Part VII—Hypothesis Tests
Chapter 19—The One-Sample Z Test

We have competing hypotheses about the value of a population parameter. It’s impossible or impractical to examine the whole population to find out which hypothesis is true so we take a random sample and see which hypothesis better supported by our sample data.

In cases where it makes sense to formulate a hypothesis that assigns a particular value to the parameter we can do a hypothesis test.

**Hypothesis Test:**
1. Formulate a null hypothesis and an alternative hypothesis.

   **Null Hypothesis:** The population parameter is a particular value and any observed sample differences can be explained by chance variation.

   **Alternative Hypothesis:** There is some other reason besides chance that explains the sample data.

2. Assume the null to be true and set up a box model based on the null.

3. Think about what you would expect to get if you randomly sampled from the null box. (In other words, what is the sampling distribution under the null hypothesis).

4. Check how likely it would have been to get our data or something even farther from the null, if the null were right. That’s called the p-value.

5. If p is very low, we say that the data support rejecting the null hypothesis.

How low is “very low”? That depends on how strong the prior evidence was that supported the null. If you’re testing whether or not a coin is fair and you have a coin that looks and feels perfectly normal, it takes some very weird, very unlikely results to make you believe it’s not a fair coin. If your coin has a strange rattle to it, you may not require such strong evidence.

The convention is to reject the null when $p < 0.05$ and call the result “significant”. When $p < 0.01$ the result is called “highly significant”. There’s no particular justification for those values but they’re very commonly used.

To try to decide how much you still believe the null hypothesis, you have to combine the data with what you knew ahead of time, just as we did in trying to decide how likely a positive ELISA result was to mean you have AIDS.
Example 1: A large lecture has 1000 students. On a midterm exam, the class average was 70 with a SD of 12. It is suggested that the early morning section has more serious students, and would therefore have higher scores. This section has 36 students and their average was 75. The question is whether this 5-point difference is real or whether it may simply be due to chance variation.

Argue by contradiction: Suppose it is simply due to chance, how likely would it be to get such a big difference?

Step 1: Formulate a null hypothesis and an alternative hypothesis
- Null hypothesis: Our observed 5-point difference is only due to chance variation, there's no particular cause for it.
- Alternative hypothesis: The difference is too large to be explained by chance; it must be due to some other cause.

Step 2: Set up a box model under the null hypothesis
Imagine we choose 36 tickets at random without replacement from a box containing 1000 tickets. Each ticket shows an exam score. The average of the box is 70, and the SD is 12.

1,000 tickets on each is written an exam score

1000 tickets or each is written an exam score

Step 3: Compute a test statistic based on the null hypothesis box model
How likely is it to draw 36 tickets at random from the null box and get such a big average? In other words, how big a difference is it in terms of SEs? Compute the Z-statistic:

\[ Z = \frac{\text{Observed mean} - \text{Expected mean}}{\text{SE}} \]

\[ Z = \frac{75 - 70}{2} = \frac{5}{2} = 2.5 \]

Step 4: Use the normal curve to find the P-value, (the observed significance level). Draw the normal curve and indicate P by shading.

Step 5: Conclusion. Small values of P are evidence against the null hypothesis; they indicate something besides chance was going on. So we reject the null when P is small. How small does P have to be to reject the null?

It's arbitrarily, there's no justification for these cut-offs, but the convention is the following:
- If P < 5%, the result is called statically significant.
- If P < 1%, the result is called highly statistically significant.

If the null was true, we'd only see this big a difference 0.61% of the time, so we conclude something besides chance is going on.
The z-test can also be used when the situation involves classifying and counting. In such cases, the box model contains zeros and ones and the SD is easily computed using the formula:

Example 2: An experiment on ESP was done at UC Davis to determine whether people thought to be clairvoyant really had ESP. A machine called the "Aquarius" used a random number generator to pick one of 4 targets. The subjects tried to guess which target the machine chose. There were 15 "clairvoyants", each of whom made 500 guesses for a total of 15 x 500 = 7,500 guesses. Out of these 2,006 were right. If the subjects had no ESP they'd be right 1/4 of the time, which would be 1/4 (7,500) = 1,875. Can the 2,006-1875 = 131 correct guesses be explained by chance variation?

a) Formulate the null hypothesis in terms of a box model.

Null: The people were just guessing. The 131 extra correct is just due to the luck of the draw.

Alternative: The 131 extra is too big to be due to chance. They can't just be guessing.

b) Compute z and P.

\[
Z = \frac{\text{Obs} - \text{Exp}}{\text{SE}_{\text{Exp}}} = \frac{2,006 - 1,875}{37} = \frac{131}{37} = 3.5
\]

131 correct huge!

What do you conclude?

P is 0.0235%. reject null something besides chance, not necessarily ESP.

P = 100 - 99.953% = 0.047%

P = 0.0235%
Example 3: In 1965, the U.S. Supreme Court decided the case of Swain v. Alabama. Swain, an African American man, was convicted in Talladega County, Alabama of raping a white woman. He was sentenced to death. The case was appealed to the Supreme Court on the grounds that there were no African Americans on the jury, even though about 26% of the adult men in Talladega County were African American.

The Supreme Court denied the appeal, on the following grounds. The jury was selected from a panel of about 100 randomly selected people, 8 of whom were African American. (They didn't serve on the 12-person jury because they were "struck", by preemptory challenges by the prosecution. Such challenges are constitutionally protected.) The presence of 8 African Americans on the panel showed "The overall percentage disparity (between 26% and 8%) is small and reflects no attempt to include or exclude a specified number of Negroes." The Supreme Court said the "small" difference could simply be due to chance.

c) Formulate the null hypothesis in terms of a box model.

[Diagram of box model with proportions and calculations]

Alt: Diff. is too big to be due to chance (Jury was not randomly selected)

b) Compute $z$ and $P$.

$z = \frac{\text{obs} - \text{exp}}{\text{SE} \%} = \frac{8\% - 26\%}{4.4\%} = -18\%$

$z = -4.1$

$P = 0.002\%$

z-score is in the critical region.

Reject null hypothesis that jury was not selected randomly.
Example 4
Suppose a large University claims that the average ACT score of their incoming freshman class is 30, but we think the University may be inflating their average. To test the University's claim, we take a simple random sample of 50 students and find their average to be only 29.1 with a SD of 4.5.

a) Formulate the null and alternative hypothesis in terms of a box model.

Null: The ave of the box (all freshmen) is 30, and the dif (29.1 - 30) is small just due to chance.
Alt: The ave in the box (all freshmen) is < 30, that's why we got such a low ave (29.1) in our sample.

b) Compute z and P.

\[ z = \frac{\text{obs} - \text{exp}}{\text{SE}_{\text{obs}}} = \frac{29.1 - 30}{0.636} = -1.42 \]

SE_{obs} = \frac{\text{SD}}{\sqrt{n}} = \frac{4.5}{\sqrt{50}} = 0.636

\[ P = 8.4\% \]

c) What do you conclude?

Cannot reject null (b/c P > 5%), conclude that it's plausible that University's claim is correct that one of all freshmen = 30
Chapter 19—The One-Sample Z Test

Summary:
• All tests of significance are based on box models.

• We assume the null hypothesis—that the difference between the observed and expected is just due to chance

• We compute z (or t) and P

• $Z$ tells us how many SE's the result is from expected and is computed by $z = (\text{observed-expected})/\text{SE}$. If the null hypothesis tells you the SD of the box, use it in computing the SD. Otherwise you have to estimate the SD from the sample.

• $P$ tells us how likely it is to get a result as extreme as or more extreme than the observed. The chance is computed assuming the null hypothesis to be true. Therefore it does NOT give us the chance that the null hypothesis is true.

• Small values of $P$ are evidence against the null. They indicate something else besides chance is causing the difference.