Solutions:

1.

The expected frequency count of male Independents is

$$\frac{\text{Row total of male} \times \text{Column total of independent}}{\text{Table total}} = \frac{400 \times 100}{1000} = 40.$$

2.

Here we test H0: The dogs have no preferences among the brands vs. H1: They have preferences.

Under H0, the expected frequency for each brand should be $150/3 = 50$.

The observed frequencies for brands A, B and C respectively are 62, 43 and 45.

Hence, $\text{Chi}^2 = \frac{(62-50)^2}{50} + \frac{(43-50)^2}{50} + \frac{(45-50)^2}{50} = 4.36$ with degrees of freedom 2.

The critical value of chi-square distribution with df=2 at 10% level is 4.605.

As $4.36 < 4.605$, we do not reject H0. So, there is not sufficient evidence at the 10% level that the dogs have preferences among the brands.

3.

From the table of margin of error for sample sizes, we see to find the answer to within 3% at the 95% confidence level, one should take a sample of size 1068.

4.

Claim I is false, because rejection at 10% level guarantees rejection at any higher levels, but does not guarantee rejection at any lower level.

Claim II is true, because p-value is the probability of the test statistic being more extreme with respect to the null hypothesis, and hence if p-value is more than the significance level, intuitively it is clear that the test statistic is not enough far from its value under the null hypothesis.

Claim III is false. The value of the test statistic does not depend on the level of the test.

So, option (B) is answer.

5.

Option (E) is the answer. If p-value of the test is less than the level, we reject H0. All other given conditions are not true.

6.

The event can occur, but only once out of 1000 trials. So, it is unlikely to occur. Option (A) is the answer.

7.

Normality, constant (error) variance and error terms with a mean of zero (from normal equation) are assumptions in the simple linear regression model, whereas variance of 1 is not an assumption.

8.

For a given hypothesis test, if we do not reject H0, and H0 is true, then no error has been committed. Type I error is rejecting a true hypothesis, type II error is accepting a false hypothesis and type III error is correctly rejecting the null hypothesis for the wrong reason.

9.

As sample size increases, the standard error (s.e.) of the estimate decreases. Confidence interval (CI) limits are directly proportional to s.e. So, with increase in sample size, CI decreases. Also, as confidence level decreases, the critical value of the test decreases. Confidence interval (CI) limits are directly proportional to the critical value. So, with decrease in sample size, CI decreases. Answer is (c).

10.

For siple linear regression, r^2 denotes the proportion of total variance explained by the regression, whereas the rest of the variation is considered due to error. As among the two models, r^2 for Model I is larger, the SSE for Model I is lower. Also, as model I explains more variability than model II, a Prediction based on Model I is likely better than a prediction based on Model II.

11.

As the data distributions is assumed to be fairly symmetric, we assume the distribution is normal. Now the interval (6,24) is sample mean +- 1 standard deviation. From the property of normal distribution, thus 68% of data are expected to fall between 6 ans 24.

12.

We reject H0 (row and column are independent) when observed value of the chi square test statistic is more than the critical value of the test. As the difference between the observed and expected frequencies decrease, the value of the test statistic decreases, and hence, the probability of concluding that the row variable is independent of the column variable increases.

13.

The normal probability plot in a multiple regression analysis plots the quantiles of the residuals against the quantiles of the corresponding theoretical normal distribution. So, if normality assumption is true, both the quantiles should be same and hence, the plot should be a straight line.

14.

True. Correlation measures the strength and direction of the linear relationship between two quantitative variables, but not non-linear relationship.

15.

A matched-pair experiment is one where subjects are randomly assigned to one of two treatments. It is a special case of a randomized block design. It can be used when the experiment has only two treatment conditions; and subjects can be grouped into pairs, based on some blocking variable. Then, within each pair, subjects are randomly assigned to different treatments.

16.

If the constant variance assumption holds, residuals should not change with x. Hence, the plot should form a horizontal band pattern.

17.

We are testing H0:$p = .23$ vs. H1:$p < .23$.

Sample size 84 is large. So, we use z-test.

Standard error $= \sqrt{.23(1 - .23)/84} = .046$.

Sample estimate $= 15/84 = .179$.

So, test statistic = (.179-.23)/.046 = -1.107.

Hence, p-value $= P(Z < -1.107) = .134$, which is more than significance level .05. So, we do not reject H0 and conclude that there is no strong evidence to claim that the new medication reduces the mortality rate

18.

Sample estimate $= 73/89 = 0.82$.

Standard error = $\sqrt{.82(1-.82)/89} = 0.04$.

Critical value = $z_{.02/2} = 2.326$.

Hence the 98% CI is $(0.820 \pm 0.04\ \ 2.326) = (0.820 \pm 0.095)$.

19.

As the p-value 0.067 is less than the significance level 0.10, H0 would be rejected.

20.

Option (A) in answer. Assumptions of independence and normality are necessary. But equality of sample variances does not imply the test statistic will be 1.

21.

No, because not every sample of the intended size has an equal chance of being selected. For simple random sample, there are both randomness and independence. For systematic sample, even if randomness is utilized, we no longer have independence.

22.

The standard deviation of the sample mean x is $\frac{5}{\sqrt{25}} = 1$.

23.

As the sample size 10 is small, we will use critical values from t distribution with df=10-1=9.

Critical value = $t_{.02/2;9} = 2.821$.

S0, 98% CI is $(\$47.52 \pm \$2.821 \times 1.59/\sqrt{10})$, i.e., $(\$47.52 \pm \$1.42)$ dollar.

24.

False. For simple linear regression, the coefficient of determination is the square of the correlation coefficient between the response and the explanatory variables, and hence always a positive quantity. So, it does not indicate whether the relationship is positive or negative.

25.

We use chi sqaure test of independence to test H0: Catching a cold and taking vitamin C are independent vs. H1: They are not independent.

Expected frequenciesr are E1=100*280/450=62.22, E2=100*110/450=24.44, E3=100*60/450=13.33, E4=350*280/450=217.78, E5=350*110/450=85.56, E6=350*60/450=46.67.

The test statistic is

$$\chi^2 = \frac{(57-62.22)^2}{62.22} + \frac{(26-24.44)^2}{24.44} + \frac{(17-13.33)^2}{13.33} + \frac{(223-217.78)^2}{217.78} + \frac{(84-85.56)^2}{85.56} + \frac{(43-46.67)^2}{46.67} = 1.99.$$

p-value $= P(\chi^2 > \chi^2_{(3-1)(2-1)} = 0.370$.

As the p-value is more than the significance level .1, we reject H0 at 10% level, and hence conclude that there is not sufficient evidence at the 10% level of a relationship between taking vitamin C and catching fewer colds, which is option (E). (B) is not true because in testing we do not say 'prove', we say 'there is sufficient evidence'.

26.

Option (D). It is correct to say that there is a 95% chance or we are 95% confident that the confidence interval we calculated contains the true difference in mean cholesterol level lowering. It is not quite correct to say that there is a 95% chance that the true difference in mean cholesterol level lowering lies within the interval. The interval is random, not the parameter (difference in mean cholesterol level lowering).

27.

A treatment is a condition or intervention which is applied to individuals.

28.

We know, $z = \frac{x-\mu}{\sigma}$.

Here, $1.20 = \frac{x-100}{15}$, which implies $x = 118$.

29.

A phenomenon or process that produces results that cannot be predicted with certainty is random.

30.

A measurement process is said to be biased if it consistently understates or overstates the true value.

31.

Option (E). All these are indicative of some correlation among the error terms. Hence, error terms are not independent.

32.

I will use a chi-square goodness of fit test, which tests how the distribution of the data fits a given hypothesized distribution. Here the hypothesized distribution is specified by the company claim. The data distribution is the distribution of cards in the bought package.

33.

For a chi-square test for independence, the null hypothesis states that the two relevant classifications are statistically independent. (D)

34.

In any normal distribution, the proportion of observations that are within 2 standard deviations of the mean is closest to 95% or 0.95.

35.

The total number of outcomes of 5 tosses of a die is 6^5.

We can choose 2 tosses (for 'Four's) out of 5 tosses in $\binom{(}{5}, 2) = 10$ ways. For each of these 10 ways, the other three tosses can result in 5^3 ways.

So, in 5 tosses the probability of getting "Four" two times is $\frac{10\times 5^3}{6^5} = 0.161$.

36.

(B) Sometimes. For rejection of null, we need to have test statistics value more than the critical value. With the increase of α value from 0.01 to 0.05, the critical value decreases. So the comparison and hence conclusion will depend on whether the later critical value is more or less than the test statistic.

37.

For CRD one-way ANOVA, within treatment variance is the error variance (SSE).

$F = \frac{SSB/df_B}{MSE/df_E}$. (SS for sum of squares, B for between)

So, as SSB increase or decreases with respect to SSE, the value of respectively increases or decreases.

Option (E) is answer.

38.

(D) None of the above. Unbiased means the expected value of the sample proportion (which is a random variable) will be equal to the population proportion (which is a parameter).

39.

In a hypothesis testing, the P-value is a probability. The P-value is defined as the probability, under the assumption of hypothesis H, of obtaining a result equal to or more extreme than what was actually observed.

40.

When comparing the variances of two normally distributed populations using independent random samples, the correct test statistic to use is F statistic. F test is dependent on the assumption of normality.

41.

In this case, the standard deviation of the sample is $\sqrt{\frac{p(1-p)}{n}}$, where p is the population proportion. So as sample size n increases, the standard deviation of the sample decreases. (A) holds. (B) and (C) do not hold.

42.

When the population is normally distributed, population standard deviation is unknown and the sample size is n = 15, the confidence interval for the population mean is based on the t distribution.

43.

Bob makes his first free throw on his fifth shot means Bob misses to make free throw in first 4 shots and makes a free throw in the 5th shot. The probability is $(1-.7)^4 \times .7 = 0.0057$.

44.

The slope of the regression line is -0.002.

45.

As the p-value 0.0012 is less than the significance level 0.05, H0 should be rejected.

46.

Here we test H0: p=0.08 vs. H1: p>0.08, where p is the population proportion of defective cars.

Sample proportion $\widehat{p} = 33/300 = 0.11$.

So test statistic $= \frac{\widehat{p}-p}{\sqrt{p(1-p)/n}} = \frac{.11-.08}{\sqrt{.08\times.92/300}} = 1.915$.

Hence, p-value is $P(Z > 11.915) = 0.028$, which is less than the significance level 0.05. So, she should reject the 8% claim. Option (B) is answer.

47.

Five percent of pregnancy lengths are larger than the 95th percentile of the normal distribution with mean 260 and sd 10, which is approximately 276 days.

48.

D is false. Null hypothesis says that row and column variable are independent.

49.

Suppose the mean setting should be m.

So, the ounces delivered, say X, follows a normal distribution with mean m and sd 0.3.

Considering to the given condition,

$P(X > 12) = .01,$

$P(Z = \frac{X-m}{.3} > \frac{12-m}{.3}) = .01$, where Z is a standard normal variate.

Now, the point above which lies 1% of the standard normal values is 2.326.

Thus, $\frac{12-m}{.3} = 2.326$, which implies $m = 11.30$.

50.

As the sample size 15 is small, we will use t distribution with df=15-1=14.

t score $= \frac{290-300}{50/\sqrt{15}} = -0.775$.

Thus, the required probability is $P(t_{14} < -0.775) = .226$.