(b) Based on your finding in (a), make a general conjecture as to the value that minimizes the *SAD* function where there is an odd number of observations.
(c) Use Excel to evaluate the *SSD* function for this new data set. At what value does the minimum occur? Is this consistent with your finding about where *SSD* was minimized with eight observations? Explain.

2.1.9 Create an Example

(a) Create a hypothetical example of ten exam scores (say, between 0 and 100 with repeats allowed) such that 90% of the scores are above the mean.
(b) Repeat (a) for the condition that the mean is roughly 40 points less than the median.
(c) Repeat (a) for the condition that the IQR equals 0 and the mean is more than twice the median.

2.1 Summary

This section has introduced you to descriptive analyses for comparing two or more groups with a *quantitative* response variable. One important point is that the basic strategy for analyzing data is the same as in Chapter 1 for a categorical response: Start with graphical displays and numerical summaries. The graphical displays that you have studied are *dotplots*, *histograms*, and *boxplots*. One complication of working with quantitative variables is that their distributions reveal several aspects of interest, and you have learned to focus on *center*, *spread*, *shape*, and *unusual observations*. You have studied numerical summaries of both center (*mean* and *median*) and spread (*standard deviation* and *interquartile range*), and you have learned that the *five-number summary* provides a quick description that is useful for comparing distributions between groups. You have also studied properties of these statistics, such as *resistance* (not being affected by outliers), the *Empirical Rule* (which provides an interpretation of standard deviation for mound-shaped data), *z-scores* (which enable you to compare measurements on different scales), and the principle of *least squares*. You have also practiced *transforming* data so that its distribution becomes more symmetric and the groups easier to compare.

2.2 STATISTICAL SIGNIFICANCE

We now turn our attention to the issue of whether a difference in the distribution of response variables between two groups is statistically significant, that is, larger than would be predicted by randomization alone if there were really no difference between the groups. The concept of statistical significance is the same as you studied earlier, and our strategy for studying it will also be the same: We will use simulations to investigate the pattern of variation that would result from randomization alone, and then we will more formally look at all possible random assignments. Although all of this is the same as in Chapter 1, one complication with quantitative response variables is that we must focus on a specific aspect of the distribution and a specific measure of that aspect. We will concentrate on comparing the centers of the distributions, and we will use both the mean and median as measures of center.

INVESTIGATION 2.2.1 SLEEP DEPRIVATION AND VISUAL LEARNING

Researchers have established that sleep deprivation has a harmful effect on visual learning. In a recent study, Stickgold, James, and Hobson (2000) investigated whether subjects could “make up” for sleep deprivation by getting a full night’s sleep in subsequent nights. This study involved randomly assigning 21 subjects (volunteers between the ages of 18 and 25) to one of two groups: One group was deprived of sleep on the night following training with a visual discrimination task, and the other group was permitted unrestricted sleep on that first night. Both groups were allowed unrestricted sleep on the following two nights and then were retested on the third day. Subjects' performance on the test was recorded as the minimum time (in milliseconds) between stimuli appearing on a computer screen for which they could accurately report what they had seen on the screen. Previous studies had shown that subjects deprived of sleep performed significantly worse the following day, but it was not clear how long these negative effects would last. The data are presented here (a negative value indicates a decrease in performance):

Sleep-deprivation group: −10.7, 4.5, 2.2, 21.3, −14.7, −10.7, 9.6, 2.4, 21.8, 7.2, 10.0
Unrestricted-sleep group: 25.2, 14.5, −7.0, 12.6, 34.5, 45.6, 11.6, 18.6, 12.1, 30.5

a. Is this an experiment or an observational study? Explain.
✎

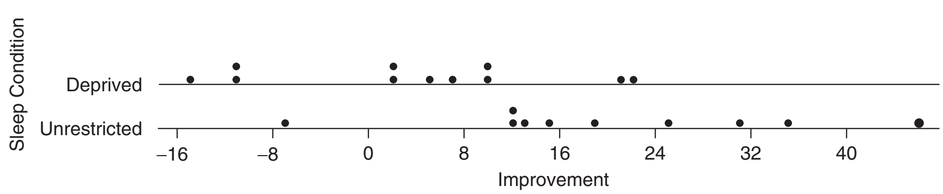
b. Identify the explanatory (EV) and response (RV) variables. Also classify each as being categorical or quantitative, and include a schematic of the experimental design.

EV: Type:

RV: Type:

Design:

The following dotplots reveal the distributions of improvements between the two groups, along with the mean and median improvements in each group:



Sleep group	Sample size	Mean improvement	Median improvement
Deprived	11	3.90	4.50
Unrestricted	10	19.82	16.55

c. Comment on what the dotplots and summary statistics reveal about the researchers' conjecture that sleep deprivation is harmful to learning improvement.
✎

d. Calculate the difference in group means and the difference in group medians (subtracting the "deprived" group's value from the "unrestricted" group's value).
✎

e. Is it *possible* that the differences seen here could have occurred just by chance variation, due to the random assignment of subjects to groups, even if there were really no effect of the sleep condition on improvement?
✎

As with the Friendly Observers experiment from Chapter 1 (Investigation 1.5.1), we again need to judge how much evidence the experimental data provide in support of the researchers' conjecture that sleep deprivation has a harmful effect on learning. We are now working with a quantitative response variable rather than a categorical one, but we will use the same basic logic of *statistical significance*: We will ask whether the observed experimental results are very unlikely to have occurred by chance variation if the explanatory variable has no effect, that is, if the two groups are interchangeable. The general technique is to *simulate a randomization test*: We will take the experimental results, randomly assign them between the two groups, see whether randomization alone produces an outcome as extreme as in the actual research study, and repeat this process a large number of times.

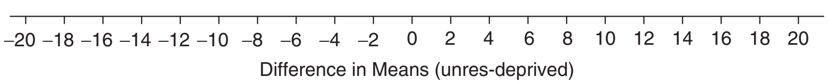
f. On each of 21 index cards write one of the learning improvement values. Then shuffle up the cards and deal out 11 of them to represent the subjects randomly assigned to the "sleep-deprived" group. Calculate the mean and median of the improvements in this group (feel free to use Minitab). Also calculate the mean and median of the improvements for the other 10 subjects, representing those who were randomly assigned to the "unrestricted-sleep" group. Then calculate the difference in group means and difference in group medians, subtracting in the same order as before.

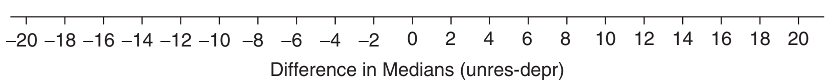
Simulated difference in means:

Simulated difference in medians:

g. Are your simulated results for the difference in means at least as extreme as the actual experimental result? How about for the difference in medians? (*Hint*: Remember that "extreme" is measured in the direction conjectured by the researchers, so in this case "extreme" considers the "unrestricted-sleep" group having this much or even more learning improvement on average than the "sleep-deprived" group.)

h. Combine results with your classmates to produce a dotplot of the difference in group means and a dotplot of the difference in group medians. How many repetitions of the random assignment were conducted altogether? How many (and what proportion) of them produced a difference in means at least as big as in the actual experiment? How many (and what proportion) of them produced a difference in medians at least as big as in the actual experiment?





Recall that these proportions from the class simulation are approximations to the *p-values*.

i. What conclusion would you draw from this simulation regarding the question of whether the learning improvements from the sleep-deprived group are *significantly* lower than those from the unrestricted-sleep group? Also explain the reasoning process by which your conclusion follows from the class simulation results.

Simulating more repetitions would provide a better understanding of how significant (i.e., unlikely to have happened by randomization alone) these experimental results are. In other words, more repetitions will enable us to approximate these *p*-values more accurately. We will again turn to technology, in particular to Minitab, to perform the simulation more quickly and efficiently.

j. Open the Minitab worksheet `SleepDeprivation.mtw` and enable commands. Notice that C1 contains the learning improvement values and C2 contains the group designation. Use Minitab to do one repetition of the random assignment of these 21 learning improvement values among the two groups by typing:

`MTB> sample 21 c2 c3`	Randomizes the contents of C2, storing results in C3
`MTB> unstack c1 c4 c5;`	Separates the values in C1 into 2 columns
`SUBC> subs c3;`	using the group labels assigned in C3
`SUBC> varnames.`	Uses the category names as column name

Determine the two group means and calculate their difference.

```
MTB> let c6 = mean(c4)-mean(c5)
```

✎

Our goal now is to repeat this randomization process many times and accumulate the results. To do this, you will write a Minitab macro. The macro will repeatedly execute the preceding commands. To store the results from each repetition, the macro will also incorporate a counter (called `k1`), which keeps track of the repetition number and also the row number in `C6` where the results of the subsequent repetitions will be stored.

k. Use a text editor (such as Notepad) to create a file consisting of the following commands (punctuation is important):

```
sample 21 c2 c3
unstack c1 c4 c5;
subs c3;
varnames.
let c6(k1)=mean(c4)-mean(c5)
let k1=k1+1
```

Save this file as "randomize.mtb" (with the quotation marks so that the mtb extension is preserved). Each time this macro is run, the counter `k1` will increase by 1 so that the difference in means is stored in a subsequent row of `C6`. Before you run the macro, you need to initialize the counter in the Session window:

```
MTB> let k1=1
```

Then choose File > Other Files > Run an Exec. Click "Select File" and then locate `randomize.mtb` on your computer. Run the macro once. You should see the contents of C2 have been randomly "shuffled" and put into C3, and the resulting difference in groups means is stored in C6. Report the difference in means that you obtained.

Simulated difference in means:

l. Now execute the macro 999 more times. The macro will continue to place resulting differences into C6. When the macro is finished, construct a dotplot or histogram of the differences in group means. Describe this empirical randomization distribution, paying particular attention to where the actual difference found in the experiment falls.

✎

m. What proportion of the differences in group means exceeds the difference in the actual study? Use Minitab to answer this by creating an indicator variable:

```
MTB> let c8=(c6<=-15.92)
MTB> tally c8
```

Column 8 shows a 1 if this inequality is met and a 0 if not.

Proportion of differences at most 15.92:

n. Reanswer (i) about the conclusion that you would draw based on this larger number of repetitions.

✎

o. Does the design of this study allow you to conclude that the reduction in learning improvement is due to the sleep deprivation? Explain.

✎

Study Conclusions

These data come from a randomized, comparative experiment. The dotplots and descriptive statistics reveal that the sleep-deprived subjects tended to have lower improvements than those permitted unrestricted sleep. To investigate whether this difference is larger than could be expected from randomization alone (assuming no real difference between the two conditions), you simulated a randomization test by assigning the 21 measurements to the two groups at random. You should have found that randomization alone rarely produced group differences as extreme as in the actual study (the *p*-value is less than .01 for comparing the means). Thus, we have fairly strong evidence that the learning improvements are genuinely lower with the sleep-deprived subjects. Moreover, because this was a randomized experiment and not an observational study, we can draw a causal conclusion that the sleep deprivation was the cause of the lower learning improvements.

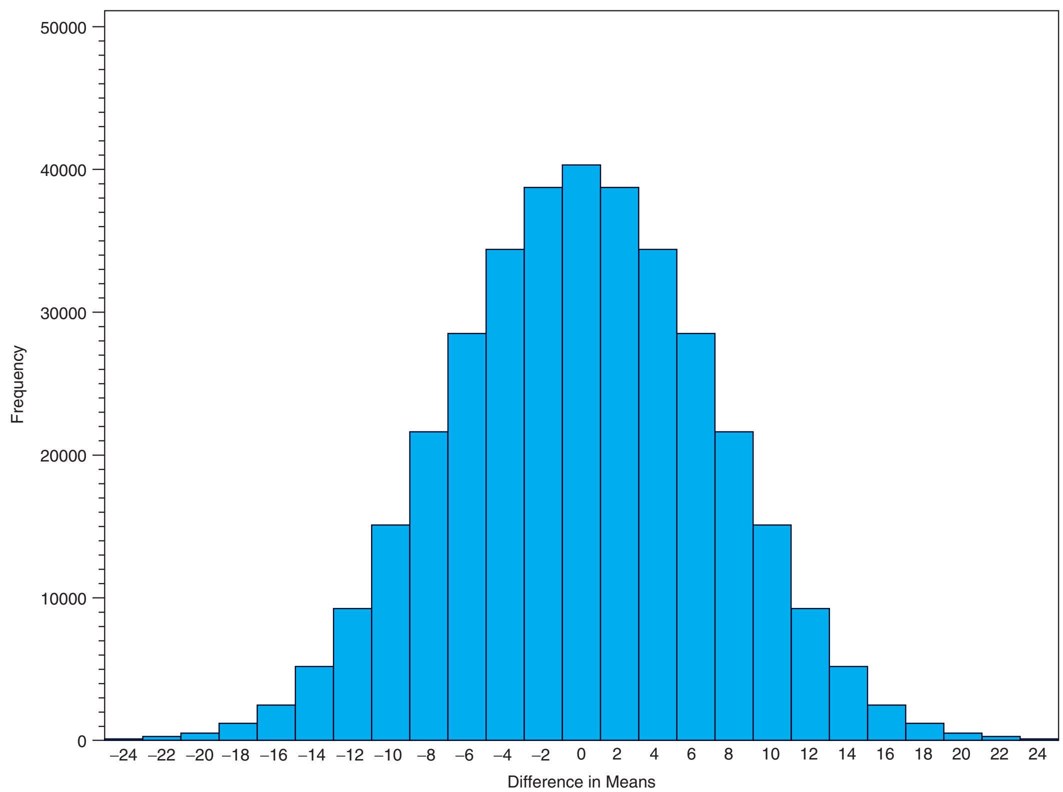
In Chapter 1 you followed up on the simulation analysis by employing a mathematical probability model to determine the *p*-value exactly. Fisher's exact test only applies to categorical variables, but the principle is the same in this study with a quantitative response. We can calculate the exact *p*-value of this randomization test by

- Listing all possible ways to randomly assign the 21 subjects to two groups of 11 and 10
- Determining the difference in group means (or medians) for all of those random assignments
- Examining the distribution of those differences, and
- Counting how many and what proportion are at least as large as the actual experiment's difference.

p. How many different random assignments are there of 21 subjects into a group of 11 and a group of 10? (*Hint*: Use a combination.)

✎

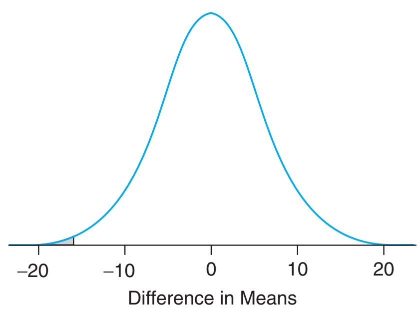
q. The following histogram (from SAS®) shows the distribution of all 352,716 differences in group means. Does this resemble the (approximate) distribution as determined by your simulation results? Explain.



✎

r. It turns out that 2,533 of the 352,716 different random assignments produce a difference in group means of at least 15.92. Use this information to determine the exact *p*-value of this randomization test. Is it close to the approximate *p*-value from your simulation?
✎

As you might expect, it can be rather tedious, even with computers, to list out all of these possible random assignments. And these sample sizes are relatively small! One shortcut is to only count how many assignments give results more extreme than the one observed, but we can often appeal to a mathematical model as well. The distribution of the differences in means from these randomizations will always produce a symmetric distribution, which, when standardized, can often be reasonably modeled by what's called a *t* distribution. You will learn more about the *t* distribution in later chapters. For now, we will tell you that the *t* distribution model approximates the *p*-value to be .0069. This provides an alternative approximation of the *p*-value, which will be particularly accurate with large sample sizes.



INVESTIGATION 2.2.2 MORE SLEEP DEPRIVATION

Reconsider the summary statistics and boxplots for the sleep deprivation study:

Sleep group	Sample size	Mean improvement	Median improvement
Deprived	11	3.90	4.50
Unrestricted	10	19.82	16.55