Chapter 23 – Comparing Two Proportions

23.1 STATE: We want to estimate the difference between the proportions of male and female Internet users who used the Internet to obtain health information in the last year. PLAN: Let p_M be the proportion of all males who have used the Internet to search for health information and p_F be the proportion of females who have done so. We want a 95% confidence interval for the difference in these proportions. SOLVE: The samples are large, with clearly more than 10 successes and failures in each sample. Assume the observations from each group can be thought of as an SRS. The sample proportions are $\hat{p}_{\mathrm{F}} = \frac{811}{1308} = 0.62$ and $\hat{p}_{\mathrm{M}} = \frac{520}{1084} = 0.4797$. The standard error of the difference is $SE = \sqrt{\frac{0.62(1-0.62)}{1308} + \frac{0.4797(1-0.4797)}{1084}} = 0.0203$. The 95% confidence interval is (0.62 – 0.4797) ± 1.96(0.0203) = 0.1403 ± 0.0398, or 0.1005 to 0.1801. CONCLUDE: We are 95% confident that between 10% and 18% more women than men have looked for health information on the Internet.

23.2 (a) Let p_{A} denote the proportion of children who developed an immune response after receiving the aerosolized vaccine and p_{I} the proportion who developed a response after receiving the injection. **(b)** Yes, assuming the samples can be thought of as SRSs, we can use the large-sample interval, since there are more than 10 successes (developing a response) and 10 failures (no response) in each group. **(c)** The sample proportions are $\hat{p}_{\mathrm{A}} = \frac{662}{775} = 0.8542$ and $\hat{p}_{\mathrm{I}} = \frac{743}{785} = 0.9465$. The standard error of the difference is $SE = \sqrt{\frac{0.8542(1-0.8524)}{775} + \frac{0.9465(1-0.9465)}{785}} = 0.015$. A 95% confidence interval is $0.8542 - 0.9465 \pm 1.96(0.015) = -0.0923 \pm 0.0294$, or -0.1217 to -0.0629. We are 95% confident that between 6.29% and 12.17% more children develop a response after receiving the injection compared to the aerosolized vaccine. **(d)** No. The interval is entirely above 6% (or below -6%), indicating the aerosolized vaccine is inferior to the injection.

23.3 STATE: We want to estimate the difference between the proportion of males and females who support physician-assisted suicide. PLAN: Let p_{M} and p_{F} denote the proportion of males and females who support physician-assisted suicide, respectively. We will construct a 99% confidence interval for the difference $p_{\mathrm{M}} - p_{\mathrm{F}}$. SOLVE: Each sample has more than 10 successes (responded "allow") and 10 failures (did not respond "allow"). Assuming we can treat each sample as an SRS, we can construct a large sample interval. The sample proportions are $\hat{p}_{\mathrm{M}} = \frac{539}{729} = 0.7394$ and $\hat{p}_{\mathrm{F}} = \frac{582}{888} = 0.6554$. The standard error is $SE = \sqrt{\frac{0.7394(1-0.7394)}{729} + \frac{0.6554(1-0.6554)}{888}} = 0.0228$. The 99% confidence interval is $0.7394 - 0.6554 \pm 2.576(0.0228) = 0.0253$ to 0.1427. CONCLUDE: We are 99% confident that between 2.53% and 14.27% more males than females support physician-assisted suicide.

23.4 (a) STATE: We want to know if blacking out the mouth region reduces a subject's ability to correctly match the dog-owner pairs. PLAN: Let p_1 be the proportion who correctly identify the pairs under the conditions of experiment 1, and let p_2 denote the proportion when the mouth is blacked out. We are testing the hypotheses $H_0 : p_1 = p_2$ versus $H_a : p_1 > p_2$. SOLVE: Assume the samples can each be thought of as an SRS. We can conduct a hypothesis test since there are more than five successes (correctly identified the pairs) and more than five failures in each sample. The sample proportions are $\hat{p}_1 = \frac{49}{61} = 0.8033$ and $\hat{p}_2 = \frac{37}{51} = 0.7255$. The pooled sample proportion is $\hat{p} = \frac{49 + 37}{61 + 51} = 0.7679$. The test statistic is $z = \frac{0.8033 - 0.7255}{\sqrt{\left(0.7679(1 - 0.7679)\left(\frac{1}{61} + \frac{1}{51}\right)\right)}} = 0.9712$. The one-sided P-value is $p = 0.166$. There is not enough evidence to conclude that blacking out the mouth region reduces a subject's ability to correctly match dog-owner pairs. **(b)** STATE: We want to know if blacking out the eye region reduces a subject's ability to correctly match the dog-owner pairs. PLAN: Let p_1 be the proportion who correctly identify the pairs under the conditions of experiment 1, and let p_2 denote the proportion when the eye region is blacked out. We are testing the hypotheses $H_0 : p_1 = p_2$ versus $H_a : p_1 > p_2$. SOLVE: Assume the samples can each be thought of as an SRS. We can conduct a hypothesis test since there are more than five successes (correctly identified the pairs) and more than five failures in each sample. The sample proportions are $\hat{p}_1 = 0.8033$ and $\hat{p}_2 = \frac{30}{60} = 0.5$. The pooled sample proportion is $\hat{p} = \frac{49 + 30}{61 + 60} = 0.6529$. The test statistic is $z = \frac{0.8033 - 0.5}{\sqrt{\left(0.6529(1 - 0.6529)\left(\frac{1}{61} + \frac{1}{60}\right)\right)}} = 3.5$. The one-sided P-value is less than 0.0005. CONCLUDE: There is strong evidence that blacking out the eye region reduces one's ability to match dog-owner pairs. **(c)** The conclusions in parts (a) and (b) imply the eye region plays a larger role in matching dogs and owners than the mouth region does.

23.5 STATE: Is helmet use less common among skiers and snowboarders with head injuries than skiers and snowboarders without head injuries? PLAN: Let p_1 and p_2 be (respectively) the proportions of injured skiers and snowboarders and the proportion of uninjured skiers and snowboarders who wear helmets, respectively. We test $H_0 : p_1 = p_2$ versus $H_a : p_1 < p_2$. SOLVE: Assume the observations from each group can be thought of as an SRS. The smallest count is 96, so the significance testing procedure is safe. We find $\hat{p}_1 = \frac{96}{578} = 0.1661$ and $\hat{p}_2 = \frac{656}{2992} = 0.2193$. The pooled proportion is $\hat{p} = \frac{96+656}{578+2992} = 0.2106$. Then for the significance test, $SE = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{578} + \frac{1}{2992}\right)} = 0.01853$. The test statistic is therefore $z = \frac{0.1661-0.2193}{0.01853} = -2.87$ and $P = 0.0021$. CONCLUDE: We have strong evidence (significant at $\alpha = 0.01$) that skiers and snowboarders with head injuries are less likely to use helmets than skiers and snowboarders without head injuries.

23.6 STATE: Does the proportion having the primary outcome differ in the treatment versus control groups? PLAN: Let p_1 be the proportion experiencing the primary outcome with the treatment and p_2 the proportion without the treatment. We test $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$. SOLVE: All counts of successes and failures are larger than five. Subjects were randomly assigned to the treatment or placebo. The sample proportions are $\hat{p}_1 = \frac{113}{3180} = 0.0355$ and $\hat{p}_2 = \frac{157}{3168} = 0.0496$. The pooled sample proportion is $\hat{p} = \frac{113 + 157}{3180 + 3168} = 0.0425$. The test statistic is $z = \frac{0.0355 - 0.0496}{\sqrt{0.0425(1 - 0.0425)\left(\frac{1}{3180} + \frac{1}{3168}\right)}} = -2.78$. The two-sided P-value is between 0.005 and 0.01. There is evidence that the proportion experiencing the primary outcome is different for those receiving the treatment compared to those without.

23.7 (a) We should not use the large-sample confidence interval because only five infants in the consumption group developed a peanut allergy. **(b)** The sample sizes become 265 and 268, with the counts of successes now 37 and 6. **(c)** Let p_{A} denote the proportion who develop a peanut allergy after avoiding peanuts and p_C the proportion after consuming peanuts. Assume the observations from each group can be thought of as an SRS. The sample proportions are $\tilde{p}_{\mathrm{A}} = \frac{37}{265} = 0.1396$ and $\tilde{p}_{\mathrm{C}} = \frac{6}{268} = 0.0224$. The standard error is $SE = \sqrt{\frac{0.1396(1 - 0.1396)}{265} + \frac{0.0224(1 - 0.0224)}{268}} = 0.0231$. The 99% plus four confidence interval for $p_{\mathrm{A}} - p_{\mathrm{C}}$ is $0.1396 - 0.0224 \pm 2.576(0.0231) = 0.0577$ to 0.1767. We are 99% confident that between 5.77% and 17.67% more infants with severe eczema, egg allergy, or both will develop a peanut allergy by 60 months after avoiding peanuts compared to those who consume it.

23.8 PLAN: We want a 95% confidence interval for the difference in *Xerospirea hartwegiana* shrubs that will resprout after being clipped and exposed to fire or not exposed to fire. This was a designed experiment, with the shrubs randomly assigned to treatment (fire exposure) or control. SOLVE: The total number of shrubs exposed to each treatment was only 12; counts of resprouted shrubs were 12 and 8 (successes), with 0 and 4 shrubs that did not resprout (failures). Because these counts are small, we need the plus four confidence interval. Let $\tilde{p}_{\mathrm{T}} = \frac{8+1}{12+2} = 0.6429$ be the estimate of resprouting for shrubs exposed to fire and $\tilde{p}_{\mathrm{C}} = \frac{12+1}{12+2} = 0.9286$ be the estimate of resprouting for control shrubs. The plus four 95% confidence interval for $p_{\mathrm{C}} - p_{\mathrm{T}}$ is $(0.9286 - 0.6429) \pm 1.96\sqrt{\frac{\tilde{p}_{\mathrm{C}}(1-\tilde{p}_{\mathrm{C}})}{14} + \frac{\tilde{p}_{\mathrm{T}}(1-\tilde{p}_{\mathrm{T}})}{14}} = 0.2857 \pm 0.2849 = 0.0008$ to 0.5706. CONCLUDE: We are 95% confident that burning reduces the percent of shrubs that will resprout by between 0.08% and 57.1%. (Control shrubs are much more likely to resprout.)

23.9 (b) H_0: $p_9 = p_{12}$ versus H_a: $p_9 > p_{12}$.

23.10 (a) $\hat{p}_9 = 0.396$ and $\hat{p}_{12} = 0.338$. $\hat{p}_9 = \frac{1374}{3470}$ and $\hat{p}_{12} = \frac{1116}{3301}$.

23.11 (b) $\hat{p} = 0.368$. $\hat{p} = \frac{1374 + 1116}{3470 + 3301}$.

23.12 (c) $z = 4.94$. $z = \frac{0.396 - 0.338}{\sqrt{0.368(1 - 0.368)\left(\frac{1}{3470} + \frac{1}{3301}\right)}}$.

23.13 (b) 0.058 ± 0.019. That is, $0.396 - 0.338 \pm 1.645\sqrt{\frac{0.396(1-0.396)}{3470} + \frac{0.338(1-0.338)}{3301}}$.

23.14 (a) z = 2.25, P < 0.02. $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.70-0.20}{\sqrt{0.45(1-0.45)\left(\frac{1}{10}+\frac{1}{10}\right)}}$ = 2.25, and the P-value is one-sided and less than 0.02.

23.15 (b) may be inaccurate because some counts of successes and failures are too small. We have only three failures in the treatment group and only two successes in the control group.

23.16 (b) 0.417 ± 0.304. $\tilde{p}_1 = \frac{7+1}{10+2}$ = 0.667, $\tilde{p}_2 = \frac{2+1}{10+2}$ = 0.25, and the margin of error is $1.645\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{12} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{12}}$ = 0.304.

23.17 (a) The four counts are 117, 53, 152, and 165, so all counts are large enough.
(b) Using the large-sample method, $\hat{p}_1 = \frac{117}{170} = 0.6882$ and $\hat{p}_2 = \frac{152}{317} = 0.4795$, and the 95% confidence interval is $\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{170} + \frac{\hat{p}_2(1-\hat{p}_2)}{317}}$ = 0.2087 ± 0.0887 = 0.12 to 0.2974. Based on these samples, between 12% and 29.7% more younger teens than older teens have posted false information in their online profiles, at 95% confidence.

23.18 (a) For the sibutramine group, $\hat{p}_1 = \frac{561}{4906}$ = 0.1143. For the control (placebo) group, $\hat{p}_2 = \frac{490}{4898}$ = 0.1. **(b)** The counts are 561, 4345, 490, and 4408—easily large enough for use of the large-sample confidence interval procedure. **(c)** Using the large-sample method, the 95% confidence interval is $\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{4906} + \frac{\hat{p}_2(1-\hat{p}_2)}{4898}}$ = 0.0143 ± 0.0122 = 0.0021 to 0.0265, or 0.2% to 2.7%.

23.19 (a) One of the counts is 0; for large-sample methods, we need all counts to be at least 10 for the confidence interval (at least 5 for the hypothesis test). Assume the mice were randomly assigned to the treatment. **(b)** After we add the two observations to each sample, the sample size for the treatment group is 35, 24 of which have tumors; the sample size for the control group is 20, 1 of which has a tumor. **(c)** $\tilde{p}_1 = \frac{23+1}{33+2}$ = 0.6857 and $\tilde{p}_2 = \frac{0+1}{18+2}$ = 0.05. The plus four 99% confidence interval is $\tilde{p}_1 - \tilde{p}_2 \pm 2.576\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{35} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{20}}$ = 0.6357 ± 0.238 = 0.3977 to 0.8737. We are 99% confident that lowering DNA methylation increases the incidence of tumors by between about 40% and 87%.

23.20 (a) Let p_1 and p_2 be (respectively) the proportions of subjects in the treatment and control groups experiencing a primary outcome. We test $H_0: p_1 = p_2$ versus $H_a: p_1 \neq p_2$. Assume the observations from each group can be thought of as an SRS. For the treatment group, $\hat{p}_1 = \frac{561}{4906}$ = 0.1143. For the control (placebo) group, $\hat{p}_2 = \frac{490}{4898}$ = 0.1. The pooled estimate is $\hat{p} = \frac{561+490}{4906+4898}$ = 0.1072. $z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{4906}+\frac{1}{4898}\right)}}$ = 2.29, and P = 0.022. We have strong evidence that the proportion of subjects on sibutramine who are suffering a primary outcome differs from those who are on the placebo. **(b)** A comparison group is important because we want to learn about the difference in rate of primary outcome due to sibutramine. A placebo should be used in order to blind patients to which group they are in and to account for any possible placebo effect.

23.21 (a) Let p_1 and p_2 be (respectively) the proportions of subjects in the music and no music groups who receive a passing grade on the Maryland HSA. We test $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$. For the music group, $\hat{p}_1 = \frac{2818}{3239}$ = 0.87. For the no music group, $\hat{p}_2 = \frac{2091}{2787}$ = 0.75. The pooled estimate is $\hat{p} = \frac{2818+2091}{3239+2787}$ = 0.815. Hence, $z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{3239}+\frac{1}{2787}\right)}} =$ 11.96 (may vary slightly due to roundoff). An observed difference of 0.87 − 0.75 = 0.12 in group proportions is much too large to be explained by chance alone, and P < 0.0001. We have overwhelming evidence [Or do we? See part (b).] that the proportion of music students passing the Maryland HSA is greater than that for the no music group. **(b)** and **(c)** This is an observational study—people who choose to (or can afford to) take music lessons differ in many ways from those who do not. Hence, we cannot conclude that music causes an improvement in Maryland HSA achievement.

23.22 We estimate the overall proportion of ninth-graders who passed the HSA test. As computed in Exercise 23.21, $\hat{p} = \frac{2818+2091}{3239+2787} = \frac{4909}{6026}$ = 0.815. A 95% confidence interval for the proportion p is given by $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{6026}}$ = 0.815 $\pm$ 0.01 = 0.805 to 0.825, or 80.5% to 82.5%.

23.23 We have at least 10 successes and 10 failures for both samples. For the music group, $\hat{p}_1 = \frac{2818}{3239}$ = 0.87. For the no music group, $\hat{p}_2 = \frac{2091}{2787}$ = 0.75. $\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{3239} + \frac{\hat{p}_2(1-\hat{p}_2)}{2787}}$ = 0.1 to 0.14, or 10% to 14%.

23.24 (a) $\hat{p}_1 = \frac{270}{1847} = 0.1462$ is the proportion of patients with complications before the restriction, and $\hat{p}_2 = \frac{170}{1639} = 0.1037$ is the proportion with complications after the restriction. We can test $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ to assess the strength of the evidence that the proportions are different. We have $\hat{p} = \frac{270+170}{1847+1639} = 0.1262$ as the pooled proportion, and we find $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{1847}+\frac{1}{1639}\right)}} = 3.77$ with P-value 0.0002. This is extremely strong evidence that the proportions are different. **(b)** This study was observational; there were no assigned treatments because all “subjects” underwent bariatric surgery—they (or their records) were examined later for complications. We cannot determine cause and effect in observational studies. **(c)** The possible other reasons for the decrease in complications are all lurking (confounding) variables; they weaken the case that the decline in complications is due to the restrictions. **(d)** Answers will vary, but the reason for a comparative control group is to eliminate lurking variables (as much as possible). We can (safely) assume that both groups experienced the same exposure to newer methods, increased surgeon experience, etc. This allows us to focus on the factor of interest—the imposition of the restrictions and whether those restrictions improved results.

23.25 (a) To test $H_0: p_M = p_F$ versus $H_a: p_M \neq p_F$, we find $\hat{p}_M = \frac{15}{106}$ = 0.1415, $\hat{p}_F = \frac{7}{42}$ = 0.1667, and $\hat{p}$ = 0.1486. Then, $SE = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{106}+\frac{1}{42}\right)} = 0.06485$, so $z = \frac{\hat{p}_\mathrm{M}-\hat{p}_\mathrm{F}}{0.06485} = -0.39$. This gives P = 0.6966, which provides virtually no evidence of a difference in failure rates. **(b)** We have $\hat{p}_\mathrm{M} = \frac{450}{3180}$ = 0.1415, $\hat{p}_\mathrm{F} = \frac{210}{260}$ = 0.1667, and $\hat{p}$ = 0.1486, but now $SE = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{3180}+\frac{1}{1260}\right)} = 0.01184$, so $z = \frac{\hat{p}_\mathrm{M}-\hat{p}_\mathrm{F}}{0.01184} = -2.13$and P = 0.0332. **(c)** We are asked to construct two confidence intervals—one based on the smaller samples of part (a) and one based on the larger samples of part (b). First, for case (a), $\hat{p}_M$ = 0.1415 and $\hat{p}_F$ = 0.1667, so a 95% confidence interval for the difference is $\hat{p}_M - \hat{p}_F \pm 1.96\sqrt{\frac{\hat{p}_\mathrm{M}(1-\hat{p}_\mathrm{M})}{106}+\frac{\hat{p}_\mathrm{F}(1-\hat{p}_\mathrm{F})}{42}}$ = −0.156 to 0.1056. We note that because there were only seven business failures in those businesses headed by women, this interval is not really appropriate (even though the hypothesis test was appropriate). For case (b), $\hat{p}_M$ = 0.1415 and $\hat{p}_F$ = 0.1667. The resulting confidence interval is $\hat{p}_M - \hat{p}_F \pm 1.96\sqrt{\frac{\hat{p}_\mathrm{M}(1-\hat{p}_\mathrm{M})}{3180}+\frac{\hat{p}_\mathrm{F}(1-\hat{p}_\mathrm{F})}{1260}}$ = −0.0491 to −0.0013.

23.26 PLAN: Let p_1 be the proportion of college graduates who have used a ride-hailing app and p_2 the proportion of non-college graduates who have used a ride-hailing app. We want to know if the proportions differ, so we test $H_0: p_1 = p_2$ versus $H_\mathrm{a}: p_1 \neq p_2$. SOLVE: Assume the observations from each group can be thought of as an SRS. All counts are much larger than five, so we can perform the hypothesis test. The sample proportions are $\hat{p}_1 = \frac{687}{2369} = 0.29$ and $\hat{p}_2 = \frac{268}{2418} = 0.1108$. The pooled sample proportion is $\hat{p} = \frac{687+268}{2369+2418} = 0.1995$. The test statistic is $z = \frac{0.29-0.1108}{\sqrt{0.1995(1-0.1995)\left(\frac{1}{2369}+\frac{1}{2418}\right)}} = 15.51$. CONCLUDE: The P-value is approximately zero, so there is overwhelming evidence that the proportion of adults who have used a ride-hailing app is different between college graduates and those without college degrees.

23.27 PLAN: Let p_1 be the proportion of males who have used a ride-hailing app and p_2 the proportion of females. We want to know if the proportions differ, so we test $H_0: p_1 = p_2$ versus $H_\mathrm{a}: p_1 \neq p_2$. SOLVE: Assume the observations from each group can be thought of as an SRS. All counts are much larger than five, so we can perform the hypothesis test. The sample proportions are $\hat{p}_1 = \frac{378}{2361} = 0.1601$ and $\hat{p}_2 = \frac{340}{2426} = 0.1401$. The pooled sample proportion is $\hat{p} = \frac{378 + 340}{2361 + 2426} = 0.15$. The test statistic is $z = \frac{0.1601 - 0.1401}{\sqrt{0.15(1 - 0.15)\left(\frac{1}{2361} + \frac{1}{2426}\right)}} = 1.94$. CONCLUDE: The *P*-value is *P* = 0.0524, so there is not strong evidence to suggest a difference in the proportion of adult males and females who have used a ride-hailing app.

23.28 PLAN: Use the same notation from Exercise 23.26. SOLVE: Assume the observations from each group can be thought of as an SRS. We can construct a large-sample 90% confidence interval because there are more than 10 counts in each group. The 90% confidence interval is $0.29 - 0.1108 \pm 1.645\sqrt{\frac{0.29(1 - 0.29)}{2369} + \frac{0.1108(1 - 0.1108)}{2418}} = 0.1606$ to 0.1978. CONCLUDE: We are 90% confident that the proportion of adults with college degrees who have used a ride-hailing app is between about 16% and 20% higher than for adults without college degrees.

23.29 PLAN: Let p_{C} be the proportion who would abstain from smoking with Chantix and p_{P} the proportion without Chantix. We want a 99% confidence interval for $p_{\text{C}} - p_{\text{P}}$. SOLVE: The sample proportions are $\hat{p}_{\text{C}} = \frac{244}{760} = 0.3211$ and $\hat{p}_{\text{P}} = \frac{52}{750} = 0.0693$. Assume the observations from each group can be thought of as an SRS. The observed counts of successes and failures are all at least 10, so we can construct a confidence interval. The standard error is $SE = \sqrt{\frac{0.3211(1 - 0.3211)}{760} + \frac{0.0693(1 - 0.0693)}{750}} = 0.0193$. The 99% confidence interval is $0.3211 - 0.0693 \pm 2.576(0.0193) = 0.2021$ to 0.3015. CONCLUDE: We are 99% confident that the proportion who abstain from smoking on Chantix is between 20.21% and 30.15% higher than those without Chantix.

23.30 (a) PLAN: Let p_1 denote the proportion who would buy the dress with the headless mannequin and p_2 the proportion with the mannequin with a head. We want to conduct a hypothesis test of the hypotheses $H_0: p_1 = p_2$ versus $H_{\text{a}}: p_1 \neq p_2$. SOLVE: The observed counts of successes and failures are all at least five. Assuming the sample can be thought of as an SRS, we can perform a hypothesis test. The sample proportions are $\hat{p}_1 = \frac{18}{53} = 0.3396$ and $\hat{p}_2 = \frac{10}{53} = 0.1887$. The pooled sample proportion is $\hat{p} = \frac{10 + 18}{53 + 53} = 0.2642$. The test statistic is $z = \frac{0.3396 - 0.1887}{\sqrt{0.2642(1 - 0.2642)\left(\frac{1}{53} + \frac{1}{53}\right)}} = 1.76$. The two-sided *P*-value is $P = 0.078$. CONCLUDE: There is not strong evidence to conclude the proportion of women who would buy the dress differs between those who viewed the dress on a mannequin with or without a head. **(b)** Based on the result in part (a), it does not matter if manufacturers use display mannequins with or without a head.

23.31 PLAN: Let p_1 denote the proportion who stop when a neutral expression was used and p_2 the proportion when smiling. We want to conduct a hypothesis test of the hypotheses $H_0: p_1 = p_2$ versus $H_{\text{a}}: p_1 < p_2$. SOLVE: The observed counts of successes and failures are all at least five. Assuming the sample can be thought of as an SRS, we can perform a hypothesis test. The sample proportions are $\hat{p}_1 = \frac{172}{400} = 0.43$ and $\hat{p}_2 = \frac{226}{400} = 0.565$. The pooled sample proportion is $\hat{p} = \frac{172+226}{400+400} = 0.4975$. The test statistic is $z = \frac{0.43 - 0.565}{\sqrt{0.4975(1-0.4975)\left(\frac{1}{400}+\frac{1}{400}\right)}} = -3.82$. From Table C, the one-sided P-value is less than 0.0005. CONCLUDE: There is evidence that a smile increases the proportion of drivers who stop for a pedestrian at a pedestrian crossing.

23.32 (a) PLAN: Let p_1 be the proportion of subjects who had made at least two serious attempts to quit smoking in the treatment group and p_2 the proportion in the placebo group. We want to conduct a hypothesis test of $H_0: p_1 = p_2$ versus $H_a: p_1 \neq p_2$. SOLVE: There are more than five successes and five failures in each group, and we can think of each group as an SRS. The sample proportions are $\hat{p}_1 = \frac{439}{760} = 0.5776$ and $\hat{p}_2 = \frac{303}{750} = 0.404$. The pooled sample proportion is $\hat{p} = \frac{439+303}{760+750} = 0.4914$. The test statistic is $z = \frac{0.5776 - 0.404}{\sqrt{0.4914(1-0.4914)\left(\frac{1}{760}+\frac{1}{750}\right)}} = 6.75$. The P-value is approximately zero. CONCLUDE: There is very strong evidence that the proportion who have made at least two serious attempts to quit smoking is not the same for the treatment and placebo groups. **(b)** PLAN: Let μ_1 be the mean number of cigarettes smoked per day for subjects receiving treatment and μ_2 the mean for the placebo group. We want to do a two-sample t test of the hypotheses $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$. SOLVE: If we assume the population of both groups is Normal and both groups can be thought of as a SRS, then we can perform a two-sample t test. The test statistic is $t = \frac{20.6 - 20.8}{\sqrt{\frac{8.5^2}{760}+\frac{8.2^2}{750}}} = -0.465$. Using Option 2, the degrees of freedom are $750 - 1 = 749$. Using Table C, the P-value is greater than 0.5. CONCLUDE: There is not evidence that the mean number of cigarettes smoked per day is different for the treatment and placebo groups.

23.33 PLAN: Let p_1 and p_2 be (respectively) the proportions of mice ready to breed in good acorn years and bad acorn years. We give a 90% confidence interval for $p_1 - p_2$. SOLVE: Assume the mice are randomly assigned to a treatment. One count is only seven, and the guidelines for using the large-sample method call for all counts to be at least 10, so we use the plus four method. We have $\tilde{p}_1 = \frac{54+1}{72+2} = 0.7432$ and $\tilde{p}_2 = \frac{10+1}{17+2} = 0.5789$, so the plus four 90% confidence interval is $\tilde{p}_1 - \tilde{p}_2 \pm 1.645\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{74}+\frac{\tilde{p}_2(1-\tilde{p}_2)}{19}} = 0.1643 \pm 0.2042 = -0.0399$ to 0.3685. CONCLUDE: We are 90% confident that the proportion of mice ready to breed in good acorn years is between 0.04 lower than and 0.37 higher than the proportion in bad acorn years.

23.34 PLAN: To answer the question about whether children who had early childhood education have a higher proportion of consistent employment, we will use a test of $H_0: p_E = p_C$ versus $H_a: p_E > p_C$ where p_E is the proportion from the intervention group who received intensive early childhood education. If warranted by the test, we will estimate the difference with a 95% confidence interval for the difference in proportions. SOLVE: Assume the observations from each group can be thought of as an SRS. All counts are larger than 10 (the smallest is 13 for the intervention group who were not consistently employed), so inference is appropriate. The sample proportions are $p_{\mathrm{E}} = \frac{39}{52} = 0.75$ and $p_{\mathrm{C}} = \frac{26}{49} = 0.5306$, and $\hat{p} = \frac{39+26}{52+49} = 0.6436$. The test statistic is $z = \frac{0.75-0.5306}{\sqrt{0.6436(1-0.6436)\left(\frac{1}{52}+\frac{1}{49}\right)}} = 2.3$, with $P = 0.0107$. We have strong evidence of a difference in the proportions in the two groups who had consistent employment. How large is the difference? The 95% confidence interval is $(0.75 - 0.5306) \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{52} + \frac{\hat{p}_2(1-\hat{p}_2)}{49}} = 0.0367$ to 0.4021. CONCLUDE: We have strong evidence that children who receive intensive early childhood education are more likely to be consistently employed at age 30. Between about 3.7% and 40.2% more children with early education are likely to be consistently employed than those who do not receive the education, at 95% confidence.

23.35 (a) This is an experiment, because the researchers assigned subjects to the groups being compared. **(b)** PLAN: Let p_1 and p_2 be (respectively) the proportions of subjects who have an RV infection for the HL+ group and the control group. We test $H_0\colon p_1 = p_2$ versus $H_a\colon p_1 < p_2$. SOLVE: Assume the observations from each group can be thought of as an SRS. We have large-enough counts (49, 67, 49, and 47) to use the large-sample significance testing procedure safely. Now, $\hat{p}_1 = \frac{49}{49+67} = 0.4224$, $\hat{p}_2 = \frac{49}{49+47} = 0.5104$, and $\hat{p} = \frac{49+49}{116+96} = 0.4623$. $SE = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{116}+\frac{1}{96}\right)} = 0.0688$. The test statistic is therefore $z = \frac{0.4224-0.5104}{0.0688} = -1.28$, for which $P = 0.1003$. CONCLUDE: We do not have enough evidence to reject the null hypothesis; there is little evidence to conclude that the proportion of HL+ users with a rhinovirus infection is less than that for non-HL+ users.

23.36 Let p_1 be the proportion who toss six or more heads under the conditions of the experiment and p_2 the proportion under the conditions of the control group. **(a)** $\hat{p}_1 = \frac{33}{61} = 0.541$ and $\hat{p}_2 = 25/67 = 0.3731$. **(b)** PLAN: Test the hypotheses $H_0\colon p_1 = p_2$ versus $H_a\colon p_1 \neq p_2$. SOLVE: Assume the observations from each group can be thought of as an SRS. There are more than five successes and five failures in each group. The pooled sample proportion is $\hat{p} = \frac{25 + 33}{67 + 61} = 0.4531$. The test statistic is $z = \frac{0.5410 - 0.3731}{\sqrt{0.4531(1 - 0.4531)\left(\frac{1}{67}+\frac{1}{61}\right)}} = 1.91$. Using Table C, the two-sided P-value is $P = 0.056$. CONCLUDE: There is moderate evidence that the proportions reporting tossing heads six or more times differ between the two groups. This provides some evidence in favor of the researchers' conjecture that banking culture favors dishonest behavior.