Past-paper walk-through

Setting Up a Test for the Difference of Two Population Proportions ▪ 6.10

eXercise

Questions in this set

- 2024 Q1
- 2019 Q4
- 2015 Q4

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 1

2024 ▪ Q1

A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

At a significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who

Question 1

would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

Solution 1

Mark scheme

- Two-sample z-test for a difference in population proportions. Let $p_{younger}$ = the true proportion of all exercise center members ages 18-55 who would be interested in online fitness classes, and p_{older} = the true proportion of all members age 56+ who would be interested. $H_0 : p_{younger} = p_{older}$, $H_a : p_{younger} \neq p_{older}$ (two-sided), with context referencing both age groups (18-55, 56+) AND the populations (all exercise center members). Conditions: random samples (170 members 18-55; 230 members 56+) support independence; the 10

Question 4

2019 ▪ Q4

Tumbleweed, commonly found in the western United States, is the dried structure of certain plants that are blown by the wind. Kochia, a type of plant that turns into tumbleweed at the end of the summer, is a problem for farmers because it takes nutrients away from soil that would otherwise go to more beneficial plants. Scientists are concerned that kochia plants are becoming resistant to the most commonly used herbicide, glyphosate. In 2014, 19.7 percent of 61 randomly selected kochia plants were resistant to glyphosate. In 2017, 38.5 percent of 52 randomly selected kochia plants were resistant to glyphosate. Do the data provide convincing statistical evidence, at the level of $\alpha = 0.05$, that there has been an increase in the proportion of all kochia plants that are resistant to glyphosate?

Solution 4

Mark scheme

- **Section 1 – Hypotheses and test procedure:** Let p_{14} and p_{17} represent the population proportions of kochia plants resistant to glyphosate in 2014 and 2017, respectively. $H_0 : p_{17} - p_{14} = 0$ tested against $H_a : p_{17} - p_{14} > 0$. The appropriate procedure is a two-sample z-test for a difference in proportions: $z = \dfrac{\hat{p}_{17} - \hat{p}_{14}}{\sqrt{\hat{p}_c(1-\hat{p}_c)/n_{17} + \hat{p}_c(1-\hat{p}_c)/n_{14}}}$, where $\hat{p}_c = \dfrac{n_{14}\hat{p}_{14} + n_{17}\hat{p}_{17}}{n_{14} + n_{17}}$ is the pooled estimate.

Solution 4

Mark scheme

- **Section 2 – Conditions and calculation:** Conditions: (1) independent random samples were taken in 2014 (n=61) and 2017 (n=52), and it is reasonable that the 2017 sample was not influenced by the 2014 sample; (2) approximate normality: pooled estimate $\hat{p}_c = \frac{61(0.197)+52(0.385)}{61+52} \approx 0.2835$; expected counts $61(0.2835) \approx 17.29$, $61(0.7165) \approx 43.71$, $52(0.2835) \approx 14.74$, $52(0.7165) \approx 37.26$, all greater than 10, so the normal approximation is valid; (3) independence (10
- **Section 3 – Conclusion:** Because the p-value (0.0135) is less than $\alpha = 0.05$, there is convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population.

Solution 4

Mark scheme

- Full 4 points requires all three sections essentially correct: correct hypotheses/test identification, correct conditions-check and calculation ($z \approx 2.21$, $p \approx 0.0135$), and a correct conclusion in context justified by comparing the p-value to α. (A chi-square test for homogeneity or a two-sided confidence interval for the difference in proportions are also accepted alternative approaches with equivalent numeric results.)

Question 4

2015 ▪ Q4

A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 1,000 adult volunteers were randomly assigned to one of two groups. Half of the volunteers were assigned to the experimental group that took a low-dose aspirin each day, and the other half were assigned to the control group that took a placebo each day. At the end of six years, 15 of the people who took the low-dose aspirin had developed colon cancer and 26 of the people who took the placebo had developed colon cancer. At the significance level $\alpha = 0.05$, do the data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers?

Solution 4

Mark scheme

- Step 1 (hypotheses): Let p_{asp} = the true proportion of adults like the volunteers who would develop colon cancer within six years if they took a low-dose aspirin daily, and p_{plac} = the corresponding proportion if they took a placebo daily. $H_0 : p_{asp} = p_{plac}$ (equivalently $p_{asp} - p_{plac} = 0$) vs $H_a : p_{asp} < p_{plac}$ (equivalently $p_{asp} - p_{plac} < 0$).

Solution 4

- Step 2 (procedure and conditions): Two-sample z-test for comparing proportions. Random-assignment condition satisfied because this was a randomized experiment (volunteers were randomly assigned to aspirin or placebo). Large-counts condition satisfied: 15 colon-cancer cases in the aspirin group, 26 in the placebo group, $500 - 15 = 485$ cancer-free in the aspirin group, and $500 - 26 = 474$ cancer-free in the placebo group - all large enough.
- Step 3 (test statistic and p-value): $\hat{p}_{asp} = 15/500 = 0.030$, $\hat{p}_{plac} = 26/500 = 0.052$, combined $\hat{p}_{combined} = (15+26)/(500+500) = 0.041$. $z = \dfrac{0.030 - 0.052}{\sqrt{0.041(1-0.041)\left(\frac{1}{500}+\frac{1}{500}\right)}} \approx -1.75$ (calculator: -1.7542). p-value $= P(Z \leq -1.75) \approx 0.0401$ (calculator: 0.0397).

Solution 4

Mark scheme

- Step 4 (conclusion): Because the p-value (≈ 0.04) is less than $\alpha = 0.05$, reject H_0. The data provide convincing statistical evidence that the proportion of adults like the volunteers who would develop colon cancer while taking daily low-dose aspirin is less than the proportion who would develop colon cancer while taking a daily placebo. Full credit needs all four steps: correctly defined/labeled hypotheses, the named procedure with both conditions checked, a correct test statistic with a p-value consistent with it, and a conclusion in context explicitly linked to comparing the p-value with α.