Practice making Tests of Significance

1. According to one investigator's model, the data are like 400 draws made at random from a large box. The null hypothesis says that the average of the box equals 50; the alternative says that the average of the box is more than 50. In fact, the data averaged out to 52.75, and the SD was 25. Compute z and P. What do you conclude?

2. An ESP experiment asks the subject to guess unseen which target was chosen randomly from 10 targets by the researcher.

   (A subject without ESP guesses as if at random; a subject with ESP has a greater than random chance of guessing correctly.)

   Suppose that in 1,000 trials, a subject scores 173 correct guesses.

   (a) Set up the null hypothesis as a box model.
   (b) The SD of the box is ____. Fill in the blank, using one of the options below, and explain briefly.
   $$\sqrt{0.1 \times 0.9} \qquad \sqrt{0.173 \times 0.827}$$
   (c) Make the z-test.
   (d) What do you conclude?

3. A coin is tossed 10,000 times, and it lands heads 5,067 times. Is the chance of heads equal to 50%? Or are there too many heads for that?

   (a) Formulate the null and alternative hypotheses in terms of a box model.
   (b) Compute z and P.
   (c) What do you conclude?

4. One large course has 900 students, broken down into section meetings with 30 students each. The section meetings are led by teaching assistants. On the final, the class average is 63, and the SD is 20. However, in one section the average is only 55. The TA argues this way:

   > If you took 30 students at random from the class, there is a pretty good chance they would average below 55 on the final. That's what happened to me—chance variation.

   Is this a good defense? Answer yes or no, and explain briefly.

5. Seven hundred and fifty draws are made at random with replacement from a box of numbered tickets; 412 are positive. Someone tells you that 50% of the tickets in the box show positive numbers. Do you believe it? Answer yes or no, and explain. (The conclusion must be justified using the statistical methodolgy of this course.)

Answers follow on the next page.

Answers.

1. SE for average ≈ 1.25, so $z \approx (52.75 - 50)/1.25 \approx 2.2$ and P is approximately the area to the right of 2.2 under the normal curve. From the table, this is about 1.4%. The difference is hard to explain as chance variation. The alternative hypothesis is looking good. (This assumes that the line is drawn at 5%.)

2. (a) Null hypothesis: the number of correct guesses is like the sum of 1,000 draws from a box with one ticket marked 1 and nine 0's.
   (b) $\sqrt{0.1 \times 0.9}$. The null hypothesis tells you what is in the box. Use it.
   (c) $z \approx (173 - 100)/9.5 \approx 7.7$, and P is tiny. SE sum $= \sqrt{0.1 \times 0.9} \times \sqrt{1000} = 0.3 \times \sqrt{1000} \approx 9.5$.
   (d) Whatever it was, it wasn't chance variation.

3. (a) Tossing the coin is like drawing at random with replacement 10,000 times from a 0–1 box, with 0 = tails and 1 = heads. The fraction of 1's in the box is unknown.
   Null hypothesis: this fraction equals 1/2. Alternative: the fraction is bigger than 1/2.
   The number of heads is like the sum of the draws.
   (b) $z = 1.34$, $P \approx 9\%$.
   (c) The coin looks to be fair.

4. The TA's null hypothesis: the scores in his section are like 90 draws from a box containing all 900 scores. (There is little difference between drawing with or without replacement, because the box is so big.) The null hypothesis specifies the average and the SD of the box, 63 and 20. The EV for the average of the draws is 63, and the SE is 3.65. So $z = (\text{obs} - \text{exp})/\text{SE} = (55 - 63)/3.65 \approx -2.2$, and $P \approx 1\%$. The TA's defense is not good.

5. No. The expected number of positives is 375, and the SE is $\sqrt{750} \times 0.5 \approx 13.69$. The observed number is 2.67 SEs above the expected, when the observed number is corrected to 411.5. If, in actuality, 50% of the tickets showed positive numbers, the chances of a result this extreme or more would have been less than 2/5 of 1%. (Here a one-sample test is appropriate.)