More Than Your Heart Desires…
…An Exploration of Blocking

Carolyn Doetsch, Ocean Lakes High School, Virginia Beach, VA
cdoetsch@worldnet.att.net
Peter Flanagan-Hyde, Phoenix Country Day School, Phoenix, AZ
pflanaga@pcds.org
Mary Harrison, Salem High School, Virginia Beach, VA
johne2@series2000.com
Josh Tabor, Wilson High School, Hacienda Heights, CA
jtabor@hlpusd.k12.ca.us
Chuck Tiberio, Wellesley High School, Wellesley, MA
tiberio@tiac.net

NCSSM Statistics Leadership Institute
July, 2000

A certain HMO has an important decision to make. To create a more efficient and profitable business, the HMO will only allow doctors to prescribe one drug for a particular purpose, even though there may be other similar drugs on the market. Two competing pharmaceutical companies have drugs that are designed to reduce cholesterol levels. Both drugs have been on the market for several years and are effective for most people.

Your job is to help the HMO determine which drug they should use, based on its effectiveness in the general population. To make sure the study is not biased, it will be double-blind. Thus, neither the subjects nor you will know the identity of the drugs, which have been prepared by an independent lab and labeled drug A and drug B.

To obtain subjects for your study, a list of clinics in the U.S. was obtained and one clinic was randomly selected. Within the clinic, nearly 50 people volunteered to be in the study. Of the volunteers, 24 were selected to participate based on their gender, age, frequency of exercise (0 = low, 2 = high), and their initial cholesterol count.

Your job is to assign them to treatments and determine which drug is more effective. To measure the success of each drug (improvement), you will subtract their final cholesterol count from their initial cholesterol count (initial – final). Thus, the higher the difference, the more effective the drug was in reducing the cholesterol of the subject.

Unfortunately, we cannot actually perform this experiment. Instead, we will simulate the experiment using data that was randomly generated to mimic a typical clinic. The data we will use is in the tables below. In a real experiment, the researcher would not know the last 4 columns until after the experiment (and then only half of the values in each of the subsequent "Reading" columns would be known, since only half the subjects will receive each treatment). Do your best to pretend that these data are not available when you assign the treatments.

Sorted by Age

Row #	Gender	Age	Exercise Level	Initial Cholesterol	Reading after A	Reading after B	Improvement using A	Improvement using B
1	1	31	1	270	244	241		
2	0	34	2	258	251	242		
3	0	35	1	274	257	249		
4	0	35	0	312	286	259		
5	1	36	2	254	243	247		
6	0	36	1	276	263	249		
7	0	37	0	304	258	268		
8	1	38	1	289	264	261		
9	0	40	2	266	242	260		
10	1	41	2	256	255	245		
11	0	41	0	301	261	250		
12	1	42	1	256	245	232		
13	1	47	2	243	241	244		
14	1	47	1	284	272	257		
15	0	49	1	266	234	241		
16	1	49	0	280	233	241		
17	1	52	0	290	247	252		
18	0	54	2	257	247	249		
19	0	54	0	302	267	264		
20	0	55	2	261	256	260		
21	1	55	0	291	248	247		
22	0	58	1	274	256	249		
23	1	60	2	236	243	237		
24	1	60	0	304	279	256		

Sorted by Initial Cholesterol

Row #	Gender	Age	Exercise Level	Initial Cholesterol	Reading after A	Reading after B	Improvement using A	Improvement using B
1	1	60	2	236	243	237		
2	1	47	2	243	241	244		
3	1	36	2	254	243	247		
4	1	41	2	256	255	245		
5	1	42	1	256	245	232		
6	0	54	2	257	247	249		
7	0	34	2	258	251	242		
8	0	55	2	261	256	260		
9	0	40	2	266	242	260		
10	0	49	1	266	234	241		
11	1	31	1	270	244	241		
12	0	35	1	274	257	249		
13	0	58	1	274	256	249		
14	0	36	1	276	263	249		
15	1	49	0	280	233	241		
16	1	47	1	284	272	257		
17	1	38	1	289	264	261		
18	1	52	0	290	247	252		
19	1	55	0	291	248	247		
20	0	41	0	301	261	250		
21	0	54	0	302	267	264		
22	0	37	0	304	258	268		
23	1	60	0	304	279	256		
24	0	35	0	312	286	259		

I. Trial Number 1: Completely Randomized Design

A. Choose 12 different random numbers from1-24 and assign those subjects to drug A. The remaining 12 subjects will be given drug B.

B. Record the improvement of each subject after treatment with his assigned drug (Improve A, Improve B) in the table below.

C. Calculate the mean improvement for each drug.

Row #	**Improvement using A**	**Row #**	**Improvement using B**
mean improvement		**mean improvement**	

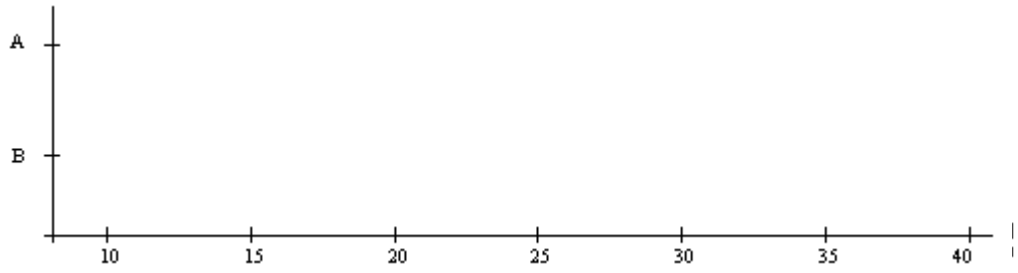
In your simulation, which drug did a better job reducing cholesterol? _____

Combine your results with the rest of the class.

How many times was A better? _____ How many times was B better? _____

What were the mean and standard deviation of the sample means? mean _____ sd _____

Sketch parallel boxplots using the axes below:



II. Trial Number 2: Randomized Complete Block—Blocked by Age

A. Use the table that is sorted by age.

B. Divide the subjects into 12 blocks of 2 (the first 2 subjects, the second 2, and so on). Note: a block is a group of people who are anticipated to have responses more alike than people in different blocks. In this case, you think that the group of younger people have similar responses to the drugs, and the group of older people have similar responses to the drugs, but not necessarily the same as the younger people.

C. Within each block, choose a random number from 1-2 and assign that subject to drug A. The remaining subject will receive drug B.

D. Record the improvement of each subject to their assigned drug in the table below.

E. Calculate the mean improvement for each drug

Row #	**Improvement using A**	**Row #**	**Improvement using B**
mean improvement		**mean improvement**	

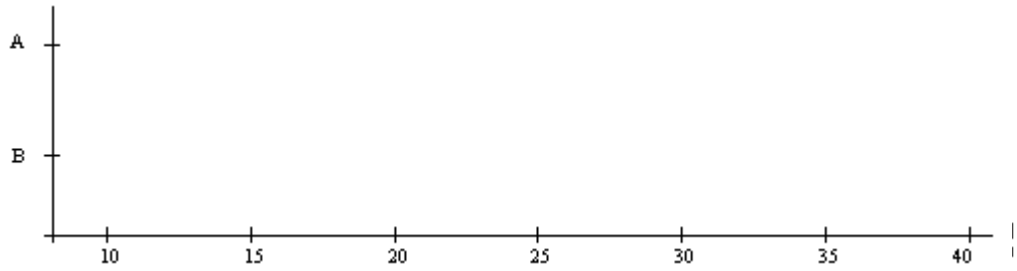
In your simulation, which drug did a better job reducing cholesterol? _____

Combine your results with the rest of the class.

How many times was A better? _____ How many times was B better? _____

What were the mean and standard deviation of the sample means? mean _____ sd _____

Sketch parallel boxplots using the axes below:



How do these results compare to Trial 1? Compare the means and standard deviations as well as the overlap in the boxplots.

What can you conclude about the effectiveness of Age as a blocking variable?

Trial Number 3: Randomized Complete Block—Blocked by Initial Cholesterol

A. Use the table that is sorted by initial cholesterol.

B. Now continue as in trial #2. Divide the subjects into 12 blocks of 2 and randomly select one from each block to receive drug A. The other receives drug B. Record improvements and calculate means.

Row #	**Improvement using A**	**Row #**	**Improvement using B**
mean improvement		**mean improvement**	

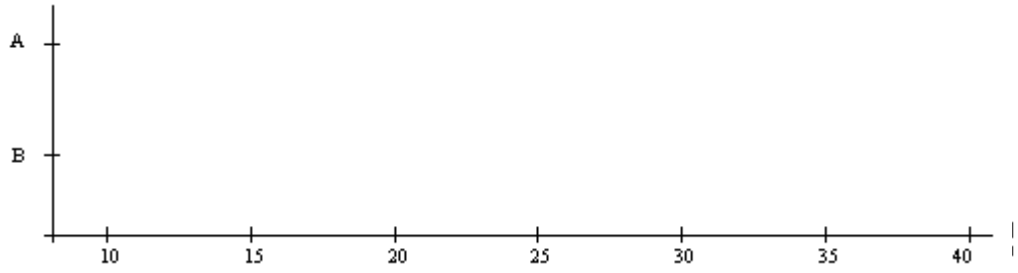
In your simulation, which drug did a better job reducing cholesterol? _____

Combine your results with the rest of the class.

How many times was A better? _____ How many times was B better? _____

What were the mean and standard deviation of the sample means? mean _____ sd _____

Sketch parallel boxplots using the axes below:



How do these results compare to Trial 1? Compare the means and standard deviations as well as the overlap in the boxplots.

How do these results compare to the previous Trial?

What can you conclude about the effectiveness of Initial Cholesterol as a blocking variable?

III. Trial Number 4: Randomized Complete Block—Blocked by Gender and Initial Cholesterol

A. Hopefully you found that blocking by initial cholesterol helped reduce the variability in the sample means for drugs A and B. To further reduce the variability, experimenters will often block on a second variable.

B. For this trial, we will block by gender and initial cholesterol. We chose gender because it often creates variability in clinical trials and we will continue to use initial cholesterol since blocking by this factor has already proven to be effective.

C. To block by these two variables, use the table that is already sorted by cholesterol. Then, starting at the top, find the first 2 subjects that have gender 0. This is the first block. The second block will be the next two subjects with gender 0, etc. The seventh block will be the first two subjects with gender 1, etc.

D. Within each block, randomly assign one of the subjects to receive drug A and assign the other to drug B.

E. Record the improvement of each subject to their assigned drug in the table below.

F. Calculate the mean improvement for each drug.

Row #	Improvement using A	Row #	Improvement using B
mean improvement		mean improvement	

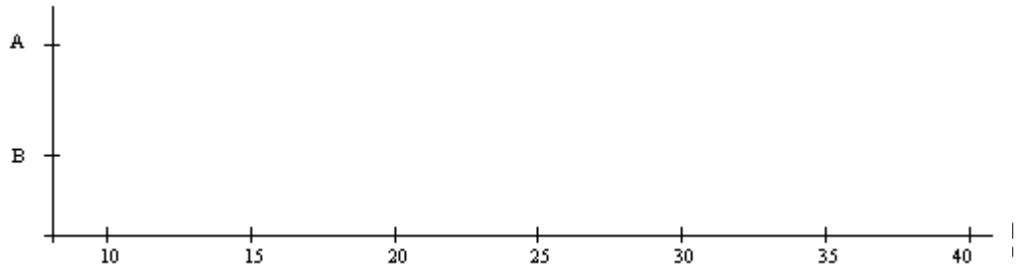
In your simulation, which drug did a better job reducing cholesterol? _____

Combine your results with the rest of the class.

How many times was A better? _____ How many times was B better? _____

What were the mean and standard deviation of the sample means? mean _____ sd _____

Sketch parallel boxplots using the axes below:



How do these results compare to Trial 1? Compare the means and standard deviations as well as the overlap in the boxplots.

How do these results compare to the previous Trial?

What can you conclude about the effectiveness of Gender and Initial Cholesterol as blocking variables?

V. Trial Number 5: Randomized Complete Block—Blocked by Exercise and Initial Cholesterol.

A. Since gender did not help reduce the variability any more than initial cholesterol by itself, we will ignore gender and now introduce exercise level as an additional blocking factor with initial cholesterol.

B. To block by these two variables, use the table that is already sorted by cholesterol. Then, starting at the top, find the first 2 subjects that have exercise level 0. This is the first block. The second block will be the next two subjects with exercise level 0, etc. The fifth block will be the first two subjects with exercise level 1, etc.

C. Within each block, randomly assign one of the subjects to receive drug A and assign the other to drug B.

D. Record the improvement of each subject to their assigned drug in the table below.

E. Calculate the mean improvement for each drug.

Row #	**Improvement using A**	**Row #**	**Improvement using B**
mean improvement		**mean improvement**	

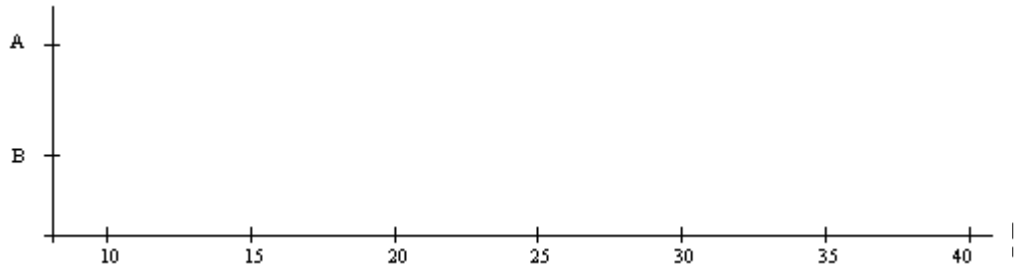
In your simulation, which drug did a better job reducing cholesterol? _____

Combine your results with the rest of the class.

How many times was A better? _____ How many times was B better? _____

What were the mean and standard deviation of the sample means? mean _____ sd _____

Sketch parallel boxplots using the axes below:



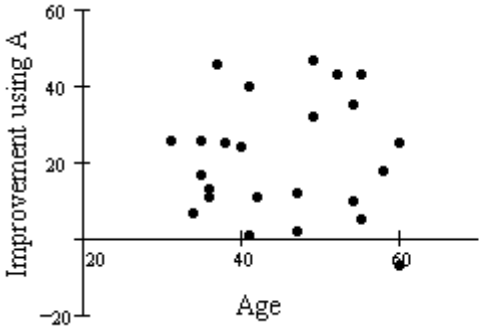
How do these results compare to Trial 1? Compare the means and standard deviations as well as the overlap in the boxplots.

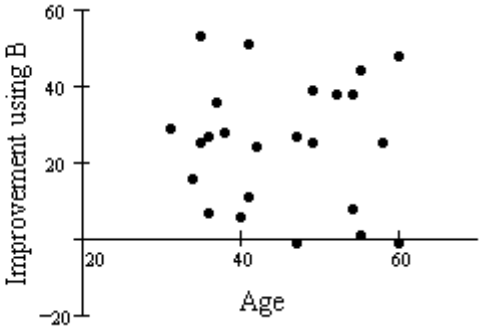
How do these results compare to the previous Trial?

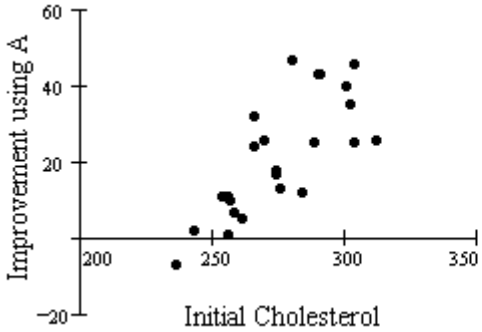
What can you conclude about the effectiveness of Exercise and Initial Cholesterol as blocking variables?

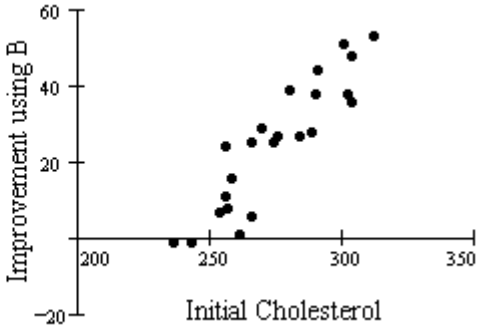
VI. When will blocking be useful?

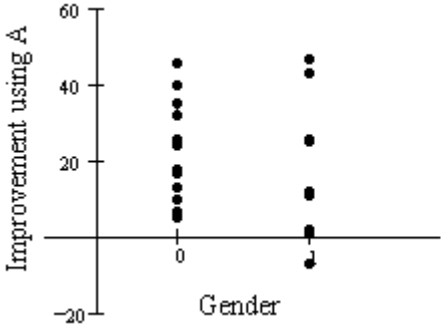
A. Sketch scatterplots of each of the 4 blocking variables vs. improvement using drug A and improvement using drug B. What do you notice about the scatterplots that helps explain the results of Trials 2-5?

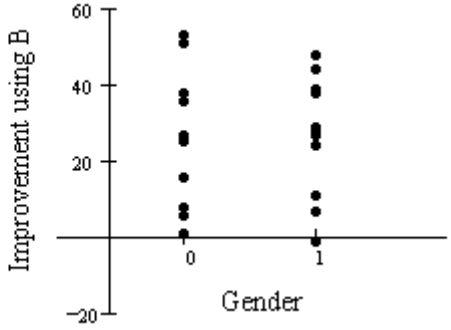


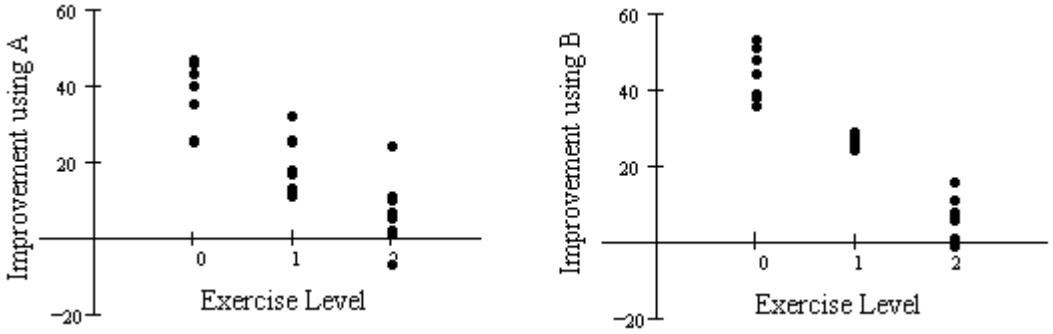












B. Are there times when you shouldn't block? Explain.

NOTE: While we can safely conclude that drug B is more effective than drug A in our 24 subjects, we cannot infer that the drugs will work the same for all the patients of the clinic, since the sample of 24 was not randomly selected. Similarly, we cannot make a conclusion about the general population with complete confidence. Although researchers frequently make this inference, it is a controversial practice.

Sorted by Age

Row #	Gender	Age	Exercise Level	Initial Cholesterol	Reading after A	Reading after B	Improvement using A	Improvement using B
1	1	31	1	270	244	241		
2	0	34	2	258	251	242		
3	0	35	1	274	257	249		
4	0	35	0	312	286	259		
5	1	36	2	254	243	247		
6	0	36	1	276	263	249		
7	0	37	0	304	258	268		
8	1	38	1	289	264	261		
9	0	40	2	266	242	260		
10	1	41	2	256	255	245		
11	0	41	0	301	261	250		
12	1	42	1	256	245	232		
13	1	47	2	243	241	244		
14	1	47	1	284	272	257		
15	0	49	1	266	234	241		
16	1	49	0	280	233	241		
17	1	52	0	290	247	252		
18	0	54	2	257	247	249		
19	0	54	0	302	267	264		
20	0	55	2	261	256	260		
21	1	55	0	291	248	247		
22	0	58	1	274	256	249		
23	1	60	2	236	243	237		
24	1	60	0	304	279	256		

Sorted by Initial Cholesterol

Row #	Gender	Age	Exercise Level	Initial Cholesterol	Reading after A	Reading after B	Improvement using A	Improvement using B
1	1	60	2	236	243	237		
2	1	47	2	243	241	244		
3	1	36	2	254	243	247		
4	1	41	2	256	255	245		
5	1	42	1	256	245	232		
6	0	54	2	257	247	249		
7	0	34	2	258	251	242		
8	0	55	2	261	256	260		
9	0	40	2	266	242	260		
10	0	49	1	266	234	241		
11	1	31	1	270	244	241		
12	0	35	1	274	257	249		
13	0	58	1	274	256	249		
14	0	36	1	276	263	249		
15	1	49	0	280	233	241		
16	1	47	1	284	272	257		
17	1	38	1	289	264	261		
18	1	52	0	290	247	252		
19	1	55	0	291	248	247		
20	0	41	0	301	261	250		
21	0	54	0	302	267	264		
22	0	37	0	304	258	268		
23	1	60	0	304	279	256		
24	0	35	0	312	286	259		