Hypothesis testing

Assess evidence about some claim concerning a population.
An outcome that would rarely happen if a claim were true is good evidence that the claim is not true.
e.g., I want to assess whether you studied for a test which consists of 10 True or False problems. Outcome: you got all the 10 problems right. I assume you did not study for the test. The probability that you can get 10 right under this assumption is $(1/2)^{10} = 0.001$ which is very small. So my assumption is not ture and I should conclude you studied for the test. Note the probability is computed using the multiplation rule under independence.
The probability that we use to determine whether an event is rare is called the **significance level** of the test, denoted by α. The commonly used values are 0.05 and 0.01 and 0.10.

Hypothesis

null hypothesis (H_0) states a parameter equals a certain value. Usually the null hypothesis is a statement of "no effect" or "no difference."

alternative (or alternate) hypothesis (H_1) states the value of the parameters differs from the value stated in H_0.

Note H_0 is just an easier way to write the words "null hypthesis". Same for H_1.

One and two sided tests

The alternative hypothesis is **one-sided** (or one-tailed) if it states that a parameter is larger than (or smaller than) the null hypothesis value.
It is **two-sided**(or two-tailed) if it states that the parameter is different from the null value (it could be either smaller or larger).
e.g., $H_0 : \mu = 0$
$H_1 : \mu > 0$ (right tailed)
$H_1 : \mu < 0$ (left tailed)
$H_1 : \mu \neq 0$ (two-tailed)

How to state null and alternative hypothesis

8.1. (p 339). Boxes of cereal are labeled as containing 20 ounces. An inspector thinks the mean weight may be less than 20 ounces.
$H_0 : \mu = 20$
$H_1 : \mu < 20$ (put what you think or wonder or suspect in H_1)
8.2: Last year the the mean rent for an apartment in a city was 800 dollars. A realtor believes the mean rent is higher this year.
$H_0 : \mu = 800$
$H_1 : \mu > 800$ (again put what you believe in H_1)
8.3. Scores on a standardized test have a mean of 70. Some modifications were made to the test and an educator believes the mean may have changed.
$H_0 : \mu = 70$
$H_1 : \mu \neq 70$ (have changed could be higher or lower, i.e., not equal to 70)
. So what you think or believe or hope determines H_1.

Type I and Type II error

	reject H_0	do not reject H_0
When H_0 is true	Type I error	Correct Decision
When H_0 is false	correct decision	Type II error

Example: H_0 : He is innocent; H_1 : He is guilty

	reject H_0 and convict him	do not reject H_0
When he is innocent	Type I error	Correct Decision
When he is guilty	correct decision	Type II error

So Type I error is wronging an innocent person and Type II error is letting go of a criminal.

The p-value method

- A **test statistic** calculated from data measures how far the data diverge from what we would expect if H_0 were true. Large values of the statistic show that the data are not consistent with H_0.
- The probability, computed assuming H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed is called the **P -value** of the test.
- The smaller the p-value, the stronger the evidence against H_0.

Test for μ: assume σ is known

$H_0 : \mu = \mu_0$
Test statitic $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = z^*$.
The P-value for
$H_1 : \mu > \mu_0$ is $P(z > z^*)$,
$H_1 : \mu < \mu_0$ is $P(z < z^*)$,
$H_1 : \mu \neq \mu_0$ is $2P(z > |z^*|)$.
example 8.13, 8.14.

Example 8.13

According to a survey, the mean height of adult men in US is 69.7 inches with $\sigma = 3$ inches. A sociologist believes the mean height of male business executives, μ, may be greater than 69.7 inches. A SRS of 100 male executives has a mean height of 69.9 inches. Assume the male executive heights also have $\sigma = 3$ inches. Can we conculde the male business executives are taller, on average than the general male population at the $\alpha = 0.05$ level?

$\bar{x} = 69.9, \sigma = 3, n = 100.$

$H_0 : \mu = 69.7$

$H_1 : \mu > 69.7$ (right tailed test)

$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{69.9 - 69.7}{0.3} = 0.67.$

P-value $= P(z > 0.67) = 0.2514$. (get this from the z chart)

P-value bigger than α. Do not reject H_0. Not enough evidence to conclude H_1.

Example 8.14

The mean satisfaction score in a large company was 74 with $\sigma = 8$. After a new telecommuting policy has been place for a while, a SRS of 80 workers gave a sample mean score of 76. Assume σ is still 8. Can we conclude the mean level of satisfaction is different since the policy change at $\alpha = 0.05$?

$\bar{x} = 76, \sigma = 8, n = 80.$

$H_0 : \mu = 74$

$H_1 : \mu \neq 74$

$z = \frac{76-74}{8/\sqrt{80}} = 2.24.$

P-value$= 2P(z > 2.24) = 2 * 0.0125 = 0.025.$

Reject H_0 at the $\alpha = 0.05$ level. We conclude that the mean score changed after the adoption of telecommuting.

The conclusion centers on H_0. We either reject H_0 (when P-value is smaller than α) or do not reject H_0 (when P-value is bigger than α). If we reject H_0, we can further say we have sufficient evidence to support H_1. If we do not reject H_1, we can further say we do not have sufficient evidence to support H_1.

Statistical Significance does not mean Practical Importance

If the P -value is as small or smaller than α , we say that the result is statistically significant at level α.
"Significant" does not mean "important." It means simply "not likely to happen just by chance".

Check your understanding exercise Prob 50 P 363

A random sample of 60 second-graders in a school are given a standard math test. The sample mean score is $\bar{x} = 52$. Assume $\sigma = 15$. The nationwide average score is 50. Do the second graders in this school have greater math skills than the nationwide average? Perform a test at $\alpha = 0.01$ level.

solution

$\bar{x} = 52, \sigma = 15, n = 60.$
$H_0 : \mu = 50$
$H_1 : \mu > 50$
Test statistic $z = \frac{52-50}{15/\sqrt{60}} = 1.03.$
p-value $= P(z > 1.03) = 1 - 0.8485 = 0.1515 > \alpha.$
Do not reject H_0. There is no sufficient evidence that the second graders in her school have better math skills than the nationwide average.
In summary, the solutions consist of 4 steps. Step 1: write down H_0, H_1. Step 2: get the test statistic value. Step 3. Get the p-value. Step 4. Make your conclusion.

Exercise

In a sample of 300 men between the age of 60 and 69, the mean height was $\bar{x} = 69.0$ inches. Assume the population standard deviation is $\sigma = 2.84$ inches. According to the National Health Reports, the mean height for US men is 69.4 inches. Public health researchers want to determine if the mean height for older men is less than the mean height of all adult men. Perform a hypothesis test to answer this question at $\alpha = 0.05$ level of significance.

Solution

Let μ be the mean height of all older men in US.
$H_0 : \mu = 69.4$
$H_1 : \mu < 69.4$
$z = \frac{69-69.4}{2.84/\sqrt{300}} = -2.44$
P-value $= P(z < -2.44) = 0.007 < 0.05$.
Reject H_0. There is sufficient evidence that the mean height of older men is less than the mean height of all adult men.

P 365, 56. The mean weight of 6 year old girls in 2004 was 49.3 pounds. A survey in 2012 reported a sample of 177 girls had an average weight of 51.9 pounds. Assume the population standard deviation $\sigma = 17$ pounds. Can we conclude the mean weight of 6 year old girls is higher in 2012 than in 2004? Use the $\alpha = 0.01$ level of significance.

Solution: $\bar{x} = 51.9, n = 177, \sigma = 17.$

$H_0 : \mu = 49.3$

$H_1 : \mu > 49.3$

$z = \frac{51.9-49.3}{17/\sqrt{177}} = 2.03$

P-value $= P(z > 2.03) = 0.02 > 0.01.$

Do not reject H_0. There is no sufficient evidence to conclude the mean weight of 6 year old girls is higher in 2012 than in 2004.