6.3 Hypothesis testing

Last section illustrated how probability can enable confidence interval estimation. We can also use probability as a means to use data to quantitatively assess the plausibility of a trial value of a parameter.

Statistical inference is using data from the sample to draw conclusions about the population.

1. Interval estimation (confidence intervals)

2. Hypothesis testing

Definition 6.3. Statistical *significance testing* is the use of data in th quantitative assessment of the plausibility of some trial value for a parameter (or function of one or more parameters).

Significance (or hypothesis) testing begins with the specification of a trial value (or **hypothesis**).

Definition 6.4. A *null hypothesis* is a statement of the form

$$\text{Parameter} = \#$$

or

$$\text{Function of parameters} = \#$$

for some # that forms the basis of investigation in a significance test. A null hypothesis is usually formed to embody a status quo/"pre-data" view of the parameter. It is denoted H_0.

Definition 6.5. An *alternative hypothesis* is a statement that stands in opposition to the null hypothesis. It specifies what forms of departure from the null hypothesis are of concern. An alternative hypothesis is denoted as H_a. It is of the form

$$\text{Parameter} \neq \# \quad \text{or} \quad \text{Parameter} > \# \quad \text{or} \quad \text{Parameter} < \#$$

Examples (testing the true mean value):

$$H_0 : \mu = \# \quad H_0 : \mu = \# \quad H_0 : \mu = \#$$
$$H_a : \mu \neq \# \quad H_a : \mu > \# \quad H_a : \mu < \#$$

Often, the alternative hypothesis is based on an investigator's suspicions and/or hopes about th true state of affairs.

The **goal** is to use the data to debunk the null hypothesis in favor of the alternative.

1. Assume H_0.
2. Try to show that, under H_0, the data are preposterous.
3. If the data are preposterous, reject H_0 and conclude H_a.

The outcomes of a hypothesis test consists of:

Example 6.11 (Fair coin). Suppose we toss a coin $n = 25$ times, and the results are denoted by $X_1, X_2, \ldots, X_{25}$. We use 1 to denote the result of a head and 0 to denote the results of a tail. Then $X_1 \sim Binomial(1, \rho)$ where ρ denotes the chance of getting heads, so $\mathrm{E}(X_1) = \rho, \mathrm{Var}(X_1) = \rho(1 - \rho)$. Given the result is you got all heads, do you think the coin is fair?

In the real life, we may have data from many different kinds of distributions! Thus we need a universal framework to deal with these kinds of problems.

6.3.1 Significance tests for a mean

Definition 6.6. A *test statistic* is the particular form of numerical data summarization used in a significance test.

Definition 6.7. A *reference (or null) distribution* for a test statistic is the probability distribution describing the test statistic, provided the null hypothesis is in fact true.

Definition 6.8. The *observed level of significance or p-value* in a significance test is the probability that the reference distribution assigns to the set of possible values of the test statistic that are at least as extreme as the one actually observed.

Based on our results from Section 6.2 of the notes, we can develop hypothesis tests for the true mean value of a distribution in various situations, given an iid sample $X_1, \ldots, X_n$ where $H_0 : \mu = \mu_0$.

Let K be the value of the test statistic, $Z \sim N(0,1)$, and $T \sim t_{n-1}$. Here is a table of p-values that you should use for each set of conditions and choice of H_a.

Situation	K	$H_a : \mu \neq \mu_0$	$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$
$n \geq 25, \sigma$ known	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$P(|Z| > K)$	$P(Z < K)$	$P(Z > K)$
$n \geq 25, \sigma$ unknown	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$P(|Z| > K)$	$P(Z < K)$	$P(Z > K)$
$n < 25, \sigma$ unknown	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$P(|T| > K)$	$P(T < K)$	$P(T > K)$

Steps to perform a hypothesis test:

1.

2.

3.

4.

5.

6.

Example 6.12 (Cylinders)**.** The strengths of 40 steel cylinders were measured in MPa. The sample mean strength is 1.2 MPa with a sample standard deviation of 0.5 MPa. At significance level $\alpha = 0.01$, conduct a hypothesis test to determine if the cylinders meet the strength requirement of 0.8 MPa.

Example 6.13 (Concrete beams)**.** 10 concrete beams were each measured for flexural strength (MPa). The data is as follows.

[1] 8.2 8.7 7.8 9.7 7.4 7.8 7.7 11.6 11.3 11.8

The sample mean was 9.2 MPa and the sample variance was 3.0933 MPa. Conduct a hypothesis test to find out if the flexural strength is different from 9.0 MPa.

6.3.2 Hypothesis testing using the CI

We can also use the $1 - \alpha$ confidence interval to perform hypothesis tests (instead of p-values). The confidence interval will contain μ_0 when there is little to no evidence against H_0 and will not contain μ_0 when there is strong evidence against H_0.

Steps to perform a hypothesis test using a confidence interval:

1.

2.

3.

4.

5.

6.

Example 6.14 (Breaking strength of wire, cont'd)**.** Suppose you are a manufacturer of construction equipment. You make 0.0125 inch wire rope and need to determine how much weight it can hold before breaking so that you can label it clearly. You have breaking strengths, in kg, for 41 sample wires with sample mean breaking strength 91.85 kg and sample standard deviation 17.6 kg. Using the appropriate 95% confidence interval, conduct a hypothesis test to find out if the true mean breaking strength is above 85 kg.

Example 6.15 (Concrete beams, cont'd)**.** 10 concrete beams were each measured for flexural strength (MPa). The data is as follows.

[1] 8.2 8.7 7.8 9.7 7.4 7.8 7.7 11.6 11.3 11.8

The sample mean was 9.2 MPa and the sample variance was 3.0933 MPa. At $\alpha = 0.01$, test the hypothesis that the true mean flexural strength is 10 MPa using a confidence interval.

Example 6.16 (Paint thickness, cont'd). Consider the following sample of observations on coating thickness for low-viscosity paint.

[1] 0.83 0.88 0.88 1.04 1.09 1.12 1.29 1.31 1.48 1.49 1.59 1.62 1.65 1.71 [15] 1.76 1.83

Using $\alpha = 0.1$, test the hypothesis that the true mean paint thickness is 1.00 mm. Note, the 90% confidence interval for the true mean paint thickness was calculated from before as $(1.201, 1.499)$.