Stat 20 Fall 2003, Quiz 2 Answers

B. M. Bolstad, bolstad@stat.berkeley.edu

Sept 22, 2003

Question 1. (*12 points*)

The Gallup polling organization asked the question "Do you approve or disapprove of the way George W. Bush is handling the situation with Iraq?" on Oct 6-8. 47% said they approved, 50% disapproved and 3% had no opinion.

(a) What is the 95% confidence interval for the proportion of people who approved of the President's handling of the situation in Iraq if 1300 people were surveyed?
Answer: First note that $\hat{p} = 0.47$ and

$$\text{SE}\left(\hat{p}\right) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.47(0.53)}{1300}} = 0.0138$$

For a 95% confidence interval $z^* = 1.96$ and so the interval should be

$$0.47 \pm 1.96(0.0138)$$

and so with 95% confidence we may say that p lies in $(0.443, 0.497)$.

Answer Aside: Note that typically Gallup determines the margin of error using the widest 95% CI possible for a given sample size (this corresponds to taking p=0.5). "For results based on this sample, one can say with 95% confidence that the maximum error attributable to sampling and other random effects is ±3 percentage points."

(b) Can you conclude that half of the US population approved of the President's handling of the situation in Iraq? explain why or why not.
Answer: Half the population agreeing would mean that $p = 0.5$. This value is not in our confidence interval in (a) therefore we cannot conclude that half the US population agreed with the Presidents handling of the situation in Iraq.

(c) Sometimes Gallup uses multiple versions of the same question with different wording. Explain why they might do this.
Answer: It is possible that different wordings can result in different responses. Obviously leading questions like "Given the Presidents request for $87 Billion dollars do you support his handling of the war in Iraq?" (which would likely decrease show a lower level of support for the President) and "The Iraqi people enjoy more freedoms today than they had before the recent action. Do you support the Presidents handling of the situation in Iraq?" (which would likely have higher support) can be immediately excluded as possibilities for clear bias. But it might be much harder to gauge the bias in "Do you agree or disagree of the way George

W. Bush is handling the situation with Iraq?" or "Do you approve or disapprove of the way the President is handling the situation with Iraq?". By asking multiple versions of the same question Gallup may see whether the differences in the wording of the question has an effect on the response.

(d) Gallup uses random telephone dialing to carry out it polling. A fixed number of phone numbers are chosen to be polled. Explain why the poll was carried out over three days rather than on a single day?
Answer: This is done to reduce the non-response rate (and any possible bias this might have). By picking a fixed set of numbers to be polled and then repeatedly trying to contact them until a response is received this bias is removed. For example, people who are less likely to be at home, such as young single adults, or people who spend a lot of time on the phone, would have a lower probability of falling into the sample than an adult American who was always at home and rarely talked on his or her phone if we just sampled on a single day.

Question 2. (*13 points*)

A clinical trial was conducted to examine the effectiveness of daily doses of aspirin in the treatment of strokes. Patients were randomized into treatment and control groups. Neither the physician nor the patient knew whether they were receiving the aspirin or a placebo tablet. After six months of treatment, the attending physicians evaluated each patient's progress as favorable or unfavorable. Of the 78 patients in the aspirin group, 63 had favorable outcomes. The control group of 77 patients had 43 patients with favorable outcomes. Let p_1 be the proportion of patients in the treatment group (aspirin) who had favorable outcomes after six months. Let p_2 be the proportion of patients in the control group (placebo) who had favorable outcomes after six months.

(a) What z^* should we use for a 98% confidence interval?
Answer: Remember that the standard normal distribution is symmetric around 0. we want z^* to be the number such that the area under the curve between $-z^*$ and z^* is 0.98. If 0.98 is in the middle of the graph then 0.02 must be split evenly into the two tails (ie 0.01 in the lower tail and 0.01) in the upper tail. This means that z* is given either from
$P(Z < z^*) = 0.99$ or $P(Z < -z^*) = 0.01$
looking in the table we find 0.99 corresponds to a z^* of 2.32

(b) Compute $\text{SE}\,(\hat{p}_1 - \hat{p}_2)$ and then the 98% confidence interval for the difference $p_1 - p_2$.
Answer: First, our two estimates of the sample proportions are $\hat{p}_1 = \frac{63}{78} = 0.8077$ and $\hat{p}_2 = \frac{43}{77} = 0.5584$. Remember that

$$\text{SE}\,(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

and so

$$\text{SE}\,(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\frac{63}{78}\left(1-\frac{63}{78}\right)}{78} + \frac{\frac{43}{77}\left(1-\frac{43}{77}\right)}{77}} = 0.0721$$

so the 98% confidence interval is given by

$$(0.8077 - 0.5584) \pm 2.32(0.0721)$$

which gives the interval $(0.082, 0.416)$.

(c) Would you conclude that using aspirin helps in the treatment of strokes?
Answer: Yes, Because 0 does not lie in the confidence interval in (b) we can assume that $p_1 - p_2$ is not 0 and therefore there is a difference between the treatment group and the control group. In addition because this interval is all positive we can conclude that the aspirin has had beneficial effect on stroke patients.

(d) Explain how blocking might be carried out in the context of this experiment. Give a brief description of how you might design this experiment (if it was to be carried out again).
Answer: In the context of this experiment we might block by using other variables which we can not control but might have an effect on the outcome. For example age, gender, history of stroke or heart attacks. In particular we would group the subjects based upon these variables into blocks. Within each block we would randomly assign the subjects to either receive the aspirin or the placebo. We could then make comparisons within each block of the difference between the treatment group and the control group. Blocking serves to reduce the variance.