Introduction to Inference

Tests of Significance

Objectives

Tests of significance

Stating hypotheses

Reasoning of significance tests

Test statistics and *P*-values

Tests for a population mean

The significance level α

Two-sided tests and confidence intervals

P-values versus fixed α

Reasoning of significance tests

We have seen that the properties of the sampling distribution of the sample mean help us estimate a range of likely values for population mean μ.
We can also rely on the properties of the sample distribution to test hypotheses.

Ex: You are in charge of quality control in your food company. You sample randomly 35 packs of cherry tomatoes, each labeled 1/2 lb. (227 g).

The average weight from your 35 boxes is 225 g. Obviously, we cannot expect boxes filled with whole tomatoes to all weigh exactly half a pound. Thus,

- Is the somewhat smaller weight simply due to chance variation?

OR

- Is it evidence that the calibrating machine that sorts cherry tomatoes into packs needs revision?

Stating hypotheses

A **test of statistical significance** tests a specific hypothesis using sample data to decide on the validity of the hypothesis.

In statistics, a **hypothesis** is a statement about the population, developed for the purpose of testing. **It is a statement about the value of a population parameter.**

What you want to know: Does the calibrating machine that sorts cherry tomatoes into packs need revision?

The same question reframed statistically: Is the population mean μ for the distribution of weights of cherry tomato packages equal to 227 g (i.e., half a pound)?

Stating hypotheses

- The statement being tested in a test of significance is called the null hypothesis, H_0. The test of significance is designed to assess the strength of the evidence against the null hypothesis. It is usually a statement of “no effect” or “no difference.”
- The alternative hypothesis is the statement we suspect is true instead of the null hypothesis. It is labeled Ha.
- Null Hypothesis:
  - It is a statement about the value of a population parameter.
  - Status quo.
  - Denoted as H_0.
- Research (Alternate) Hypothesis:
  - It is a statement about the value of a population parameter.
  - What the researcher is trying to prove.
  - Opposite of the H_0.
  - Denoted as Ha or H_1.

One-sided and two-sided tests

- A **two-tail** or **two-sided test** of the population mean has these null and alternative hypotheses:

  H_0: μ = [a specific number]

  H_a: $\mu \neq$ [a specific number]

- A **one-tail** or **one-sided test** of a population mean has these null and alternative hypotheses:

  H_0: μ = [a specific number]

  H_a: $\mu <$ [a specific number] OR

  H_0: μ = [a specific number]

  H_a: $\mu >$ [a specific number]

Stating hypotheses

Weight of cherry tomato packs:

H_0 : μ = 227 g (μ is the average weight of the population of packs)

H_a : $\mu \neq$ 227 g (μ is either larger or smaller)

Possible Results from Testing

- Reject the **Null Hypothesis**

- Do Not Reject the **Null Hypothesis** or Failure to Reject the **Null Hypothesis**

Running a test of significance is a balancing act between the chance α of making a **Type I error** and the chance β of making a **Type II error.** Reducing α reduces the power of a test and thus increases β.

	H_0 true	H_a true
REJECT H_0	Type I error	Correct decision
DO NOT REJECT H_0	Correct decision	Type II error

Sample Evidence: Test statistics

- The test statistic is based on the statistic that estimates the parameter.

- Because Normal calculations require standardized variables, we use as our test statistic the standardized sample mean.

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- This random variable has the standard Normal distribution N(0, 1).

Statistical Significance

- In statistics, a result is called **statistically significant** if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level, *alpha*.

The *P*-value

The packaging process has a known standard deviation, s = 7 g.

$H_0 : \mu = 227$ g

$H_a : \mu \neq 227$ g

The average weight from your 35 random boxes is 225 g.

What is the probability of drawing a random sample such as yours if H_0 is true?

Tests of statistical significance quantify the chance of obtaining a particular random sample result **if the null hypothesis were true**. This quantity is the ***P*-value.**

This is a way of assessing the "believability" of the null hypothesis given the evidence provided by a random sample.

Interpreting a *P*-value

Could random variation alone account for the difference between the null hypothesis and observations from a random sample?

- A small *P*-value implies that random variation because of the sampling process alone is not likely to account for the observed difference.

- With a small *P*-value we **reject H_0.** The true property of the population is **significantly** different from what was stated in H_0.

Thus, small *P*-values are strong evidence AGAINST H_0.

But how small is small…?

The significance level, α

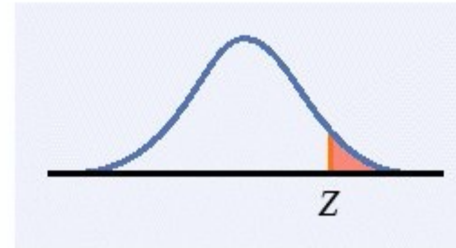
We compare the *p*-value with a level that we regard as decisive called the significance level, *α*. This value is decided arbitrarily before conducting the test.

- If the *P*-value is equal to or less than *α* (**$p \le \alpha$**), then we **reject H_0**.
- If the *P*-value is greater than α (**$p > \alpha$**), then we **fail to reject H_0**.

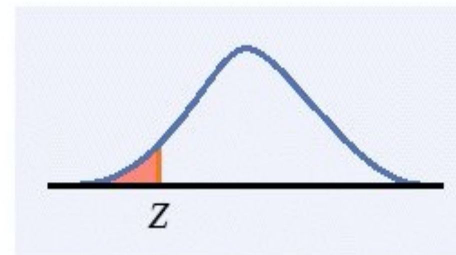
P-value in one-sided and two-sided tests

One-sided (one-tailed) test

$H_a: \mu > \mu_0$ is $P(Z \geq z)$

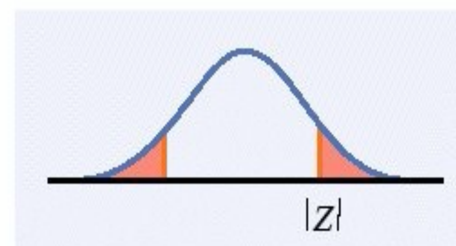


$H_a: \mu < \mu_0$ is $P(Z \leq z)$



Two-sided (two-tailed) test

$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$



To calculate the *p*-value for a two-sided test, use the symmetry of the Normal curve. Find the *p*-value for a one-sided test and double it.

Does the packaging machine need revision?

- $H_0 : \mu = 227$ g versus $H_a : \mu \neq 227$ g
- What is the probability of drawing a random sample such as yours if H_0 is true?
- Use alpha = 0.05

$$\bar{x} = 225\text{g} \quad s = 7\text{g} \quad n = 35$$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{225 - 227}{7/\sqrt{35}} = -1.69$$

The area under the standard Normal curve

to the left of z is 0.0455.

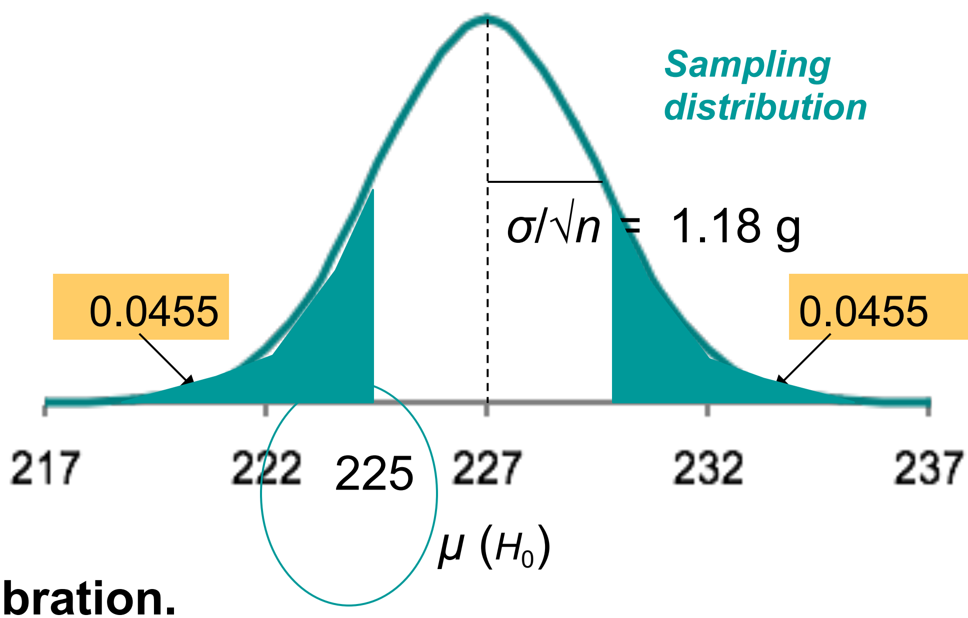
Thus, P-value = 2*0.0455 = 0.0910.

The probability of getting a random

sample average so different from

μ is not unlikely enough to reject H_0.

➔The machine does not need recalibration.



Result and Conclusion

- Result: Do not Reject H_0
- Conclusion: There is not enough evidence, at the 0.05 level of significance, to conclude that the average weight (Mu) is not exactly equal to 227g.
- Therefore, based on the evidence, the machine does not need recalibration.

- What would be a consequence of incurring type I error?
- What would be a consequence of incurring type II error?

Steps of Hypothesis Testing

1. Formulate the **null hypothesis** (commonly, that the observations are the result of pure chance) and the **alternative hypothesis** (commonly, that the observations show a real effect combined with a component of chance variation).
2. Identify a **test statistic** that can be used to assess the truth of the null hypothesis.
3. Compute the **p-value**, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. **The smaller the p-value, the stronger the evidence against the null hypothesis**.
4. Compare the p-value to an acceptable significance value (the alpha value). **If the p-value is less than or equal to alpha**, then the observed effect is statistically significant, the null hypothesis is ruled out in favor of the alternative hypothesis.