thalamic nucleus. These projections, mainly from the basolateral nuclei, enable the amygdala to influence prefrontal cortex processing, most commonly linked to learning, working memory, and decision-making.

Fourth, are projections to the brainstem sensory and motor areas, originating primarily in the central and medial amygdalar nuclei. The central nucleus is also the recipient of major sensory information arriving from the brainstem. These efferents provide a 'feedback' to the sensory areas in the brainstem, and allow the amygdala direct access to the autonomic areas necessary for the expression of emotional and motivational responses.

Fifth, the amygdala projects back to the cortical sensory areas from which it receives inputs, thus providing a route for a feedback on sensory processing. In addition, it sends substantial, topographic inputs to various regions of the hippocampal formation, which are critical for learning and memory.

Clearly, the amygdala is well placed for a critical role in the stimulus-reinforcement type of associative learning, it receives sensory information from fundamentally all sensory modalities, and in turn, is in a position to influence a variety of motor systems via its wide ranging efferents. Furthermore, it is increasingly clear that distinct functional subsystems exist within the amygdala, and they each have their unique set of projections. An understanding of the underlying principles of the functional and anatomical organization of the amygdala is one of the most challenging questions that the field is facing today.

4. *Models of Amygdalar Organization*

In the 1980s and 1990s two models of amygdalar organization have been proposed. The first model suggests that part of the amygdala (central and medial nuclei) together with the substantia innominata and the bed nuclei of the stria terminalis, form a structural and functional unit—'the extended amygdala.' This hypothesis is based on cytoarchitectural, histochemical, and connectional similarities between the three parts of the continuum (De Olmos and Heimer 1999). Although this model fosters better understanding of the organization of central and medial parts, it does not provide a place for the rest of the amygdala.

Another more comprehensive model of amygdalar organization was proposed based on current embryological, neurotransmitter, connectional and functional data. This model argues that the amygdala is neither a structural nor a functional unit, but rather an arbitrarily defined collection of cell groups in the cerebral hemispheres, originally based on cytoarchitectonics. This suggests that it is more useful to place the various amygdalar cell groups within the context of the major divisions of the cerebral hemispheres—cortex and basal ganglia—and then to define the topographical organization of functionally defined systems within these divisions. Thus, various parts of the amygdala can be classified as belonging to one of the three distinct telencephalic groups: caudal olfactory cortex, specialized ventromedial extension of the striatum, or ventromedial expanse of the claustrum. Functionally they belong to the olfactory, autonomic, or frontotemporal cortical systems, respectively (Swanson and Petrovich 1998).

Future research hopefully will delineate how the dynamics of information flow through the different amygdalar subsystems contribute to its different functions.

See also: Emotion, Neural Basis of; Fear: Potentiation of Startle; Fear: Psychological and Neural Aspects; Motivation, Neural Basis of; Reinforcement: Neurochemical Substrates

Bibliography

Aggleton J P (ed.) 1992 *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction.* Wiley-Liss, New York

De Olmos J S, Heimer L 1999 The concepts of the ventral striatopallidal system and extended amygdala. *Annals of New York Academy of Science* **877**: 1–32

Kaada B R 1972 Stimulation and regional ablation of the amygdaloid complex with reference to functional representations. In: Eleftheriou B E (ed.) *The Neurobiology of the Amygdala*. Plenum Press, New York

Johnston J B 1923 Further contributions to the study of the evolution of the forebrain. *Journal of Comparative Neurology* **35**: 337–482

Klüver H, Bucy P C 1939 Preliminary analysis of functions of the temporal lobes in monkeys. *Archives of Neurological Psychiatry* **42**: 979–1000

Swanson L W, Petrovich G D 1998 What is the amygdala? *Trends in Neurosciences* **21**: 323–31

Weiskrantz L 1956 Behavioral changes associated with the ablation of the amygdaloid complex in monkeys. *Journal of Comparative Physiological Psychology* **49**: 381–91

G. D. Petrovich

Analysis of Variance and Generalized Linear Models

Analysis of variance (ANOVA) models are models that exploit the grouping structure in a set of data and lend themselves to the examination of main effects and interactions. These models are often referred to as being hierarchical in that it makes no sense to test whether main effects can be dropped from a model that includes that factor in an interaction, and it makes no sense to test whether a lower order interaction can be dropped from a model that includes any higher order interaction involving all of the same factors. The modeling ideas also extend to generalized linear models (GLIMs).

1. One-way Analysis of Variance

One-way ANOVA models identify groups of observations using a separate mean for each group. For example, the observations can be the age of suicide categorized by mutually exclusive ethnic groups. Suppose there are a groups with N_i observations in each group. The jth observation in the ith group is written y_{ij} with expected value $E(y_{ij}) = \mu_i$. The one-way analysis of variance model is

$$y_{ij} = \mu_i + \varepsilon_{ij} \tag{1}$$

$i = 1, \ldots, a$, $j = 1, \ldots, N_i$ where the ε_{ij}s are unobservable random errors with mean 0, often assumed to be independent, normally distributed with variance σ^2, written $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The ANOVA is a procedure for testing whether the groups have the same mean values μ_i. It involves obtaining two statistics. First, the mean squared error (MSE) is an estimate of the variance σ^2 of the individual observations. Second, the mean squared groups ($MSGrps$) is an estimate of σ^2 when the means μ_i are all the same and an estimate of σ^2 plus a positive number when the μ_i are different. The ratio of these two statistics, $MSGrps/MSE$ should be approximately 1 if the μ_is are all the same and tends to be larger than 1 if the μ_is are not all the same. Under the normality assumptions for the ε_{ij}s, when the μ_is are all the same the exact value of $MSGrps/MSE$ is random and follows an F distribution with $a-1$ degrees of freedom in the numerator and $n - a$ degrees of freedom in the denominator. Here $n = N_1 + \cdots + N_a$ is the total number of observations in the data. If the observed value of $MSGrps/MSE$ is so much larger than 1 as to be relatively inconsistent with it coming from the F distribution, one concludes that the μ_is must not all be the same.

In the special balanced case where $N_i = N$ for all i, the computations are particularly intuitive as analyzing variances. To find the MSE, simply find the sample variance within each group, i.e., compute the sample variance from $y_{i1}, \ldots, y_{iN}$, then average these a numbers to get MSE. This provides an estimate of σ^2. To get $MSGrps$, first compute the sample mean for each group, say $\bar{y}_i$. Then compute the sample variance of the $\bar{y}_i$s and multiply by N to get $MSGrps$. $MSGrps$ estimates $\sigma^2 + Ns^2_\mu$ where s^2_μ is a new parameter consisting of the sample variance of the unknown parameters $\mu_1, \ldots, \mu_a$. The μ_is are all equal if and only if $s^2_\mu = 0$. $MSGrps/MSE$ estimates $1 + Ns^2_\mu/\sigma^2$.

Terminology varies considerably for these concepts. The mean squared error is also called the mean squared residual and the mean squared within (groups). The mean squared groups is also called the mean squared treatments (because often the groups are identified as different treatments in an experiment) and the mean squared between (groups).

An alternative but equivalent model used with one-way ANOVA is

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{2}$$

$i = 1, \ldots, a$, $j = 1, \ldots, N_i$. Here μ is a grand mean and the α_is are differential effects for the groups. The problem with this model is that it is overparameterized, i.e., the μ and α_i parameters are not identifiable. Even if you know the μ_is, there is an infinite number of ways to define the μ and α_i parameters. In fact, you can arbitrarily pick a value for any one of the μ or α_i parameters and still make them agree with any set of μ_is. Without including extraneous side conditions that have nothing to do with the model, it is impossible to estimate any of the μ and α_i parameters. It is, however, possible to estimate some functions of them, like values $\mu + \alpha_i$, or contrasts among the α_is like $\alpha_1 - \alpha_2$. Linear functions of the μ and α_i parameters for which linear unbiased estimates exist are called estimable functions. See also *Statistical Identification and Estimability.*

The F test can also be viewed as testing the full one-way ANOVA model against the reduced model

$$y_{ij} = \mu + \varepsilon_{ij}.$$

The reduced model can be viewed as either dropping the subscript i from μ_i in model (1) or as dropping the α_is from model (2). In either case, the reduced model does not allow for separate group effects. The F statistic comes from the error terms of the two models

$$F = \frac{SSE(\text{Red.}) - SSE(\text{Full})}{dfE(\text{Red.}) - dfE(\text{Full})} \Big/ MSE(\text{Full}) = \frac{MSGrps}{MSE}.$$

Extensions of ANOVA to more general situations depend crucially on the idea of testing full models against reduced models; see *Linear Hypothesis.*

2. Two-way Analysis of Variance

Two-way ANOVA can be thought of as a special case of one-way ANOVA in which the groups have a two-factor structure that we want to exploit in the analysis. For example, Everitt (1977) discusses 97 ten-year-old school children who were cross-classified by two factors, first, the risk of their home environment: not at risk (N) or at risk (R), and then the adversity of their school conditions: low, medium, or high. This defines six groups, each a combination of a home environment and a school condition. To illustrate the modeling concepts, suppose the dependent variable y

is the score on a test of verbal abilities. In general, we would write a model

$$y_{gk} = \mu_g + \varepsilon_{gk} \tag{3}$$

where $g = 1, \ldots, G$ indicates the different groups and $k = 1, \ldots, N_g$ indicates the observations within the group. In the specific example, $G = 6$ and $(N_1, N_2, N_3, N_4, N_5, N_6) = (17, 8, 18, 42, 6, 6)$.

To use the two-factor structure, we begin by identifying the factors mumerically. Let $i = 1,2$ indicate home environment: not at risk ($i = 1$), at risk ($i = 2$). Let $j = 1, 2, 3$ identify school adversity: low ($j = 1$), medium ($j = 2$), high ($j = 3$). In general, let $i = 1, \ldots, a$ denote the levels of the first factor and $j = 1, \ldots, b$ denote levels of the second factor. Without changing anything of substance in the model, we can rewrite the ANOVA model (3) as

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \tag{4}$$

$i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, N_{ij}$. All we have done is replace the single index for the groups g with two numbers ij that are used to identify the groups. This is still a one-way ANOVA model and can be used to generate an F test of whether all the μ_{ij}s are equal. In our example, $a = 2$, $b = 3$, and $(N_{11}, N_{21}, N_{12}, N_{22}, N_{13}, N_{23}) = (17, 8, 18, 42, 6, 6)$. Note that with a two factor model the number of groups is $G = ab$ and $E(y_{ijk}) = \mu_{ij}$.

The one-way ANOVA model (4) is often rewritten in an overparameterized version that includes numerous unidentifiable parameters. The overparameterized model is called the two-way ANOVA with interaction model, and is written

$$y_{ijk} = \mu + \alpha_i + \eta_j + (\alpha\eta)_{ij} + \varepsilon_{ijk}. \tag{5}$$

Here μ is a grand mean, the α_is are differential effects for the first factor, the η_js are differential effects for the second factor, and the $(\alpha\eta)_{ij}$s are called interaction effects. In reality, all of the μ, α_i, and η_j parameters are extraneous. If we drop all of them, we get the model $y_{ijk} = (\alpha\eta)_{ij} + \varepsilon_{ijk}$, which is just the one-way ANOVA model (4) with the μ_{ij}s relabeled as $(\alpha\eta)_{ij}$s. In model (5), without the introduction of side conditions that have nothing to do with the model, it is impossible to estimate or conduct any test on any function of the parameters that does not involve the $(\alpha\eta)_{ij}$ parameters. Since main effect parameters are extraneous when interactions are in the model, computer programs that purport to give tests for main effects after fitting interactions are really testing some arcane function of the interaction parameters that is determined by a choice of side conditions used by the program.

The interesting aspect of the two-way ANOVA with interaction model is that it suggests looking at the two-way ANOVA without interaction model

Table 1
Mean verbal test scores, additive model

	μ_{ij}	j (School) 1 (Low)	2 (Medium)	3 (High)
i	1 (N)	110	105	95
(Home)	2 (R)	100	95	85

Table 2
Mean verbal test scores, interaction model

	μ_{ij}	j (School) 1 (Low)	2 (Medium)	3 (High)
i	1 (N)	110	105	95
(Home)	2 (R)	100	95	80

$$y_{ijk} = \mu + \alpha_i + \eta_j + \varepsilon_{ijk} \tag{6}$$

This model is not equivalent to the one-way ANOVA model, but it includes nontrivial group effects. It amounts to imposing a restriction on the μ_{ij}s that

$$\mu_{ij} = \mu + \alpha_i + \eta_j \tag{7}$$

for some μ, α_is, and η_js. This indicates that the group effects μ_{ij} have a special structure in which the group effect is the sum of a grand mean and differential effects for each factor, hence model (6) is referred to as an additive model. The two-way ANOVA without interaction model is still overparameterized, so none of the individual parameters are estimable, but typically contrasts in the α_is and η_js are estimable as well as values $\mu + \alpha_i + \eta_j$.

To illustrate the modeling concepts, suppose that in our example the group means for the verbal test scores take the values shown in Table 1. These μ_{ij}s satisfy the additive model (7) so they do not display interaction. For example, take $\mu = 0$, $\alpha_1 = 110$, $\alpha_2 = 100$, $\eta_1 = 0$, $\eta_2 = -5$, $\eta_3 = -15$. Note that μ, the α_is, and the η_js are not identifiable (estimable) because there is more than one way to define them that is consistent with the μ_{ij}s. For example, we can alternatively take $\mu = 100$, $\alpha_1 = 5$, $\alpha_2 = -5$, $\eta_1 = 5$, $\eta_2 = 0$, $\eta_3 = -10$. However, identifiable functions of the parameters include $\mu + \alpha_i + \eta_j$, $\alpha_1 - \alpha_2$, $\eta_1 - \eta_3$, and $\eta_1 + \eta_2 - 2\eta_3$. Identifiable functions take on the same values for any valid choices of μ, the α_is, and the η_js. For example, the effect of changing from a not at risk home situation to an at risk home situation is a decrease of $\alpha_1 - \alpha_2 = 10$ points in mean verbal test score. Similarly, the effect of changing from a low school adversity situation to a high school adversity situation is a decrease of $\eta_1 - \eta_3 = 15$ points in mean verbal test score.

The beauty of the additive model (6) is that there is one number that describes the mean difference between students having not at risk home status and students having at risk home status. The difference is a drop of 10 points *regardless* of the school adversity status. This number can legitimately be described as the effect of going from not at risk to at risk. (Note that unless the students were assigned randomly to their home conditions, this does not imply that changing a student's status from at risk to not at risk will *cause*, on average, a 10 point gain.) Similarly, the effect of school adversity does not change with the home status. For example, going from low to high school adversity induces a 15 point drop regardless of whether the students are at risk or not at risk. Statistically, tests for main effects are tests of whether any of the α_is are different from each other and whether any of the η_js are different from each other, in other words whether the home statuses are actually associated with different mean verbal test scores and similarly for the school adversities.

The existence of interaction is simply any structure to the μ_{ij}s that cannot be written as $\mu_{ij} = \mu + \alpha_i + \eta_j$ for some μ, α_is, and η_js. For example, consider Table 2. In this case, the relative effect of having at risk home status for low or medium adversity schools is a drop of 10 points in mean test score, however for highly adverse schools, the effect of an at risk home status is a drop of 15 points. Unlike cases where the additive model (7) holds, the effect of at risk home status depends on the level of school adversity, so there is no one number that can characterize the effect of at risk home status. It makes no sense to consider the effect of home conditions without specifying the school adversity. Similarly, the effects of school adversity change depending on home status. For example, going from low to high school adversity induces a 15 point drop for students whose homes are not at risk, but a 20 point drop for at risk students. Again, there is no one number than can characterize the change from low to high school adversity, so there is no point in considering this change without specifying the home risk status. One moral of this discussion is that when interaction exists, there is no point in looking at tests of main effects, because main effects are essentially meaningless. If there is no one number that describes the effect of changing from not at risk to at risk, what could $\alpha_1 - \alpha_2$ possibly mean?

In practice, analyses such as these are conducted on estimates of the μ_{ij}s, and the estimates are subject to variability.

A primary use of model (6) is to test whether this special group effects structure fits the data. Model (6) is used as a reduced model and is tested against the full model (5) that includes interaction. This test is referred to as a test of interaction. In particular, the interaction test is not a test of whether the $(\alpha\eta)_{ij}$s are all zero, it is a test of whether every possible definition of the $(\alpha\eta)_{ij}$s must be consistent with the two-way without interaction model. The easiest way to think about this correctly is to think about testing whether one can simply drop the $(\alpha\eta)_{ij}$s from model (5).

If the two-way without interaction model (6) fits the data, it is interesting to see if any special case (reduced model) also fits the data. Two obvious choices are fitting a model that drops the second factor (school) effects

$$y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk} \tag{8}$$

and a model that drops the first factor (home) effects

$$y_{ijk} = \mu + \eta_j + \varepsilon_{ijk} \tag{9}$$

If model (8) fits the data, there is no evidence that the second factor helps explain the data over and above what the first factor explains. Similarly, if (9) fits the data, there is no evidence that the first factor helps explain the data over and above what the second factor explains.

Given model (8) with only the first factor effect, we can evaluate whether the data fit a reduced model without that effect,

$$y_{ijk} = \mu + \varepsilon_{ijk} \tag{10}$$

This model comparison provides a test of whether the first factor is important in explaining the data when the second factor is ignored. Similarly we can start with model (9) and compare it to model (10). If model (10) fits, neither factor helps explain the data.

Through fitting a series of models, we can test whether there is evidence for the interaction effects, test whether there is evidence for the second factor effects η_j when the first factor α_i effects are included in the model, test whether there is evidence for the second factor effects η_j when the first factor α_i effects are not included in the model, and perform two similar tests for the importance of the α_i effects. In the special case when the N_{ij}s are all the same, the test for α effects including ηs, that is, the test of model (6) vs. model (9), and the test for α effects ignoring ηs, that is, the test of model (8) vs. model (10), are identical. Similarly, the two tests for η effects are identical. While this identity greatly simplifies the analysis, it only occurs in special cases and is not generally applicable. Moreover, it does not extend to generalized linear models such as log-linear models and logistic regression. The appropriate way to think about these issues is in terms of model comparisons.

3. *Higher-order Analysis of Variance*

The groups in a one-way ANOVA can also result from combining the levels of three or more factors. Consider a three-way cross-classification of Everitt's 97 students by further classifying them into students displaying or not displaying deviant classroom behavior. This determines $G = 12$ groups. The one-way ANOVA model $y_{gm} = \mu_g + \varepsilon_{gm}$, $g = 1, \ldots, G$, $m = 1, \ldots, N_g$ can be rewritten for three factors as $y_{ijkm} = \mu_{ijk} + \varepsilon_{ijkm}$ where the three factors are identified by $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$ so that $G = abc$, the group sample sizes are N_{ijk}, and $E(y_{ijkm}) = \mu_{ijk}$. In the example, we continue to use i and j to indicate home and school conditions, respectively, and now use $k = 1,2$ to indicate classroom behavior with $k = 1$ being nondeviant.

In its most overparameterized form, a three-way ANOVA model is

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk} + (\alpha\eta\gamma)_{ijk} + \varepsilon_{ijkm} \quad (11)$$

This includes a grand mean μ, main effects for each factor α_is, η_js, γ_ks, two-factor interactions $(\alpha\eta)_{ij}$s, $(\alpha\gamma)_{ik}$s, $(\eta\gamma)_{jk}$s, and a three-factor interaction $(\alpha\eta\gamma)_{ijk}$s. In model (11), dropping the redundant terms, that is, everything except the three-factor interaction terms, gives a one-way ANOVA model $y_{ijkm} = (\alpha\eta\gamma)_{ijk} + \varepsilon_{ijkm}$. The primary interest in model (11) is that it suggests a wealth of reduced models to consider.

The first order of business is to test whether the three factor interaction is necessary, i.e., test the full model (11) against the reduced model without the three-factor interaction

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk} + \varepsilon_{ijkm} \quad (12)$$

Focusing on only the important parameters, model (12) is equivalent to $y_{ijkm} = (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk} + \varepsilon_{ijkm}$. In fact, even this version of the model is overparameterized. Model (11) is equivalent to the one-way ANOVA model, so if the three-factor interaction test is significant, there is no simplifying structure to the treatments. Probably the best one can do is to think of the problem as a one-way ANOVA and draw whatever conclusions are possible from the groups.

If model (12) fits the data, we can seek simpler models or try to interpret model (12). Conditioning on the levels of one factor, simplifies interpretations. For example, if $i = 1$, $y_{1jkm} = \mu + \alpha_1 + \eta_j + \gamma_k + (\alpha\eta)_{1j} + (\alpha\gamma)_{1k} + (\eta\gamma)_{jk} + \varepsilon_{1jkm}$ or $y_{1jkm} = [\mu + \alpha_1] + [\eta_j + (\alpha\eta)_{1j}] + [\gamma_k + (\alpha\gamma)_{1k}] + (\eta\gamma)_{jk} + \varepsilon_{1jkm}$. This is simply a two-factor model with interaction. If we change to $i = 2$, we again get a two-factor interaction model, one that has different main effects, but has the same interaction terms as when $i = 1$. For example, in Table 3 μ_{ijk}s satisfy model (12), that is, $\mu_{ijk} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk}$ for some definitions of the parameters.

The key point in Table 3 is that, regardless of school adversity, the relative effect of not being at risk is always one point higher for nondeviants than for deviants. Thus, for high adversity the nondeviant home difference is 101–99 which is one point higher than the deviant home difference 99–98.

Table 3
No three-factor interacton

μ_{ijk}		Adversity of School (j)					
		Low		Medium		High	
	Home (i)	N	R	N	R	N	R
Classroom	Nondeviant	101	100	101	99	101	99
Behavior (k)	Deviant	100	100	100	99	99	98

Table 4
Two two-factor interactions

μ_{ijk}		Adversity of School (j)					
		Low		Medium		High	
	Home (i)	N	R	N	R	N	R
Classroom	Nondeviant	101	100	101	99	100	97
Behavior (k)	Deviant	100	98	100	97	99	95

Table 5
Rearrangement of Table 4

μ_{ijk}		Home (i)					
		N			R		
	School (j)	L	M	H	L	M	H
Classroom	Nondeviant	101	101	100	100	99	97
Behavior (k)	Deviant	100	100	99	98	97	95

If model (12) is adequate, the next step is to identify which of the two factor interaction terms are important. For example, we can drop out the $(\eta\gamma)_{jk}$ term to get

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + \varepsilon_{ijkm} \quad (13)$$

Dropping unimportant parameters, this is equivalent to $y_{ijkm} = (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + \varepsilon_{ijkm}$. To interpret model (13), condition on the factor that exists in both interaction terms. In our example, this is home conditions. The resulting model for $i = 1$ is a no interaction model $y_{1jkm} = [\mu + \alpha_1] + [\eta_j + (\alpha\eta)_{1j}] + [\gamma_k + (\alpha\gamma)_{1k}] + \varepsilon_{1jkm}$ with a similar no interaction model for $i = 2$ but with different main effects. For example, suppose the m_{ij}s are given in Table 4. Rearranging Table 4 to group together categories with i fixed gives Table 5. For fixed home conditions, there is no interaction. For not at risk students, deviant behavior is associated with a one point drop and high school adversity is associated with a one point drop. For at risk students, deviant behavior is associated with a 2 point drop and, relative to low school adversity, medium and high adversities are associated with a one point and a three point drop, respectively.

Table 6
One two-factor interaction

μ_{ijk}		Adversity of School (j)					
		Low		Medium		High	
	Home (i)	N	R	N	R	N	R
Classroom	Nondeviant	101	101	101	100	101	100
Behavior (k)	Deviant	100	100	100	99	100	99

Alternatively, we could eliminate both the $(\alpha\gamma)_{ik}$s and $(\eta\gamma)_{jk}$s from model (12) to get

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + \varepsilon_{ijkm} \qquad (14)$$

which is equivalent to $y_{ijkm} = \gamma_k + (\alpha\eta)_{ij} + \varepsilon_{ijkm}$. We can think of this as a model for no interaction in a two-factor analysis in which one factor is indicated by the pair *ij* and the other factor is indicated by *k*. In the example, this means there is a main effect for classroom behavior plus an effect for each of the six combinations of home and school conditions. In particular, consider Table 6. Here, deviant behavior is always associated with a one point drop but there is interaction between home and school. For low school adversity, there is no effect of home conditions but for medium or high adversity, being at risk is associated with a one point drop.

From model (14L50), we could then fit a model that involves only the main effects

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + \varepsilon_{ijkm} \qquad (15)$$

Model (15) is equivalent to $y_{ijkm} = \alpha_i + \eta_j + \gamma_k + \varepsilon_{ijkm}$. If model (15) is adequate, we could consider models that successively drop out the α_is, the η_js, and the γ_ks.

Fitting this sequence of successively smaller models, (11) then (12) then (13) then (14) then (15), then three models with the main effects successively dropped out, provides one tool for analyzing the data. There are 36 different orders possible for sequentially dropping out the two-factor interactions and then the main effects. In the balanced case of $N_{ijk} = N$ for all i, j, k, the order of dropping the effects does not matter, so one can construct an ANOVA table to examine each of the individual sets of effects, that is, main effects, two-factor interactions, three-factor interaction. For unbalanced cases, one would need 36 different ANOVA tables, one for each sequence, so ANOVA tables are almost never examined except in balanced cases. (Two-factor models only generate two distinct ANOVA tables). For unbalanced cases, instead of looking at the ANOVA tables, it is more convenient to simply report the *SSE* and *dfE* for all of the relevant models using a notation that identifies models by only their important parameters. For four or more factors, the number of potentially interesting models becomes too large to evaluate all of them.

After deciding on one or more appropriate models, the μ_{ijk}s can be estimated subject to the model and the models interpreted subject to the variability of the estimates.

To interpret a three-factor interaction, recall that just as a two-factor interaction exists when the effect of one factor changes depending on the level of the other factor, one can think of a three-factor interaction as a two-factor interaction that changes depending on the level of the third factor. However, it may be more productive to think of three-factor interaction in a three-factor model as simply specifying a one-way ANOVA. As with two-factor models, it makes little sense to test the main effects of a factor when that factor is involved in an important interaction. Similarly, it makes little sense to test that a lower-order interaction can be dropped from a model that includes a higher-order interaction involving all of the same factors.

A commonly used generalization of the ANOVA models is to allow some of the parameters to be unobservable random variables (random effects) rather than fixed unknown parameters. See *Hierarchical Models: Random and Fixed Effects.*

4. Generalized Linear Models (GLIM)

ANOVA is best suited for analyzing normally distributed data. These are measurement data for which the random observations are symmetrically distributed about their mean values. Generalized linear models (GLIMs) use similar linear structures to analyze other kinds of data, such as count data and time to event data.

We can rethink two-way ANOVA models as independent observations y_{ijk} normally distributed with mean m_{ij} and variance σ^2, write $y_{ijk} \sim N(m_{ij}, \sigma^2)$. The interaction model (5) has

$$m_{ij} = \mu + \alpha_i + \eta_j + (\alpha\eta)_{ij}$$

The no interaction model (6) has

$$m_{ij} = \mu + \alpha_i + \eta_j$$

Now consider data that are counts y_{ij} of some event. Assume the counts are independent Poisson random variables with means $\mathrm{E}(y_{ij}) = m_{ij}$. The use of two subscripts indicates that the data are classified by two factors. For Poisson data the m_{ij}s must be nonnegative, so $\log(m_{ij})$ can be both positive and negative. Linear models naturally allow both positive and negative mean values, so it is natural to model the $\log(m_{ij})$s with linear models. For example, we might use the log-linear model

$$\log(m_{ij}) = \mu + \alpha_i + \eta_j + (\alpha\eta)_{ij}$$

or, without interaction,

$$\log(m_{ij}) = \mu + \alpha_i + \eta_j$$

These log-linear models are also appropriate for multinomial data and independent groups of multinomial data.

Similarly, if the data y_{ij} are independent binomials with N_{ij} trials and probability of success p_{ij}, then we can analyze the proportions $\hat{p}_{ij} \equiv y_{ij}/N_{ij}$ with $\mathrm{E}(\hat{p}_{ij}) = m_{ij} \equiv p_{ij}$. Probabilities are defined to be between 0 and 1, the odds $p_{ij}/(1-p_{ij})$ take positive values, and the log-odds can take any positive or negative value. It is natural to write linear models for the log-odds such as

$$\log[p_{ij}/(1-p_{ij})] = \mu + \alpha_i + \eta_j + (\alpha\eta)_{ij}.$$

or, without interaction,

$$\log[p_{ij}/(1-p_{ij})] = \mu + \alpha_i + \eta_j$$

More generally, in GLIMs we create linear models for $g(m_{ij})$ using any strictly monotone function $g(\cdot)$, called a link function. The linear structures used for one-way ANOVA, two-way ANOVA, and higher-order ANOVA models all apply to GLIMs, and appropriate models can be found by comparing the fit of full and reduced models similar to methods for unbalanced ANOVA. In particular, cross-classified tables of counts often involve many factors.

We now examine in more depth ANOVA-type models for binomial data. Reconsider Everitt's 97 children cross-classified by the risk of their home environment and adversity of their school conditions. Rather than studying their verbal test score performance conditional on their classroom behavior as we did when discussing three-factor ANO we now examine models for whether the children display deviant or nondeviant behavior. In particular, we model the probability of a student falling into the deviant behavior category given their membership in the six home–school groups, write the probability of deviant behavior when in the *ij* group as p_{ij} and model the log-odds with analysis of variance type models,

$$\log[p_{ij}/(1-p_{ij})] = \mu_{ij}$$

The overparameterized model

$$\log[p_{ij}/(1-p_{ij})] = \mu + \alpha_i + \eta_j + (\alpha\eta)_{ij}$$

is equivalent to the original model. Neither of these models really accomplishes anything because they fit a separate parameter to every home–school category, so the models place no restrictions on the observations.

Table 7
Log-odds : additive model

		j (School)		
	μ_{ij}	1 (Low)	2 (Medium)	3 (High)
i	1 (N)	−2	−1	1
(Home)	2 (R)	0	1	3

Table 8
Log-odds : nonadditive model

		j (School)		
	μ_{ij}	1 (Low)	2 (Medium)	3 (High)
i	1 (N)	−2	−1	1
(Home)	2 (R)	−1	1	4

The additive model

$$\log[p_{ij}/(1-p_{ij})] = \mu + \alpha_i + \eta_j$$

constitutes a real restriction on the parameters.

Suppose the log-odds, the μ_{ij}s, satisfy Table 7. These have the structure of an additive model. In this case the log-odds for deviant behavior are two larger for at risk homes than for not at risk homes, e.g., $2 = 0-(-2) = 1-(-1) = 3-1$. This means that the odds are $e^2 = 7.4$ times larger for at risk homes. In particular, the log-odds of deviant behavior for low adversity schools and not at risk homes is -2 so the odds are $O_{11} = e^{-2} = 0.135$ and the probability if $p_{11} = O_{11}/[1+O_{11}] = 0.12$. Similarly, the odds and probability of deviant behavior for low adversity schools and at risk homes are $O_{21} = e^0 = 1$ and $p_{21} = 0.5$. The change in odds is $0.135 \times 7.4 = 1$. The 7.4 fold increase in odds is the same for all school conditions. Similarly, comparing low to high school adversities, the difference in log-odds is $3 = 1-(-2) = 3-0$, so the odds of deviant behavior are $e^3 = 20$ times greater in highly adverse schools, regardless of the home situation. The odds of deviant behavior in highly adverse schools and not at risk homes is $O_{13} = e^1 = 2.7$ which is 20 times greater than $O_{11} = 0.135$, the odds for low adversity schools and not at risk homes.

An example of a nonadditive model has log-odds as shown in Table 8. For low adversity schools the log-odds only differ by 1, for medium adversity schools they differ by 2, and for high adversity schools they differ by 3. Thus the effect of home conditions on the log-odds depends on the level of school adversity.

5. *Conclusions*

The fundamental ANOVA model is the one-way model. It specifies different mean values for different groups. When the groups are identified as combinations of two or more factors, models incorporating main effects and interactions become a useful device for examining the underlying structure of the data. Appropriate models can be identified by fitting sequences of successively smaller models and using general testing procedures to identify the models within each sequence that fit well. Having identified a model, it is important to interpret what that model suggests about the underlying data structure. ANOVA models, the sequential model fitting procedures, and the interpretations apply to both balanced and unbalanced data and to generalized linear models. The sequential model fitting simplifies in balanced ANOVA allowing for it to be summarized in an ANOVA table.

See also: Multivariate Analysis: Discrete Variables (Logistic Regression); Simultaneous Equation Estimates (Exact and Approximate), Distribution of

Bibliography

Agresti A 1990 *Categorical Data Analysis*. Wiley, New York
Christensen R 1996a *Plane Answers to Complex Questions: The Theory of Linear Models*, 2nd edn. Springer-Verlag, New York
Christensen R 1996b *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. Chapman and Hall, London
Christensen R 1997 *Log-linear Models and Logistic Regression*, 2nd edn. Springer-Verlag, New York
Everitt B S 1977 *The Analysis of Contingency Tables*. Chapman and Hall, London
Fienberg S E 1980 *The Analysis of Cross-classified Categorical Data*, 2nd edn. MIT Press, Cambridge, MA
Graybill F A 1976 *Theory and Application of the Linear Model*. Duxbury Press, North Scituate, MA
Hosmer D W, Lemeshow S 1989 *Applied Logistic Regression*. Wiley, New York
Lee Y, Nelder J A 1996 Hierarchical generalized linear models, with discussion. *Journal of the Royal Statistical Society, Series B* **58**: 619–56
McCullagh P, Nelder J A 1989 *Generalized Linear Models*, 2nd edn. Chapman and Hall, London
Nelder J A 1977 A reformation of linear models. *Journal of the Royal Statistical Society, Series A* **140**: 48–63
Scheffé H 1959 *The Analysis of Variance*. Wiley, New York
Searle S R 1971 *Linear Models*. Wiley, New York
Seber G A F 1977 *Linear Regression Analysis*. Wiley, New York

R. Christensen

Analytic Induction

Analytic induction (AI) is a research logic used to collect data, to develop analysis, and to organize the presentation of research findings. Its formal objective is causal explanation, a specification of the individually necessary and jointly sufficient conditions for the emergence of some part of social life. AI calls for the progressive redefinition of the phenomenon to be explained (the explanandum) and of explanatory factors (the explanans), such that a perfect (sometimes called 'universal') relationship is maintained. Initial cases are inspected to locate common factors and provisional explanations. As new cases are examined and initial hypotheses are contradicted, the explanation is reworked in one or both of two ways. The definition of the explanandum may be redefined so that troublesome cases either become consistent with the explanans or are placed outside the scope of the inquiry; or the explanans may be revised so that all cases of the target phenomenon display the explanatory conditions. There is no methodological value in piling up confirming cases; the strategy is exclusively qualitative, seeking encounters with new varieties of data in order to force revisions that will make the analysis valid when applied to an increasingly diverse range of cases. The investigation continues until the researcher can no longer practically pursue negative cases.

1. *The Methodology Applied*

Originally understood as an alternative to statistical sampling methodologies, 'analytic induction' was coined by Znaniecki (1934), who, through analogies to methods in chemistry and physics, touted AI as a more 'scientific' approach to causal explanation than 'enumerative induction' that produces probabilistic statements about relationships. After a strong but sympathetic critique by Turner (1953), AI shed the promise of producing laws of causal determinism that would permit prediction. The methodology subsequently became diffused as a common strategy for analyzing qualitative data in ethnographic research. AI is now practiced in accordance with Znaniecki's earlier (1928), less famous call for a phenomenologically grounded sociology. It primarily continues as a way to develop explanations of the interactional processes through which people develop homogeneously experienced, distinctive forms of social action.

The pioneering AI studies centered on turning points in personal biographies, most often the phase of commitment to behavior patterns socially defined as deviant, such as opiate addiction (Lindesmith 1968), embezzlement (Cressey 1953), marijuana use (Becker 1953), conversion to a millenarian religious sect (Lofland and Stark 1965), abortion seeking (Manning 1971), and youthful theft (West 1978, a rare study focusing more on desistance than onset). Studies at the end of the twentieth century have addressed more situationally specific and morally neutral phenomena. These include occupational perspectives exercised in