12.3.

H_0: The vehicle mileage is normally distributed
H_A: The vehicle mileage is not normally distributed

Vehicle mileage is a continuous variable. Thus, to determine whether the distribution is normally distributed, students will need to break the data into classes. There is no one correct set of intervals. In chapter 2 the rule of thumb for developing a grouped data frequency distribution is to use between 5 and 20 groups. One possible approach is to break the data into six classes. Each class would be one standard deviation in width – three classes below the mean and three classes above the mean.

To find the classes and observed frequencies first determine the sample mean and sample standard deviation:

$$\bar{x} = \frac{\sum x}{100} = \frac{1{,}384}{100} = 13.84$$

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = 4.34$$

Classes	Observed Frequency
< 5.16	1
5.16 < 9.50	14
9.50 < 13.84	36
13.84 < 18.18	32
18.18 < 22.52	14
> 22.52	3
	100

The next step is to convert the upper limit of the first class to a z value as follows

$$= \frac{x - \bar{x}}{s} = \frac{5.16 - 13.84}{4.34} = -2.00$$

From the standard normal table,

$$P(z < -2.00) = .5000 - .4772 = .0228$$

The expected frequency for this class then is .0228 x 100 = 2.28

The upper limit of the second class corresponds to $z = -1.00$. From the standard normal table:

$$P(-2.00 < z < -1.00) = (.5000 - .3413) - .0228 = .1359$$

The expected frequency of the second class is .1359 x 100 = 13.59

In a like manner, find the expected frequencies for each class shown as follows:

Classes	Observed Frequency	Expected Frequency
< 5.16	1	2.28
5.16 < 9.50	14	13.59
9.50 < 13.84	36	34.13
13.84 < 18.18	32	34.13
18.18 < 22.52	14	13.59
> 22.52	3	2.28
	100	100

Recalling that the expected cell frequencies need to be 5 or greater, the first and last classes need to be collapsed into their adjacent classes giving

The chi-square goodness of fit test is

Class	Observed Frequency	Normal Distributio n Probability	Expected Frequenc y	(O-E)^2/E
< 9.5	16	0.1587	15.87	0.0477
9.5 < 13.84	35	0.3413	34.13	0.1025
13.84 < 18.18	32	0.3413	34.13	0.1329
> 18.18	17	0.1587	15.87	0.0805

The calculated chi-square statistic is 0.3636. Since the calculated chi-square test statistic is less than the critical chi-square test statistic for an alpha value of 0.01 and 1 degrees of freedom (= k –1 - # parameters estimated = 4 – 1 – 2 = 1), 6.6349, we do not reject the null hypothesis and we conclude that the vehicle mileage is normally distributed. Notice, the class boundaries were constructed using $\bar{x} \pm cs$; c = 1,2, and 3. The lower two classes were combined as were the upper two classes to obtain expected values greater than 5.

12.7.

a. H_0: The ATM usage distribution is Poisson distributed with a mean equal to 5 per hour.
H_A: The ATM usage distribution is not Poisson distributed with a mean equal to 5 per hour.

b. The first step is to construct the frequency distribution from the data along with the expected frequencies based on the null hypothesis that the distribution is Poisson with mean = 5.

x	Observed Frequency	Poisson Probability	Expected Frequency
0	0	0.00673795	0.3369
1	0	0.03368973	1.6845
2	0	0.08422434	4.2112
3	0	0.1403739	7.0187
4	1	0.17546737	8.7734
5	3	0.17546737	8.7734
6	2	0.14622281	7.3111
7	12	0.10444486	5.2222
8	17	0.06527804	3.2639
9	8	0.03626558	1.8133
10	1	0.01813279	0.9066
11	4	0.00824218	0.4121
12	2	0.00343424	0.1717
13 and Over	0	0.00201885	0.1009
	50	1	50

Cells need to be combined since all expected cell frequencies should be 5 or more. This is shown as follows:

ATM Usage	Observed	Poisson Probability	Expected	$(O-E)^2/E$
2 or less	0	0.1247	6.2326	6.2326
3	0	0.1404	7.0187	7.0187
4	1	0.1755	8.7734	6.8874
5	3	0.1755	8.7734	3.7992
6	2	0.1462	7.3111	3.8583
7	12	0.1044	5.2222	8.7966
8 or more	32	0.1334	6.6686	96.2244

The calculated chi-square test statistic is 6.2326 + 7.0187 + 6.8874 +…+ 96.2244 = 132.8171

The critical value of the chi-square test statistic with 7-1 = 6 degrees of freedom and alpha = 0.10 is 10.6446. Since the calculated value of 132.8171 is greater than the critical value of 10.6446, we reject the null hypothesis and conclude that the ATM usage rate is not Poisson distributed with a mean of 5.00 per hour.

12.8.

a. H_0: The pipe diameter is normally distributed with a mean of 2.00 and a standard deviation of 0.10.

H_A: The pipe diameter is not normally distributed with a mean of 2.00 and a standard deviation of 0.10.

One way to group the data into classes to test this hypothesis is to form six classes of one standard deviation in width. The observed frequency distribution is:

Classes	Frequency	Normal Probability	Expected Frequencies
< 1.70	0	0.0013	0.039
1.70 < 1.80	0	0.0215	0.645
1.80 < 1.90	3	0.1359	4.077
1.90 < 2.00	14	0.3413	10.239
2.00 < 2.10	7	0.3413	10.239
2.10 < 2.20	6	0.1359	4.077
2.20 < 2.30	0	0.0215	0.645
2.30 >	0	0.0013	0.039

Collapse cells so that the expected frequencies are 5 or more. Not, we would need to collapse the data into only two classes to achieve this. If we use four classes, two of the expected frequencies are very close to five.

The chi-square goodness of fit test is:

Class	Observed Frequency	Normal Distribution Probability	Expected Frequency	(O-E)^2/E
Under 1.9	3	0.15865526	4.7596577 9	0.65055
1.9 to under 2.0	14	0.34134474	10.240342 2	1.380327 6
2.0 to under 2.1	7	0.34134474	10.240342 2	1.025338 6
2.1 and over	6	0.15865526	4.7596577 9	0.323226 8

The calculated chi-square statistic is 0.65055 + 1.3803276 + 1.0253386 + 0.3232268 = 3.379443. Since the calculated chi-square test statistic is less than the critical chi-square test statistic for an alpha value of 0.01 and 3 degrees of freedom, 11.3449, we do not reject the null hypothesis and we conclude that the pipe diameter may be normally distributed with a mean of 2.00 inches and a standard deviation of 0.10. Notice, we have two expected values less than 5, but since we do not reject the null hypothesis we don't have to adjust the cell sizes.

b. Based on the test, we have no reason to conclude that the company is not meeting its product specification.

12.13.

H_0: Response to question 1 is independent of the response to question 2.
H_A: Response to question 1 is not independent of the response to question 2.
$\alpha = .05$

The expected frequencies are calculated by multiplying the row total by the column total and then dividing by the grand total. For instance for the YES, YES cell, we get (17 x 17)/30 = 7.37

Observed Frequencies			
Question 1	Question 2		
	Yes	No	Total
Yes	6	11	17
No	7	6	13
Total	13	17	30
Expected Frequencies			
Question 1	Question 2		
	Yes	No	Total
Yes	7.37	9.63	17
No	5.63	7.37	13
Total	13	17	30

The calculated chi-square is:

$$\chi^2 = \sum\sum\frac{(o-e)^2}{e} = \frac{(6-7.37)^2}{7.37} + \frac{(11-9.63)^2}{9.63} + \frac{(7-5.63)^2}{5.63} + \frac{(6-7.37)^2}{7.37} = 1.037$$

The critical value of the chi-square test statistic for $\alpha = 0.05$ and (2-1)(2-1) =1 d.f. is 3.8415. Since the calculated value of 1.0377 < 3.8415, we do not reject the null hypothesis and conclude that the response to question 1 is independent of the response to question 2.

12.14.

H_0: Response to question 1 is independent of the response to question 2.
H_A: Response to question 1 is not independent of the response to question 2.

Alpha = 0.05

Observed Frequencies			
Care?	Smoke?		
	Yes	No	Total
Yes	11	139	150
No	29	21	50
Total	40	160	200
Expected Frequencies			
Care?	Smoke?		
	Yes	No	Total
Yes	30	120	150
No	10	40	50
Total	40	160	200

$$\chi^2 = \sum\sum\frac{(o-e)^2}{e} = \frac{(11-30)^2}{30} + \frac{(139-120)^2}{120} + \frac{(29-10)^2}{10} + \frac{(21-40)^2}{40} = 60.167$$

The calculated chi-square test statistic is 12.033 + 3.008 + 36.100 + 9.025 = 60.167. The critical value of the chi-square test statistic for alpha = 0.05 and 1 d.f. is 3.8415. Since the calculated value of 60.167 > 3.8415, we do reject the null hypothesis and conclude that the response to question 1 is not independent of the response to question 2.

The p-value for this test is smaller than 0.01. Small p-values provide strong evidence to reject the null hypothesis.

12.20.

a. H_0: The number of transactions is independent of the marital status of the customer.
H_A: The number of transactions is not independent of marital status of the customer.
$\alpha = 0.05$

Decision rule: If the calculated value of the chi-square test statistic is greater than the critical value of the chi-square test statistic having d.f. = (r-1)(c-1) = (4-1)(5-1) = 12 and alpha = 0.05 = 21.0261, reject the null hypothesis. Otherwise, do not reject the null hypothesis.

The observed and expected cell frequencies are

Observed Frequencies	**Number of Transactions**					
Marital Status	**0-10**	**11-20**	**21-30**	**31-40**	**Over 40**	**Total**
Single	13	23	19	20	11	**86**
Married	6	15	33	45	27	**126**
Divorced	4	19	22	20	15	**80**
Other	2	11	8	5	2	**28**
Total	**25**	**68**	**82**	**90**	**55**	**320**
Expected Frequencies	**Number of Transactions**					
Marital Status	**0-10**	**11-20**	**21-30**	**31-40**	**Over 40**	**Total**
Single	6.71875	18.275	22.0375	24.1875	14.78125	**86**
Married	9.84375	26.775	32.2875	35.4375	21.65625	**126**
Divorced	6.25	17	20.5	22.5	13.75	**80**
Other	2.1875	5.95	7.175	7.875	4.8125	**28**
Total	**25**	**68**	**82**	**90**	**55**	**320**

The calculated value of the chi-square test statistic is 28.43554. The critical value of the chi-square test statistic is 21.0261. Since the calculated value is greater than the critical value, we reject the null hypothesis and conclude that marital status is not independent of the number of transactions.

b. Yes, there are two cells with an expected value less than 5.00. We could combine the categories *Divorced* and *Other* to get a larger expected cell frequency. The new contingency table becomes

Observed Frequencies	**Number of Transactions**					
Marital Status	**0-10**	**11-20**	**21-30**	**31-40**	**Over 40**	**Total**
Single	13	23	19	20	11	**86**
Married	6	15	33	45	27	**126**
Divorced/Other	6	30	30	25	17	**108**
Total	**25**	**68**	**82**	**90**	**55**	**320**
Expected Frequencies	**Number of Transactions**					
Marital Status	**0-10**	**11-20**	**21-30**	**31-40**	**Over 40**	**Total**
Single	6.71875	18.275	22.0375	24.1875	14.78125	**86**
Married	9.84375	26.775	32.2875	35.4375	21.65625	**126**
Divorced/Other	8.4375	22.95	27.675	30.375	18.5625	**108**
Total	**25**	**68**	**82**	**90**	**55**	**320**

The calculated chi-square statistic is now 23.94657. Because the calculated value of the test statistic is greater than the critical value of 15.5073 (alpha = 0.05 and 8 d.f.), we reject the null hypothesis and conclude that marital status and number of transactions are not independent.

Note, students may also choose to collapse the number of transactions variable which will produce a different contingency table and the calculated chi-square value.

c. Because we rejected the null hypothesis when one expected cell frequency was smaller than 5.00, we should combine categories and rerun the test. In this example we combined *Divorced* and *Other* into one category. The small expected cell frequencies can inflate the calculated chi-square value and lead to an increased chance of making a Type I error. If the null hypothesis is not rejected, you do not need to worry when the expected cell frequencies drop below 5.00.