Two of the comparisons result in a p-value that is less than 0.0083 — the comparison of group 1 with group 4, and the comparison of group 2 with group 4. Therefore, we conclude that the mean systolic blood pressures of both nonsmokers and current smokers are lower than the mean systolic blood pressure of tobacco chewers.

Exercise 8 [18]

a. To test the null hypothesis that the mean LDL cholesterol levels are identical for each of the four populations, we must first calculate estimates of the within-groups and between-groups variances. Note that

[9 pts]

$$\begin{aligned} s_W^2 &= \frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2+(n_3-1){s_3}^2+(n_4-1){s_4}^2}{n_1+n_2+n_3+n_4-4} \\ &= \frac{(72)(1.62)^2+(104)(1.43)^2+(239)(1.24)^2+(1079)(1.31)^2}{73+105+240+1080-4} \\ &= 1.75. \end{aligned}$$

Since

$$\begin{aligned} \bar{x} &= \frac{n_1\bar{x}_1+n_2\bar{x}_2+n_3\bar{x}_3+n_4\bar{x}_4}{n_1+n_2+n_3+n_4} \\ &= \frac{73(6.22)+105(5.81)+240(5.77)+1080(5.47)}{73+105+240+1080} \\ &= 5.58, \end{aligned}$$

we have that

$$\begin{aligned} s_B^2 &= \frac{n_1(\bar{x}_1-\bar{x})^2+n_2(\bar{x}_2-\bar{x})^2+n_3(\bar{x}_3-\bar{x})^2+n_4(\bar{x}_4-\bar{x})^2}{4-1} \\ &= \frac{73(0.64)^2+105(0.23)^2+240(0.19)^2+1080(-0.11)^2}{4-1} \\ &= 19.06. \end{aligned}$$

Therefore, the test statistic is

$$\begin{aligned} \mathrm{F} &= \frac{s_B^2}{s_W^2} \\ &= \frac{19.06}{1.75} \\ &= 10.89. \end{aligned}$$

For an F distribution with $4-1=3$ and $1498-4=1494$ df, $p < 0.001$. Therefore, we reject the null hypothesis.

[2] b. We conclude that mean LDL cholesterol level is not the same for each of the four groups.

[2] c. In order to use the one-way analysis of variance technique, the four populations must be at least approximately normally distributed, and their variances must all be the same.

[5] d. It is necessary to take an additional step in this analysis. We have concluded that the mean LDL levels are not the same for all four groups, but we cannot yet say which group means differ from which others. In order to do this, we need to use the Bonferroni method of multiple comparisons. (It would tell us that patients with intermittent claudication and those with minor asymptomatic disease have higher mean LDL cholesterol levels than patients with no disease.)

Exercise 6 [18]

[3] a) For samples of size 25, the distribution of sample proportions has mean 0.328, has standard deviation $\sqrt{0.328(1-0.328)/25} = 0.094$, and is approximately normal given that n is sufficiently large.

[5] b) Since $np = 25(0.328) = 8.2$ and $n(1-p) = 25(1-0.328) = 16.8$ are both greater than 5, we can apply the normal approximation. We are interested in the area under the standard normal curve that lies to the right of $p = 0.45$. Solving for z,

$$\begin{aligned} z &= \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \\ &= \frac{0.45-0.328}{0.094} \\ &= 1.30. \end{aligned}$$

The area that lies to the right of $z = 1.30$ is 0.097. Therefore, approximately 9.7% of the samples will have a sample proportion of 0.45 or larger.

[5] c) In this case, we are interested in the area under the curve that lies to the left of $p = 0.20$. Solving for z,

$$\begin{aligned} z &= \frac{0.20-0.328}{0.094} \\ &= -1.36. \end{aligned}$$

The area that lies to the left of $z = -1.36$ is 0.087. Therefore, about 8.7% of the samples have a sample proportion of 0.20 or smaller.

[5] d) For the standard normal distribution, the value $z = -1.28$ cuts off the lower 10% of the curve. Writing

$$\begin{aligned} z &= -1.28 \\ &= \frac{\hat{p}-0.328}{0.094} \end{aligned}$$

and solving for $\hat{p}$, we find that $\hat{p} = (-1.28)(0.094) + 0.328 = 0.21$. As a result, the value $p = 0.21$ cuts off the lower 10% of the distribution.

Exercise 7 [17]

[6] a. A point estimate for p is

$$\begin{aligned}\hat{p} &= \frac{15}{27} \\ &= 0.556.\end{aligned}$$

Since $n\hat{p} = 27(0.556) = 15$ and $n(1-\hat{p}) = 27(0.444) = 12$, the sample size is large enough to justify the use of the normal approximation. Therefore, an approximate 95% confidence interval for p is

$$\left(0.556 - 1.96\sqrt{\frac{0.556(1-0.556)}{27}},\ 0.556 + 1.96\sqrt{\frac{0.556(1-0.556)}{27}}\right)$$

or

$$(0.369,\ 0.743).$$

We are 95% confident that these limits cover the true population proportion p.

[2] b. The null hypothesis of the test is

$$H_0: \mu = 0.328.$$

[2] c. The alternative hypothesis is

$$H_A: \mu \neq 0.328.$$

[5] d. The test statistic is

$$\begin{aligned}z &= \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \\ &= \frac{0.556 - 0.328}{\sqrt{0.328(1-0.328)/27}} \\ &= 2.52.\end{aligned}$$

Therefore, $p = 2(0.006) = 0.012$. Since $p > 0.01$, we are unable to reject the null hypothesis.

[2] e. We conclude that for children with an oral cleft, there is no evidence that the proportion of mothers who smoked during pregnancy is different from the proportion of mothers who smoked for children with other types of malformations. (Note: If the test were being conducted at the 0.05 level of significance, we would reject H_0 and conclude that the proportion is higher than 32.8%.)

[5] extra credit

f. In this case, $p_0 = 0.328$ and $p_1 = 0.250$. Since $\alpha = 0.01$ for a two-sided test and $\beta = 0.10$, we have that $z_{\alpha/2} = 2.58$ and $z_\beta = 1.28$, and

$$\begin{aligned}n &= \left[\frac{2.58\sqrt{p_0(1-p_0)} + 1.28\sqrt{p_1(1-p_1)}}{p_1 - p_0}\right]^2 \\ &= \left[\frac{2.58\sqrt{0.328(1-0.328)} + 1.28\sqrt{0.250(1-0.250)}}{0.250 - 0.328}\right]^2 \\ &= 512.3.\end{aligned}$$

A sample of size 513 would be required.