1. No. The expected number of positives is 250, and the SE is $\sqrt{500} \times 0.5 \approx 11$. The observed number is 2.4 SEs above the expected. (Here, a one-sample test is appropriate.)

2. (a) The SE for the difference is 5.9%, so $z = 2.4/5.9 \approx 0.4$; looks like chance.

   (b) The SE for the difference is 1.8; the observed difference is 4.9; so $z = 4.9/1.8 \approx 2.7$ and $P \approx 0.3$ of 1%. The difference looks real.

   Comment. There is more information in the sample average than in the number of positive terms, at least for this example.

3. There are two samples, you need to make a two-sample z-test. Model: there are two boxes. The 2005 box has a ticket for each person in the population, marked 1 for those who would rate clergymen "very high or high," and 0 otherwise. The 2005 data are like 1000 draws from the 1985 box. The 2000 box is set up the same way. The null hypothesis says that the percentage of 1's in the 2005 box is the same as in the 2000 box. The alternative hypothesis says that the percentage of 1's in the 2005 box is smaller than the percentage of 1's in the 2000 box.

   The SD of the 2005 box is estimated from the data as $\sqrt{0.54 \times 0.46} \approx 0.50$. On this basis, the SE for the 2005 number is $\sqrt{1000} \times 0.50 \approx 16$: the number of respondents in the sample who rate clergymen "very high or high" is 540, and the chance error in that number is around 16. Convert the 15 to percent, relative to 1000. The SE for the 2005 percentage is estimated as 1.6%. Similarly, the SE for the 2000 percentage is about 1.5%.

   The SE for the difference is computed from the square root law (p. 502) as

   $$\sqrt{1.6^2 + 1.5^2} \approx 2.2\%.$$

   The observed difference is $54 - 60 = -6\%$. On the null hypothesis, the expected difference is 0%. So $z = (\text{obs} - \text{exp})/\text{SE} = -6/2.2 \approx -2.7$, and $P \approx 3/1000$. The difference is real. What the cause is, the test cannot say.

   Comment. Either a one-sided or a two-sided test can be used. Here, the distinction is not so relevant: for discussion, see chapter 29.

4. You need more information. The method of section 2 does not apply because you do not have two independent samples. The method of sections 3–4 does not apply because you observe two responses for each person. See p.517.

5. This is like the radiation-surgery example in section 4. Each subject has two possible responses, one to item A and one to item B. The investigators only observe one of the two, chosen at random. To make the test, pretend you have two independent random samples. With item A, the percentage who answer "yes" is 46%; the SE for this percentage is 3.5%. With item B, the percentage is 88% and the SE is 2.4%. The difference between the percentages is $46\% - 88\% = -42\%$. The SE for the difference is conservatively estimated as $\sqrt{3.5^2 + 2.4^2} \approx 4.2\%$. So $z = -42/4.2 = -10$. The framing of the question makes a difference. That is what the experiment tells you.

6. This is just like the previous exercise. In the calculator group, 7.2% get the right answer; in the pencil-and-paper group, 23.6%. The SEs are 1.6% and 2.7%. The difference between the percentages is -16.4%, and the SE for the difference is conservatively estimated as $\sqrt{1.6^2 + 2.7^2} \approx 3.1\%$. So $z = -16.4/3.1 \approx -5$, and $P \approx 0$. The difference is real. (Students who used the calculator seemed to forget what the arithmetic was all about.)

7. (a) This is like the radiation-surgery example in section 4. (Also see review exercises 5 and 6 above.) The SE for the treatment percent is 2.0%. The SE for the control pecent is 4.0%. The SE for the difference is 4.5%. The observed difference, 1.1%, is only 0.24 of an SE. This could easily be due to chance. Income support was a good idea that didn't work.

   (b) This is like example 4. The SE for the treatment average is 0.7 weeks; for the control average, 1.4 weeks; for the difference, 1.6 weeks. The observed difference is −7.5 weeks. So $z = -7.5/1.6 \approx 4.7$ and $P \approx 0$. The difference is real. Income support makes the released prisoners work less, which might explain the findings in part (a).

8. The data can be summarized as follows:

	Prediction Request	Request only
Predicts	22/46	NA
Agrees	14/46	2/46

   (a) This is like the radiation-surgery example in section 4. The two percentages are 47.8% and 4.4%. The SEs are about 7.4% and 3.0%, respectively. The difference is 43.4% and the SE for the difference is 8%. So $z = 43.4/8 \approx 5.4$ and $P \approx 0$. The difference is real. People overestimate their willingness to do volunteer work.

   (b) The two percentages are 30.4% and 4.4%. The SEs are about 6.8% and 3.0%, respectively. The difference is 26% and the SE for the difference is 7.4%. So $z = 26/7.4 = 3.5$ and $P \approx 2/10{,}000$. The difference is real. Asking people to predict their behavior changes what they will do.

   (c) Here, a two-sample z-test is not legitimate. There is only one sample, and two responses for each person in the sample. Both responses are observed, so the method of section 4 does not apply. The responses are correlated, so the method of example 3 does not apply. See p.517.

   Comment. In parts (a) and (b), the number of draws is small relative to the number of tickets in the box. So there is little difference between drawing with or without replacement, and little dependence between the treatment and control averages. See pp.510, 517.

9. This is like the radiation-surgery example in section 4. In the positive group, the percentage accepted is $28/53 \times 100\% \approx 52.8\%$; in the negative group, 14.8%. The SEs are 6.9% and 4.8%. The difference is 38% and the SE for the difference is 8.4%. So $z = 38/8.4 \approx 4.5$ and $P \approx 0\%$. There is a big difference between the two groups, and the difference cannot be explained by chance. Journals prefer the positive articles.

10. This test is not legitimate. There is dependence between the first-borns and second-borns.

11. (i) The expected value for the difference between the average score in the 2004 sample and the average score in the 1990 sample equals the expected value for the average score in the 2004 sample, minus the expected value for the average score in the 1990 sample. (ii) The expected value for the average score in the 2004 sample is the average of the 2004 box; likewise, the expected value for the average score in the 1990 sample is the average of the 1990 box. (iii) If you put the previous points together, the expected value for the difference between the averages of the two samples equals the difference between the averages of the two boxes. (iv) The null hypothesis says that the difference between the averages of the two boxes is 0. That is why 0 is the right benchmark in the numerator of the z-statistic.