Think It Through

If H_0: identical population distributions for the three groups were true, the Kruskal-Wallis test statistic would have an approximate chi-squared distribution with $df = 2$. Table 15.8 reports that the test statistic is H = 0.72. The P-value is the right-tail probability above 0.72. Table 15.8 reports this as 0.696, about 0.7. It is plausible that GPA is independent of dating group. Table 15.8 shows that the sample median GPAs are not very different, and since the sample sizes are small, these sample medians do not give much evidence against H_0.

Insight

If the P-value had been small, to find out which pairs of groups significantly differ, we could follow up the Kruskal-Wallis test by a Wilcoxon test to compare each pair of dating groups. Or, we could find a confidence interval for the difference between the population medians for each pair.

> *Try Exercise 15.8*

Comparing Matched Pairs: The Sign Test

Chapter 10 showed that it's possible to compare groups using either *independent* or *dependent* samples. So far in this chapter, the samples have been independent. When the subjects in the two samples are matched, such as when each treatment in an experiment uses the same subjects, the samples are dependent. Then we must use different methods.

For example, the tanning experiment from Examples 1 and 2 could have used a crossover design instead: The participants get a tan using one treatment, and after it wears off they get a tan using the other treatment. The order of using the two treatments is random. For each participant, we observe which treatment gives the better tan. That is, we make comparisons by pairing the two observations for the same participant.

For such a matched pairs experiment, let p denote the population proportion of cases for which a particular treatment does better than the other treatment. Under the null hypothesis of identical treatment effects, $p = 0.50$. That is, each treatment should have the better response outcome about half the time. (We ignore those cases in which each treatment gives the *same* response.) Let n denote the sample number of pairs of observations for which the two responses differ. For large n, we can use the z test statistic to compare the sample proportion $\hat{p}$ to the null hypothesis value of 0.50. (See the margin Recall box.) The P-value is based on the approximate standard normal sampling distribution.

A test that compares matched pairs in this way is called a **sign test**. The name refers to how the method evaluates for each matched pair whether the difference between the first and second response is *positive* or *negative*.

Recall

From Section 9.2, to test H_0: $p = 0.50$ with sample proportion $\hat{p}$ when $n \geq 30$, the test statistic is

$$z = (\hat{p} - 0.50)/se,$$

with $se = \sqrt{(0.50)(0.50)/n}$. ◄

SUMMARY: Sign Test for Matched Pairs

1. **Assumptions:** Random sample of matched pairs for which we can evaluate which observation in a pair has the better response.
2. **Hypotheses:** H_0: Population proportion $p = 0.50$ who have better response for a particular group
   H_a: $p \neq 0.50$ (two-sided) or H_a: $p > 0.50$ or H_a: $p < 0.50$ (one-sided)
3. **Test statistic:** $z = (\hat{p} - 0.50)/se$, as shown in margin recall box.
4. **P-value:** For large samples ($n \geq 30$), use tail probabilities from standard normal. For smaller n, use binomial distribution (discussed in Example 7).
5. **Conclusion:** Report the P-value and interpret in context.