Z-test Worksheet

October 2026

1 Earthquakes

We will be talking about earthquakes with seismic events of MB > 4.0 (MB is body-wave magnitude, a measure of size of an earthquake). The events occurred in a cube near Fiji since 1964. The 200 observations we are working with are a random sample of all earthquakes (with MB > 4.0) in that area with an average MB of 4.62. Let's assume the researcher who collected this data is trying to show that, if an earthquake in and around Fiji has an MB > 4, then the earthquakes have an average magnitude greater than 4.5.

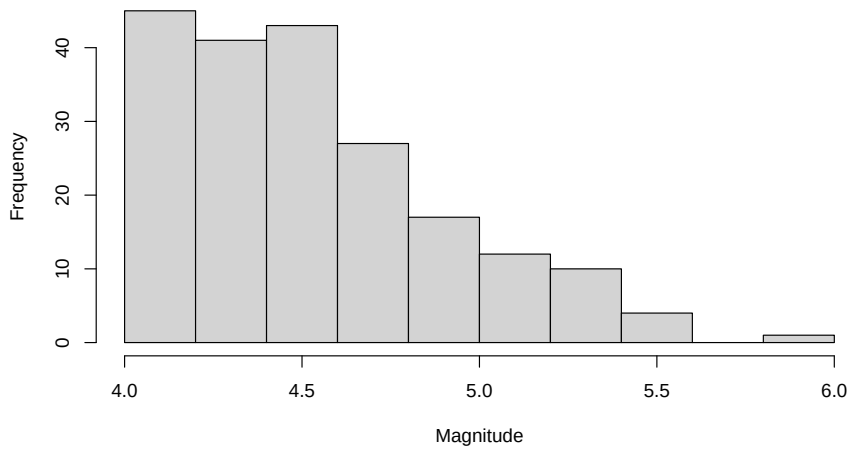
1. Please state the null and alternative hypothesis. Note here we are using averages and not proportions.

   H_0: $\mu \leq 4.5$

   H_A: $\mu > 4.5$

2. Below is the histogram of our sampled earthquakes. On the side, please describe the histogram and include the number of modes, if it's symmetric or skewed (or neither), and if their our outliers.

   Unimodal, one outlier, skewed right



We will now need to check the assumptions....

3. First, was the sample randomly collected?

   Yes, it says a random sample

4. Next up is the population normally distributed or is the sample size large assumption

1. Is the population normally distributed? How did you come to this conclusion?

   No, the sample distribution (which should look like the population distribution) is not bell-shaped so we think the population is not bell-shaped

2. Is the sample size sufficiently large?

   Yes, it is. 200 observations is more than enough for the Central Limit Theorem to kick in

3. Finally, is the sample independent and identically distributed assumption

   1. Is there a reason to believe the earthquakes would form subpopulations? Yes actually. There are three differenet types of earthquakes (as per the resident seismologist on campus) which behave slightly differently. More so, Fiji is actually in the middle of a hotbed of activity that has all three earthquakes.

   2. Is there a reeason to believe that one earthquake will affet another earthquake? Yes, again as per the professor on campus. Earthquakes come in both spatial and temporal groups where one affects the others.

      SIDE NOTE: This is what Box was talking about where it's critical for a statistician to know the science behind the data. I couldn't have answered those questions that competently without a researchers input. We, statisticians, are support players but to support others we need to have a basic idea of what is happening. Become statisticians and you will learn a little about a lot of things.

3. Please write down the sampling distribution for the sample mean of the earthquake's magnitude for this Z-test. We believe that the variance of the population, $\sigma^2 = .25$ and that we have a sample size of 200.

   Assuming we passed our assumptions, and noting that our null hypothesis is mu is less than OR EQUAL to 4.5, our sample mean is distrbuted as....

$$\bar{X} \sim \ N(\mu, \sigma^2/n)$$

   which can be rewrote as....

$$\bar{X} \sim \ N(4.5, .25/200)$$

4. Please standardize the sample mean we observed using the distribution you found in question 4, which will produce the test statistic. Hint: The formula is $\frac{observed\ -\ mean}{st.dev.}$

   The mean and standard deviation I get from the above normal distribution. Note our st. dev. here will be the square root of our variance from above so

$$st.\ dev. \ = \ \sqrt{.25/200} \ = \ .0354$$

   The "observed" value is the sample mean we actually saw (4.62). So our test statistic is....

$$\frac{4.62 - 4.5}{.0354} = 3.39$$

5. Please find the probability associated with your answer to question 5. Hint: use the normal calculator online again.

   The p-value is approximately 0. My machine has it at .00034 so about 3 in every 10,000 samples taken, if the null hypothesis is true, will have a sample mean as or more high than 4.62

6. Write down your decision for the test. Please remember to state the strength of evidence. Reject/Fail to reject is wrong so do not write it (see notes for a discussion).

   Our p-value is low. As such, we have very strong evidence that the true mean MB of the earthquakes is above 4.5 MB.

2 Temp.

One of my best friends is from New England and has a theory: Fahrenheit is superior to Celcius because it has a natural interpretation. 0 degrees is about as cold as it gets and 100 is about as hot as it gets in both the Midwest and New England which means you can think of it as "percent hot". Let's take his claim one step further: If Fahrenheit is percent hot, then the average annual temp is 50 degrees. Let's test this and see if we have evidence that it is different from 50.

To do this we will use data collected, every year, in New Haven, Conn. from 1912 to 1971, for a total of 60 observations and it is known that the population variance is 8 (the units are degrees F^2 but don't worry too much about that). The sample mean from this data was 51.16 degrees F.

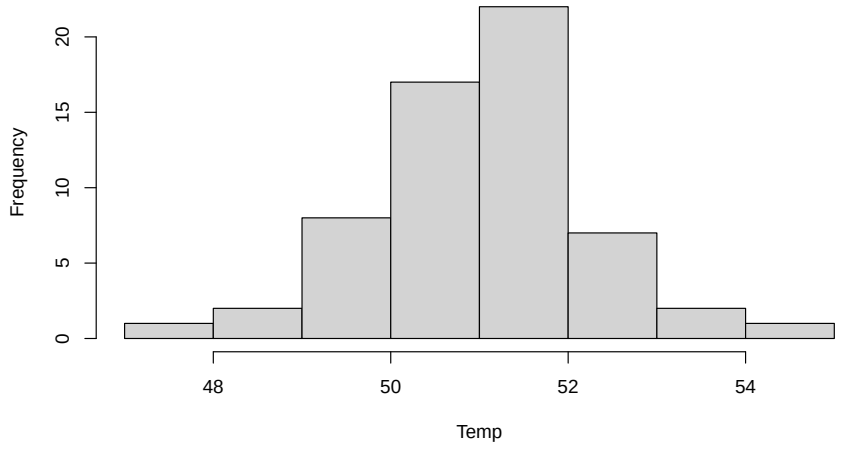
1. Please write down the null and alternative hypothesis.

   H_0: $\mu = 50$

   H_A: $\mu \neq 50$

2. Below is the histogram of temps recorded. Please, off on the side, describe the number of modes, if it's symmetric or skewed, and if there are outliers or not.

   Unimodal, no outliers, symmetrical-ish. This is probably a normal distribution



3. Was the data randomly collected?

   No, it was taken every single year at the exact same location.

4. Is the population normally distributed OR is the sample size large? How did you come to this conclusion?

   Yes, the sample is normally distributed so we expect the population to be normally distributed (the pop should look like the sample, the sample is usually all we actually know). We don't need to invoke the

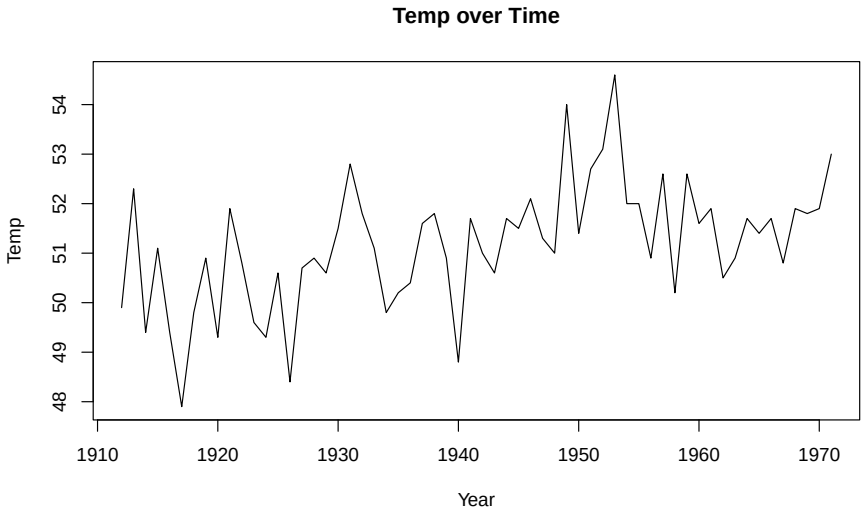
central limit theorem (large n) but we can and could pass this assumption that way (symmetric without outliers need only about 20 observations).

5. Is the data independent and identically distributed? HINT: use the graph below which shows the average yearly temp over time to answer this one.

   No, it's not IID and in fact fails both conditions.

   It is not independent because there is a correlation over time (eg if the last five years have all been hot we expect his year to be hot).

   It also lacks identically distributed. The reason is that, on the graph, the early years are mostly below (temp wise) the later years (which average a higher temp). So depending on where we are at on the graph I would either guess higher or lower. Big picture then is that they are not identically distributed.



6. Let's pretend we pass all of our assumptions. Please state the sampling distribution for the sample mean for average temp in New Haven.

$$\bar{X} \sim \ N(\mu, \sigma^2/n)$$

   which can be rewrote as....

$$\bar{X} \sim \ N(50, 8/60)$$

   where 8 is the population variance, 60 is the sample size, and 50 is the hypothesized mean.

7. Please standardize the sample mean we observed using the distribution you found in question 6, which will produce the test statistic. Hint: The formula is $\frac{observed\ -\ mean}{st.dev.}$

$$\frac{51.16\ -\ 50}{\sqrt{(8/60)}}\ =\ 3.177$$

8. Please find the probability associated with your answer to question 7. Hint: use the normal calculator online again.

   My computer is returning a p-value of .00149. Note that we want the two tailed option.

9. Write down your decision for the test. Please remember to state the strength of evidence. Reject/Fail to reject is wrong so do not write it (see notes for a discussion). Bonus question: Does the difference matter?

   Our p-value is considered "strong" evidence so.....We have strong evidence the mean temperature in New England and the Upper Midwest (cough New Haven Conn cough) is different from 50 degrees F.

   No, the difference really doesn't matter. Most humans can't tell the difference between two temps that are off by only a degree and some change. This is a moment where the ability to detect a difference doesn't mean the difference is relevant.

   Side Note 1: This is a time where what we sampled (temps in New Haven) and what we want to talk about (New England and the upper Midwest) don't align perfectly. Our results aren't generalizable to that large of a geographic area. Instead, if we only talk about New Haven our results and discussion become more serious and relevant. From the sampling notes slide deck we would say there is severe "undercoverage"

   Side Note 2: The sample I knew was going to be normally distributed before I ever looked at it, why? Very good! The data is actually the average yearly temp in New Haven. So each data point is already an average of 365 days which is a large enough number for the central limit theorem to kick in even if they lack IID.