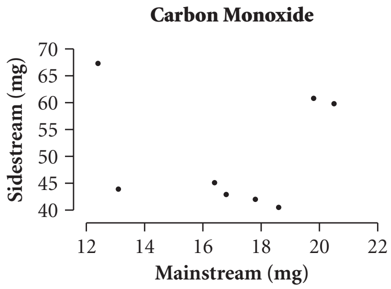
other brands, but here, also, the justification does not come from the logic of random samples.

The two "conditions" being compared, sidestream versus mainstream, cannot be assigned as in a true experiment, so inference about cause, of the sort that can be made from experimental data, is not possible here.

The only statistical inference you can make here is to rule out a pattern caused just by chance: Is the observed difference too big to have occurred just by chance? In this kind of inference, use *P*-values as a kind of ruler for measuring the size of the observed difference.

d. A scatterplot shows no evidence of a positive correlation (between sidestream and mainstream readings) of the sort that indicates that a paired *t*-test will give much lower *P*-values than a two-sample *t*-test. Brand H stands out because of its high sidestream and low mainstream carbon monoxide.



e. A stemplot for the differences is strongly skewed toward the high values.

```
2|19 42 61 87   Key: 2|19 = 21.9
3|08 93
4|10
5|49
```

Because the sample size is only 8, you should expect the skewness to lead to a capture rate that is less than advertised. It follows that the *P*-value will be artificially (and incorrectly) low.

f. Here are summaries.

	Sidestream	Mainstream	Side − Main
n	8	8	8
$\bar{x}$	50.29	16.93	33.36
s	10.54	2.93	11.04

The paired *t*-statistic is $t = 8.54$ on 7 *df*, which gives a two-tailed *P*-value far below 0.0001.

The two-sample *t*-statistic is $t = 8.629$ on 8.1 degrees of freedom, with a *P*-value far below 0.0001. The *P*-values of < 0.0001 are similar, which was predicted in parts b and d.

g. The evidence is so strong, "you don't need a weatherman to know which way the wind blows."

For the eight brands tested, the carbon monoxide levels in sidestream smoke are much higher than the levels in mainstream smoke. A difference this large cannot be reasonably attributed to chance.

AP Sample Test

AP1. C. All of these methods should reduce the margin of error. However, decreasing the confidence level will not only lead to a smaller calculated margin of error due to the smaller value of t^*, it will also decrease her confidence that the interval contains the true angle.

AP2. D. While a log transformation would decrease the value of the mean, making the mean smaller is not a reason for using a log transformation. (See also page 604 of the student book.)

AP3. C. Because s is less than σ in more than half of all samples, t will be larger in absolute value than z in more than half of all samples.

AP4. B. Looking at $df = 14$, the (two-sided) *P*-value that corresponds to this t is about 0.015, so the null hypothesis can be rejected at the 0.05 level but not the 0.01 level.

AP5. E. The data should be paired and analyzed using a confidence interval for the mean difference.

AP6. B. Do not use a transformation to try to eliminate a single outlier when there is no skewness in the other values.

AP7. A. The low *P*-value indicates that the difference in mean scores that was actually observed would be unlikely if there's no real difference between the population means before and after.

AP8. E. Since the distribution of their data fits conditions for inference, the *t*-procedure is probably accurate enough for the subpopulation of dragons they actually observed.

AP9. **a.** No mention is made of whether the treatments were randomly assigned to subjects. Both distributions are skewed right with outliers. One group has 43 subjects, enough that the shape should not be a concern. However, the other group has 35 subjects and at least three outliers. It is worth trying to find a transformation that would make the distribution more symmetric.

b. As discussed in part a, using a transformation on the data would be better. The larger concern, however, is that the researchers pooled the variances when the standard deviations are not close to each other. They should not have pooled.

c. *State your hypotheses:*

H_0: $\mu_1 = \mu_2$, where μ_1 is the mean time it would take all 78 subjects to see the diamond if none of them had been shown a drawing of the diamond and μ_2 is the mean time it would have taken them to see the diamond if they had all been shown the drawing

H_a: $\mu_1 \neq \mu_2$

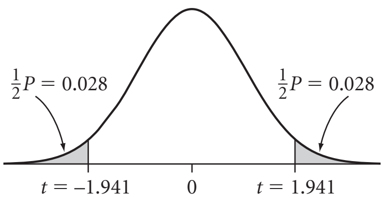
Calculate the test statistic, find the P-value, and draw a sketch: The pooled variance is the weighted average of the two sample variances. The sample variances are $8.0854^2 \approx 65.374$ and $4.8017^2 \approx 23.056$. The weighted average of these is $\frac{43}{78} \cdot 65.374 + \frac{35}{78} \cdot 23.056 \approx 46.385$. The pooled standard deviation, then, is $\sqrt{46.385} \approx 6.811$ or 6.8149 using the two-sample t-test and no rounding.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(8.560 - 5.551) - 0}{6.811\sqrt{\frac{1}{43} + \frac{1}{35}}} \approx 1.941$$

The calculator reports 76 degrees of freedom, and the two-sample t-test gives a t-statistic of 1.9395.

Note: The denominator is on the AP Exam formula sheet as the standard deviation of the difference in means in the "special case" where two population standard deviations are equal. The estimate of σ is the weighted average of the sample variances.



You get a P-value of around 0.056.

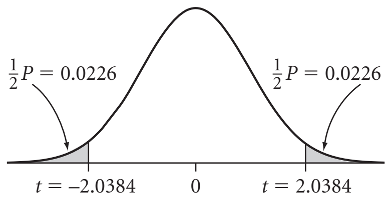
Write your conclusion in context, linked to your computations: Because the P-value is greater than the stated $\alpha = 0.05$, do not reject the null hypothesis. You do not have statistically significant evidence that those who saw a drawing of the diamond before the test had a different mean time to see the diamond than those who did not see the drawing of the diamond beforehand.

d. Repeating the test with unpooled variances, the test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(8.560 - 5.551) - 0}{\sqrt{\frac{8.085^2}{43} + \frac{4.802^2}{35}}} \approx 2.038$$

where the calculator reports 70.04 degrees of freedom. The P-value is 0.045. In this case the null hypothesis would have been rejected.



e. Pooling the variances when the standard deviations are not comparable makes it harder to reject the null hypothesis. In other words, it reduces the power of the test.

AP10. a. There is still no mention of randomization. However, the distributions are now fairly symmetric with no outliers. If treatments were randomly assigned, conditions are now met for inference.

b. *State your hypotheses:*

H_0: $\mu_1 = \mu_2$, where μ_1 is the mean of the logs of the time it would take all 78 subjects to see the diamond if none of them had been shown a drawing of the diamond and μ_2 is the mean of the logs of the time it would have taken them to see the diamond if they had all been shown the drawing

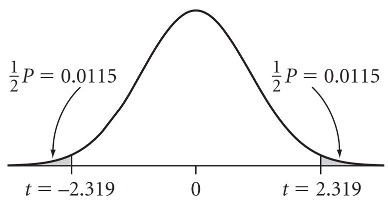
H_a: $\mu_1 \neq \mu_2$

Compute the test statistic, find the P-value, and draw a sketch: The pooled variance is the weighted average of the sample variances: $s_p^2 = \frac{43}{78} \cdot 0.1249 + \frac{35}{78} \cdot 0.1261 \approx 0.1254$, so $s_p \approx 0.3542$.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(0.7904 - 0.6034) - 0}{0.3542\sqrt{\frac{1}{43} + \frac{1}{35}}} \approx 2.319$$

where the calculator reports 76 degrees of freedom. The P-value is 0.0231.



Write your conclusion in context, linked to your computations: Because the P-value is less than $\alpha = 0.05$ stated in AP9, reject the null hypothesis. You have statistically significant evidence that the means of the log of the times required to see the diamond is different if the people see a drawing of the diamond previously than if they do not.

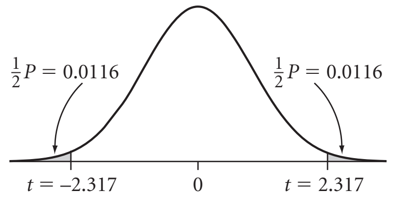
c. *Compute the test statistic, find the P-value, and draw a sketch:*

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(0.7904 - 0.6034) - 0}{\sqrt{\frac{0.3534^2}{43} + \frac{0.3552^2}{35}}}$$

$$\approx 2.318$$

where the calculator reports 72.67 degrees of freedom and a *P*-value of 0.0233.



d. There was little effect. The *P*-value was slightly lower with the pooled variance, so there was a very small increase in power.