Part IV Review Exercises

IV.1 (a) We label each subject using labels 01, 02, 03, …, 44, and then enter the partial table of random digits and read two-digit groups. The labels 00 and 45 to 99 are not used in this example, so we ignore them. We also ignore any repeats of a label, since that subject is already assigned to a group. We need to pick 22 subjects in this way to have the regular chips first (the other 22 subjects will have the fat-free chips first), but here we pick only the first 5. The first two-digit group is 19, so the subject with label 19 is in the regular chips first group. The second two-digit group is 22 and the third two-digit group is 39, so the subjects with labels 22 and 39 are also in the regular chips first group. The fourth two-digit group is 50, which we ignore. The next two two-digit groups are 34 and 05, so the subjects with labels 34 and 05 are in the regular chips first group. (b) Since we want to compare the amounts of regular and fat-free chips eaten and each woman serves as her own control, we will use a paired *t* test. The hypotheses are $H_0: \mu_{\text{diff}} = 0$ and $H_a: \mu_{\text{diff}} \neq 0$, where "diff" = (weight in grams of regular potato chips eaten) – (weight in grams of fat-free potato chips eaten).

IV.2 Step 1: **Hypotheses**

$H_0:$ There is no difference in response between walking and resting flies.

$H_a:$ There is some difference in response between walking and resting flies.

Step 2: **Conditions** The expected cell counts are

	Response to Vibrate?	
	Yes	No
Fly was walking	$\frac{38\times 64}{96} = 25.\overline{3}$	$\frac{58\times 64}{96} = 38.\overline{6}$
Fly was resting	$\frac{38\times 32}{96} = 12.\overline{6}$	$\frac{58\times 32}{96} = 19.\overline{3}$

All expected cell counts are greater than 1 and no expected cell counts are less than 5 (the smallest expected cell count is 12.7).

Step 3: **Calculations**

- **Test statistic**

$$\begin{aligned} \mathrm{X}^2 &= \sum \frac{(O-E)^2}{E} \\ &= 9.2807 + 6.0805 + 18.5614 + 12.1609 \\ &= 46.0835 \end{aligned}$$

- ***P*-value** Under the null hypothesis, the test statistic has a χ^2 distribution with 1 degree of freedom. The *P*-value is less than 0.0005 (technology gives us 1.13×10^{-11}).

Step 4: **Interpretation:** Because the expected cell counts are all large, the *P*-value from Table E will be quite accurate. There is strong evidence to reject H_0 $\left(\mathrm{X}^2 = 46.0835,\ \text{df} = 1, P\text{-value} < 0.0005\right)$ and conclude that resting flies respond differently than flies that are walking.

IV.3

```
Stem-and-leaf of Diabetic and Normal Mice  N  = 24
Leaf Unit = 0.1

          Diabetic Mice |    | Normal Mice
                      0 |  1 |
                        |  2 |
                        |  3 |
                        |  4 | 9
                        |  5 |
                        |  6 | 6
                      4 |  7 | 277
                     85 |  8 | 4578
                      8 |  9 | 145
                 854310 | 10 | 04
                      7 | 11 |
                        | 12 | 6
                      6 | 13 | 38
                    874 | 14 | 5
                      2 | 15 |
                      4 | 16 |
                      2 | 17 |
                     42 | 18 |
                     62 | 19 |
                        | 20 |
                        | 21 |
                      6 | 22 |
```

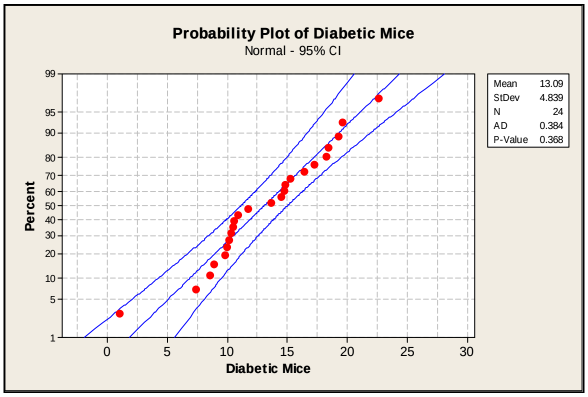
It appears that the potentials for diabetic mice are, in general, higher than those for normal mice. The potentials for diabetic mice appear to be right-skewed, while the potentials for normal mice are somewhat symmetric (although possibly slightly right-skewed). There appears to be one low outlier (1.05) for the diabetic mice—although it does not satisfy the 1.5 × *IQR* rule, this value is far removed from the rest of the data.

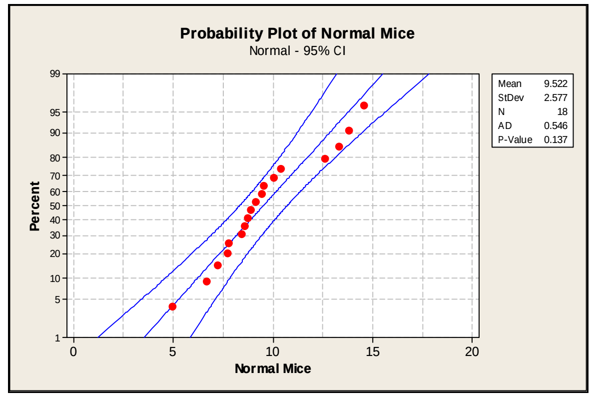
(b) <u>Step 1:</u> **Hypotheses** We want to compare the mean difference in electrical potentials of diabetic mice and normal mice. Because the researchers did not conjecture in advance that the potentials would be higher in diabetic mice, we use the two-sided alternative:

$$\begin{matrix} H_0 : \mu_{\text{diabetic}} = \mu_{\text{normal}} \\ H_a : \mu_{\text{diabetic}} \neq \mu_{\text{normal}} \end{matrix} \quad \text{or, equivalently,} \quad \begin{matrix} H_0 : \mu_{\text{diabetic}} - \mu_{\text{normal}} = 0 \\ H_a : \mu_{\text{diabetic}} - \mu_{\text{normal}} \neq 0 \end{matrix}$$

<u>Step 2:</u> **Conditions** Since both population standard deviations are unknown, we should use a two-sample *t* test if the conditions are met.

- **SRS** We will treat both samples as if they were SRSs from their respective populations—we have no reason to believe that these samples are not representative of their respective populations.
- **Normality** Normal probability plots show the one unusually low value for the diabetic mice, but are otherwise fairly linear. It appears that the two populations from which these mice came are plausibly Normal.





- **Independence** The two samples are independent since a mouse cannot have diabetes and not have diabetes at the same time.

Step 3: **Calculations**

The Minitab printout reports the results of the two-sample *t* procedure.

```
Two-sample T for Diabetic Mice vs Normal Mice

                N   Mean  StDev  SE Mean
Diabetic Mice  24  13.09   4.84     0.99
Normal Mice    18   9.52   2.58     0.61


Difference = mu (Diabetic Mice) - mu (Normal Mice)
Estimate for difference:  3.56736
95% CI for difference:  (1.21573, 5.91900)
T-Test of difference = 0 (vs not =): T-Value = 3.08  P-Value = 0.004  DF = 36
```

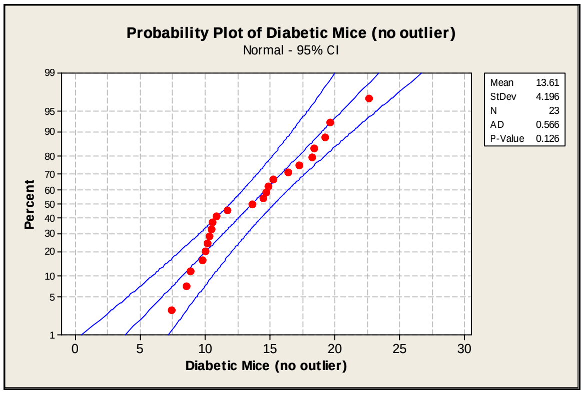
- **Test statistic** The two-sample *t* statistic has the value $t = 3.08$.
- ***P*-value** The degrees of freedom are 36, and the *P*-value is 0.004.

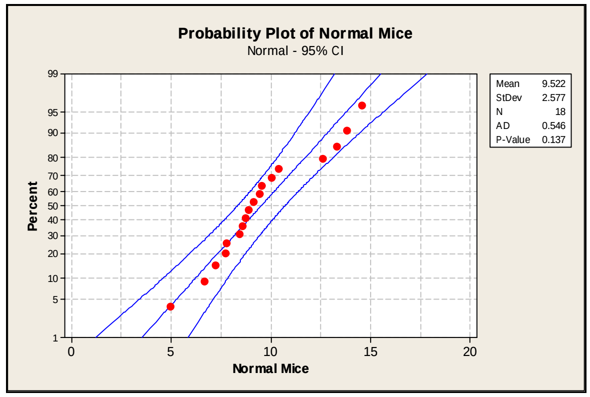
Step 4: **Interpretation** The low *P*-value ($P = 0.004$) provides strong evidence against H_0. We reject H_0 at any reasonable significance level and conclude that the mean potentials differ in diabetic and normal mice. The 95% confidence interval provides additional information; namely, that we are 95% confident that the actual difference in mean potentials is between 1.2 and 5.9 millivolts higher in diabetic than normal mice.

(c) Here are the revised steps 2, 3, and 4 for the significance test from part (b) with the unusually low value (1.05) from the diabetic mice removed.

Step 2: **Conditions** Since both population standard deviations are unknown, we should use a two-sample *t* test if the conditions are met.

- **SRS** We will treat both samples as if they were SRSs from their respective populations—we have no reason to believe that these samples are not representative of their respective populations.
- **Normality** The Normal probability plots no longer show the unusual value for the diabetic mice, and are fairly linear in shape. It appears that the populations are plausibly Normal.





- **Independence** The two samples are independent since a mouse cannot have diabetes and not have diabetes at the same time.

Step 3: **Calculations**
The Minitab printout reports the results of the two-sample *t* procedure.

```
Two-sample T for Diabetic Mice (no outlier) vs Normal Mice

                     N   Mean  StDev  SE Mean
Diabetic Mice       23  13.61   4.20     0.87
Normal Mice         18   9.52   2.58     0.61

Difference = mu (Diabetic Mice (no outlier)) - mu (Normal Mice)
Estimate for difference:  4.09082
95% CI for difference:  (1.93287, 6.24878)
T-Test of difference = 0 (vs not =): T-Value = 3.84  P-Value = 0.000  DF = 37
```

- **Test statistic** The two-sample *t* statistic has the value $t = 3.84$.
- ***P*-value** The degrees of freedom are 37, and the *P*-value is 0.000.

Step 4: **Interpretation** The extremely low *P*-value ($P = 0.000$) provides strong evidence against H_0. We reject H_0 at any reasonable significance level and conclude that the mean potentials differ for diabetic and normal mice. From the reported confidence interval, we can be 95% confident that the actual difference in mean potential is between 1.9 and 6.25 millivolts higher in diabetic than normal mice.

The unusually low value does not affect the conclusion.

IV.4 (a) Step 1: **Parameter** The population is all 17-year-olds who are still in school. The parameter of interest is *p*, the actual proportion of these students who have at least one college graduate parent.
Step 2: **Conditions** We should use *z* procedures to estimate *p* if the conditions are satisfied.

- **SRS** The sample was taken using a multistage design, but we are told that the "overall effect is quite similar to an SRS of 17-year-olds who are still in school."
- **Normality** We check the counts of "successes" and "failures":

$$n\hat{p} = 1014 \geq 10 \text{ and } n\left(1-\hat{p}\right) = 1144 \geq 10$$

- **Independence** There are at least (10)(2158) = 21,580 teens aged 17 who are still in school.

Step 3: **Calculations** A 99% confidence interval for *p* is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.47 \pm 2.576\sqrt{\frac{0.47(0.53)}{2158}}$$
$$= 0.47 \pm 0.028 = (0.442, 0.498)$$

Step 4: **Interpretation** We can say with 99% confidence that the true proportion of teens aged 17 who are still in school is between 0.442 and 0.498.

(b) I would expect that 17-year-olds who had dropped out of school would be less likely to have at least one college graduate parent. Because of this, I would expect my estimate in part (a) to be too high.

IV.5 Step 1: **Hypotheses** We want to compare the mean mathematics scores of males and females in the population of all 17-year-old students. Because we are asked whether male or female scores differ on the mathematics portion of the NAEP, we use the two-sided alternative:

$$\begin{array}{l} H_0 : \mu_{\mathrm{W}} = \mu_{\mathrm{M}} \\ H_a : \mu_{\mathrm{W}} \neq \mu_{\mathrm{M}} \end{array} \quad \text{or, equivalently,} \quad \begin{array}{l} H_0 : \mu_{\mathrm{W}} - \mu_{\mathrm{M}} = 0 \\ H_a : \mu_{\mathrm{W}} - \mu_{\mathrm{M}} \neq 0 \end{array}$$

Step 2: **Conditions** Since both population standard deviations are unknown, we should use a two-sample *t* test if the conditions are met.

- **SRS** We are told that the sampling used in the NAEP approximates simple random sampling.
- **Normality** We do not have the data, but our sample sizes are so large that two-sample *t* procedures should be robust even if the population distributions aren't close to Normal.
- **Independence** The two samples are independent since a 17-year-old student is either a male or a female.

Step 3: **Calculations**

- **Test statistic** The two-sample *t* statistic is

$$t = \frac{\bar{x}_W - \bar{x}_M}{\sqrt{\frac{s_W^2}{n_W} + \frac{s_M^2}{n_M}}} = \frac{305 - 308}{\sqrt{(0.9)^2 + (1.0)^2}} = \frac{-3}{1.345} = -2.23$$

- ***P*-value** There are 1035 degrees of freedom, so we will use the df = 1000 row. Table C shows that the upper tail probability lies between 0.01 and 0.02, so the *P*-value for this two-sided significance test is between 0.02 and 0.04.

Step 4: **Interpretation** The low *P*-value ($0.02 < P < 0.04$) provides some evidence against H_0. We reject H_0 at the 0.05 significance level and conclude that the mean mathematics scores are different for 17 year old females than for 17 year old males.

IV.6 (a) Randomly assign 20 infants to the PBM group, and 19 infants to each of the NLCP, PL-LCP, and TG-LCP groups. Each infant will receive the supplement assigned to him/her. We then compare the development of the infants in each of the four groups.

(b) Step 1: **Hypotheses** We want to use X^2 to compare the distribution of gender in the four treatment groups. Our hypotheses are

H_0: There is no difference in gender among the groups.

H_a: There is some difference in gender among the groups.

Step 2: **Conditions** To use the chi-square test for homogeneity of populations:

- The data must come from independent SRSs from the population of interest. We are willing to treat the subjects in the four groups as SRSs from their respective populations since the subjects were randomly assigned to treatment groups.
- All expected cell counts are greater than 1, and no more than 20% are less than 5. The smallest expected cell count is 8.636.

  Expected cell counts:

	PBM	NLCP	PL-LCP	TG-LCP
Female	10.909	10.364	10.364	10.364
Male	9.091	8.636	8.636	8.636

Step 3: **Calculations**

- **Test statistic**

$$\begin{aligned} X^2 &= \sum \frac{(O-E)^2}{E} \\ &= 0.0008 + 0.0390 + 0.0390 + 0.1795 + 0.0009 + 0.0468 + 0.0468 + 0.2154 \\ &= 0.5682 \end{aligned}$$

  We have 3 degrees of freedom for this statistic, since we have $r = 2$ genders and $c = 4$ groups: df = $(r - 1)(c - 1) = (2 - 1)(4 - 1) = 3$.
  Under the null hypothesis that the randomization worked and the gender distributions are the same in all 4 groups, the test statistic X^2 has a χ^2 distribution with df = 3.
- ***P*-value** To obtain the *P*-value, look at the df = 3 row in Table E. The calculated value $X^2 = 0.5682$ lies below the critical value for probability 0.25. The *P*-value is therefore greater than 0.25. (TI-84 gives $P = 0.904$).

Step 4: **Interpretation** Because the expected cell counts are large enough, the *P*-value from Table E will be accurate. There is not enough evidence to reject H_0 ($X^2 = 0.5682$, df = 3, $P >$ 0.25), so we conclude that the randomization worked—that there is no gender difference among the groups.

IV.7 (a) This is an observational study, since researchers did not determine whether babies had VLBW or not.

(b) Step 1: **Hypotheses** We want to test

$$H_0 : p_{\text{VLBW}} = p_{\text{normal}}$$

$$H_a : p_{\text{VLBW}} < p_{\text{normal}}$$

Where p_{VLBW} and p_{normal} are the proportions of very low birth weight and normal babies, respectively, who graduate from high school.

Step 2: **Conditions** If the conditions are satisfied, we should perform a two proportion *z* test.

- **SRS** We must assume that these two samples of babies are representative of their respective populations.
- **Normality** Assuming $H_0 : p_{\text{VLBW}} = p_{\text{normal}}$ is true, our best estimate for the proportion of babies in either population that will graduate from high school is

$$\hat{p}_C = \frac{179+193}{242+233} = 0.783$$

Since $n_{\text{VLBW}}\hat{p}_C = 189.5$, $n_{\text{VLBW}}(1-\hat{p}_C) = 52.5$, $n_{\text{normal}}\hat{p}_C = 182.4$, and $n_{\text{normal}}(1-\hat{p}_C) = 50.6$, we are safe using Normal approximation.

- **Independence** The data come from two independent samples of babies.

Step 3: **Calculations**

- **Test statistic** The two proportion z statistic is

$$z = \frac{\hat{p}_{\text{VLBW}} - \hat{p}_{\text{normal}}}{\sqrt{\hat{p}_C(1-\hat{p}_C)\left(\dfrac{1}{n_{\text{VLBW}}}+\dfrac{1}{n_{\text{normal}}}\right)}} = \frac{0.740-0.828}{\sqrt{0.783(0.217)\left(\dfrac{1}{242}+\dfrac{1}{233}\right)}} = -2.33$$

- ***P*-value** From Table A, the area to the left of z = –2.33 under the standard Normal curve is 0.0099.

Step 4: **Interpretation** Since the P-value, 0.0099, is smaller than any of our standard significance levels (α = 0.10, 0.05, 0.01), there is strong evidence to reject H_0. We conclude that the graduation rate of VLBW babies is lower than the graduation rate of normal birth weight babies.

(c) Step 1: **Parameter** We want to estimate the difference in mean IQ scores between VLBW and non-VLBW men.

Step 2: **Conditions**

- **SRS** These samples are not SRSs of their populations, but we will assume that they are representative of their respective populations.
- **Normality** The sample sizes (n = 113 and n = 106) are large enough to ensure the robustness of the two sample t procedures, even if the two population distributions aren't close to Normal.
- **Independence** Men were either born VLBW or not, so the two samples are independent.

Step 3: **Calculations** A 90% confidence interval for the difference in mean IQ scores between VLBW and non-VLBW men is

$$(\bar{x}_1-\bar{x}_2) \pm t^*\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = (87.6-94.7)\pm 1.660\sqrt{\frac{15.1^2}{113}+\frac{14.9^2}{106}}$$
$$= -7.1 \pm 3.37 = (-10.47, -3.73)$$

using df = 100.

Step 4: **Interpretation** We are 90% confident that the mean IQ score for VLBW men is between 3.73 and 10.47 points below the mean IQ score for non-VLBW men.

IV.8 (a) t procedures will be reasonably accurate for these data since we have very large samples. The central limit theorem will provide the Normality we need.

(b) Step 1: **Parameter** The population of interest is female mice. We want to estimate the mean endurance μ of female mice swimming.

Step 2: **Conditions** We will use the one-sample t procedures if the conditions are met.

- **SRS** We are not told that this is an SRS of female mice. We will assume that this sample of female mice is representative of the population.

- **Normality** We do not have the data for swimming endurance. However, since the sample size is large, we are safe using *t* procedures even if the population distribution isn't close to Normal.
- **Independence** We must assume that there are at least (10)(162) = 1620 female mice. The swim times of the individual female mice should be independent.

Step 3: **Calculations** The mean endurance for female mice is $\bar{x} = 11.4$ minutes. The confidence interval formula is $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$. We use the *t* distribution with df = 162 – 1 = 161. There is no row corresponding to 161 degrees of freedom in Table C, so we will use df = 100. At the 95% confidence level, the critical value is $t^* = 1.984$. So the 95% confidence interval for *μ* is

$$\begin{aligned} \bar{x} \pm t^* \frac{s}{\sqrt{n}} &= 11.4 \pm 1.984 \frac{26.09}{\sqrt{162}} \\ &= 11.4 \pm 4.07 \\ &= (7.33, 15.47) \end{aligned}$$

Step 4: **Interpretation** We are 95% confident that the true mean endurance of female mice swimming is between 7.33 and 15.47 minutes.

(c) The confidence interval for mean endurance of female mice swimming is almost entirely above the confidence interval for mean endurance of male mice swimming. However, it is more appropriate to look at the confidence interval for the difference between the means. We are told that a 95% confidence interval for the difference in mean endurances of male and female mice swimming is (0.5, 8.9). Since zero is not contained in this interval, it appears that the mean endurance of female mice is greater than the mean endurance of male mice by between 0.5 and 8.9 minutes.

IV.9 (a) For each additional kilogram of nitrogen per hectare of land area per year, plant species richness decreases by 0.408, on average. (b) About 55% of the variation in plant species richness is explained by the linear relationship with the amount of nitrogen deposited. (c) The null hypothesis is $H_0: \beta = 0$ and the alternative hypothesis is $H_a: \beta \neq 0$. This very low *P*-value says that we should reject the null hypothesis and conclude that there is statistically significant evidence of a linear relationship between plant species richness and amount of nitrogen deposited.

IV.10 (a) Step 1: **Hypotheses** We want to draw conclusions about p_{20}, the proportion of married taxpayers who were offered a 20% match and opened an IRA, and p_{50}, the proportion of married taxpayers who were offered a 50% match and opened an IRA. We hope to show that a higher proportion of those offered a 50% match opened IRAs, so we have a one-sided alternative:

$$\begin{aligned} H_0&: p_{20} = p_{50} \\ H_a&: p_{20} < p_{50} \end{aligned} \quad \text{or, equivalently,} \quad \begin{aligned} H_0&: p_{20} - p_{50} = 0 \\ H_a&: p_{20} - p_{50} < 0 \end{aligned}$$

Step 2: **Conditions** Since we are testing a claim about two population proportions, we should use a two-proportion *z* test if the conditions are satisfied.

- **SRS** Our subjects are two groups in a randomized comparative experiment. The random assignment should allow researchers to attribute any significant difference in the proportion of married taxpayers who open an IRA to the amount matched. We are viewing each group as a representative sample from the hypothetical population of married taxpayers who could be offered these matching opportunities. Since the subjects were not randomly selected from a larger population, however, this may limit the researchers' ability to generalize the findings of the study.

- **Normality** The combined proportion of married taxpayers who opened an IRA is

$$\hat{p}_C = \frac{240+456}{1780+1831} = \frac{696}{3611} = 0.1927$$

Using this value, we find that

$$n_1\hat{p}_C = (1780)(0.1927) = 343.0 \qquad n_2\hat{p}_C = (1831)(0.1927) = 352.8$$
$$n_1(1-\hat{p}_C) = (1780)(0.8073) = 1437.0 \quad n_2(1-\hat{p}_C) = (1831)(0.8073) = 1478.2$$

which are all 5 or larger.

- **Independence** Due to the random assignment, these two groups of married taxpayers can be viewed as independent samples.

Step 3: **Calculations**

- **Test statistic** The z test statistic is

$$\begin{aligned} z &= \frac{\hat{p}_{20} - \hat{p}_{50}}{\sqrt{\hat{p}_C(1-\hat{p}_C)\left(\frac{1}{n_{20}} + \frac{1}{n_{50}}\right)}} \\ &= \frac{0.1348 - 0.2490}{\sqrt{0.1927(0.8073)\left(\frac{1}{1780} + \frac{1}{1831}\right)}} \\ &= \frac{-0.1142}{0.01313} \\ &= -8.70 \end{aligned}$$

- ***P*-value** The one-sided P-value is the area under the standard Normal curve to the left of –8.70. This area is close to zero.

Step 4: **Interpretation** Since the P-value is near zero, the results are significant at any reasonable significance level. There is strong evidence that a higher proportion of married taxpayers open an IRA when offered a 50% match than when offered a 20% match.

(b) Step 1: **Parameters** We want to draw conclusions about μ_{20}, the mean contribution of married taxpayers who are offered a 20% match and open an IRA, and μ_{50}, the mean contribution of married taxpayers who are offered no match and open an IRA. We want to construct a 95% confidence interval for the mean added contribution (among taxpayers who decide to contribute) due to offering a 20% match over no match.

Step 2: **Conditions**

- **SRS** Our subjects are two groups in a randomized comparative experiment. The random assignment should allow researchers to attribute any significant difference in the proportion of married taxpayers who open an IRA to the amount matched. We are viewing each group as a representative sample from the hypothetical population of married taxpayers who could be offered these matching opportunities. Since the subjects were not randomly selected from a larger population, however, this may limit the researchers' ability to generalize the findings of the study.

- **Normality** The sample sizes are large enough (n = 49 and n = 240) to use 2 sample t procedures even if the population distributions aren't close to Normal.
- **Independent Samples** Due to the random assignment, these 2 groups of married taxpayers can be viewed as independent samples.

Step 3: **Calculations** A 95% confidence interval for the difference in mean contribution (among taxpayers who decided to contribute) due to offering a 20% match over no match is

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \left(1723 - 1549\right) \pm 2.021\sqrt{\frac{1332^2}{240} + \frac{1652^2}{49}}$$
$$= 174 \pm 507.6 = \left(-333.6, 681.6\right)$$

using df = 40.

Step 4: **Interpretation** We are 95% confident that the mean contribution for those who are offered a 20% match and who open an IRA is between \$333.60 lower than and \$681.60 higher than the mean contribution for those who are offered no match and who open an IRA.

(c) The results in part (a) are more useful to answer H&R Block's original question. The significance test in part (a) led us to conclude that a higher proportion of married taxpayers who are offered a 50% match will choose to open an IRA than married taxpayers who are offered a 20% match.

The confidence interval in part (b) provides an estimate of the difference in the mean amount contributed to IRA accounts among married taxpayers who receive no matching contributions and those who receive a 20% match. In other words, part (b) only examines how much people will contribute to an IRA, not what proportion will choose to open an IRA.

It appears that more families would contribute to an IRA if the money they invest was matched by their employer or another group.

(d) We cannot generalize the results of this study to all married U.S. taxpayers because there may be a systematic difference between married taxpayers in the St. Louis area who use H&R Block and other married tax payers. The individuals in this study were not a random sample from a larger population of interest.