Center for Quality and Applied Statistics
Kate Gleason College of Engineering
Rochester Institute of Technology

Technical Report 2004–6

August 16, 2004

The APD Method of Understanding Variances and Expected Mean Squares

Joseph G. Voelkel
Center for Quality and Applied Statistics
Rochester Institute of Technology
jgvcqa@rit.edu

The APD Method of Understanding Variances and Expected Mean Squares

by

Joseph G. Voelkel

Abstract

The usual formulas for the sample variance and main-effect mean squares can be rewritten as a sum of squares of all pairwise differences (APD) of random variables. Under common model assumptions, such random-variable pairs are uncorrelated, and this leads to a simpler, more intuitive, approach to finding associated expectations. We illustrate this with a variety of examples, and show how this approach can provide insight into ANOVA models. We also show how this approach can be used for solving other problems, including finding expected mean squares for interaction terms, expected value of s^2 in finite-population sampling, and $\mathrm{Var}\,(s^2)$ both in the i.i.d. case and an extension to the non-identically-distributed case.

1 The APD Method

The APD (all paired differences) method is based on an algebraic relationship between the sample variance of a set of data and the squared differences of each distinct pair

Key Words and Phrases: finite-population sampling; restricted and unrestricted forms; ANOVA

of data values. The relationship may be expressed in any of the following ways:

$$\begin{aligned}
s^2 &= \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 / (n-1) && (1)\\
&= (1/4) \sum_{j=1}^{n} \sum_{i=1}^{n} (X_i - X_j)^2 \Big/ \binom{n}{2} && (2)\\
&= (1/2) \sum_{i<j} (X_i - X_j)^2 \Big/ \binom{n}{2} && (3)\\
&= \sum_{i<j} (X_i - X_j)^2 / n(n-1). && (4)
\end{aligned}$$

The usefulness of the equivalence of (1) with the other formulas will soon be seen. To prove it, first start with (2). Summing over each i and each j is equivalent to summing over the three sets $\{i < j\}$, $\{i = j\}$, and $\{i > j\}$. The middle set contributes nothing to the sum, while the first and last sets contribute equally. This shows that (2) is equivalent to (3), and hence (4). To show that (2) is equivalent to (1), write (2) as

$$\begin{aligned}
&(1/4) \sum_{j=1}^{n} \sum_{i=1}^{n} \left(X_i - \bar{X} + \bar{X} - X_j\right)^2 \Big/ \binom{n}{2}\\
= \;& (1/4) \sum_{j=1}^{n} \sum_{i=1}^{n} \left[\left(X_i - \bar{X}\right)^2 + \left(X_j - \bar{X}\right)^2 + 2\left(X_i - \bar{X}\right)\left(X_j - \bar{X}\right)\right] \Big/ \binom{n}{2}\\
= \;& (1/4) \left[n \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + n \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2\right] \Big/ \binom{n}{2}\\
= \;& \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 / (n-1).
\end{aligned}$$

2 Some Fundamental Examples of the Use of APD

The first example of the use of APD is to gain insight into the population and sample variances. The former, σ^2, is equal to one-half of the expected squared difference between two values—$\mathrm{E}\left[(X_i - X_j)^2\right]/2$. However, using (3), and noting that the number of terms in the sum is $\binom{n}{2}$, we see that s^2, is simply one-half of the average squared distances among all the distinct pairs of the data. So the result that hold in expectation also holds on a sample-to-sample basis.

A more important application of the APD method is that it arguably provides both a simpler and more insightful approach to finding expectations involving terms that are "s^2" in nature than the commonly used methods. We start by showing that s^2 is unbiased for σ^2 under the usual assumptions. We use the notation

$$\{X_i\} \sim uncorr\left(\mu_i, \sigma_i^2\right),\ i = 1, \ldots, n$$

to mean that $X_1, X_2, \ldots, X_n$ are uncorrelated random variables with $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$. We omit subscript limits when they are obvious.

Example 1 *If* $\{X_i\} \sim uncorr\left(\mu, \sigma^2\right),\ i = 1, \ldots, n,$ *then* $\mathrm{E}\left(s^2\right) = \sigma^2$.

Proof. The assumptions directly imply that $\mathrm{E}\left((X_i - X_j)^2\right) = \mathrm{Var}(X_i - X_j) + [\mathrm{E}(X_i - X_j)]^2 = 2\sigma^2$, so, using (3),

$$\begin{aligned} \mathrm{E}\left(s^2\right) &= (1/2)\sum_{i<j} \mathrm{E}\left[(X_i - X_j)^2\right] / \binom{n}{2} \\ &= (1/2)\sum_{i<j} 2\sigma^2 / \binom{n}{2} \\ &= \sigma^2. \end{aligned}$$

The elegance of the method here is that *the covariances of all terms under investigation* (here, X_i and X_j) *are zero.* This circumvents the annoying task that the standard proof entails of either (1) handling covariances (here, between X_i and $\bar{X}$) or expanding $\left(X_i - \bar{X}\right)^2$ into terms of the form $X_i X_j / n$. However, even if the covariances between X_i and X_j were non-zero, the APD approach could use these covariances directly and simply, as we show in Example 4.

We next investigate what happens when the variances of the $\{X_i\}$, and then the means of the $\{X_i\}$, may differ.

Example 2 *If* $\{X_i\} \sim uncorr\left(\mu, \sigma_i^2\right),\ i = 1, \ldots, n,$ *then* $\mathrm{E}\left(s^2\right) = \sum_{i=1}^n \sigma_i^2 / n$.

Proof. We have

$$\begin{aligned}
\mathrm{E}\left(s^2\right) &= (1/2)\sum_{i<j}\mathrm{E}\left[(X_i - X_j)^2\right]/\binom{n}{2} \\
&= (1/2)\sum_{i<j}\left(\sigma_i^2 + \sigma_j^2\right)/\binom{n}{2} \\
&= (1/2)\sum_{i=1}^{n}(n-1)\,\sigma_i^2/\binom{n}{2} \\
&= \sum_{i=1}^{n}\sigma_i^2/n,
\end{aligned}$$

where the third line is obtained by noting that each σ_i^2 appears in exactly $(n-1)$ terms in the sum.

Example 3 *If* $\{X_i\} \sim$ *uncorr* (μ_i, σ^2), $i = 1, \ldots, n$, *then* $\mathrm{E}\left(s^2\right) = \sigma^2 + \sum_{i=1}^{n}\left(\mu_i - \bar{\mu}\right)^2/(n-1)$.

Proof. Because

$$\begin{aligned}
\mathrm{E}\left[(X_i - X_j)^2\right] &= \mathrm{Var}\left(X_i - X_j\right) + \left[\mathrm{E}\left(X_i - X_j\right)\right]^2 \\
&= 2\sigma^2 + \left(\mu_i - \mu_j\right)^2,
\end{aligned}$$

then

$$\begin{aligned}
\mathrm{E}\left(s^2\right) &= (1/2)\sum_{i<j}\left[2\sigma^2 + \left(\mu_i - \mu_j\right)^2\right]/\binom{n}{2} \\
&= \sigma^2 + \sum_{i=1}^{n}\left(\mu_i - \bar{\mu}\right)^2/(n-1),
\end{aligned}$$

where the second term on the last line was found by using the algebraic relation from (3) to (1) applied to the $\{\mu_i\}$ instead of the $\{X_i\}$. All three examples could be subsumed under a general one of different means and variances, but the above approach seems more instructive.

Next we examine the finite population case.

Example 4 . *If* $\{X_i\}$, $i = 1, \ldots, n$, *is a random sample from a population of size* N, *with population variance* σ^2, *then* $\mathrm{E}\left(s^2\right) = \sigma^2 N/(N-1)$.

Proof. Without loss of generality, assume the population mean μ is zero. To use the APD method, we need to find, for $i \neq j$,

$$\begin{aligned} \mathrm{E}\left((X_i - X_j)^2\right) &= \mathrm{Var}(X_i) + \mathrm{Var}(X_j) - 2\,\mathrm{E}(X_i X_j) \\ &= 2\sigma^2 - 2\,\mathrm{E}(X_i X_j). \end{aligned}$$

But

$$\begin{aligned} \mathrm{E}(X_i X_j) &= \mathrm{E}(\mathrm{E}(X_i X_j \,|X_j)) \\ &= \mathrm{E}(X_j\,\mathrm{E}(X_i \,|X_j)) \\ &= \mathrm{E}\left(X_j\left(\frac{N\mu - X_j}{N-1}\right)\right) \\ &= \mathrm{E}\left(X_j\left(\frac{-X_j}{N-1}\right)\right) \\ &= -\frac{\sigma^2}{N-1}. \end{aligned}$$

The third line in this last equation is simply "average of all X's except for X_j." This shows in a direct way the difference between sampling from finite and infinite populations. So, $\mathrm{E}\left((X_i - X_j)^2\right) = 2\sigma^2(1 + 1/(N-1)) = 2\sigma^2 N/(N-1)$, and hence $\mathrm{E}(s^2) = \sigma^2 N/(N-1)$. Once again, we have avoided the excess algebra required by the customary proof—see, e.g., Cochran (1977)—while at the same time providing insight into the problem. Note, for example, that $\mathrm{E}\left((X_i - X_j)^2\right)$ is larger here than in the infinite-population case, which makes sense—crudely speaking, sampling one unit below the mean implies that the next unit is more likely to be above the mean.

3 Simple ANOVA Examples

We use the notation of Lorenzen and Anderson (1993) and make the usual assumptions to find expected mean squares:

1. The ε's under consideration are assumed to be random variables with mean 0 and variance σ^2.

2. Random variables with the same letter(s) appearing on the right side of the equations have means of 0 and a common variance, denoted by using the letter(s) as a subscript. For example, if $\{B_j\}$ are random variables, they have mean 0 and variance σ_B^2.

3. All random variables on the right side of the equation are uncorrelated with each other unless specifically stated otherwise. For example, if there are n ε_i's and J B_j's, the $n+J$ random variables are uncorrelated.

We use the standard "bar" and "dot" notation to indicate averaging.

Example 5 *(Fixed effects, one-way ANOVA). Let*

$$Y_{ij} = \mu + A_i + \varepsilon_{j(i)},\ i = 1, \ldots, I,\ j = 1, \ldots, J,$$

where the $\{A_i\}$ *are subject to the restriction* $\sum A_i = 0$*. Find* $\mathrm{E}\,[MS_A]$*.*

Solution. For this model, $MS_A = J\sum_{i=1}^{I}\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 / (I-1)$. By the assumptions, it follows that $\left\{\bar{Y}_{i.}\right\} \sim uncorr\,(\mu + A_i, \sigma^2/J)$, so the results of Example 3 can be used. Define

$$Q[A] = \sum_{i=1}^{I} A_i^2 / (I-1)\,.$$

Then, because $\bar{A}_. = 0$,

$$\begin{aligned}
\mathrm{E}\,[MS_A] &= J\sum_{i=1}^{I}\mathrm{E}\left[\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2\right] / (I-1) \\
&= J\left(\sigma^2/J + \sum_{i=1}^{I}\left(A_i - \bar{A}_.\right)^2 / (I-1)\right) \\
&= \sigma^2 + JQ[A].
\end{aligned}$$

(It is instructive to derive this using the idea of Example 3 directly on the $\left\{\bar{Y}_{i.}\right\}$ rather than simply performing the substitution as here.)

Example 6 *(Random effects, one-way ANOVA). Let*

$$Y_{ij} = \mu + A_i + \varepsilon_{j(i)},\ i = 1, \ldots, I,\ j = 1, \ldots, J,$$

where the $\{A_i\}$ *are random (with mean* 0*, variance* σ_A^2*, and so on, as noted earlier). Find* $\mathrm{E}\,[MS_A]$*.*

Solution. Again, $MS_A = J\sum_{i=1}^{I}\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 / (I-1)$. We will find $\mathrm{E}\,[MS_A]$in two different ways. First, because $\bar{Y}_{i.} = \mu + A_i + \bar{\varepsilon}_{.(i)}$, then $\left\{\bar{Y}_{i.}\right\} \sim uncorr\,(\mu, \sigma_A^2 + \sigma^2/J)$. The results of Example 1 can be used directly, leading to

$$\begin{aligned}
\mathrm{E}\,[MS_A] &= J\left(\sigma_A^2 + \sigma^2/J\right) \\
&= J\sigma_A^2 + \sigma^2.
\end{aligned}$$

Another approach to the problem is to use the relation $\mathrm{E}[MS_A] = \mathrm{E}[\mathrm{E}[MS_A | \{A_i\}]]$. This conditioning on the $\{A_i\}$ is insightful. For example, if factor A represents an "operator" effect, where I operators have been selected at random from a much larger group, then $\mathrm{E}[MS_A | \{A_i\}]$ indicates the EMS for *this* group of operators (fixed effects for *these* $\{A_i\}$), while the second expectation is over *all* operators (random effects). We can use the fixed-effects example as a basis, but noting that the restriction "$\sum A_i = 0$" no longer applies because of random sampling. This leads to

$$\begin{aligned}
\mathrm{E}[MS_A] &= \mathrm{E}\left[\mathrm{E}\left[J\sum_{i=1}^{I}\left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 / (I-1) \mid \{A_i\}\right]\right] \\
&= \mathrm{E}\left[\sigma^2 + J\sum_{i=1}^{I}\left(A_i - \bar{A}_{.}\right)^2 / (I-1)\right] \\
&= \sigma^2 + J\sigma_A^2.
\end{aligned}$$

These results apply when sampling $\{A_i\}$ from an infinite population. If they are sampled from a finite population, we can find $\mathrm{E}[MS_A]$ in the above by using the finite population correction factor as shown in Example 4.

4 Two-way ANOVA with Fixed Effects and Sampling

Consider a design intended to study a part dimension. Say the design has two crossed factors, associated with temperatures $\{A_i\}$ and pressure $\{B_j\}$, where both factors are fixed, and that each of these IJ treatment combinations are replicated K times. For each replicate L parts are randomly selected from the large number produced for that run and all $IJKL$ parts are then measured in a random order. The natural ANOVA model is

$$Y_{ijkl} = \mu + A_i + B_j + AB_{ij} + C_{k(ij)} + \varepsilon_{l(ijk)},$$

with $\sum A_i = \sum B_j = 0$, $\sum_j AB_{ij} = 0$ for each i and for each j, and the $\left\{C_{k(ij)}\right\}$ and the $\left\{\varepsilon_{l(ijk)}\right\}$ are random.

Example 7 *Find* $\mathrm{E}(MS_B)$.

Solution. We start with $MS_B = IKL\sum_{j=1}^{J}\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2 / (J-1)$. Because $\bar{Y}_{.j..} = \mu + B_j + \overline{C}_{.(.j)} + \bar{\varepsilon}_{.(.j.)}$, then $\left\{\bar{Y}_{.j..}\right\} \sim uncorr\,(\mu + B_j, \sigma_C^2/IK + \sigma^2/IKL)$, so Example 3 applies:

$$\begin{aligned}
\mathrm{E}\left[MS_B\right] &= IKL\sum_{j=1}^{J}\mathrm{E}\left[\left(\bar{Y}_{.j..}-\bar{Y}_{....}\right)^2\right]/\left(J-1\right)\\
&= IKL\left(\sigma_C^2/IK+\sigma^2/IKL+Q\left[B\right]\right)\\
&= \sigma^2+L\sigma_C^2+IKLQ\left[B\right].
\end{aligned}$$

Example 8 *Find* $\mathrm{E}\left(MS_{AB}\right)$.

Solution. We show three approaches here. First, we show and then use the well-known relation

$$SS_{AB}=SS_{A+B+AB}-SS_A-SS_B.$$

Using poor but obvious notation, write $\bar{Y}_{ij..}$ as $\bar{Y}_{ij}$, $\bar{Y}_{....}$ as $\bar{Y}$ etc., for brevity. Then

$$\begin{aligned}
SS_{AB} &= KL\sum_{ij}\left(\bar{Y}_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}\right)^2 &(5)\\
&= KL\sum_{ij}\left(\bar{Y}_{ij}-\left(\bar{Y}+\bar{Y}_{i.}-\bar{Y}+\bar{Y}_{.j}-\bar{Y}\right)\right)^2\\
&= KL\sum_{ij}\left(\left(\bar{Y}_{ij}-\bar{Y}\right)-\left(\bar{Y}_{i.}-\bar{Y}\right)-\left(\bar{Y}_{.j}-\bar{Y}\right)\right)^2 &(6)\\
&= KL\sum_{ij}\left(\bar{Y}_{ij}-\bar{Y}\right)^2-JKL\sum_{i}\left(\bar{Y}_{i.}-\bar{Y}\right)^2-IKL\sum_{j}\left(\bar{Y}_{.j}-\bar{Y}\right)^2\\
&= SS_{A+B+AB}-SS_A-SS_B.
\end{aligned}$$

(Certain cross-product sums in (6) are zero and others simplify.) Treat each of the three SS's separately. For the first,

$$\left\{\bar{Y}_{ij}\right\}\sim uncorr\left(\mu+A_i+B_j+AB_{ij},\sigma_C^2/K+\sigma^2/KL\right),$$

so $s_{ij}^2\equiv\sum_{ij}\left(\bar{Y}_{ij}-\bar{Y}\right)^2/\left(IJ-1\right)$ has expectation

$$\begin{aligned}
&\sigma_C^2/K+\sigma^2/KL+\sum_{ij}\left(A_i+B_j+AB_{ij}\right)^2/\left(IJ-1\right)\\
=\ &\sigma_C^2/K+\sigma^2/KL+\left(J\left(I-1\right)Q\left[A\right]+I\left(J-1\right)Q\left[B\right]+dfQ\left[AB\right]\right)/\left(IJ-1\right)
\end{aligned}$$

where $Q[AB] = \sum_{ij} AB_{ij}^2/df$ and $df \equiv (I-1)(J-1)$. The second and third values are derived from the previous example. This results in an $\mathrm{E}[MS_{AB}]$ of

$$
\begin{aligned}
& \mathrm{E}\left[s_{ij}^2\right](IJ-1)KL/df - \mathrm{E}[MS_A](I-1)/df - \mathrm{E}[MS_B](J-1)/df \\
= \; & \sigma_C^2(IJ-1)L/df + \sigma^2(IJ-1)/df \\
& + (J(I-1)Q[A] + I(J-1)Q[B] + dfQ[AB])KL/df \\
& - \left(\sigma^2 + L\sigma_C^2 + JKLQ[A]\right)(I-1)/df \\
& - \left(\sigma^2 + L\sigma_C^2 + IKLQ[B]\right)(J-1)/df \\
= \; & \sigma^2 + L\sigma_C^2 + KLQ[AB].
\end{aligned}
$$

We have once again avoided working with correlated random variables. However, the elegance of the technique now is obscured by the algebra.

A second way to solve the problem is to use the model to rewrite the sums of squares as

$$
\begin{aligned}
SS_{AB} &= KL\sum_{ij}\left(\bar{Y}_{ij} - \left(\bar{Y} + \bar{Y}_{i.} - \bar{Y} + \bar{Y}_{.j} - \bar{Y}\right)\right)^2 \\
&= KL\sum_{ij}\left[AB_{ij} + \left(\bar{C}_{.(ij)} - \bar{C}_{.(i.)} - \bar{C}_{.(.j)} + \bar{C}_{.(..)}\right) + \right. \qquad (7) \\
& \qquad \left.\left(\bar{\varepsilon}_{.(ij.)} - \bar{\varepsilon}_{.(i..)} - \bar{\varepsilon}_{.(.j.)} + \bar{\varepsilon}_{.(...)}\right)\right]^2.
\end{aligned}
$$

Regarding this as a sum of three sets of terms, the sums of the cross-product terms disappear upon taken expectations. It is easy to show, using the expansion idea in (5), that if $\{X_{ij}\} \sim uncorr\left(\mu, \sigma^2\right),\ i = 1, \ldots, I,\ \ j = 1, \ldots, J$, that

$$
\begin{aligned}
& \mathrm{E}\left(\sum_{ij}\left(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}\right)^2\right) \\
= \; & \mathrm{E}\left(\sum_{ij}\left(X_{ij} - \bar{X}_{..}\right)^2 - J\sum_i\left(\bar{X}_{i.} - \bar{X}_{..}\right)^2 - I\sum_j\left(\bar{X}_{.j} - \bar{X}_{..}\right)^2\right) \\
= \; & (IJ-1)\sigma^2 - (I-1)\sigma^2 - (J-1)\sigma^2 \\
= \; & (I-1)(J-1)\sigma^2.
\end{aligned}
$$

Applying this to the MS version of (7) yields

$$
\mathrm{E}(MS_{AB}) = KL\left(Q[AB] + \sigma_C^2/K + \sigma^2/KL\right),
$$

as before. This is a bit more direct but still not very satisfying.

The third approach, which is most strongly in the spirit of APD and which provides the most insight into the problem, involves extending the APD idea to two-way sets of numbers. Consider data $\{X_{ij},\ i=1,\ldots,I,\ j=1,\ldots,J\}$. The two-way APD identity is

$$\begin{aligned}
&\sum_i \sum_j \left(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}\right)^2 \\
= \ &\frac{1}{4IJ} \sum_i \sum_j \sum_{i'} \sum_{j'} \left[(X_{ij} - X_{ij'}) - (X_{i'j} - X_{i'j'})\right]^2,
\end{aligned}$$

where i and i' range from 1 to I and j and j' range from 1 to J. (Note that the differences of the differences, $(X_{ij} - X_{ij'}) - (X_{i'j} - X_{i'j'})$, are precisely how one would calculate a two-factor interaction if the first factor occured at levels i and $i^{'}$ and the second at levels j and $j^{'}$). To prove this identity, replace X_{ij} in the second line by the group $G_{ij} = X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}$, with similar replacements for $X_{ij'}$, $X_{i'j}$, and $X_{i'j'}$. Note that $(G_{ij} - G_{ij'}) - (G_{i'j} - G_{i'j'}) = (X_{ij} - X_{ij'}) - (X_{i'j} - X_{i'j'})$. An expansion of $(G_{ij} - G_{ij'} - G_{i'j} + G_{i'j'})^2$ shows that the cross products terms disappear in the summation. Because, e.g., $\sum_i \sum_j \sum_{i'} \sum_{j'} G_{ij}^2 = IJ \sum_i \sum_j G_{ij}^2$, the proof is complete.

Now, returning to the original example, let $H_{ij} = AB_{ij} + \bar{C}_{.(ij)} + \bar{\varepsilon}_{.(ij.)}$. Then

$$\begin{aligned}
SS_{AB} &= KL \sum_{ij} \left(\bar{Y}_{ij} - \left(\bar{Y} + \bar{Y}_{i.} - \bar{Y} + \bar{Y}_{.j} - \bar{Y}\right)\right)^2 \\
&= KL \frac{1}{4IJ} \sum_i \sum_j \sum_{i'} \sum_{j'} \left[\left(\bar{Y}_{ij} - \bar{Y}_{ij'}\right) - \left(\bar{Y}_{i'j} - \bar{Y}_{i'j'}\right)\right]^2 \\
&= KL \frac{1}{4IJ} \sum_i \sum_j \sum_{i'} \sum_{j'} \left[(H_{ij} - H_{ij'}) - (H_{i'j} - H_{i'j'})\right]^2 \\
&= KL \frac{1}{4IJ} \sum_i \sum_j \sum_{i'} \sum_{j'} \Big(\left[(AB_{ij} - AB_{ij'}) - (AB_{i'j} - AB_{i'j'})\right]^2 \\
&\quad + \left[\left(\bar{C}_{\cdot(ij)} - \bar{C}_{\cdot(ij')}\right) - \left(\bar{C}_{\cdot(i'j)} - \bar{C}_{\cdot(i'j')}\right)\right]^2 \\
&\quad + \left[\left(\bar{\varepsilon}_{.(ij.)} - \bar{\varepsilon}_{.(ij'.)}\right) - \left(\bar{\varepsilon}_{.(i'j.)} - \bar{\varepsilon}_{.(i'j'.)}\right)\right]^2\Big),
\end{aligned}$$

where the last equality holds because the cross-product terms disappear in the summation.

Note that any squared group equals 0 if $i = i'$ or $j = j'$. For a fixed value of i and j, there are $I + J - 1$ such terms, so there are $IJ - I - J + 1 = (I-1)(J-1)$ remaining terms. Because of this, the first group in this last equality is simply $KL\,(I-1)\,(J-1)\,Q\,[AB]$.

When we take expectations of the second group, for a fixed value of i and j, there are once again $I+J-1$ such 0 terms. For each of the remaining $IJ-I-J+1=(I-1)(J-1)$ terms

$$\begin{aligned} & \mathrm{E}\left(\left[\left(\bar{C}_{\cdot(ij)}-\bar{C}_{\cdot(ij')}\right)-\left(\bar{C}_{\cdot(i'j)}-\bar{C}_{\cdot(i'j')}\right)\right]^2\right) \\ = \; & 4\,\mathrm{Var}\left(\bar{C}_{\cdot(ij)}\right)=4\sigma_C^2\,/\,K. \end{aligned}$$

So, the sum over i and j is $KL\frac{1}{4IJ}\sum_i\sum_j(I-1)(J-1)\,4\sigma_C^2\,/\,K=L(I-1)(J-1)\,\sigma_C^2$. Similarly, the last group's expectation is $(I-1)(J-1)\,\sigma^2$, so

$$\mathrm{E}\left(MS_{AB}\right)=KLQ\left[AB\right]+L\sigma_C^2+\sigma^2.$$

This proof shows more clearly (1) how the interaction term really consists of the more intuitive "differences of differences," (2) where the $(I-1)(J-1)$ term arises, and (3) how terms such as "K" arise in the expectation.

This can be extended to higher-order interactions. For example, the summation part of the SS associated with a three-way interaction is 1/8 of the sum of the "differences of differences of differences."

5 Unrestricted and Restricted Forms of ANOVA Models

The APD method also leads to insight into different so-called *forms* of ANOVA models, where the forms are restricted and unrestricted. See, e.g., Lorenzen and Anderson (1993). Consider a model with two crossed factors, say machines $\{A_i\}$ and operators $\{B_j\}$. Assume that the machine factor is fixed, but the operator factor is random. Suppose that each of the IJ machine/operator combinations measures a random selection of K parts, in a random order. Then the natural ANOVA model is

$$Y_{ijk}=\mu+A_i+B_j+AB_{ij}+\varepsilon_{k(ij)}, \tag{8}$$

with $\sum A_i=0$, $\mathrm{Var}\left(B_j\right)=\sigma_B^2$. But what should be said about the $\{AB_{ij}\}$? In the *unrestricted form of the model*, the $\{AB_{ij}\}$ are fully uncorrelated with each other and have $\mathrm{Var}\left(AB_{ij}\right)=\sigma_{AB}^2$.

Example 9 *In the unrestricted form of the model (8), find* $\mathrm{E}\left(MS_B\right)$.

Solution. For the model (8), $MS_B=IK\sum_{j=1}^{J}\left(\bar{Y}_{.j.}-\bar{Y}_{...}\right)^2/\left(J-1\right)$. Because $\bar{Y}_{.j.}=\mu+B_j+\overline{AB}_{.j}+\bar{\varepsilon}_{.(.j)}$, then $\left\{\bar{Y}_{.j.}\right\}\sim$ *uncorr* $\left(\mu,\sigma_B^2+\sigma_{AB}^2/I+\sigma^2/IK\right)$, so

Example 1 applies:

$$\begin{aligned}
\mathrm{E}\left[MS_B\right] &= IK\sum_{j=1}^{J}\mathrm{E}\left[\left(\bar{Y}_{.j.}-\bar{Y}_{...}\right)^2\right]/\left(J-1\right)\\
&= IK\left(\sigma_B^2+\sigma_{AB}^2/I+\sigma^2/IK\right)\\
&= IK\sigma_B^2+K\sigma_{AB}^2+\sigma^2.
\end{aligned}$$

In the *restricted form of the model*, the nature of the $\{AB_{ij}\}$ changes. We now require that the $\{AB_{ij}\}$ sum to zero for every level of the random factor: $\sum_{i=1}^{a} AB_{ij} = 0$ for each j. This may be a reasonable restriction, because it can be considered as forcing the fixed effects (of machines) to sum to zero for each level of the random factor (operator). Mathematically, this is equivalent to re-labeling AB_{ij} in the unrestricted form of (8) as AB'_{ij} and defining AB_{ij} in the restricted form of (8) as $AB_{ij} = AB'_{ij} - \overline{AB}'_{.j}$ (and see Schwarz (1993)).

Example 10 *In the restricted form of the model (8), find* $\mathrm{E}\left(MS_B\right)$.

Solution. Again, $MS_B = IK\sum_{j=1}^{J}\left(\bar{Y}_{.j.}-\bar{Y}_{...}\right)^2/\left(J-1\right)$. However, now $\bar{Y}_{.j.} = \mu + B_j + \overline{AB}_{.j} + \bar{\varepsilon}_{.(.j)} = \mu + B_j + \bar{\varepsilon}_{.(.j)}$, so $\left\{\bar{Y}_{.j.}\right\} \sim uncorr\left(\mu, \sigma_B^2 + \sigma^2/IK\right)$, yielding:

$$\begin{aligned}
\mathrm{E}\left[MS_B\right] &= IK\sum_{j=1}^{J}\mathrm{E}\left[\left(\bar{Y}_{.j.}-\bar{Y}_{...}\right)^2\right]/\left(J-1\right)\\
&= IK\left(\sigma_B^2+\sigma^2/IK\right)\\
&= IK\sigma_B^2+\sigma^2.
\end{aligned}$$

There appears to be no consensus as to which form of the model is more reasonable. For an idea at some issues that need to be considered, see McLean, Sanders, and Stroup (1991).

6 APD to find $\mathrm{Var}\left(s^2\right)$

The APD method can also be used to find more complex expressions. We show here how to find $\mathrm{Var}\left(s^2\right)$. The solution requires a bit of effort, but its straightforward nature typifies the APD method.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a finite fourth moment. Without loss of generality, let $\mathrm{E}\left(X_i\right) = 0$. Let $\mathrm{Var}\left(X_i\right) = \mathrm{E}\left(X_i^2\right) = \sigma^2$ and $\mathrm{E}\left(X_i^4\right) = \mu_4$.

# terms	ij with	Name	Expectation
n	ii or jj or $i'i'$	Zero	0
2	ij or ji	Direct	$2\mu_4 + 6\sigma^4$
$4(n-2)$	ij', $i'j$, ji', or $j'i$	Indirect	$\mu_4 + 3\sigma^4$
$(n-2)(n-3)$	$i'j'$	Independent	$4\sigma^4$

Table 1: Terms in Summation of s^4

Then $\operatorname{Var}(s^2) = \mathrm{E}(s^4) - (\mathrm{E}(s^2))^2 = \mathrm{E}(s^4) - \sigma^4$. To find $\mathrm{E}(s^4)$, we use (2) to write

$$16\binom{n}{2}^2 s^4 = \left[\sum_{j=1}^{n}\sum_{i=1}^{n}(X_i - X_j)^2\right]^2. \tag{9}$$

Before simplifying, the sum in the brackets has n^2 terms—we will call this the *inner sum*—and squaring this produces n^4 terms—the *outer sum*—of the form $(X_a - X_b)^2(X_c - X_d)^2$. Some terms are 0, and others include either statistically independent or dependent cross-product parts. To understand this in the general case, represent $(X_i - X_j)^2$ as ij. Then all n^2 terms in the inner sum can be written as a $n \times n$ matrix, as shown in (10) for the case $n = 5$. (The bolded and italicized terms will be explained shortly.)

$$\mathbf{M} = \begin{bmatrix} 11 & 12 & \mathit{13} & \mathit{14} & \mathit{15} \\ 21 & 22 & \mathit{23} & \mathit{24} & \mathit{25} \\ \mathit{31} & \mathit{32} & 33 & \mathbf{34} & \mathbf{35} \\ \mathit{41} & \mathit{42} & \mathbf{43} & 44 & \mathbf{45} \\ \mathit{51} & \mathit{52} & \mathbf{53} & \mathbf{54} & 55 \end{bmatrix}. \tag{10}$$

Then (9) can be found by multiplying each of the n^2 cells in the matrix by each of the n^2 cells in the same matrix and summing the result. This can be written as $\mathbf{1}'(\mathbf{M} \otimes \mathbf{M})\mathbf{1}$, where $\mathbf{1}$ is an $n^4 \times 1$ vector of 1's and $\otimes$ is the Kronecker product. But we will solve this in a direct fashion.

There are n diagonal elements in $\mathbf{M}$, each of which contribute nothing to the outer sum. So $n(n^2) = n^3$ terms in the outer sum are 0. Now consider an off-diagonal element of $\mathbf{M}$, say $ij = 12$. There are $n^2 - n$ such elements. Let i' (or j') denote a generic value of the index that is equal to neither i nor j. So here, i' =3, 4, or 5, and similarly for j'. Each such ij will be involved in n^2 terms in the outer sum, and these terms in turn can be subdivided as shown in the first three columns of Table 1.

To continue the example with $ij = 12$, consider the matrix in (10) with respect to Table 1. There are five Zero terms (diagonals, 11 through 55), two Direct ones (12 and 21), twelve Indirect ones, italicized, and six Independent ones, bolded. By expanding the terms, it is straightforward to find the expected value for each term—these results appear in the last column of Table 1. By adding up all these terms for a particular ij, multiplying by $n(n-1)$ (the number of such ij combinations), and simplifying, we have

$$\begin{aligned} \mathrm{E}\left(s^4\right) &= \frac{n(n-1)\left(4(n-1)\mu_4 + 4\left(n^2-2n+3\right)\sigma^4\right)}{16\binom{n}{2}^2} \\ &= \frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4 \end{aligned}$$

so

$$\begin{aligned} \mathrm{Var}\left(s^2\right) &= \frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4 - \sigma^4 \\ &= \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\sigma^4. \end{aligned}$$

If we define the coefficient of kurtosis to be $\gamma_2 = \mu_4/\sigma^4 - 3$, we can rewrite this as

$$\mathrm{Var}\left(s^2\right) = \sigma^4\left(\frac{\gamma_2}{n} + \frac{2}{n-1}\right),$$

a result that appears, but is not derived, in Scheffé (1959, p. 83).

The results for the more general case, where $\mathrm{Var}(X_i) = \sigma_i^2$ and $\mathrm{E}\left(X_i^4\right) = \mu_{4i}$, can also be found. For example, with the APD approach above, it is easy to show that the expectations in Table 1 respectively become 0, $\mu_{4i} + \mu_{4j} + 6\sigma_i^2\sigma_j^2$, $\mu_{4i} + \sigma_i^2\sigma_j^2 + \sigma_i^2\sigma_{j'}^2 + \sigma_j^2\sigma_{j'}^2$ (for ij with ij'), and $\left(\sigma_i^2 + \sigma_j^2\right)\left(\sigma_{i'}^2 + \sigma_{j'}^2\right)$. Because the elements of these terms are either of the form μ_{4i} or $\sigma_i^2\sigma_j^2$ $(j \neq i)$, the linearity and symmetry of the formula lead to a solution of

$$\mathrm{Var}\left(s^2\right) = \frac{\overline{\mu_{4.}}}{n} + \frac{3-n}{n(n-1)}\left(\overline{\sigma_.^2}\right)^2,$$

where $\overline{\mu_{4.}} = \sum \mu_{4i}/n$ and $\overline{\sigma_.^2} = \sum \sigma_i^2/n$.

References

Cochran, W. G. (1977), *Sampling Techniques* (2nd ed.), New York: Wiley.

Lorenzen, T. J., and Anderson, V. L. (1993), *Design of Experiments: A No-Name Approach*, New York: Marcel Dekker.

McLean, R. A., Sanders, W. L., and Stroup, W. (1991), "A unified approach to mixed linear models," *The American Statistician*, 45, 54–64.

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley.

Schwartz, C. J. (1993), "The mixed-model ANOVA: The truth, the computer packages, the books. Part I: Balanced data," *The American Statistician*, 47, 48–58. (Comment: (1993), 47, 316. Reply: (1995), 49, 324.)