the most beetles, followed by C and then B.

---

SAS Program

---

```

* bark_beetle_experiment.sas;
options pageno=1 linesize=80;
goptions reset=all;
title "One-way ANOVA for bark beetle trapping experiment";
data bark_beetle;
    input treat $ count;
    * Apply transformations here;
    y = log10(count);
    datalines;
A   373
A   126
A   255
A   138
A   379
B    25
B    64
B    62
B    71
B    54
C   449
C   249
C    69
C   199
C    84
;
run;
* Print data set;
proc print data=bark_beetle;
run;
* Plot means, standard errors, and observations;
proc gplot data=bark_beetle;
    plot y*treat=1 / vaxis=axis1 haxis=axis1;
    symbol1 i=std1mjt v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;

```

```
* One-way ANOVA with all fixed effects;
proc glm data=bark_beetle;
    class treat;
    model y = treat;
    * Calculate means for each group;
    means treat;
    output out=resids p=pred r=resid;
run;
goptions reset=all;
title "Diagnostic plots to check ANOVA assumptions";
* Plot residuals vs. predicted values;
proc gplot data=resids;
    plot resid*pred=1 / vaxis=axis1 haxis=axis1;
    symbol1 v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;
* Normal quantile plot of residuals;
proc univariate noprint data=resids;
    qqplot resid / normal waxis=3 height=4;
run;
quit;
```

---

---

 SAS Output
 

---

One-way ANOVA for bark beetle trapping experiment 1  
 09:32 Tuesday, August 31, 2010

Obs	treat	count	y
1	A	373	2.57171
2	A	126	2.10037
3	A	255	2.40654
4	A	138	2.13988
5	A	379	2.57864
6	B	25	1.39794
7	B	64	1.80618
8	B	62	1.79239
9	B	71	1.85126
10	B	54	1.73239
11	C	449	2.65225
12	C	249	2.39620
13	C	69	1.83885
14	C	199	2.29885
15	C	84	1.92428

One-way ANOVA for bark beetle trapping experiment 2  
 09:32 Tuesday, August 31, 2010

The GLM Procedure

Class Level Information

Class	Levels	Values
treat	3	A B C

Number of Observations Read	15
Number of Observations Used	15

One-way ANOVA for bark beetle trapping experiment 3  
 09:32 Tuesday, August 31, 2010

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.14818176	0.57409088	8.60	0.0048
Error	12	0.80114853	0.06676238		
Corrected Total	14	1.94933029			

R-Square	Coeff Var	Root MSE	y Mean
0.589013	12.30880	0.258384	2.099182

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	2	1.14818176	0.57409088	8.60	0.0048

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	2	1.14818176	0.57409088	8.60	0.0048

One-way ANOVA for bark beetle trapping experiment 4  
09:32 Tuesday, August 31, 2010

## The GLM Procedure

Level of treat	N	Mean	Std Dev
A	5	2.35942757	0.22948244
B	5	1.71603276	0.18282085
C	5	2.22208543	0.33793710

---



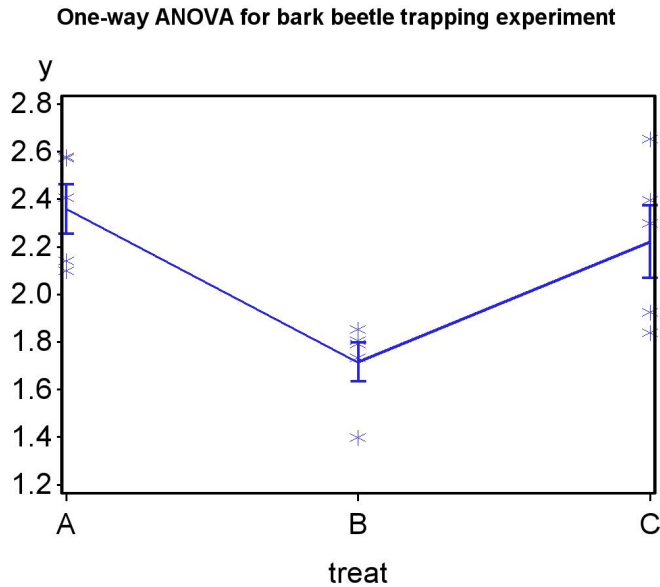


Figure 11.7: Means and standard errors for three treatments

### 11.2.5 One-way ANOVA for Example 2 - sample calculation

We will conduct an  $F$  test for our second data set, involving a study of bark beetles trapped at five different sites ( $a = 5$ ) selected at random from a collection of sites, with five traps per site ( $n = 5$ ). This implies a random effects model, and we are therefore interested in testing  $H_0 : \sigma_A^2 = 0$  vs.  $H_1 : \sigma_A^2 > 0$ . Some preliminary calculations for the  $F$  test are shown in Table 11.2. We first find the mean  $\bar{Y}_i$  for each site, then calculate the grand mean as the average of the site means:

$$\bar{Y} = \frac{\sum_{i=1}^a \bar{Y}_i}{a} \quad (11.32)$$

$$= \frac{2.0120 + 2.4700 + 2.2460 + 2.6960 + 1.1940}{5} \quad (11.33)$$

$$= \frac{10.6180}{5} = 2.1236. \quad (11.34)$$

We then have

$$SS_{among} = n \sum_{j=1}^a (\bar{Y}_{i.} - \bar{\bar{Y}})^2 \quad (11.35)$$

$$= 5 [(2.0120 - 2.1236)^2 + \dots + (1.1940 - 2.1236)^2] \quad (11.36)$$

$$= 5 [0.0125 + 0.1200 + 0.0150 + 0.3276 + 0.8642] \quad (11.37)$$

$$= 6.6965 \quad (11.38)$$

We next calculate  $MS_{among}$ :

$$MS_{among} = \frac{SS_{among}}{a-1} = \frac{6.6965}{5-1} = 1.6741. \quad (11.39)$$

$$(11.40)$$

Now we find  $SS_{within}$ , first calculating  $(Y_{ij} - \bar{Y}_{i.})^2$  for the observations in each group and then summing these for each group (see Table 11.2). Summing these quantities in turn across all groups, we obtain

$$SS_{within} = 0.1598 + 0.4730 + 0.3419 + 0.7600 + 0.0459 = 1.7806. \quad (11.41)$$

$$(11.42)$$

We then have

$$MS_{within} = \frac{SS_{within}}{a(n-1)} = \frac{1.7806}{5(5-1)} = 0.0890. \quad (11.43)$$

$$(11.44)$$

Calculating the test statistic  $F_s$ , we obtain

$$F_s = \frac{MS_{among}}{MS_{within}} = \frac{1.6741}{0.0890} = 18.8101, \quad (11.45)$$

$$(11.46)$$

with  $df_1 = a - 1 = 4 - 1 = 4$  and  $df_2 = a(n - 1) = 5(5 - 1) = 20$ . From Table F, we find  $P < 0.001$ . The variance among sites is highly significant ( $F_{4,12} = 18.8101, P < 0.001$ ).

### 11.2.6 One-way ANOVA for Example 2 - SAS demo

We can carry out the  $F$  test as well as estimate the variance components ( $\sigma_A^2$  and  $\sigma^2$ ) for the random effects model using SAS. The first section of the program involving the `data` step and `gplot` graph is similar to the fixed effects program. The next section of the program fits the random effects model to the data and conducts the  $F$  test, using `proc mixed` (see listing below). As before, the `class` statement tells SAS that the variable `site` is the one that defines different groups in the ANOVA. Now recall that the model for random effects one-way ANOVA is given by the equation

$$Y_{ij} = \mu + A_i + \epsilon_{ij}. \quad (11.47)$$

Note that  $A_i$  corresponds to `site` in the bark beetle study. In `proc mixed`, fixed effects in the model are placed in a `model` statement, while any random effects are listed in a `random` statement (SAS Institute Inc. 2014a). Because our random effects model only has one random effect, `site`, this is listed in the `random` statement. There are no fixed effects in this model, so the `model` statement lists nothing after the equals sign. The option `ddfm=kr` specifies a general method of calculating the degrees of freedom that works well under many circumstances, including more complicated models.

```
* One-way ANOVA with random effects - F test;
proc mixed method=type3 data=bark_beetle;
    class site;
    model y = / ddfm=kr;
    random site;
run;
* One-way ANOVA with random effects - variance components;
proc mixed cl data=bark_beetle;
    class site;
    model y = / ddfm=kr outp=resids;
    random site;
run;
```

Why is `proc mixed` invoked twice in this program? The first one generates the  $F$  statistic for testing  $H_0 : \sigma_A^2 = 0$  vs.  $H_1 : \sigma_A^2 > 0$ , using the option `method=type3`. This is not the default in `proc mixed`, which appears more designed to estimate the variance components in random effects (Littell et al. 1996). If we drop this option, as in the second `proc mixed` statement, we

get only these estimates and no  $F$  test. Confidence intervals for the variance components are requested using the `ci` option. The variance components estimated in the second `proc mixed` using a version of maximum likelihood, the preferred method of estimating these quantities.

The complete SAS program and output are listed below. The variance among sites is highly significant ( $F_{4,12} = 18.77$ ,  $P < 0.0001$ ). The second call to `proc mixed` provide estimates and confidence intervals for the two variance components and confidence intervals. We have  $\hat{\sigma}_A^2 = 0.3174$  for which the 95% confidence interval is (0.1093, 3.1458), and  $\hat{\sigma}^2 = 0.0893$  with confidence interval (0.0523, 0.1863). From these results, we see that the variance among sites is considerably greater than the variance within sites ( $0.3174 > 0.0893$ ).

---

SAS Program

---

```
* bark_beetle_random.sas;
options pageno=1 linesize=80;
goptions reset=all;
title "One-way ANOVA for bark beetle sampling study";
data bark_beetle;
    input site $ count;
    * Apply transformations here;
    y = log10(count);
    datalines;
1  137
1  101
1  113
1   48
1  155
2  156
2  165
2  652
2  179
2  757
3  278
3  197
3   95
3  395
3   83
4 2540
4   613
4   200
4   251
4   390
5    18
```

```
5 16
5 11
5 21
5 14
;
run;
* Print data set;
proc print data=bark_beetle;
run;
* Plot means, standard errors, and observations;
proc gplot data=bark_beetle;
    plot y*site=1 / vaxis=axis1 haxis=axis1;
    symbol1 i=std1mjt v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;
* One-way ANOVA with random effects - F test;
proc mixed method=type3 data=bark_beetle;
    class site;
    model y = / ddfm=kr;
    random site;
run;
* One-way ANOVA with random effects - variance components;
proc mixed cl data=bark_beetle;
    class site;
    model y = / ddfm=kr outp=resids;
    random site;
run;
goptions reset=all;
title "Diagnostic plots to check ANOVA assumptions";
* Plot residuals vs. predicted values;
proc gplot data=resids;
    plot resid*pred=1 / vaxis=axis1 haxis=axis1;
    symbol1 v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;
* Normal quantile plot of residuals;
proc univariate noprint data=resids;
    qqplot resid / normal waxis=3 height=4;
run;
quit;
```

---

## SAS Output

One-way ANOVA for bark beetle sampling study 1  
 14:44 Tuesday, September 20, 2011

Obs	site	count	y
1	1	137	2.13672
2	1	101	2.00432
3	1	113	2.05308
4	1	48	1.68124
5	1	155	2.19033
6	2	156	2.19312
7	2	165	2.21748
8	2	652	2.81425
9	2	179	2.25285
10	2	757	2.87910
11	3	278	2.44404
12	3	197	2.29447
13	3	95	1.97772
14	3	395	2.59660
15	3	83	1.91908
16	4	2540	3.40483
17	4	613	2.78746
18	4	200	2.30103
19	4	251	2.39967
20	4	390	2.59106
21	5	18	1.25527
22	5	16	1.20412
23	5	11	1.04139
24	5	21	1.32222
25	5	14	1.14613

One-way ANOVA for bark beetle sampling study 2  
 14:44 Tuesday, September 20, 2011

## The Mixed Procedure

## Model Information

Data Set	WORK.BARK_BEETLE
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	Type 3

Residual Variance Method      Factor  
 Fixed Effects SE Method      Kenward-Roger  
 Degrees of Freedom Method    Kenward-Roger

Class Level Information

Class	Levels	Values
site	5	1 2 3 4 5

Dimensions

Covariance Parameters	2
Columns in X	1
Columns in Z	5
Subjects	1
Max Obs Per Subject	25

Number of Observations

Number of Observations Read	25
Number of Observations Used	25
Number of Observations Not Used	0

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
site	4	6.706318	1.676580	Var(Residual) + 5 Var(site)
Residual	20	1.786777	0.089339	Var(Residual)

Type 3 Analysis of Variance

Source	Error Term	Error DF	F Value	Pr > F
site	MS(Residual)	20	18.77	<.0001
Residual	.	.	.	.

One-way ANOVA for bark beetle sampling study 3  
 14:44 Tuesday, September 20, 2011

The Mixed Procedure

Covariance Parameter  
Estimates

Cov Parm	Estimate
site	0.3174
Residual	0.08934

Fit Statistics

-2 Res Log Likelihood	25.1
AIC (smaller is better)	29.1
AICC (smaller is better)	29.7
BIC (smaller is better)	28.3

One-way ANOVA for bark beetle sampling study 4  
 14:44 Tuesday, September 20, 2011

The Mixed Procedure

Model Information

Data Set	WORK.BARK_BEETLE
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Kenward-Roger
Degrees of Freedom Method	Kenward-Roger

Class Level Information

Class	Levels	Values
site	5	1 2 3 4 5



## Dimensions

Covariance Parameters	2
Columns in X	1
Columns in Z	5
Subjects	1
Max Obs Per Subject	25

## Number of Observations

Number of Observations Read	25
Number of Observations Used	25
Number of Observations Not Used	0

## Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	46.39671929	
1	1	25.08857565	0.00000000

Convergence criteria met.

One-way ANOVA for bark beetle sampling study 5  
14:44 Tuesday, September 20, 2011

## The Mixed Procedure

## Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
site	0.3174	0.05	0.1093	3.1458
Residual	0.08934	0.05	0.05229	0.1863

## Fit Statistics

-2 Res Log Likelihood	25.1
AIC (smaller is better)	29.1

AICC (smaller is better)	29.7
BIC (smaller is better)	28.3

---

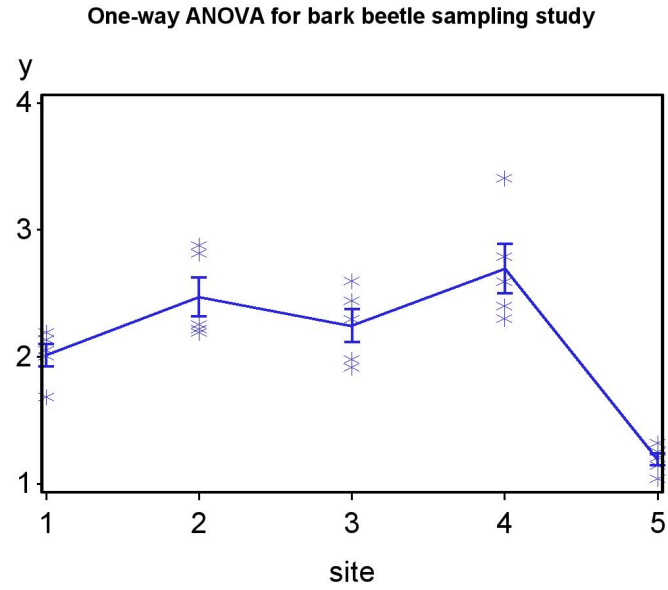


Figure 11.8: Means and standard errors for five study sites

## 11.3 Maximum likelihood estimates

This section sketches how the parameters in one-way ANOVA can be estimated using maximum likelihood. Recall that the likelihood for a random sample of three observations ( $Y_1 = 4.5, Y_2 = 5.4, Y_3 = 5.3$ ) from a normal distribution (see Chapter 8) was of the form

$$L(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(4.5-\mu)^2}{\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(5.4-\mu)^2}{\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(5.3-\mu)^2}{\sigma^2}}. \quad (11.48)$$

We found maximum likelihood estimates of the normal distribution parameters by maximizing this quantity with respect to  $\mu$  and  $\sigma^2$ .

Suppose now we have a data set that can be modeled using the fixed effects one-way ANOVA model, in particular

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}. \quad (11.49)$$

This model has a number of parameters to estimate, such as  $\mu, \alpha_i$  for  $i = 1, 2, \dots, a$ , and  $\sigma^2$ . What would the likelihood function look like for these data? Consider the first group for the bark beetle experiment (Example 1), for which we have  $Y_{11} = 2.576, Y_{12} = 2.10, Y_{13} = 2.41, Y_{14} = 2.14$ , and  $Y_{15} = 2.58$ . For the first group the model assumes that  $Y_{1j} \sim N(\mu + \alpha_1, \sigma^2)$ , and so the likelihood would be

$$L_1 = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(2.576 - (\mu + \alpha_1))^2}{\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(2.10 - (\mu + \alpha_1))^2}{\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(2.41 - (\mu + \alpha_1))^2}{\sigma^2}} \quad (11.50)$$

$$\times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(2.14 - (\mu + \alpha_1))^2}{\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(2.58 - (\mu + \alpha_1))^2}{\sigma^2}}. \quad (11.51)$$

The likelihood  $L_2$  for the second group would be similar, except that  $Y_{2j} \sim N(\mu + \alpha_2, \sigma^2)$ , and  $L_3$  similarly defined. The overall likelihood would then be defined as

$$L(\mu, \alpha_1, \alpha_2, \alpha_3, \sigma^2) = L_1 \times L_2 \times L_3. \quad (11.52)$$

Finding the maximum likelihood estimates involves maximizing this quantity with respect to the parameters  $\mu, \alpha_1, \alpha_2, \alpha_3$ , and  $\sigma^2$ . The likelihood for

designs with any number of treatment groups and replicates would be similar. Using a bit of calculus to find the maximum, it can be shown that the maximum likelihood estimates of these parameters, in general, are

$$\hat{\mu} = \bar{Y}, \quad (11.53)$$

$$\hat{\alpha}_i = \bar{Y}_i - \bar{Y}, \quad (11.54)$$

and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2}{a(n-1)} = MS_{within}. \quad (11.55)$$

(McCulloch & Searle 2001). These estimators seem quite reasonable. They use the grand mean of the data,  $\bar{Y}$ , to estimate the grand mean  $\mu$  of the model, and the difference between the  $i$ th group mean and the grand mean,  $\bar{Y}_i - \bar{Y}$ , to estimate the deviation from the group mean  $\alpha_i$ . Note that  $\hat{\sigma}^2$  is equal to  $MS_{within}$ , which we have already encountered in our ANOVA calculations.

Suppose now we have a data set suited to the random effects model, in particular

$$Y_{ij} = \mu + A_i + \epsilon_{ij}. \quad (11.56)$$

This model has three parameters to be estimated:  $\mu$ ,  $\sigma_A^2$ , and  $\sigma^2$ . The likelihood for this model is more complex because of the random effect  $A_i$ , but one can show that the maximum likelihood estimators of these parameters are

$$\hat{\mu} = \bar{Y}, \quad (11.57)$$

$$\hat{\sigma}_A^2 = \frac{MS_{among} - MS_{within}}{n}, \quad (11.58)$$

and

$$\hat{\sigma}^2 = MS_{within}. \quad (11.59)$$

An intuitive explanation of the formula for  $\hat{\sigma}_A^2$  is that  $MS_{among}$  incorporates variance from both  $A_i$  and  $\epsilon_{ij}$ , while  $MS_{within}$  only has  $\epsilon_{ij}$ . Subtracting  $MS_{within}$  from  $MS_{among}$  leaves only the variance due to  $A_i$ , so that the numerator of this expression estimates  $n\sigma_A^2$ . We then divide by  $n$  to obtain an estimate of  $\sigma_A^2$ .

Suppose that for an unusual data set we obtain  $MS_{among} < MS_{within}$ , implying a negative estimate of  $\hat{\sigma}_A^2 = 0$  according to the above equation. An inherent feature of maximum likelihood is that it restricts variance components to plausible values (McCulloch & Searle 2001), so in this case it would

simply say that  $\hat{\sigma}_A^2 = 0$ , the smallest possible nonnegative value. This would be reflected in the SAS output for `proc mixed`, which would report that the variance component in question was zero. The estimates presented here are actually obtained using a variant of maximum likelihood called restricted maximum likelihood or REML. This method is the default in SAS, and has some theoretical advantages over straight maximum likelihood (McCulloch and Searle 2001).

## 11.4 *F* test as a likelihood ratio test

The *F* test in one-way ANOVA can be derived as a likelihood ratio test, similar to the development of the *t* test in Chapter 10. We first find the maximum likelihood estimates of various parameters under  $H_1$  vs.  $H_0$ , where the parameters under consideration are the ANOVA model parameters. Recall that the observations in the fixed effects model are described as

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (11.60)$$

where  $\mu$  is the grand mean,  $\alpha_i$  is the effect of the *i*th treatment, and  $\epsilon_{ij} \sim N(0, \sigma^2)$ . This is the statistical model under the alternative hypothesis, where  $\alpha_i \neq 0$  for some *i*. Under  $H_0$  : all  $\alpha_i = 0$ , the model reduces to just

$$Y_{ij} = \mu + \epsilon_{ij}. \quad (11.61)$$

We would need to find the maximum likelihood estimates under both  $H_1$  (see previous section) and  $H_0$ , as well as  $L_{H_0}$  and  $L_{H_1}$ , the maximum height of the likelihood function under  $H_0$  and  $H_1$ . We would then use the likelihood ratio test statistic

$$\lambda = \frac{L_{H_0}}{L_{H_1}}. \quad (11.62)$$

It can be shown that there is a one-to-one correspondence between  $-2 \ln(\lambda)$  and  $F_s$  in one-way ANOVA, and so the *F* test is actually a likelihood ratio test (McCulloch & Searle 2001). A similar argument can be made to justify the *F* test for the random effects model. Like all likelihood ratio tests, large values of the test statistic  $-2 \ln(\lambda)$  or  $F_s$  indicate a lower value of the likelihood under  $H_0$  relative to  $H_1$ , and thus a poorer fit of the  $H_0$  model.

## 11.5 One-way ANOVA and two-sample $t$ tests

There is an alternative to one-way ANOVA when there are only two groups to be compared, the two-sample  $t$  test. Let  $\mu_1$  be the mean of the first group and  $\mu_2$  the second one, and suppose that the two groups have the same variance  $\sigma^2$  and sample size  $n$ . We are interested in testing  $H_0 : \mu_1 = \mu_2$  vs.  $H_1 : \mu_1 \neq \mu_2$ , to determine if there are differences in the means of the two groups. Under  $H_0$ , the test statistic

$$T_s = \frac{(\bar{Y}_1. - \bar{Y}_2.)}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \sim t_{2(n-1)}. \quad (11.63)$$

Here  $\bar{Y}_1.$  and  $\bar{Y}_2.$  are the sample means for each group, and  $s_1^2$  and  $s_2^2$  the sample variances. For a Type I error rate of  $\alpha$ , the acceptance region of the test would be the interval  $(-c_{\alpha,2(n-1)}, c_{\alpha,2(n-1)})$ , where  $c_{\alpha,2(n-1)}$  is determined using Table T (see Chapter 10). We would reject  $H_0$  if it falls on the edge or outside this interval. There are also versions of this test statistic for unequal sample sizes.

Although a two-sample  $t$  test is often used for comparing two groups, in the form above it is equivalent to the  $F$  test in one-way ANOVA. To see this, note that  $T_s^2 = F_s$  for one-way ANOVA with two groups. It can also be shown that the acceptance and rejection regions are the same for the two tests. Unlike ANOVA, though, a two-sample  $t$  test can also be used for one-tailed alternative hypotheses, such as  $H_1 : \mu_1 > \mu_2$  or  $H_1 : \mu_1 < \mu_2$ . The procedure is similar to one-sample  $t$  tests for one-tailed alternatives (see Chapter 10).

### 11.5.1 Two-sample $t$ test for Example 1 - SAS demo

We can illustrate this test by comparing treatment A and B in the Example 1 study, deleting the data for the third treatment. See SAS program and output below. The `data` and `proc gplot` portions of the program are similar to our previous one-way ANOVA code. The two-sample  $t$  test is carried using `proc ttest` (SAS Institute Inc. 2014a), with the `class` statement indicating the variable that codes for different groups (`treat`), while the `var` statement designates the dependent variable (`y`). We see there is a highly significant difference between treatment A and B ( $t_8 = 4.90, P = 0.0012$ ), with treatment A catching more beetles than B (Fig. 11.9).

---

SAS Program

---

```
* bark_beetle_experiment_ttest.sas;
options pageno=1 linesize=80;
goptions reset=all;
title "Two-sample t test for bark beetle trapping experiment";
data bark_beetle;
    input treat $ count;
    * Apply transformations here;
    y = log10(count);
    datalines;
A   373
A   126
A   255
A   138
A   379
B    25
B    64
B    62
B    71
B    54
;
run;
* Print data set;
proc print data=bark_beetle;
run;
* Plot means, standard errors, and observations;
proc gplot data=bark_beetle;
    plot y*treat=1 / vaxis=axis1 haxis=axis1;
    symbol1 i=stdimjt v=star height=2 width=3;
    axis1 label=(height=2) value=(height=2) width=3 major=(width=2) minor=none;
run;
* Two-sample t test;
proc ttest data=bark_beetle;
    class treat;
    var y;
run;
quit;
```

---

## SAS Output

Two-sample t test for bark beetle trapping experiment 1  
 16:16 Thursday, May 22, 2014

Obs	treat	count	y
1	A	373	2.57171
2	A	126	2.10037
3	A	255	2.40654
4	A	138	2.13988
5	A	379	2.57864
6	B	25	1.39794
7	B	64	1.80618
8	B	62	1.79239
9	B	71	1.85126
10	B	54	1.73239

s

Two-sample t test for bark beetle trapping experiment 2  
 16:16 Thursday, May 22, 2014

## The TTEST Procedure

Variable: y

treat	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	5	2.3594	0.2295	0.1026	2.1004	2.5786
B	5	1.7160	0.1828	0.0818	1.3979	1.8513
Diff (1-2)		0.6434	0.2075	0.1312		

treat	Method	Mean	95% CL Mean	Std Dev
A		2.3594	2.0745 2.6444	0.2295
B		1.7160	1.4890 1.9430	0.1828
Diff (1-2)	Pooled	0.6434	0.3408 0.9460	0.2075
Diff (1-2)	Satterthwaite	0.6434	0.3382 0.9486	

treat	Method	95% CL Std Dev
A		0.1375 0.6594
B		0.1095 0.5253
Diff (1-2)	Pooled	0.1401 0.3975
Diff (1-2)	Satterthwaite	



Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	8	4.90	0.0012
Satterthwaite	Unequal	7.6194	4.90	0.0014

## Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	4	1.58	0.6704

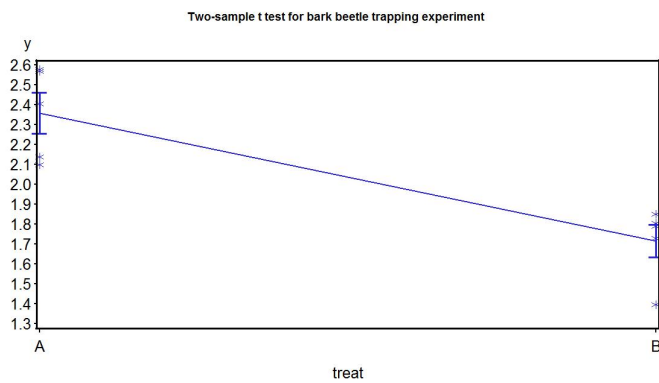


Figure 11.9: Means and standard errors for treatment A and B

## 11.6 References

- McCulloch, C. E. & Searle, S. R. (2001) *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc., New York, NY.
- Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. (1996) *The SAS System for Mixed Models*. SAS Institute Inc., Cary, NC.
- SAS Institute Inc. (2014a) *SAS/STAT 13.2 Users Guide*. SAS Institute Inc., Cary, NC.
- SAS Institute Inc. (2014b) *SAS/GRAPH 9.4: Reference, Third Edition*. SAS Institute Inc., Cary, NC.
- Winer, B. J., Brown, D. R. & Michels, K. M. (1991) *Statistical Principles in Experimental Design*, 3rd edition. McGraw-Hill, Inc., Boston, MA.

## 11.7 Problems

1. A doctor conducts an experiment in which men are placed on four different diets, consisting of a standard weight loss regimen (a control treatment) and three new diets (Diets 1, 2, 3). The weight losses (lbs) after six months are given in the following table.

Control	Diet 1	Diet 2	Diet 3
19.5	20.0	20.8	25.9
20.5	16.4	17.4	25.9
16.6	11.9	16.7	25.8
19.3	22.1	16.8	22.5

- (a) Test whether there is a significant difference among the four treatments using one-way ANOVA, using manual calculations. Report the  $P$  value and discuss the significance of the test, and then interpret the results of the experiment. Show all your calculations.
  - (b) Repeat the analysis using SAS and `proc glm`. Attach your program and output.
2. An experiment was conducted on the fecundity of a predatory insect reared on an artificial diet using four different concentrations of the preservative sorbic acid: (1) no sorbic acid, (2) 0.1% sorbic acid, (3) 0.2% sorbic acid, and (4) 0.5% sorbic acid. Twenty insects were reared at each concentration and the fecundity of the resulting adults measured. See table below.

Treatment	Observations
No sorbic acid	87, 124, 105, 87, 100, 89, 95, 79, 102, 112 92, 87, 115, 96, 111, 90, 86, 92, 109, 76
0.1% sorbic acid	105, 94, 97, 94, 83, 97, 107, 99, 104, 83 101, 71, 100, 75, 87, 106, 88, 99, 90, 74
0.2% sorbic acid	73, 94, 81, 83, 100, 98, 76, 91, 68, 82 92, 105, 76, 82, 95, 96, 101, 89, 92, 67
0.5% sorbic acid	83, 54, 86, 76, 74, 81, 79, 72, 80, 78 70, 83, 83, 85, 90, 70, 85, 94, 82, 75

Test whether there is a difference among the four treatments using one-way ANOVA and SAS. Interpret the results of this analysis, providing a  $P$  value and discussing the significance of the test. Using a graph,

explain what happens to fecundity as the concentration of sorbic acid changes.