Answers

1. (a) The expected count for the "independent–trust media yes" cell is

$$\frac{(41)(67)}{170} = 16.159.$$

   (b) The chi-square statistic is
$\chi^2 = 3.594 + 5.525 + 2.042 + 3.140 + 0.402 + 0.618 = 15.321.$

   (c) Since all the expected counts are at least 5, it is appropriate to use a chi-square distribution.

   (d) The one with degrees of freedom df $= 2$.

   (e) He would place into one bag 170 slips of paper containing party affiliation information—66 slips with the word "Democrat", 63 with "Republican", and 41 with "Independent". In a different bag he would place 170 slips with 103 saying "does not trust" and 67 saying "trusts". After shaking up the contents of the bag, he would make simultaneous draws (without replacement) from each bag, using the information from the two slips to add to the observed count of one of our six cells. When the two-way table was complete, he would calculate the χ^2-statistic for that table.

2. (a) Among females we have

$$n\hat{p}_f = (1923)\frac{124}{1923} = 124 \qquad \text{and} \qquad n(1-\hat{p}_f) = 1799.$$

   Among males we have

$$n\hat{p}_m = (1236)\frac{61}{1236} = 61 \qquad \text{and} \qquad n(1-\hat{p}_m) = 1175.$$

   All of these quantities are at least 10.

   (b) Our hypotheses: $\mathbf{H}_0\colon\ p_f - p_m = 0$, $\mathbf{H}_a\colon\ p_f - p_m \neq 0$. For this test of significance we compute the pooled proportion: $\hat{p} = \dfrac{124+61}{1923+1236} = 0.0586$. We then have z-score

$$z = \frac{p_f - p_m}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_f} + \dfrac{1}{n_m}\right)}} = \frac{0.0645 - 0.0494}{\sqrt{0.0586(1-0.586)\left(\dfrac{1}{1923} + \dfrac{1}{1236}\right)}} \doteq 1.763.$$

   (c) It is

$$(\hat{p}_f - \hat{p}_m) \pm z^* \sqrt{\frac{p_f(1-p_f)}{n_f} + \frac{p_m(1-p_m)}{n_m}} = (0.0645 - 0.0494) \pm 2.576\sqrt{\frac{0.0645(1-0.0645)}{1923} + \frac{0.0494(1-0.0494)}{1236}},$$

   or $[0.0105, 0.0197]$.

3. (a) The mean is 34.70. The standard deviation is 21.70.

   (b) Using the theoretical t distribution with df $= 49$, we obtain the critical value $t^* = 2.010$. This gives us 95% CI

$$34.7 \pm (2.01)\frac{21.7}{\sqrt{50}}, \qquad \text{or} \qquad [28.53, 40.87].$$

   (c) It may be convenient to use consecutive shoppers all at one register, but the possible sources of bias include

   - the fact that shoppers who are all in the store at a similar time of day may fundamentally be different than other types of shoppers.
   - the possibility that many of the cases/shoppers chose this particular register for a *reason* and, if so, perhaps the source of this reason also sets them apart from other shoppers. For instance, perhaps the register was in a "quick shopping lane", and perhaps receipts are for smaller amounts when shoppers are in a hurry.

4. (a) On a t-distribution with 7 degrees of freedom, the right-tail area beyond $t = 2.1$ is approximately 0.037. Since the alternative hypothesis is 2-tailed, we double this to get P-value 0.074. We fail to reject the null hypothesis, as this P-value is greater than 0.05.

   (b) Any evidence for the one-sided alternative hypothesis $\mathbf{H}_a\colon\ \mu > 15$ would produce a positive standardized t-score, not a negative one. Even without checking, we can say the area to the right of $t = -2.94$ is larger than 0.5, which means we fail to reject the null hypothesis.

   (c) The area to the left of $z = -1.35$ on a standard normal distribution is approximately 0.089. At the 10% level, we reject the null hypothesis in favor of the alternative.

5. (a) The mean is $\overline{x} = \dfrac{448(70.42) + 91(71.21) + 51(80.51)}{448 + 91 + 51} \doteq 71.41.$

   (b) We have

$$\begin{aligned} \text{SSG} &= n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + n_3(\overline{x}_3 - \overline{x})^2 \\ &= 448(70.42 - 71.41)^2 + 91(71.21 - 71.41)^2 + 51(80.51 - 71.41)^2 \doteq 4666. \end{aligned}$$

   (c)

Source	Degrees of Freedom	Sum of Squares	Mean Square	F
Groups	2	4666	2333	7.14
Error	587	191745.4	326.65	
Total	589	196411.4		

   (d) The sample sizes in each group are all at least 30 (the smallest is $n_3 = 51$), which indicates the rule of thumb for having a normal sampling distribution for group means is met. The ratio of largest sample standard deviation to smallest is $18.83/14.58$, which is well shy of 2. So, it seems reasonable to use an F-distribution when assessing P-value.

   (e) The null hypothesis—that there is no association between worker classification (explanatory) and SCI (response), is manifested as "all group means are equal", or

$$\mathbf{H}_0\colon\ \mu_U = \mu_S = \mu_M.$$

   The alternative is that at least one group mean is different than the others. Consulting the theoretical F-distribution with dfs for numerator and denominator 2 and 587, respectively, the amount of right-tailed area beyond $F = 7.14$ is approximately 0.00086, which is a significant result at the $\alpha = 0.01$ level. We reject the null hypothesis in favor of the alternative, that there is a difference in population means for at least one pair of two groups.

6. Our hypotheses, written explicitly, are

$$\mathbf{H}_0\colon\ p_V = 0.07,\ p_O = 0.15,\ p_S = 0.35,\ p_N = 0.43,$$

vs. the alternative that at least one of these proportions is different than what it is proposed to be. The total number of cases is $n = 12 + 36 + 83 + 79 = 210$. Our observed and expected counts are as specified in the table

Classification	Very often	Often	Sometimes	Never
observed frequencies	12	36	83	79
expected counts ($= 210p_i$)	14.7	31.5	73.5	90.3

We will do a chi-square goodness-of-fit test. The test statistic is

$$\chi^2 = \frac{(12 - 14.7)^2}{14.7} + \frac{(36 - 31.5)^2}{31.5} + \frac{(83 - 73.5)^2}{73.5} + \frac{(79 - 90.3)^2}{90.3} \doteq 3.7807.$$

Using the χ^2 theoretical distribution calculator with df $= 3$, the right-tailed area beyond the value 3.7807 is approximately 0.286. We do not have a specified value for α, but at all the common levels of α we would fail to reject the null hypothesis. This data is consistent with the belief that students at the institution in question fit the 2005 national profile.