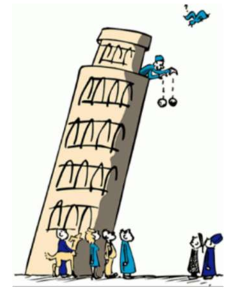
8 - Hypothesis Testing



In science, a *hypothesis* is a statement that you can test by experimentation.

Example Galileo Galilei (1564-1642) tested the hypothesis that *any pair of objects dropped from the same height fall to the Earth in equal time.*

To test this, Galileo dropped balls of various sizes off the Tower of Pisa. Each time, Galileo saw that the balls hit the ground at the same time.

In statistics, a *hypothesis* is:

e.g., The mean mass of a bag of Old Dutch Ketchup chips is 255g.

The evidence we use to test a statistical hypothesis is:

- To test a claim about a population mean μ, we use ______ from sample data.
- To test a claim about a population proportion p, we use ______ from sample data.
- To test a claim about a population variance σ^2, we use ______ from sample data.

> In statistics, a *hypothesis test* is a formal procedure that uses *sample statistics* to test a claim about the corresponding *population parameter*.

Hypothesis testing is based on the principle:

> Sample data only provides adequate support for a hypothesis if that level of support is not likely to happen *by chance*.

Example Suppose I flip a coin 6 times and I get 5 "heads." Can I conclude that the coin is biased towards heads?

Answer: No. The probability of getting 5 *or more* heads for a fair coin is: $\frac{6+1}{2^6} = 0.1094$

Since $0.1094 > 5\%$, we say that the evidence is *not statistically significant* at a 5% significance level. The coin is not biased towards heads, as far as we can tell.

Rejection Region Method for μ

Example (Two-Tailed) The network engineer claims that the mean time to transfer a 1 Gb file across a network is 12.44 seconds. You test this claim by sending a 1 Gb file across the network at 10 random times, recording the following transfer times. Test the engineers claim at a 5% significance level.

12.0	13.5	11.8	12.4	13.8
11.9	12.2	12.9	13.0	12.3

Step 1: (State the claim) – Which population parameter? What is the claim?

Step 2: (Hypotheses) – Record the *null hypothesis* and the *alternative hypothesis*

Step 3: (Test Statistic) - Calculate the appropriate test statistic

Step 4: (Rejection Region) – What values of the test statistic would make us reject H_0?

Step 5: (Decision) – Reject H_0 or Fail to Reject H_0?

Step 6: (Conclusion) – State your conclusion using the wording of the original claim.

Example (Left-Tailed) One week later, the network engineer claims that the mean time to transfer a 1 Gb file across a network has been reduced *below* 10 seconds. You test this claim by sending a 1 Gb file across the network at 5 random times, recording the following transfer times. Test the engineers claim at a 5% significance level.

9.0	8.5	10.1	8.7	9.5

p-value Method for μ

Definition The p-value of a given sample statistic is the probability of *randomly* obtaining a sample statistic as extreme as that given value, assuming the null hypothesis is true. (Here "extreme" means *in the direction specified in the alternative hypothesis*.)

$$\boxed{\text{We reject } H_0 \text{ if the } p\text{-value is below } \alpha.}$$

Example (Right-Tailed) A BCIT study aims to determine the average value of the variable:

$$X = \text{amount of time a BCIT student works (at a job) each week}$$

The study collects data for a random sample of students. Sample statistics are given:

$$n = 100; \;\; \bar{X} = 11.2; \;\; s = 4.5$$

Test the claim (at the 5% level) that the mean μ is *greater* than 10 hours per week.

Inferences about Population p

Inferences about p are based on the principle that:

Sample proportion $\hat{p}$ is *normally distributed* as long as $np \geq 5$ and $nq \geq 5$.

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Example Seven out of the 46 past US presidents have been left-handed. In the general population, 10% of individuals are left-handed. Test the claim that US presidents are drawn from a population that is more likely to be left-handed than the general population.

Example (Wald Conf. Interval) Suppose a poll of $n = 1000$ randomly selected Canadian voters indicates that $X = 512$ will vote for the Blue political party.

a. Determine a 95% confidence interval for the population proportion of Canadians who will vote for the Blue political party.

b. Test the hypothesis that *half* of voters will vote for the Blue political party. Use the traditional method.

Notes on Steps 1 and 2 (Claim and Hypotheses)

- The *null* hypothesis always involves *equality*. It gives a *default* answer to the claim/question.

- The *alternative* hypothesis always involves *inequality*.

- Sample statistics (i.e., $\hat{p}$, $\bar{X}$, s^2) *never* appear in the hypotheses.

Example Translate each of the following claims into appropriate hypotheses.

a) The mean age of a BCIT student is 24.2 years.

b) This coin has at least a 50% chance of coming up heads.

c) The average Canadian earns less than $100 000.

d) The average Canadian has at most 15.5 years of education.

Notes on Step 6 (Stating a Conclusion)

If the original claim was H_0 and we *reject* H_0:

"The evidence is strong enough to reject the claim that …"

If the original claim was H_0 and we *fail to reject* H_0:

"The evidence is not strong enough to reject the claim that …"

If the original claim was H_1 and we *reject* H_0:

"The evidence is strong enough to accept the claim that …"

If the original claim was H_1 and we *fail to reject* H_0:

"The evidence is not strong enough to accept the claim that …"

Errors in Hypothesis Testing

Type 1 Error - rejecting the null hypothesis when it is actually true.

Type 2 Error - failing to reject the null hypothesis when it is actually false.

Example In a Canadian court of law, you are presumed *innocent* until proven guilty.

H_0:

H_1:

Type 1 error would be

Type 2 error would be

		The Truth	
		H_0 is true	H_0 is false (H_1 is true)
Decision	Fail to Reject H_0		
	Reject H_0		

How can we reduce the probability of making a Type 1 error?

How can we reduce the probability of making a Type 2 error?