

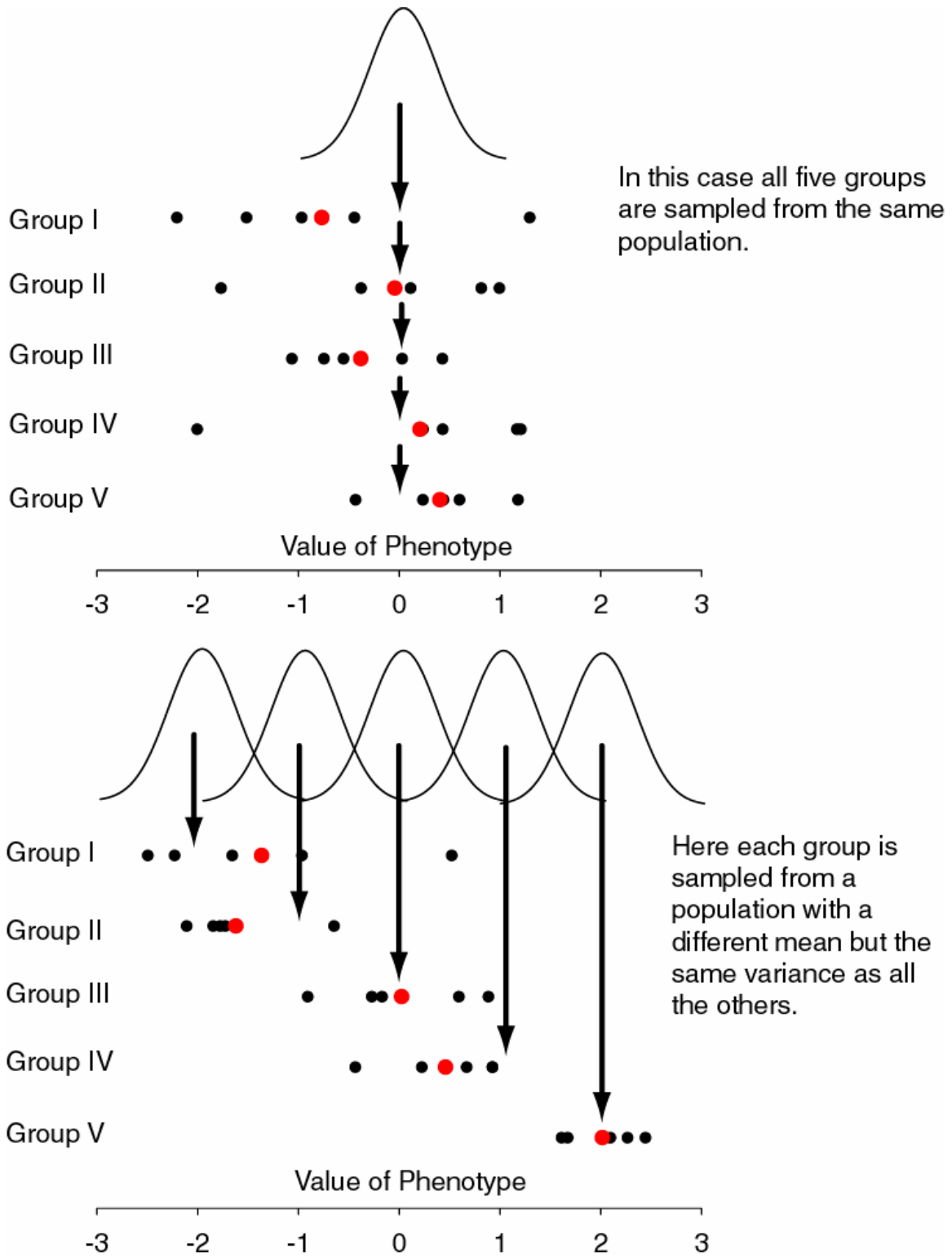
## Analysis of Variance

The analysis of variance, or ANOVA as it is often referred to, is a particularly powerful way to analyze experimental data. It is worthwhile to note that Sir Ronald Fisher, who also made numerous contributions to the theory of evolutionary biology, developed this type of statistical analysis. For example, one typical type of problem that ANOVA is well suited for is testing the effects of different fertilizers on plant growth. Random samples of plants would be treated with different fertilizers, and possibly no fertilizer, but otherwise treated the same. The growth of these plants over a fixed time period would be measured. Then one would use ANOVA to determine if there are significant differences between the plants that received fertilizers and plants that received water alone. If there is a significant effect of fertilizers on plant growth, one could further determine if some fertilizers are better than others.

In a genetic context we might be interested in determining if a particular phenotype is affected by genetic variation or is affected only the environment. If both genes and the environment affect the trait then we might like to estimate the relative magnitude of each of the factors. Suppose we were to examine a phenotype that is only affected by the environment. If we sampled progeny from five families that had been raised in identical environments we might obtain results similar to the top of figure 8. Within each group the variance in phenotype is roughly the same. The red circles show the mean of each group and we can see that the red circles show less variation than the black circles. In fact simple statistical theory tells us that when all the samples come from the same population then the variation of the mean values will be about  $1/5^{\text{th}}$  the variation of the individual values. ANOVA gives formal ways of examining this relationship. If we

performed an ANOVA on the set of data in the top of figure 8 we would have concluded that there are no significant differences between the five groups and thus your genetic background (family origin) does not affect this phenotype.

In the bottom of figure 8 we see quite a different set of results. Here the five groups (families) in fact come from populations with different mean values (these are the genetic effects). Once again if we examine the variation within a group it is the same for each of the five. This is the common environmental variation. Now when we examine the variation between the mean values it is clearly larger than it was in the top of figure 8. It also turns out that it is much larger than  $1/5^{\text{th}}$  of the variation seen within each group. These observations when processed by the ANOVA will let us conclude that family of origin (or our genes) does affect this phenotype. Furthermore, the ANOVA will give us a numerical estimate of relative effects of genes compared to the environment. These types of estimates are crucial for making predictions about the course of selection on quantitative genetic traits.



**Figure 1.** The types of samples used in the analysis of variance. In the top part of the figure samples are derived from a single population. The samples are divided into five groups each with five observations (black circles). A red circle represents the mean of each group. In the lower figure the five groups each come from five different populations with different means. If we

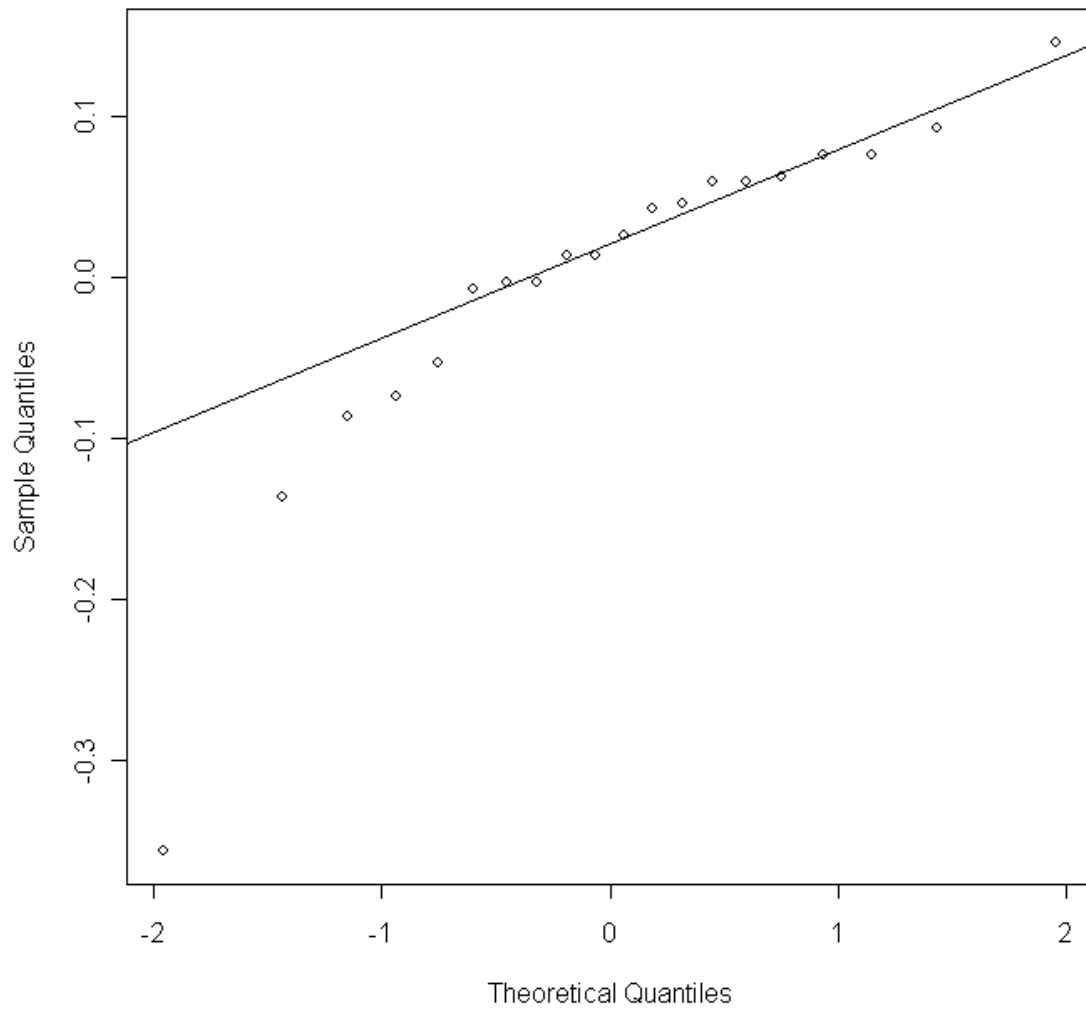
examine the variation that exists within each of the groups it is roughly the same in the top five groups and in the bottom five groups. However, if we examine the variation between mean values it is clear that this variation is greater in the bottom five means compared to the top five means.

---

```
> axauc<- read.table("axauc-sample.data")
> axauc
  population rep  viabilty
1          ax   1 0.9166667
2          ax   1 0.5000000
3          ax   1 0.9333333
4          ax   1 0.9500000
5          ax   1 0.8500000
6          ax   2 0.8666667
7          ax   2 0.8500000
8          ax   2 0.9000000
9          ax   2 0.8833333
10         ax   2 0.7500000
11         auc   1 0.7333333
12         auc   1 0.5833333
13         auc   1 0.6333333
14         auc   1 0.5333333
15         auc   1 0.5833333
16         auc   2 0.4166667
17         auc   2 0.5666667
18         auc   2 0.4666667
19         auc   2 0.6166667
20         auc   2 0.5666667

> axauc$rep<- factor(axauc$rep,ordered=FALSE)
> anova.ax1<-aov(viabilty~population+rep,axauc)
> summary (anova.ax1)
              Df  Sum Sq Mean Sq F value    Pr(>F)
population    1  0.36450  0.36450  28.1659 5.793e-05 ***
rep            1  0.00556  0.00556   0.4293  0.5211
Residuals    17  0.22000  0.01294
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> qqnorm(anova.ax1$residuals)
> qqline(anova.ax1$residuals)
```

Normal Q-Q Plot



```
> anova.ax2<-aov(asin(sqrt(viability))~population+rep,axauc)
> summary(anova.ax2)
              Df Sum Sq Mean Sq F value    Pr(>F)
population    1  0.51617  0.51617  29.6095 4.398e-05 ***
rep           1  0.00999  0.00999   0.5729  0.4595
Residuals    17  0.29636  0.01743
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> qqnorm(anova.ax2$residuals)
> qqline(anova.ax2$residuals)
```

