Chapter 26. Tests of Significance

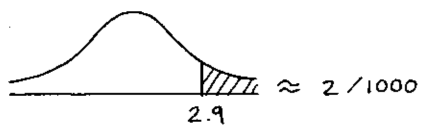
1. (a) True (p.479).

   (b) False. The null says it's chance, the alternative says it's real (pp.477–78).

2. (a) The data are like 3800 draws made at random with replacement from the box | ?? 0's ?? 1's |, with 1 = red.

   (b) Null: the fraction of 1's in the box is 18/38, or 47.4%.
   Alt: the fraction of 1's in the box is more than 18/38.

   (c) The expected number of reds (computed using the null) is 1800. The SD of the box (also computed using the null) is nearly 0.5, so the SE for the number of reds is $\sqrt{3800} \times 0.5 \approx 31$. So z = (obs − exp)/SE = (1890 − 1800)/31 ≈ 2.9, and $P \approx 2/1000$.



   (d) Yes.

Comments. (i) This problem is about the number of reds. In the formula for the z-statistic, *obs*, *exp*, and *SE* all refer to the number of reds. The expected, as always, is computed from the null. In this problem, the null gives the composition of the box, so the SD is computed from the null; it is not estimated from the data (p.485).

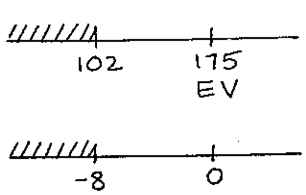
(ii) This problem, and several others below, can be done using one-sided or two-sided tests. The distinction does not matter here; it is discussed in chapter 29.

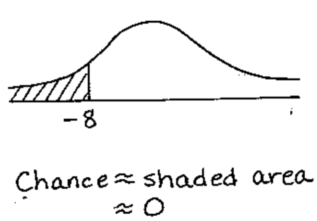
3. Null hypothesis: the data are like 200 draws from the box | 1 1 1 0 |, with 1 = blue and 0 = white. The expected number of blues is 150, and the SE is about 6, so z = (obs − exp)/SE = (142 − 150)/6 ≈ −1.3 and P ≈ 10%. This is marginal, could be chance.

4. The TA's null: the scores in his section are like 30 draws at random 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 = (obs − exp)/SE = (55 − 63)/3.65 ≈ −2.2, and $P \approx 1\%$. The TA's defense is not good.

5. The box has one ticket for each freshman at the university, showing how many hours per week that student spends at parties. So there are about 3000 tickets in the box. The data are like 100 draws from the box. The null hypothesis says that the average of the box is 7.5 hours. The alternative says that the average is less than 7.5 hours. The observed value for the sample average is 6.6 hours. The SD of the box is not known, but can be estimated from the data as 9 hours. On this basis, the SE for the sample average is estimated as 0.9 hours. Then z = (obs − exp)/SE ≈ (6.6 − 7.5)/0.9 = −1. The difference looks like chance.

6. (a) In the best case for Judge Ford, we are tossing a coin 350 times, and asking for the chance of getting 102 heads or fewer. The expected number of heads is 175 and the SE is a little over 9, so $z \approx (102 - 175)/9 \approx -8$. The chance is about 0.





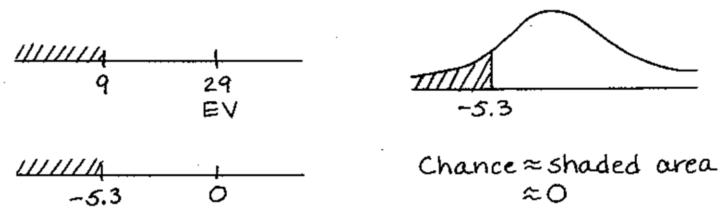
(b) 100 draws are made at random without replacement from a box with 102 1's and 248 0's, where 1 = woman and 0 = man. The expected number of 1's is 29. If the draws are made with replacement, the SE for the number of 1's is $\sqrt{100} \times \sqrt{0.29 \times 0.71} \approx 4.54$. The correction factor is

$$\sqrt{\frac{350 - 100}{350 - 1}} \approx 0.846.$$

The SE for the number, when drawing without replacement, is

$0.846 \times 4.54 \approx 3.8.$

So $z \approx (9 - 29)/3.8 \approx -5.3$. The chance is about 0.



(c) Judge Ford was not choosing at random; he was excluding women.

Comment. After Hans Zeisel made a statistical analysis of Judge Ford's procedures, the percentage of women jurors went up quite dramatically; see note 15 to chapter 26.

7. Disagree. There are 580 + 442 = 1022 subjects. With a coin, the expected number in the control group is 511, and the SE is about 16, so the chance of getting 442 or fewer in the control group is practically 0. Both patients and doctors know that if you turn up on an odd day of the month, you get the therapy. There may be a considerable temptation to enroll more patients on odd days, and that seems to be what happened.

   Comment. The danger is that the patients enrolled on the odd days will be different from the ones enrolled on the even days. For example, the doctors may tend to enroll relatively healthy patients—who need the therapy less—on the even days. If the control group starts off healthier than the treatment group, the study is biased against the treatment. Tossing a coin is wiser.

8. Model: there is one ticket in the box for each person in the county, age 18 and over. The ticket shows that person's educational level. The data are like 1000 draws from the box.

   Null: the average of the box is 13 years.
   Alt: the average of the box isn't 13 years.

   The expected value for the average of the draws is 13 years, based on the null. The SD of the box is unknown (there is no reason the spread in the county should equal the spread in the nation), but can be estimated as 5 years—the SD of the data. On this basis, the SE for the sample average is estimated as 0.16 years. The observed value for the sample average is 14 years, so

   $z = (\text{obs} - \text{exp})/\text{SE} = (14 - 13)/0.16 \approx 6,$

   and $P \approx 0$. This is probably a rich, suburban county, where the educational level would be higher than average.

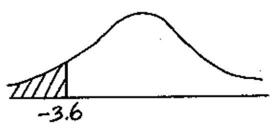
9. Something is wrong. The EV for the sum is 20; the SE for the sum is 4. The average of 144 sums should be around 20; the SE for the average is 0.33. The observed value is 3.4 SEs away from expected. That is too many SEs.

You can also get the EV and SE another way: the average of 144 sums is like the sum of $144 \times 100 = 14{,}400$ draws from the original box, divided by 144.

10. Null: the 3 Sunday numbers are like 3 draws made at random (without replacement) from a box containing all 25 numbers in the table. The average of these numbers is nearly 436, and their SD is just about 40. The EV for the average is 436, and the SE is 22; we are using the correction factor here. The 3 Sunday numbers average about 357, so

$$z = (\text{obs} - \text{exp})/\text{SE} = (357 - 436)/22 \approx -3.6,$$

and $P \approx 2/10{,}000$.



Comments. (i) Many deliveries are induced, and some are surgical, so obstetricians really can influence the timing.
(ii) In all, there are $25!/(3!\times 22!) = 2300$ samples of size 3. The 3 Sunday numbers are 344, 377, 351. The number for Saturday, August 6, is also 377; and all the other numbers in the table are larger. Therefore, exactly 2 samples have an average equal to the Sunday average, and no sample average is smaller. The exact significance probability is $2/2300 \approx 9/10{,}000$, compared to $2/10{,}000$ from the curve.

11. (a) The box has one ticket for each household in Atlanta in June 2003. If the income is over $52,000, the ticket is marked 1; otherwise, 0. The null hypothesis says that the percentage of 1's in the box is 50%; the alternative, that percentage of 1's in the box is bigger than 50%. The data are like 750 draws from the box.
    (b) The SE for the number of 1's in the sample is $\sqrt{750} \times \sqrt{0.50 \times 0.50} \approx 13.7$; the SE for the percentage of 1's in the sample is $13.7/750$, or 1.83%. So $z = 6/1.83 \approx 3.3$, and P is very small.
    (c) Median household income went up.

12. (a) There are 59 pairs, and in 52 of them, the treatment animal has a heavier cortex. On the null hypothesis, the expected number is $59 \times 0.5 = 29.5$ and the SE is $\sqrt{59} \times 0.5 \approx 3.84$. So 52 is nearly 6 SEs above average, and the chance is close to 0. Inference: treatment made the cortex weigh more.
    (b) The average is about 36 milligrams and the SD is about 31 milligrams. The SE for the average is 4 milligrams, so $z = 36/4 = 9$ and $P \approx 0$. (This is like the tax example in section 1.) Inference: treatment made the cortex weigh more.
    (c) This blinds the person doing the dissection to the treatment status of the animal. It is a good idea, because it prevents bias; otherwise, the technician might skew the results to favor the research hypothesis.