

When Does Blocking Help?

Teacher Notes, Part I

The purpose of blocking is frequently described as “reducing variability.” However, this phrase carries little meaning to most beginning students of statistics. This activity, consisting of three rounds of simulation, is designed to illustrate what reducing variability really means in this context. In fact, students should see that a better description than “reducing variability” might be “attributing variability”, or “reducing *unexplained* variability”.

The activity can be completed in a single 90-minute class or two classes of at least 45 minutes. For shorter classes you may wish to extend the simulations over two days. It is important that students understand not only what to do but also why they do what they do.

Background

Here is the specific problem that will be addressed in this activity:

A set of 24 dogs (6 of each of four breeds; 6 from each of four veterinary clinics) has been randomly selected from a population of dogs older than eight years of age whose owners have permitted their inclusion in a study. Each dog will be assigned to exactly one of three treatment groups. Group “Ca” will receive a dietary supplement of calcium, Group “Ex” will receive a dietary supplement of calcium and a daily exercise regimen, and Group “Co” will be a control group that receives no supplement to the ordinary diet and no additional exercise. All dogs will have a bone density evaluation at the beginning and end of the one-year study. (The bone density is measured in Hounsfield units by using a CT scan.) The goals of the study are to determine (i) whether there are different changes in bone density over the year of the study for the dogs in the three treatment groups; and if so, (ii) how much each treatment influences that change in bone density.

Mechanics of the Simulations

The activity consists of three separate simulations, each involving its own particular process for allocating the dogs to the treatment groups. Comparison of results across the three simulations should lead to a clearer understanding of when blocking may be used effectively, when it is not useful, and how a researcher might begin to analyze data from a blocked design. In order that students may more fully understand the importance of the various factors (variables) in this scenario, each student will play the role of a dog. As would be the case in reality, each dog is of a particular breed (which never changes) and is from a particular veterinary clinic (which also never changes). If your class has fewer than 24 students, some students should play the roles of more than one dog.

Students will obtain from the teacher cards that specify characteristics of each dog. Prior to beginning this activity, prepare the cards for your class. See the Appendix for details and

Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.

black-line masters.

Initial Discussion

Explain to students that they will be playing the roles of dogs, each student being a particular dog, and that they will each have specific traits based on several characteristics, such as breed. Explain that a numerical score will summarize their responses in the study, with positive values representing increases in bone density (as measured in Houndfield units) and negative values representing decreases in bone density. Also explain that once the students have determined which dog they will be, they will select for themselves the cards describing characteristics inherent to each individual dog. You, the teacher, will allocate the study treatments, playing the role of the researcher.

Have each student select a specific dog by name from the table in the appendix, so that each dog is represented exactly once. The students should note their dog's breed and clinic. Then have each student select one card of each of the following colors, corresponding to the characteristics of the dog whose name they selected: Magenta (Dog), Blue (Breed), Canary (Clinic), and Orange (Other Sources). Each student will need three of the White (Total Response) cards, one for each round of simulation. Stress to the students that they should **not** share the numerical values on their own cards with other students.

Now that students have their cards, explain the general meaning of the various cards, but do not mention the specific numbers associated with any particular cards. Stress once more that students should not know each other's numbers.

The number on the Magenta (Dog) card represents the average change in bone density for the dogs in this study; it applies to every experimental unit (dog) in the study.

The Blue cards give identical numbers within a given breed, but differ from breed to breed, representing the fact that some breeds are more prone to changes in bone density than other breeds.

Likewise, dogs from the same clinic will get the same number on their Canary cards, but dogs from different clinics will get different numbers.

The Orange cards represent the underlying variability among individual dogs due to factors not explicitly identified by other cards. These might include such factors as health history, age, diet, *etc.* Have students use the rule on the orange card to generate and record their own individual value.

The Teal cards (which have not yet been distributed) represent the influences on bone density due to the treatments. The researcher (teacher) will assign these cards to the dogs (students) based on the selected experimental design.

Round 1—Completely Randomized Design

For a completely randomized design the researcher avoids explicitly identifying any external sources of variability, instead controlling for such factors by randomly allocating treatments to experimental units.

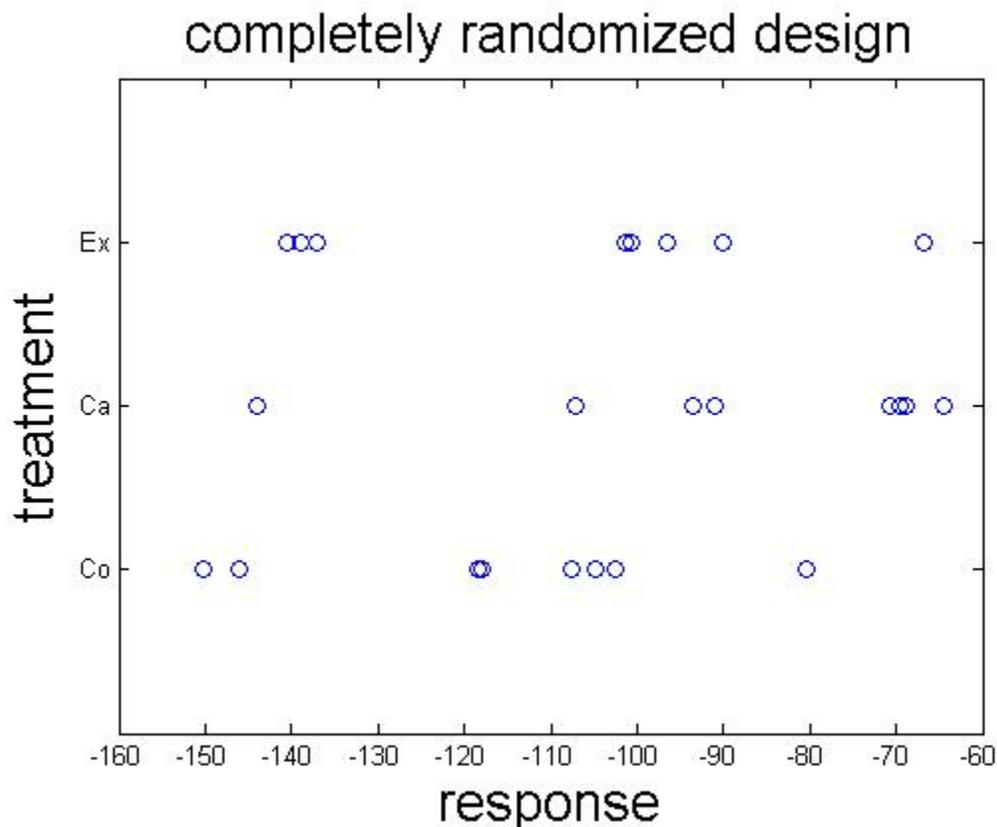
Thoroughly mix the Teal (Treatment) cards and distribute the cards to the class, one card per student. Again, ask students not to share the values they receive.

When all students have all six cards, have them complete, individually, their Total Response (white) cards for this round by adding the numbers from their other five cards. This number represents the change in bone density for the dog they represent at the end of the study.

Collect from students only their individual total scores and display these totals for the class in a table organized by treatment. (For example, place data for Treatment Co into one row, Treatment Ca into another row, and Treatment Ex into the final row.)

Examine the table of raw data to determine with students an appropriate scale for parallel dot plots. Then construct a class display consisting of parallel dot plots, one for each treatment group. Below is an example of what the table and dot plots might look like. (In the table the numbers have been rounded to the nearest integer, but the dot plots were made using software that preserved greater decimal precision.)

Ex	-101	-101	-96	-90	-67	-140	-137	-139
Ca	-107	-93	-91	-71	-64	-70	-69	-144
Co	-118	-118	-102	-108	-105	-80	-150	-146



Discuss the plots. Is any clear overall difference in the change in bone density across treatments apparent from the display? (Key questions: Is there a difference in the centers of the three distributions? How does variability within each treatment group affect the ability to see differences in overall health from group to group?)

Following completion of the discussion of Round 1, leave the table and graphical display so that they may be compared to similar tables and graphs from the remaining two rounds. Then collect the Treatment cards from all students (they should retain their Dog, Breed, Clinic, and Other Sources cards) and have students discard their Total Response cards from Round 1.

Round 2—Randomized Block Design, blocked by Breed

For a randomized block design the researcher identifies a possible external source of variability and controls for that factor by forming groups (called blocks) of experimental units in such a way that units within any single block are as alike as possible with respect to the response variable being measured. All treatments are then randomly assigned within each block. In this round of the simulation Breed plays the role of external source of variability, so blocks should be formed so that only one breed appears in any particular block.

Separate the Teal (Treatment) cards into four stacks, each stack containing exactly two of each of the three treatments: Co, Ca, and Ex. After the stacks of cards are ready, ask for all Akitas to stand or hold up their hands. Shuffle one stack of six treatment cards and randomly distribute them to those six students, one card per student. Repeat the assignment of treatments to each of the other breeds in a similar fashion. Again, ask students not to share the values they receive.

When all students have their new Treatment cards, have them complete, individually, their Total Response (white) cards for Round 2 by adding the numbers from their other four cards to their new Treatment value. This new number represents the overall change in bone density for the dog they represent at the end of the study blocked by breed.

As you did in Round 1, collect and tabulate students' individual total scores. For this round, however, organize the table by both treatment and breed. (For example, each row might represent a separate treatment, with columns grouped by breed.) Here is what your table might look like:

	Akitas		Beagles		Collies		Dalmatians	
Ex	-104	-101	-90	-87	-67	-60	-140	-132
Ca	-107	-108	-92	-98	-70	-71	-140	-143
Co	-115	-110	-103	-105	-80	-79	-151	-154

Discuss with students the following plan for analysis of these new data.

It should be apparent from the table that responses differ substantially from breed to breed. This breed-to-breed variability contributed to the overlapping displays in Round 1, and if parallel dot plots (by treatment) were made from Round 2's raw data, similar overlapping would be apparent. But we still want to compare results for Treatment Co to results for Treatment Ca to results for Treatment Ex, taking into account the obvious variability from breed to breed.

What really matters here is how much a particular treatment moves a particular dog's response away from the average response for all dogs of that breed (averaged across all treatments). Since treatments were assigned across each breed separately, the average Total Response for a given breed (averaged across all treatments) is easy to

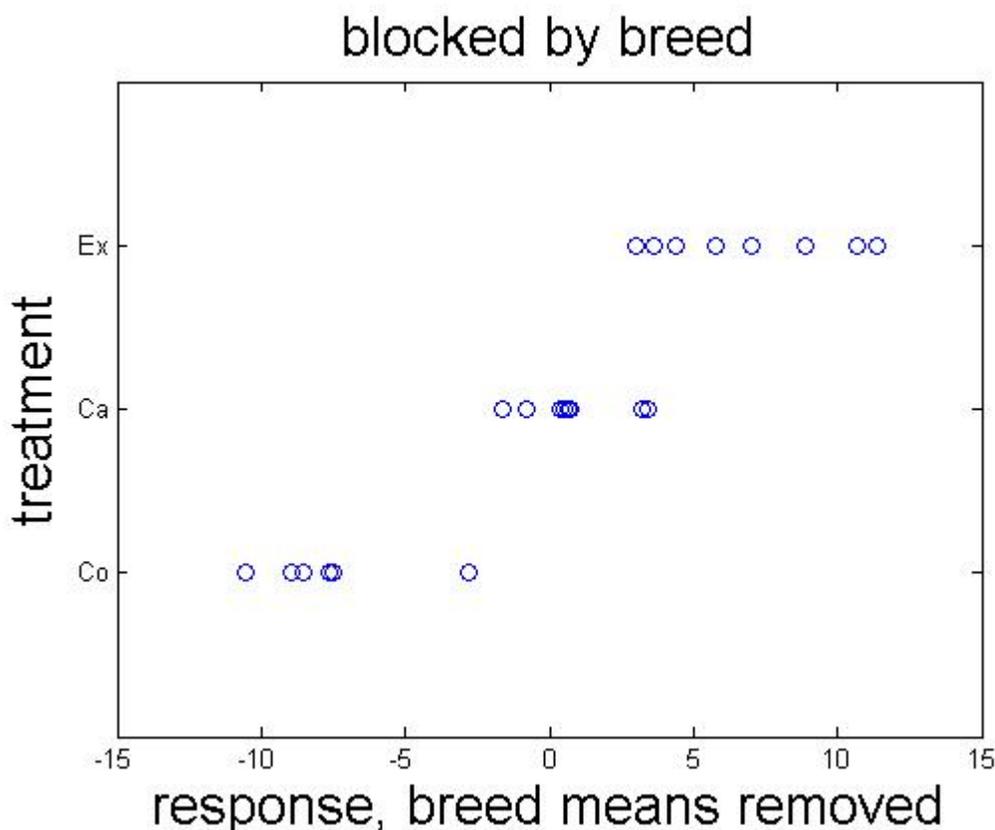
compute.

We can “remove” the breed-to-breed variability by subtracting from each dog’s Total Response value the average of the Total Response values for its respective breed. For example, if the six Akitas had an average Total Response of 121.3, then subtract 121.3 from each Akita’s individual score, leaving three new numbers that reveal the influence of the three treatments. In essence, this subtraction “re-centers” each breed’s data at 0 so that differences due to treatments become more visible.

Have students representing each breed determine their breed-average Total Response and carry out the subtraction discussed above. Construct a new breed-by-treatment table, this time for the differences obtained by these subtractions. The table below shows what the result might look like.

		Akitas		Beagles		Collies		Dalmatians	
Raw data	Ex	-104	-101	-90	-87	-67	-60	-140	-132
	Ca	-107	-108	-92	-98	-70	-71	-140	-143
	Co	-115	-110	-103	-105	-80	-79	-151	-154
	Breed averages	-108		-96		-71		-143	
data with breed variability removed	Ex	4	7	6	9	4	11	3	11
	Ca	1	0	4	-2	1	0	3	0
	Co	-7	-2	-7	-9	-9	-8	-8	-11

Examine the new table of differences to determine with students an appropriate scale for parallel dot plots. Then construct a class display consisting of parallel dot plots, one for each treatment group. The plot below is an example.



Discuss the plots. Is any clear overall difference in the mean change in bone density across treatments apparent from the display? Can you estimate the average amounts by which the two treatments Ca and Ex improve bone density compared with the control Co?

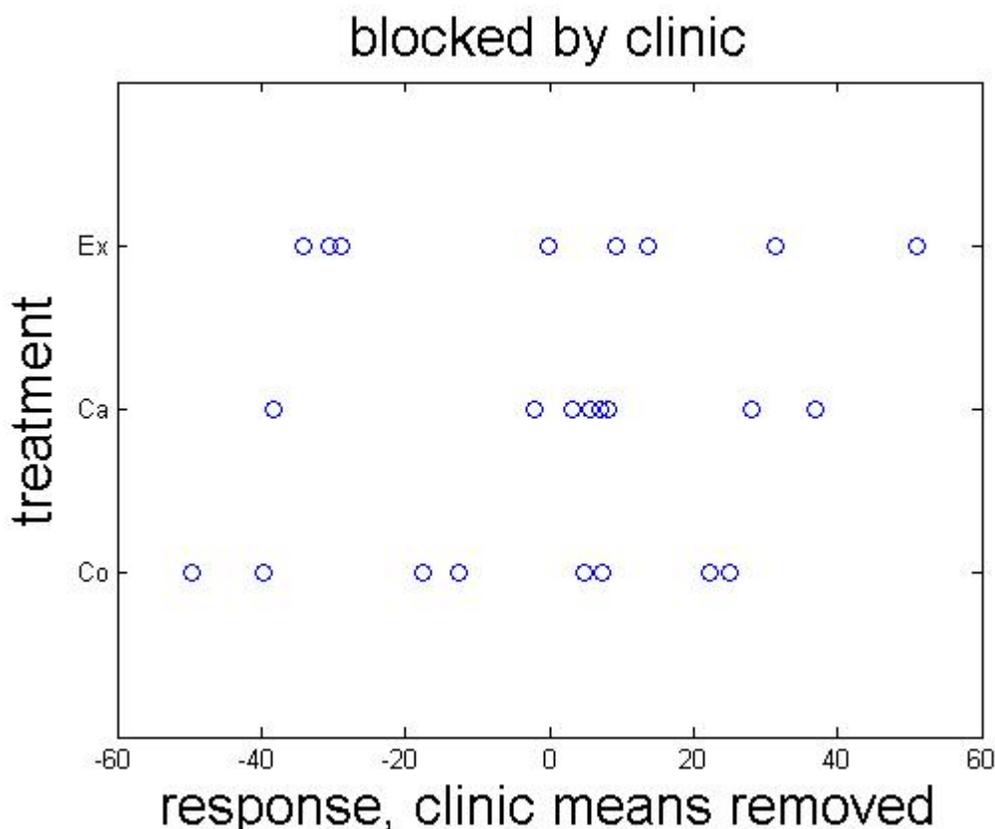
Following completion of the discussion of Round 2, leave the graphical displays from Rounds 1 and 2 so that they may be compared to a similar graph from Round 3. Then again collect the Treatment cards from all students (while they retain their Dog, Breed, Clinic, and Other Sources cards) and have students discard their Total Response cards from Round 2.

Round 3—Randomized Block Design, blocked by Clinic

Repeat Round 2, this time blocking on Clinic instead of Breed. Thus each block will consist of six dogs from the same clinic, and the teacher will distribute the Treatment cards so that each Clinic gets exactly two of each of the three treatments.

Complete the analysis (tables and graphs) as in Round 2. Of course, this time Clinic will define the rows of the table and averages by clinic will need to be subtracted in order to carry out the re-centering. The table and plot might look like those below.

		Paw Prince	Pooch Palace	Treehouse	Barking Lot
Raw data	Ex	-103 -89	-60 -140	-88 -66	-136 -140
	Ca	-98 -95	-105 -108	-70 -136	-108 -69
	Co	-81 -153	-104 -151	-115 -110	-101 -81
	Breed averages	-103	-111	-98	-106
data with clinic variability removed	Ex	0 14	51 -29	10 32	-30 -34
	Ca	5 8	6 3	28 -38	-2 37
	Co	22 -50	7 -40	-17 -12	5 25



Once more, discuss the plots. Is any clear overall difference in the mean change in bone density across treatments apparent from the display?

(In the example tables and plots included here, both the completely randomized design and the design blocked by clinic show what might be a slight difference in responses among the three treatments, favoring Ex over Ca and Ca over Co. You may or may not see a similar pattern with your class. In any case, it is not obvious that the apparent difference is not a mere chance variation due to noise. In the design blocked by breed, however, the difference is quite unmistakable.)

Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.

Wrap-Up

The original problem was to determine whether the treatments differed in their overall effects and, if so, to estimate how much they differed. Have students look back at the graphs from the three rounds of simulation and discuss the following two questions.

In which simulation round was it easiest to discern the effects of the three treatments?

How did the characteristics of the variables and the design of the allocation work together to make that round work best?

The main point of the activity is that blocking on a factor associated with large variation in the response variable permits the researcher to see differences among treatments more clearly. Blocking on a factor associated with small variation in the response variable provides little help.

When Does Blocking Help?

Teacher Notes: Part II

In the first part of this document, “Teacher Notes: Part I”, we presented an activity that you can use to demonstrate to your students (1) how a block design is implemented, (2) when blocking is helpful, and (3) how it reduces unwanted “noise” in the response variable by attributing some of it to a specific source (the blocking variable). In this second part of the document, we will explore blocking in more detail by writing mathematical models for data and discussing actual statistical practice. The material in this section is for your edification as a teacher and is beyond what students need to understand in an introductory statistics course such as AP Statistics.

Some preliminary notation

In the next section we will discuss mathematical models for data. Let’s begin by introducing some notation that is common in mathematical statistics texts.

When we write a random variable followed by a tilde (\sim), the meaning of the tilde is “has the following distribution”. For example, “ $Y \sim N(\mu, \sigma)$ ” means Y is a random variable that has a normal distribution with mean μ and standard deviation σ .¹

I was expecting the footnote to say that some references use normal, mean, variance and that you should make sure that you know which convention is being followed when you encounter this notation.

When we add a subscript to a random variable and then write a tilde followed by the letters “iid”, the meaning is “independent and identically distributed thus”. For example, $Y_i \sim \text{iid } N(\mu, \sigma)$ means that Y_1, Y_2, \dots, Y_n are all independent and all are normally distributed with mean μ and standard deviation σ .

¹ Many distribution families, like the normal distributions, have members identified by parameter values, and it is common to give those values in parentheses following a few letters denoting the distribution family. *Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.*

A simple mathematical model

If n observations Y_1, Y_2, \dots, Y_n are randomly sampled from a population that is normally distributed with a mean of μ and standard deviation of σ , then we could write $Y_i \sim \text{iid } N(\mu, \sigma)$. But we could also write the following:

$$Y_i = \mu + \varepsilon_i, \text{ where } \varepsilon_i \sim \text{iid } N(0, \sigma).$$

The equation above is a *mathematical model* of our data. The reason this written form may be preferable to the earlier $Y_i \sim \text{iid } N(\mu, \sigma)$ is that it decomposes the data into two parts: an overall mean, and variability about that mean.

In the dogs activity, i is an index counting all the dogs from $i = 1$ up to $i = 24$; Y_i is the change in bone density for dog i over the course of the study; μ is the mean change in bone density for all 24 dogs; ε_i is how much the change in bone density of dog i differs from the mean μ ; and σ is the “typical” magnitude.²

Including treatment effects

Let’s consider the study described in the activity. Dogs become susceptible to decreased bone density as they age, and we want to see whether either a dietary supplement of calcium, or a supplement of calcium combined with an exercise regimen will help stop or slow the decrease in bone density. In the first part of the activity the students simulated an experiment with three treatments and a *completely randomized design*. The three treatments groups are “Ca” for calcium alone, “Ex” for calcium with an exercise regimen, and “Co” for control, which receives neither.

After we administered the treatments and simulated our data, we wanted to see whether the distributions of the response variable in the three groups were distinctly different from one another. If they were, then we might reasonably conclude that the differences were due to the dietary supplements being administered, since the dogs were otherwise treated the same,

² ε_i is often called the “error” associated with dog i , or with measurement i . This can easily mislead students and those new to statistics. It is not an “error” in the sense that anything was done wrong, nor that the measurement of the response variable Y_i was inaccurate. There are historical reasons why the word “error” was attached to ε_i , but in the models we are looking at in this document, we should think of ε_i as “other sources of variability”—that is, the sum of everything that contributes to the difference between Y_i and μ that hasn’t already been accounted for in the model. ε_i is also sometimes called the “noise” term, which draws on the analogy of “signal” (the thing we want to detect) and “noise” (sources of variability that are unaccounted for), which risks drowning out the signal.

Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.

and the random allocation of dogs to treatments means that there shouldn't be systematic wholesale differences between the dogs in one group compared to those in another.³

Now let's add a "treatment effect" to our mathematical model. We will use the Greek letter tau (τ) because tau is a Greek "t", and "t" is the first letter in the word "treatment".

$$Y_{i,j} = \mu + \tau_j + \varepsilon_{i,j}, \text{ where } \varepsilon_{i,j} \sim \text{iid } N(0, \sigma).$$

An additional subscript, j , has been added to the response variables Y_i . This is an index of the three treatment groups, so we have $j = 1, 2, 3$. But there are only 8 dogs in each treatment group, so the meaning of the index i must change from being an index of all 24 dogs to an index of the 8 dogs within a treatment group: $i = 1, 2, \dots, 8$.

Before continuing, let's consider a single dog as an example. Let's suppose that $j = 2$ corresponds to the Ca (calcium alone) treatment group, and let's look carefully at the dog we've identified as dog number 5 within that group:

$$Y_{5,2} = \mu + \tau_2 + \varepsilon_{5,2}, \text{ where } \varepsilon_{5,2} \sim N(0, \sigma).$$

This says that among the 8 dogs in the Ca group, dog number 5 had a response variable (change in bone density over the course of the study) equal to $Y_{5,2}$. The equation also indicates that the response can be decomposed into three components. First, there's the overall mean μ , which is the same for all the dogs in the study. Since we're primarily interested in comparing treatment effects, μ isn't very interesting to us. The second component is the treatment effect τ_2 , which gets added on to the mean. All 8 of the dogs in the Ca treatment group have this effect as a component of their response variable.

Finally, there's the "noise" component, $\varepsilon_{5,2}$, which we're modeling as normally distributed with an unknown standard deviation σ .

Since μ is the same for all the dogs, then all of the *variability* in the full data set comes from the three different values of τ_j and the 24 different values of $\varepsilon_{i,j}$. The τ_j component may be thought of as a "signal" we want to detect, and the $\varepsilon_{i,j}$ component is "noise" that is making the signal difficult to detect.

³ There is a third possible cause of a wholesale groupwise difference in responses compared to another group: the difference may be due solely to chance. Statistical techniques exist to determine whether such a chance is plausible or not, but they are not the focus of this activity. This activity is designed in such a way that it may be done early in the course, well before students are introduced to formal inference. But such techniques should not be necessary if either the three groups' distributions almost completely overlap (in which case there cannot be any significant differences between the treatment groups) or they are quite visibly distinct (in which case there will certainly be significant differences between the treatment groups). This activity was designed so that when a completely randomized design is implemented, you will almost surely see no group differences, and when a design is implemented that blocks on dog breed, clear group differences will be apparent. *Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.*

Here's the issue that we want to address by blocking: the noise may be so large that it drowns out the signal. We've constructed a model that includes two sources of variability in response variables, but our actual data don't come broken down into these components: we only get to see the whole responses $Y_{i,j}$. From the collection of those responses, we want to *estimate* the τ_j 's—in particular, we want to see whether they're different from one another—but that may be hard to do if there's a lot of noise in the data.

And in fact, this is precisely what did happen in the dog activity when we performed a completely randomized design. The parallel dot plots of the response variables show a lot of variability within each treatment group: that's due to the noise term. Now we know that in fact there is a treatment effect, because we built it into our data: that was the teal cards that said -8 for the Co group, +2 for the Ca group, and +6 for the Ex group. But knowing that a treatment effect is present because you *put* it there is not the same as *finding* a treatment effect in the data. In practice, with nothing to go on but the responses $Y_{i,j}$, it may not be clear that any of the τ_j 's are different from one another if the data are very noisy.

Blocking on breed

Now let's add a fourth component to our response variable: a block effect. Let's suppose that the bone densities of different breeds of dogs decrease at different rates. In particular, let's suppose that each of the four breeds of dogs in our study has a "breed effect" that modifies the response variable by adding or subtracting a quantity that is particular to that breed. Since larger values of the response variable correspond to slower bone deterioration in dogs, a large "breed effect" indicates that the breed generally has good bone health.

We will use the Greek letter beta (β) for the block effect because the Greek beta is a "b", and "b" is the first letter in the word "block".

$$Y_{i,j,k} = \mu + \tau_j + \beta_k + \varepsilon_{i,j,k}, \text{ where } \varepsilon_{i,j,k} \sim \text{iid } N(0, \sigma).$$

As before when we added a term to our model, we have changed slightly the meaning of the index i . Now k is an index counting from 1 to 4, and β_k is the "block effect" of breed k . j is an index counting from 1 to 3, and τ_j is, as before, the effect of treatment j . Now the index i only counts from 1 to 2, indexing the two dogs in each treatment-block combination.

For example, there are two Dalmatians in the control group. If $k = 4$ is the index for Dalmatians and $j = 1$ is the index for the control, then $Y_{1,1,4}$ and $Y_{2,1,4}$ are the changes in bone density for those two Dalmatians.

Notice that in the previous model, we had $\varepsilon_{i,j}$ representing "other sources of variability in the response variable", and "other" simply meant anything other than the treatment. In the dogs activity, when we used a completely randomized design the noise was so great that the

Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.

response variables plotted separately for the three treatment groups showed no clear differences. That didn't necessarily mean that there were no differences, only that if there were any, they were being drowned out by noise—variability due to sources other than the treatment.

But our new block design allows us to attribute some of that variability to a particular source: dog breed. In doing so, we potentially account for some of the variability and can effectively “filter it out”, which is what we did in the activity when we subtracted the breed averages from the response variables and plotted the remainders grouped by treatment. Then the three treatment groups showed clear differences.

Blocking on clinic

In the third part of the activity, we blocked on clinic; that is, we randomly allocated treatments to dogs within each clinic. The model corresponding to this design is the same as the one in the last section, only this time the index k counts from 1 to 4 for the four different clinics, not for the four different breeds. So $Y_{1,1,4}$ is the change in bone density for the first of the two dogs in the control group who visits clinic 4, and $Y_{2,1,4}$ is the change in bone density for the other one. The model still looks like this:

$$Y_{i,j,k} = \mu + \tau_j + \beta_k + \varepsilon_{i,j,k}, \text{ where } \varepsilon_{i,j,k} \sim \text{iid } N(0, \sigma).$$

But what happens when we block on clinic? We saw in the activity that subtracting the clinic averages from the response variables and plotting the differences by treatment group didn't help us distinguish a difference between the three treatments. Why not?

The reason is that subtracting the block average—be it breed or clinic or something else—is a way of explaining part of the variability in the noise term by attributing it to a particular source. And the greater that variability that gets thus explained, the smaller will be the noise that remains when we plot the responses with the estimated block effect removed. Since the breed effects have a wide range—from about -30 to +30—and the clinic effect have a narrow range—from about -3 to +3—then separating the breed effect from the noise term will reduce the unexplained variability more than separating the clinic effect.

If you were about to conduct this study and had to decide whether to block on breed or on clinic, how would you decide? You would pick the one that you thought contributed the greater variability to the change in bone density. Another way of saying this is that you would block on the one that you thought had the greater association with bone density. In our simulation activity, dog breed contributed a greater amount of variability than did clinic, so it was the one that proved more useful as a block.

The point of blocking

The point of blocking is to take what would otherwise be considered undesirable noise and attribute some of it to a particular source—the blocking variable. If this is a source of much of the variability in the response variable, then blocking on it should explain a lot of that variability, leaving less unexplained variability—noise—behind. And with less noise, the treatment effects—the signal—should be easier to detect.

In our activity, we had the students “build” their response variables out of different components, but they couldn’t see what other students had written on their cards, so it wasn’t obvious how much *variability* each card type was contributing to the whole group. In the first part of the activity, when we implemented a completely randomized design, the breed effect and the clinic effect were lumped in with the “other sources of variability”, and the noise turned out to be so great that you couldn’t really tell the difference between the treatments.

Then we blocked on dog breed and found that when we took it into account by subtracting breed averages from the data, the differences in responses among the three treatment groups was much more evident.⁴ This was due to the fact that different breeds of dogs had very different natural rates of decrease in bone density.

Finally, we blocked on clinic and found that taking this into account did little to reduce unexplained variability, because the clinic contribution to the original noise term was relatively small.

More on “effects”

In the activity, the sum of the four blue breed cards was 0, as was the sum of the four canary clinic cards, as well as the sum of the three teal treatment cards. This was not an accident: the activity was designed this way. Because of that, the second dot plot—the one of responses with breed averages subtracted—showed three distinct clusters of points centered on precisely the three numbers that students had on their teal cards.

Recall that our mathematical model for the block design is

$$Y_{i,j,k} = \mu + \tau_j + \beta_k + \varepsilon_{i,j,k}, \text{ where } \varepsilon_{i,j,k} \sim \text{iid } N(0, \sigma).$$

When we introduced this model, we called τ_j and β_k “treatment effects” and “block effects” respectively, but we haven’t yet really specified what τ_j and β_k really *are*. They are the amounts by which the responses of dogs in different treatment groups and blocks,

⁴ In subtracting breed averages from the data, we were effectively removing estimates of both the τ_j ’s and the overall mean μ . Only removal of the former was useful in explaining *variability*, since μ was the same for every dog. But the average response over breed j happens to be $\mu + \tau_j$.

Blocking activity by Landy Godbold, Dan Teague, Floyd Bullard, and Chris Olsen, in consultation with Stu Hunter, Roxy Peck, Jackie Dietz, and Bob Hayden, July 2007. To be used and shared freely for teaching statistics.

respectively, differ from the overall mean μ . For this reason, the sum of the treatment effects and the block effects *must* sum to zero. If they didn't, then the overall mean wouldn't be μ .

For this reason, the word “effect” is a bit misleading, because it seems to imply, for example, that giving a dog calcium will have the “effect” of decreasing its rate of bone density decay by about 2 Houndsville units per year. (The teal Ca card had the “effect” of +2 on it, and the second dot plot also showed the filtered Ca responses as centered on about +2.) But in fact, if we want to know by how much calcium helps a dog's bones, we should really be looking at the *differences* between the treatment effects when compared with that of the control group. Since the control group received a -8 (and this was also visible in the second dot plot), and the Ca and Ex groups received +2 and +6 respectively, then the actual improvements gained by applying those two treatments compared to a no-calcium, no-exercise control are +10 and +14 respectively. (In statistical language, these are called *contrasts*.)

“Interactions”

We assumed in this activity—both when we “built” the data from the cards and when we analyzed the data—that the decrease in dog bone density over the year of the study was essentially additive with respect to breed and treatment. For example, being a Collie got you a +32 compared with the average dog in the study, and getting calcium and exercise got you a +6. But what if different breeds of dogs don't all respond differently to the different treatments?

This leads to the idea of *interactions*. Even though the idea is not part of the AP curriculum, it is on many students' minds, so I'll address it briefly here. Suppose the dogs in this study had not all responded the same to the treatments. For example, suppose that Dalmatians benefited greatly from exercise and received a +18 if they fell in that group, but Akitas benefited far less and received only a +5 if they fell in that group. The simple block model doesn't include different treatment effects for different breeds. A richer model includes interaction terms, and has many more parameters. But the design of a study meant to detect interactions would be essentially the same as the second one we did—that is, treatments would be randomly allocated to dogs within their breed rather than all at random.

Statisticians don't all agree on the use of the word “block”. To some statisticians, it is to be reserved strictly for variables inherent to the experimental units that *do not interact* with treatments of interest. Such statisticians view blocks as nothing more than undesirable noise to be estimated and subtracted out so that the “signal” of the treatment effects will be clearer.

Other statisticians take a broader view and consider “blocks” to include those inherent variables that may interact with the treatments. The difference is linguistic only. Both groups agree that if there is the possibility of interactions between an inherent variable and the treatments, then the design of the study should permit looking for those interactions.

Such a design involves grouping experimental units according to that inherent variable and randomly allocating treatments to units within those groups.⁵

The matched pairs design

You may sometimes hear that a matched pairs design is a special case of blocking. This is true. It is done when it is possible to pair experimental units according to some variable that is thought to contribute to the response variable, but which is not expected to interact with the two treatments.

For example, if you wanted to compare two different antiperspirants, you might decide to apply it to volunteers' arms and measure the amount of perspiration on the arms after a prescribed amount of physical activity. Different people would be expected to perspire different amounts, but you wouldn't expect a person's left arm to perspire differently from his right arm, nor would you expect an antiperspirant to have different effects on a person's left and right arms. So this is a good study on which to use a matched pairs design, with each person being given one antiperspirant on one arm (selected at random) and the other antiperspirant on the other arm.

After measuring the amount of perspiration at the end of the study, you could estimate the "block effect" for each individual by averaging his two measures, and then account for it by subtracting it from each observation. That is what we did with the dogs. But notice that if

the two measurements on an individual are x and y , and then we subtract $\frac{x+y}{2}$ from both

of them, the two values become $\frac{x-y}{2}$ and $\frac{-x+y}{2}$, opposites of one another. The

difference between *these* two values is still exactly $x-y$. So "filtering out" the block effect and then looking at the differences to see whether the resulting values are distinct from one another is really equivalent to looking at the differences $x-y$ and seeing whether they're distinct from zero. You can't do that with more than two treatments, nor if your design incorporates interactions. But if you are interested in comparing only two treatments, and there is a variable you can pair (*i.e.*, block) on that you think contributes to variability in the response variable but which does not interact with the two treatments, then pairing on that variable will enable to reduce the unexplained variability in the data—perhaps quite considerably—so that the variability due to the two treatments may be easier to see.

⁵ Here is a very technical note: In order to look for interactions between "blocks" and treatments, you would need replications within block-treatment combinations. We had that in our dog activity, since we had 2 dogs of each breed receiving each treatment. Had we had only one of each breed receiving each treatment, there would have been no way to tell how much of the variability in response between two dogs of the same breed was due just to noise and how much was due to the fact that the dog's breed responds differently to different treatments. On the other hand, if we wanted to block on some variable to reduce the unexplained variability in the response, but we anticipated *no* interactions at all and we truly didn't want even to look for them, then the best experimental design to use would assign exactly treatment exactly *once* per block. This is for technical reasons that have to do with degrees of freedom. They will not be discussed in this document, but interested teachers can easily find discussions of interactions in many statistics texts; for example, *The Statistical Sleuth*, by Ramsey and Schafer, or *Statistics for Experimenters*, by Box, Hunter, and Hunter.

Actual practice

Let us pause to consider how the dog study might actually be conducted were it to be done in practice. We used breeds and clinics as potential blocking variables in our activity because we had a pedagogical purpose: to demonstrate that when an intrinsic variable is strongly associated with the response variable, it may make sense to block on that variable so that much of the variability in the response variable can be accounted for. But if an intrinsic variable is only very weakly associated with the response variable, then blocking may not give you much.

That was our pedagogical purpose. But what if researchers *really* wanted to consider the three treatments Ex, Ca, and Co on a group of 24 dogs. Would they perform this study using a blocked design, blocking on breed? Probably they would not. If breed were thought to affect bone density, it would probably be because some other factor—like the dog’s size—had an effect on bone density, and dogs of the same breed tend to be about the same size. If we take that to be true, then breed in our activity was serving as a proxy for the dogs’ size.

If you wanted to block directly on the size of the dog, there would be no need to have any particular number of dogs in each of any particular number of breeds. And we also might reasonably suppose (as was posited in our activity) that a dog’s veterinary clinic has little to do with its bone density. So we might begin our study by enlisting in our study 24 dogs belonging to owners who agree to let their dogs participate.

Next we might weigh all the dogs and order them according their weight. The heaviest three could be a block, then the next heaviest three, and so on down to the lightest three, forming eight groups of 3 dogs. Within each group, the dogs are all about the same size, so we would call these “blocks”.

Then we randomly allocate “Ex”, “Ca”, and “Co” to the dogs in each block, being sure that in each block exactly one dog gets each treatment. We apply the treatments and after one year, measure the change in bone density for each dog. The analysis of the resulting data is slightly beyond the AP curriculum, since there are three treatments (with two treatments, you could perform a matched pairs analysis), but after doing this dogs activity students should at least understand that such a block design would permit researchers to account for much of the change in bone density that is due to the size of the dog.

Appendix

Preparing the cards:

- 24 Dog cards (suggested color Magenta, for Main)
- 6 of each of the 4 Breed cards (suggested color Blue)
- 6 of each of the 4 Clinic cards (suggested color Canary)
- 8 of each of the 3 Treatment cards (suggested color Teal)
- 24 Other Sources of Variability cards (suggested color Orange)
- 72 Total Response cards (suggested color White)

The dog list and black-line masters for each card appear on the following pages.

The dogs in the study:

<i>Dog Name</i>	<i>Breed</i>	<i>Clinic</i>
Elmer	Akita	Treehouse
Bernie	Akita	Barking Lot
Queenie	Akita	Pooch Palace
Sugar	Akita	Paw Prince
Jock	Akita	Pooch Palace
Curly	Akita	Treehouse
Rocky	Beagle	Paw Prince
Happy	Beagle	Pooch Palace
Nico	Beagle	Barking Lot
Alex	Beagle	Treehouse
Pepper	Beagle	Paw Prince
Max	Beagle	Paw Prince
Buster	Collie	Barking Lot
Newton	Collie	Pooch Palace
Lad	Collie	Treehouse
Sparky	Collie	Treehouse
Julius	Collie	Paw Prince
Cinnamon	Collie	Barking Lot
Spot	Dalmatian	Barking Lot
Euclid	Dalmatian	Pooch Palace
Rex	Dalmatian	Pooch Palace
Archie	Dalmatian	Barking Lot
Euler	Dalmatian	Paw Prince
Lucy	Dalmatian	Treehouse

Copy 24 of these on Magenta paper:

Dog

-105

Copy 24 of these on Orange paper:

Other Sources

Randnormal (0, 2)

Copy 72 of these on White paper:

**Total
Response**

D+B+C+T+O =

Copy 6 of these on Blue paper:

Breed: Akita

-2

Copy 6 of these on Blue paper:

Breed: Beagle

9

Copy 6 of these on Blue paper:

Breed: Collie

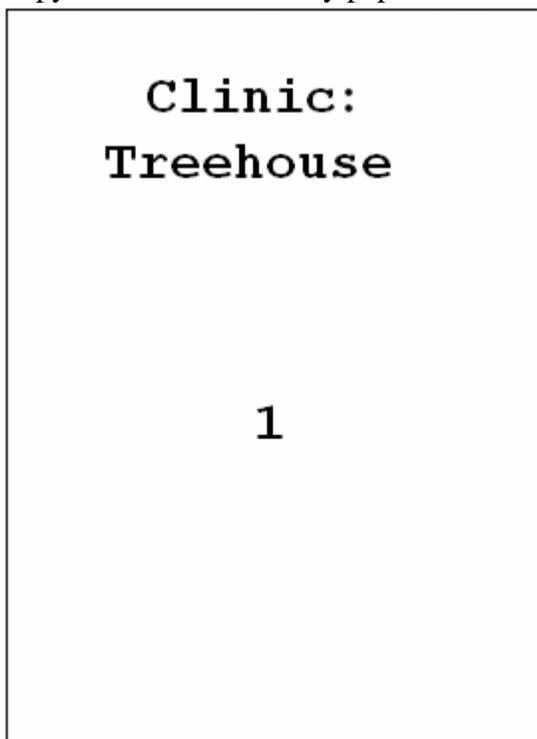
32

Copy 6 of these on Blue paper:

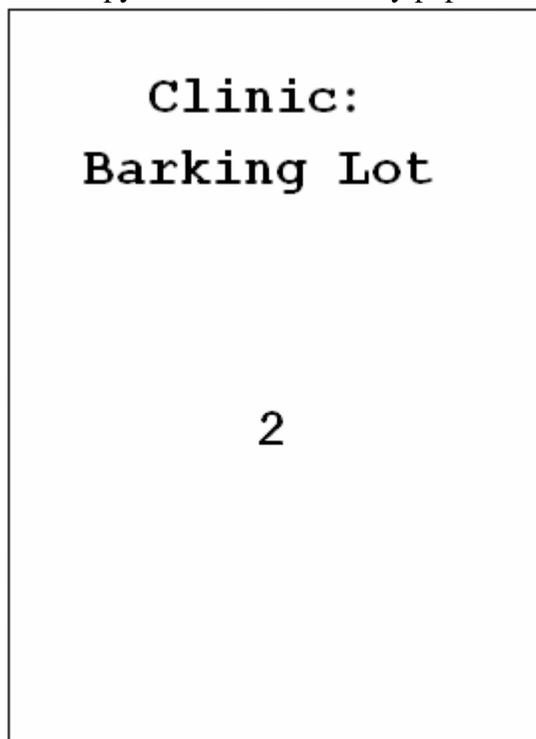
Breed: Dalmation

-39

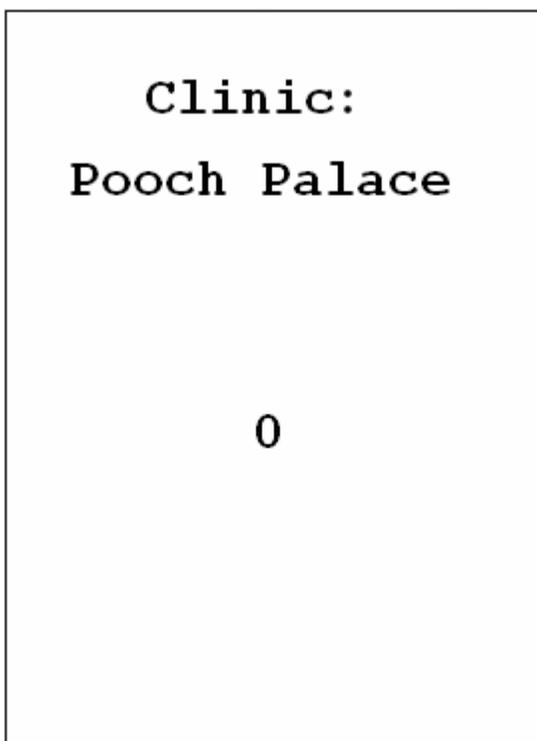
Copy 6 of these on Canary paper:



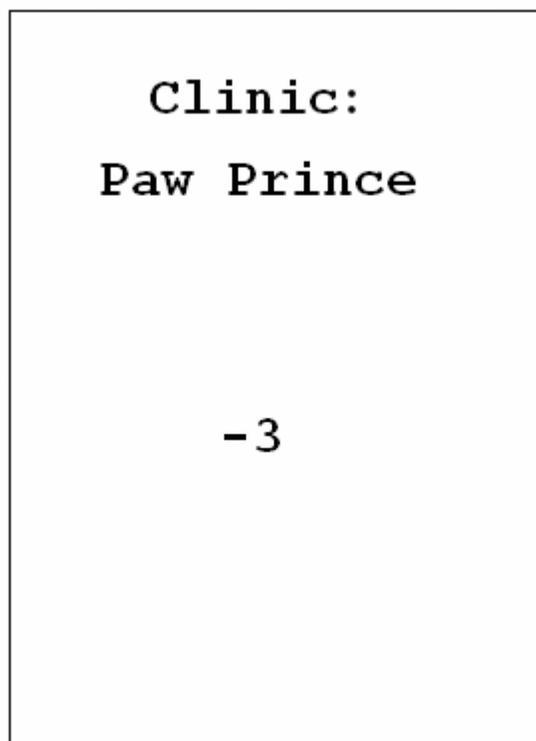
Copy 6 of these on Canary paper:



Copy 6 of these on Canary paper:



Copy 6 of these on Canary paper:



Copy 8 of these on Teal paper:

Treatment
Co

-8

Copy 8 of these on Teal paper:

Treatment
Ca

2

Copy 8 of these on Teal paper:

Treatment
Ex

6