1.3 Hypothesis Testing

Looking at the data plotted in Figure 1.1, the annual mean temperature of Atlanta appears to be increasing in the recent decades following 1980s. However, there is also significant year-to-year variations so the temperature is not simply getting warmer each year. *Hypothesis Testing* allows us to make an objective statement about the data.

Is Atlanta getting warmer?

First I want to compare the average temperature in 2000s to the long-term average. Averaging the temperatures for year 2000 to 2009 ($N = 10$), we get 62.64 degrees. As we calculated before, the long-term mean temperature is 61.82 degrees. So there is a difference of +0.82 degrees between the two averages. The real question is; Is there a big enough of a difference to say that the Atlanta is significantly warmer in 2000s than the long-term average?

The temperature randomly varies year to year, so we must define the confidence level for what we mean by "significantly" warmer. For example, 95% confidence level means that we allow one in twenty chance (5%) that we reject the null hypothesis wrongly. Ok, then what is the hypothesis? Formally we formulate the null hypothesis (**H0**) and then the alternative (**H1**) to the null hypothesis.

H0: The average temperature from 2000-2009 is NOT significantly warmer than the long-term average.

H1: The average temperature from 2000-2009 is significantly warmer than the long-term average.

We then perform a test to see if the data can accept or reject the null hypothesis (**H0**) with the stated confidence level. In this particular example, the problem is set up as an one-tail test. In one tail test, we are considering the possibility of the data in the one side (tail) of the distribution only. We already know the 2000-2009 is somewhat warmer than the long term mean, so we don't have to consider the possibility of Atlanta getting cooler.

Alternatively, we can also formulate the problem as a two-tail test as follows.

H0: The average temperature from 2000-2009 is NOT significantly different from the long-term average.

H1: The average temperature from 2000-2009 is significantly different from the long-term average.

In this case, we do not discriminate whether 2000-2009 is warmer or cooler than the long term mean. We are simply trying to determine whether the 2000-2009 is significantly different, either warmer or cooler, than the long term mean. Then, the problem can be treated as the two tail test.

In order to test these hypothesis, we need to know what is the expected distribution of the average temperature from 2000-2009, and evaluate whether a difference of +0.82 degrees lies within the envelope of the confidence interval.

Student's t-distribution

We are dealing with a small sample (N=10) and we cannot use the normal distribution for this small sample. This generally applies to the sample number of $N < 30$. In this case we use the Student's t-distribution to calculate the expected distribution for the mean of the small sample. The Student's t-distribution depends on the degree of freedom (N-1), and several examples are plotted in Figure 1.10. In MATLAB, there is a function (tpdf) that calculates values of the probability density function for Student's t-distribution. The degree of freedom (d.f.) is set to N-1, so in our case it is 9.

The 95% confidence interval is defined as the range of t-values that are consistent with the null hypothesis with a 0.95 probability. In another words, we are allowing ourselves a 5% probability that the random sampling causes the t-value to be outside of this range, thus rejecting the null hypothesis wrongly. If the t-value of the data is in the confidence interval, we cannot reject the null hypothesis.

To calculate the t-statistic of the data, we first calculate the sample mean ($\overline{x}$) and standard deviation (s) of the small sample. The t-statistic can then be calculated

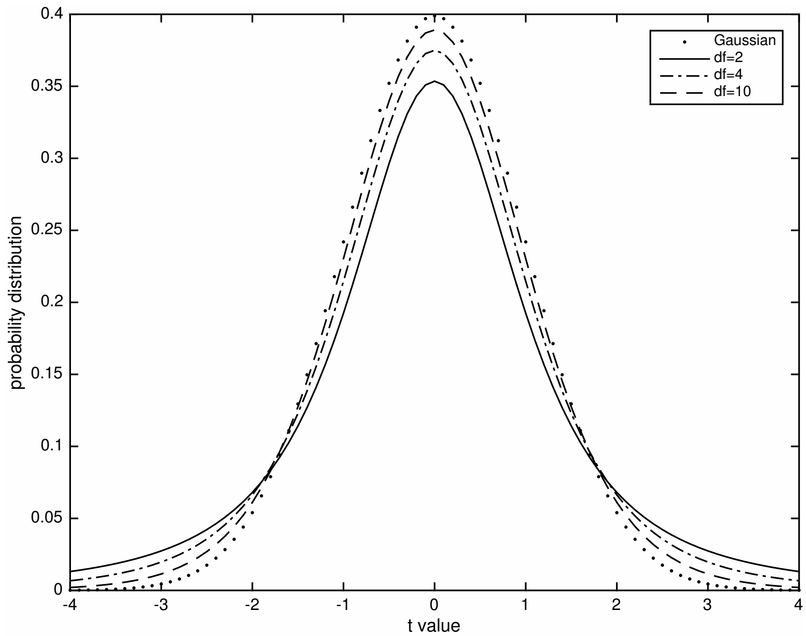

Figure 1.10: Student's t-distribution for a few different degree of freedom (d.f.). As the d.f. increases, the distribution becomes closer to Gaussian (dot). As the d.f. decreases, the distribution becomes flatter.

as

$$t = \frac{\overline{x} - \mu}{s/\sqrt{N-1}}. \tag{1.17}$$

Let's observe the form of the t-statistic. It is proportional to the difference between the sample and true mean. It is also inversely proportional to the ratio between the sample standard deviation and the square root of the sample size minus 1, which measures the standard error of the mean. The standard error of the mean is proportional to the sample standard deviation, and it shrinks according to the square root of the sample size. Thus, the value of t can be large if the sample and true mean are different, N is large and s is small. For the case of Atlanta's temperature from 2000 to 2009, we have $(\overline{x} - \mu)$ is +0.82 degrees, and its standard deviation is 0.772 degrees. So the value of t statistic is $t = (+0.82)/(0.772/3) = 3.187$.

The confidence interval depends on whether we are performing one-tail or two-tail test. Consider the null hypothesis that the mean temperature from 2000s is NOT significantly warmer than the long-term mean. In this case, we apply the one-tail test as shown in Figure 1.11a. The 95% confidence interval covers the region up to the 95 percentile of the Student's t-distribution. We can calculate the critical value (t_{crit}) for the envelope of the 95% confidence level.

$$0.95 = CDF(t_{crit}, d.f. = 9) \tag{1.18}$$

In MATLAB, there is an inverse function for the cumulative distribution function (tinv).

```
>> tcrit = tinv(.95,9);
```

and we get $t_{crit} = 1.8331$. If the t-value of the data falls into the region of $t < t_{crit}$, we cannot reject the null hypothesis. Since the t-value of the data is 3.187, we

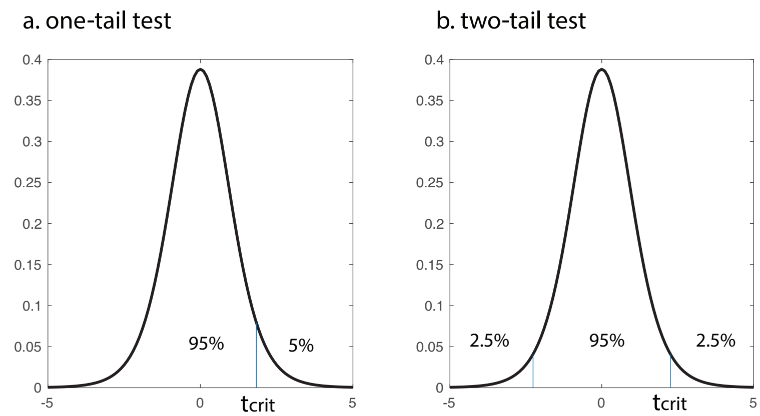

Figure 1.11: Critical interval and the critical t-values of (a) one-tail and (b) two-tail test with the d.f. = 9.

can indeed reject the null hypothesis and confirm that the 2000-2009 is significantly warmer than the long term mean.

Alternatively, consider the null hypothesis that the mean temperature from 2000s is NOT significantly different from the long term mean. This is a two-tail test, and the 95% confidence interval covers the region from 2.5 percentile to 97.5 percentile of the Student's t-distribution as shown in Figure 1.11b. Using the inverse function for the cumulative distribution function (tinv), we get,

```
>> tcrit = tinv(.975,9);
```

and we get $t_{crit} = 2.262$. Since the t-value of the data is

3.187, we can also reject the null hypothesis and confirm that the 2000-2009 is significantly different from the long term mean.

Comparing the two cases, the one tail test has a smaller critical t-value. This is because the one-tail test neglects the possibility that temperature can be lower than the long term mean. For the case of 2000s temperature, it is acceptable to make this assumption.

Is Atlanta getting cooler?

Let's consider another example. I want to compare the average temperature in 2000s to an earlier period, say 1990s. Averaging the temperatures for year 2000 to 2009, we get 62.64 degrees. Doing the same for the period of 1990 to 1999, we get 63.14 degrees. From 1990s to 2000s, it appears as if Atlanta has cooled down by 0.5 degrees, but is this a big enough of difference to say that the temperature in 2000s are significantly cooler than 1990s? Let's test the following hypothesis.

H0: The average temperatures from the two periods (1990s and 2000s) are not significantly different from one another.

H1: The average temperatures from the two periods are significantly different.

In this case, the relevant t-statistic is

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \tag{1.19}$$

where σ is defined as $\sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}$. Based on the data, the t-statistic is -1.13.

In this case, the d.f. is $N_1 + N_2 - 2 = 18$. Then we can calculate the critical value for the t-statistic. Using the tinv function of MATLAB, t_{crit} is 2.10. In this case, the t-statistic is within the envelope of the confidence level. Thus we cannot reject the null hypothesis. We conclude that the mean temperature from 2000s is not significantly different from the 1990s.

Alternatively, we can examine the one-tail test as follows.

H0: The average temperatures from 2000s is NOT significantly cooler than that of 1990s.

H1: The average temperatures from 2000s is significantly cooler than that of 1990s.

Then we can calculate the critical value for the t-statistic. Using the tinv function of MATLAB, t_{crit} is 1.73. Again, the t-statistic is within the envelope of the confidence level. Thus we cannot reject the null hypothesis. We conclude that the mean temperature from 2000s is not significantly cooler than the 1990s.

In summary, we generally take five steps to perform the hypothesis testing.

1. State the confidence level.
2. Define your null hypothesis and its alternative hypothesis.
3. State the statistic used.
4. Determine the critical region.
5. Evaluate whether or not the data is within or outside of the critical region.

Exercises

1. Repeat the two examples discussed in this section with the confidence level of 99%. Make sure you write down each of the five steps in the hypothesis testing.
2. Calculate the annual mean values of surface air temperature based on the NCEP reanalysis (air.mon.mean.nc). Choose a specific location and make a time-series plot.
3. At the chosen location, perform the t-test to determine whether the temperature during 2000-2015 is significantly warmer than the long-term average temperature. Write down each of the five steps.
4. **HW3** Publish the MATLAB script that performs all of the activity above, and submit it as a report in the PDF format.