Hypothesis Testing: Examples

Learning Objective: Explain the logic behind and the process of hypotheses testing. In particular, explain what the p-value is and how it is used to draw conclusions.

Example: 2

A certain prescription allergy medicine is supposed to contain an average of 245 parts per million (ppm) of a certain chemical. If the concentration is higher than 245 ppm, the drug will likely cause unpleasant side effects, and if the concentration is below 245 ppm, the drug may be ineffective. The manufacturer wants to check whether the mean concentration in a large shipment is the required 245 ppm or not. To this end, a random sample of 64 portions from the large shipment is tested, and it is found that the sample mean concentration is 250 ppm with a sample standard deviation of 12 ppm. Let's analyze this example according to the four steps of hypotheses testing we outlined on the previous page:

1. **Stating the claims:**

   - **Claim 1:** The mean concentration in the shipment is the required 245 ppm.
   - **Claim 2:** The mean concentration in the shipment is not the required 245 ppm.

   Note that again, claim 1 basically says: "There is nothing unusual about this shipment, the mean concentration is the required 245 ppm." This claim is challenged by the manufacturer, who wants to check whether that is, indeed, the case or not.

2. **Choosing a sample and collecting data:**

A sample of n = 64 portions is chosen and after summarizing the data it is found that the sample concentration is $\overline{x} = 250$ and the sample standard deviation is s = 12.

Is the fact that $\overline{x} = 250$ is different from 245 strong enough evidence to reject claim 1 and conclude that the mean concentration in the whole shipment is not the required 245? In other words, do the data provide strong enough evidence to reject claim 1?

3. **Assessing the evidence:**

   In order to assess whether the data provide strong enough evidence against claim 1, we need to ask ourselves the following question: If the mean concentration in the whole shipment were really the required 245 ppm (i.e., if claim 1 were true), how surprising would it be to observe a sample of 64 portions where the sample mean concentration is off by 5 ppm or more (as we did)? It turns out that it would be extremely unlikely to get such a result if the mean concentration were really the required 245. There is only a probability of 0.0007 (i.e., 7 in 10,000) of that happening. (Do not worry about how this was calculated at this point.)

4. **Making conclusions:**

   Here, it is pretty clear that a sample like the one we observed is extremely rare (or extremely unlikely) if the mean concentration in the shipment were really the required 245 ppm. The fact that we **did** observe such a sample therefore provides strong evidence against claim 1, so we reject it and conclude with very little doubt that the mean concentration in the shipment is not the required 245 ppm.

Do you think that you're getting it? Let's make sure, and look at another example.

Example: 3

Is there a relationship between gender and combined scores (Math + Verbal) on the SAT exam?

Following a report on the College Board website, which showed that in 2003, males scored generally higher than females on the SAT exam (http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2003/pdf/2003CBSVM.pdf), an educational researcher wanted to check whether this was also the case in her school district. The researcher chose random samples of 150 males and 150 females from her school district, collected data on their SAT performance and found the following:

Males

n	mean	standard deviation
150	1025	212

Females

n	mean	standard deviation
150	1010	206

Again, let's see how the process of hypothesis testing works for this example:

1. **Stating the claims:**

    - **Claim 1:** Performance on the SAT is not related to gender (males and females score the same).

    - **Claim 2:** Performance on the SAT is related to gender - males score higher.

    Note that again, claim 1 basically says: "There is nothing going on between the variables SAT and gender." Claim 2 represents what the researcher wants to check, or suspects might actually be the case.

2. **Choosing a sample and collecting data:**

    Data were collected and summarized as given above.

    Is the fact that the sample mean score of males (1,025) is higher than the sample mean score of females (1,010) by 15 points strong enough information to reject claim 1 and conclude that in this researcher's school district, males score higher on the SAT than females?

3. **Assessment of evidence:**

    In order to assess whether the data provide strong enough evidence against claim 1, we need to ask ourselves: If SAT scores are in fact not related to gender (claim 1 is true), how likely is it to get data like the data we observed, in which the difference between the males' average and females' average score is as high as 15 points or higher? It turns out that the probability of observing such a sample result if SAT score is not related to gender is approximately 0.29 (Again, do not worry about how this was calculated at this point).

4. **Conclusion:**

    Here, we have an example where observing a sample like the one we observed is definitely not surprising (roughly 30% chance) if claim 1 were true (i.e., if indeed there is no difference in SAT scores between males and females). We therefore conclude that our data does not provide enough evidence for rejecting claim 1.

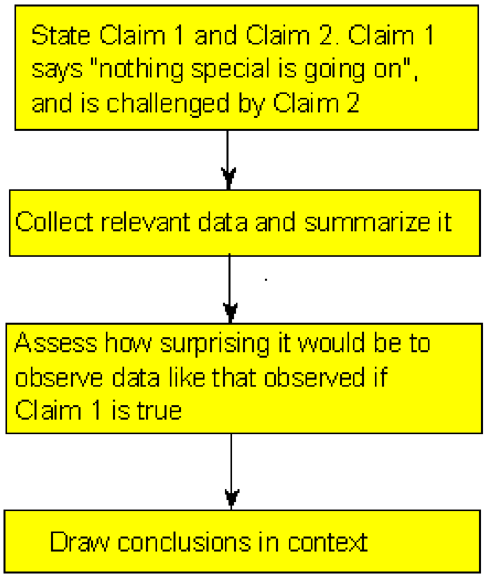
Comment

Go back and read the conclusion sections of the three examples, and pay attention to the wording. Note that there are two type of conclusions:

- "The data provide enough evidence to reject claim 1 and accept claim 2"; or
- "The data do not provide enough evidence to reject claim 1."

In particular, note that in the second type of conclusion **we did not say:** "I accept claim 1," but only "**I don't have enough evidence to reject claim 1**." We will come back to this issue later, but this is a good place to make you aware of this subtle difference.

Hopefully by now, you understand the logic behind the statistical hypothesis testing process. Here is a summary:



Learn By Doing

1/1 point (graded)

For many years "working full-time" has meant 40 hours per week. Nowadays it seems that corporate employers expect their employees to work more than this amount. A researcher decides to investigate this hypothesis.

Claim 1: The average time full-time corporate employees work per week is 40 hours.

Claim 2: The average time full-time corporate employees work per week is more than 40 hours.

To substantiate his claim, the researcher randomly selects 250 corporate employees and finds that they work an average of 47 hours per week with a standard deviation of 3.2 hours.

In order to assess the evidence, we need to ask:

- how likely it is in a sample of 250 we will find that the mean number of hours per week corporate employees work is as high as 47.
- how likely it is that the true mean number of hours per week corporate employees work is 40.
- how likely it is that the true mean number of hours per week corporate employees work is more than 40.
- how likely it is that in a sample of 250 we will find that the mean number of hours per week corporate employees work is as high as 47 if the true mean is 40. ✔

Answer
Correct:
Indeed, in hypothesis testing, in order to assess the evidence, we need to find how likely is it to get data like those observed assuming that claim 1 is true.

Submit

Learn By Doing

1/1 point (graded)
According to the Center for Disease Control (CDC), roughly 21.5% of all high-school seniors in the United States. have used marijuana. (Comments: The data were collected in 2002. The figure represents those who smoked during the month prior to the survey, so the actual figure might be higher). A sociologist suspects that the rate among African-American high school seniors is lower, and wants to check that. In this case, then,

> **Claim 1:** The rate of African-American high-school seniors who have used marijuana is 21.5% (same as the overall rate of seniors).
>
> **Claim 2:** The rate of African-American high-school seniors who have used marijuana is lower than 21.5%.

To check his claim, the sociologist chooses a random sample of 375 African-American high school seniors, and finds that 16.5% of them have used marijuana.

In order to assess this evidence, we need to find:

- how likely it is that the true rate (i.e., the rate among all African-American high-school seniors) is 21.5%.

- how likely it is that the true rate is lower than 21.5%.
- how likely it is that in a sample of 375 we'll find that as low as 16.5% have used marijuana.
- how likely it is that in a sample of 375 we'll find that as low as 16.5% have used marijuana, when the true rate is actually 21.5%. ✔

Answer
Correct:
Indeed, in hypothesis testing, in order to assess the evidence we need to find how likely it is to get data like those observed assuming that claim 1 is true?

Submit

Did I Get This

1/1 point (graded)
The most commonly accepted tradition is that college students will study 2 hours outside of class for every hour in class. This means 30 hours/week for a full-time student taking 15 units (hours of class). An educator suspects that this figure is different now than in the past.

> **Claim 1:** The average time full-time college students study outside of class per week is 30 hours.
>
> **Claim 2:** The average time full-time college students study outside of class per week is not 30 hours.

To substantiate her claim, the educator randomly selects 1,500 college students and finds that they study an average of 27 hours per week with a standard deviation of 1.7 hours.

In order to assess the evidence, we need to determine:

- how likely it is to observe a mean number of hours of studying outside of class per week that is different from 30 hours per week.
- how likely it is in a random sample of 1,500 students to observe that the average number of hours spent per week studying outside of class is at most 27 hours, or at least 33 hours.
- how likely it is in a random sample of 1,500 students to observe that the mean amount of hours of studying outside of class per week is 27 hours or less.

- [x] how likely it is in a random sample of 1,500 students to observe students studying an average of at most 27 or at least 33 hours per week outside of class, if the mean number is actually 30 hours per week. ✔

- [ ] how likely it is that the mean number of hours per week spent by students studying outside of class is 30.

Answer

Correct:

Indeed, in hypothesis testing, in order to assess the evidence, we need to find how likely it is to get data like those observed assuming that claim 1 is true.

Submit