Outline

1) Hypothesis tests

2 power and size of the test

3) size α test for μ (σ known)

4) size α test for μ (σ unknown)

5) size α test for σ

6) size α test for $\mu_1 - \mu_2$ (σ_1, σ_2 unknown)

7) size α test for σ_1/σ_2 (μ_1, μ_2 known)

8) Neyman-Pearson lemma and the likelihood ratio test

Hypothesis tests

- the objective of statistics often is to make inferences about unknown population parameters based on information contained in sample data

- these inferences are phrased either as estimates of the respective parameters or as tests of hypotheses about their values

- testing a hypothesis requires making a decision when comparing the observed sample with theory

- how do we decide whether the sample disagrees with the scientist's hypothesis?

Hypothesis tests

- when should we reject the hypothesis, when should we accept it, and when should we withhold judgment?

- what is the probability that we will make the wrong decision and consequently be led to a loss?

Example 1:

Many measurements of the viscosity of a certain oil indicate that the underlying distribution is $N(0.45, (0.03)^2)(\mu = 0.45, \sigma = 0.03)$.

Suppose on the basis of one further observation X (a sample of size 1) we wish to decide whether or not we think the quality of oil (as measured by μ) has changed, assuming that the variability (as measured by σ) is unchanged; that is, is the observed value of X consistent with a mean of $\mu = 0.45$ or is it likely that $\mu \neq 0.45$?

Thus we wish to test the hypothesis

$$H_0 : \mu = 0.45$$

this is the **null hypothesis** against the **alternative hypothesis**

$$H_1 : \mu \neq 0.45$$

Example 1:

A test of H_0 against H_1 could be:

reject H_0 (accept H_1) if $X > 0.5088$ or $X < 0.3912$ and accept H_0 if $0.3912 \leq X \leq 0.5088$. A test specifies a **rejection region**, a subset of the possible values of X, in this case the union of the intervals $(-\infty, 0.3912)$ and $(0.5088, \infty)$

Hypothesis tests

The errors in hypothesis testing can be summarised in the following table

Errors	H_0 true	H_1 true
Accept H_0	Correct decision Probability $1-\alpha$	Type II error probability β
Reject H_0	Type I error Probability= α (size)	Correct decision probability=$1-\beta$ (power)

- The **size** (or significance level) of a test = P(type I error)

$$
\begin{aligned}
&= P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \\
&= P(\text{reject } H_0 | H_0 \text{ is true}) \\
&= \text{the probability that } X \text{ will be in the rejection} \\
&\quad \text{region, calculated under the assumption} \\
&\quad \text{that } H_0 \text{ is true}
\end{aligned}
$$

Hypothesis tests

- If H_1 is true then the **power** of a test =P(reject H_0)

$$\begin{aligned} &= 1 - P(\text{accept } H_0 | H_1 \text{is true}) \\ &= 1 - P(\text{Type II error}) \end{aligned}$$

- in general (i.e. either H_0 or H_1 is true),

$$\text{power=P(reject } H_0)$$

- Thus we like tests which have small size (close to zero) and large power (close to 1)

Example 1:

Returning to the viscosity example: $X \sim N(\mu, (0.03)^2)$. The test with rejection region $X < 0.3912$ or $X > 0.5088$ has size equal to

$$\begin{aligned}
& P(X < 0.3912 \text{ or } > 0.5088 | H_0 \text{ true}) \\
= \; & P(X < 0.3912 | \mu = 0.45) + P(X > 0.5088 | \mu = 0.45) \\
= \; & P(\tfrac{X-0.45}{0.03} < \tfrac{0.3912-0.45}{0.03} | X \sim N(0.45, (0.03)^2)) \\
& + P(\tfrac{X-0.45}{0.03} > \tfrac{0.5088-0.45}{0.03} | X \sim N(0.45, (0.03)^2)) \\
= \; & P(Z < -1.96) + P(Z > 1.96), \text{ where } Z = \tfrac{X-0.45}{0.03} \sim N(0,1) \\
& = 0.05
\end{aligned}$$

since 1.96 is the upper 2.5% point of $N(0,1)$

the power of the above test (when $\mu = 0.55$) is P(reject $H_0 | H_1$ true)

$$\begin{aligned}
= \; & P(X < 0.3912 | \mu = 0.55) + P(X > 0.5088 | \mu = 0.55) \\
= \; & P(\tfrac{X-0.55}{0.03} < \tfrac{0.3912-0.55}{0.03} | X \sim N(0.55, (0.03)^2)) \\
& + P(\tfrac{X-0.55}{0.03} > \tfrac{0.5088-0.55}{0.03} | X \sim N(0.55, (0.03)^2)) \\
= \; & P(Z < -5.29) + P(Z > -1.37) = 0.91
\end{aligned}$$

since, when $\mu = 0.55, Z = \frac{X - 0.55}{0.03} \sim N(0, 1)$

Example 2:

This example provides a visual example of power functions, rather than an algebraic one. Edward Scruggs is a candidate for a football tipping panel at a local newspaper. The newspaper will decide his worthiness for the panel based on the number of correct choices he makes in 70 consecutive football games. Let

$$p = P(\text{Edward correctly chooses the winner on any one game})$$

The newspaper must decide when

$$H_0 : p = 0.5 \text{ (Edward is guessing)}$$

$$H_1 : p > 0.5 \text{ (Edward is a football expert)}$$

If the newspaper sets

$$\text{size} = \text{P(Type I error)} \;\approx 0.05$$

then calculations based on the binomial distribution can be used to show that the power function is as shown (in class).

Sampling on the Normal distribution

suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$

Size α tests for μ(σ known)

- pivotal function:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- for testing

$$H_0 : \mu = \mu_0$$

the test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

$Z \sim N(0,1)$ when H_0 is true.

Sampling on the Normal distribution

- if the alternative is

$$H_1 : \mu \neq \mu_0$$

  reject H_0 if $Z < -z_{1-\alpha/2}$ or $> z_{1-\alpha/2}$

- note size=P(reject $H_0 | H_0$ is true)

$$\begin{aligned} &= P(Z < -z_{1-\alpha/2} | \mu = \mu_0) \\ &\quad + P(Z > z_{1-\alpha/2} | \mu = \mu_0) \\ &= \alpha/2 + \alpha/2 = \alpha \end{aligned}$$

- if $H_1 : \mu > \mu_0$, reject H_0 if $Z > z_{1-\alpha}$

- if $H_1 : \mu < \mu_0$, reject H_0 if $Z < -z_{1-\alpha}$

Sampling on the Normal distribution

size α tests for μ (σ unknown)

- pivotal function

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

- for testing

$$H_0 : \mu = \mu_0$$

the test statistic is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$T \sim t_{n-1}$ when H_0 is true.

- if $H_1 : \mu \neq \mu_0$, reject H_0 if $T < t_{n-1,\alpha/2}$ or $> t_{n-1,1-\alpha/2}$

Sampling on the Normal distribution

- for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, reject H_0 if $T > t_{n-1,1-\alpha}$

- for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, reject H_0 if $T < t_{n-1,\alpha}$

Size α tests for σ

- pivotal function

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Sampling on the Normal distribution

- for testing

$$H_0 : \sigma = \sigma_0 (\sigma_0 \text{ specified})$$

  the test statistic is

$$V = \frac{(n-1)S^2}{\sigma^2}$$

  $V \sim \chi^2_{n-1}$ when H_0 is true

- if $H_1 : \sigma \neq \sigma_0$, reject$H_0$ if $V < \chi^2_{n-1,\alpha/2}$ or $> \chi^2_{n-1,1-\alpha/2}$

- for testing $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma > \sigma_0$, rejectH_0 if $V > \chi^2_{n-1,1-\alpha}$

- for testing $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma < \sigma_0$, rejectH_0 if $V < \chi^2_{n-1,\alpha}$

Two sample problem

suppose

$$X_1, X_2, \ldots, X_{n_1}, \text{ from } N(\mu_1, \sigma_1^2)$$

$$Y_1, Y_2, \ldots, Y_{n_2}, \text{ from } N(\mu_2, \sigma_2^2)$$

we assume the samples are independent

size α tests for $\mu_1 - \mu_2$ (σ_1, σ_2 unknown)

the common variance case: $\sigma_1 = \sigma_2 = \sigma$

- pivotal function

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p(\frac{1}{n_1} + \frac{1}{n_2})^{1/2}} \sim t_{n_1+n_2-2}$$

Two sample problem

- for testing $H_0 : \mu_1 = \mu_2$, the test statistic is
$$T = \frac{\bar{X} - \bar{Y}}{S_p(\frac{1}{n_1} + \frac{1}{n_2})^{1/2}}$$
$T \sim t_{n_1+n_2-2}$ when H_0 is true

- if $H_1 : \mu_1 \neq \mu_2$, reject H_0 if $T < t_{n_1+n_2-2,\alpha/2}$ or $> t_{n_1+n_2-2,1-\alpha/2}$

- for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$, reject H_0 if $T > t_{n_1+n_2-2,1-\alpha}$

- for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$, reject H_0 if $T < t_{n_1+n_2-2,\alpha}$

Two sample problem

Example 3: Data on cube compressive strength (N/mm^2) for concrete specimens made with pulverised fuel-ash miz. Assume common variance

Age			
7 days	$n_1 = 68$	$\bar{x} = 26.99$	$s_1 = 4.89$
28 days	$n_2 = 74$	$\bar{y} = 35.76$	$s_2 = 6.43$

Test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$

where $\mu_1(\mu_2)$ is the mean 7 days (28 days) strength. Solution in lecture.

Two sample problem

the paired observation case

- pivotal function

$$\frac{\bar{Z} - (\mu_1 - \mu_2)}{S_Z/\sqrt{n}} \sim t_{n-1}$$

- for testing

$$H_0 : \mu_1 = \mu_2$$

  the test statistic is

$$W = \frac{\bar{Z}}{S_Z/\sqrt{n}}$$

$W \sim t_{n-1}$ when H_0 is true

- if $H_1 : \mu_1 \neq \mu_2$, reject H_0 if $W < t_{n-1,\alpha/2}$ or $> t_{n-1,1-\alpha/2}$

Two sample problem

- for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$, reject H_0 if $W > t_{n-1,1-\alpha}$

- for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$, reject H_0 if $W < t_{n-1,\alpha}$

size α tests for σ_1/σ_2 (μ_1, μ_2 known)

- pivotal function

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

Two sample problem

- for testing $H_0 : \sigma_1 = \sigma_2$, the test statistic is

$$Z = \frac{S_1^2}{S_2^2}$$

$Z \sim F_{n_1-1,n_2-1}$ when H_0 is true

- if $H_1 : \sigma_1 \neq \sigma_2$, reject H_0 if $Z < F_{n_1-1,n_2-1,\alpha/2}$ or $> F_{n_1-1,n_2-1,1-\alpha/2}$

- for testing $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 > \sigma_2$, reject H_0 if $Z > F_{n_1-1,n_2-1,1-\alpha}$

- for testing $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 < \sigma_2$, reject H_0 if $Z < F_{n_1-1,n_2-1,\alpha}$

Two sample problem

- **Example 4:** data on serum ferritin (mg/litre) in 28 elderly males gave $S_1 = 52.6$ and for 26 young males, $S_2 = 84.2$, does this suggest that the variance of serum ferritin is smaller in elderly males than in young males?

  Solution in lecture.

Most powerful tests

- a hypothesis is **simple** if it specifies one value for the parameter ($H_0 : \mu = 0.45$ is simple) and **composite** otherwise ($H_1 : \mu \neq 0.45$ is composite)

- **the Neyman-Pearson Lemma** Suppose that we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative hypothesis $H_1 : \theta = \theta_1$, based on a random sample $Y_1, Y_2, \ldots, Y_n$ from a distribution with parameter θ.

  Let R be the rejection region defined by

$$R = \{(Y_1, Y_2, \ldots, Y_n) : \frac{L(\theta_1)}{L(\theta_0)} \geq k\}$$

where $L(\theta)$ denotes the likelihood of the sample when the value of the parameter is θ, and the value k is chosen so that the test has size α.

Then R is the rejection region of the most powerful size α test of H_0 versus H_1, i.e., if R^* is the rejection region of any other size α test then

$$\begin{aligned} & P((Y_1, Y_2, \ldots, Y_n) \in R | \theta = \theta_1) \\ > \; & P((Y_1, Y_2, \ldots, Y_n) \in R^* | \theta = \theta_1) \end{aligned}$$

The Neyman-Pearson lemma basically says that among all tests with a given probability of a type I error, the likelihood ratio test minimizes the probability of a type II error.

Most powerful tests

- suppose $X_1, X_2, \ldots, X_n$ iid $f_X(x;\theta), \theta \in \Theta \subset R$. We wish to test $H_0 : \theta = \theta_0$ (a specified constant) against $H_1 : \theta \in \Theta_1 (\subset \Theta)$, where Θ_1 contains more than 1 point.

  Let θ_1 be a point in Θ_1. If the rejection region of the most powerful size α test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ does not depend on θ_1, then that test is the **uniformly most powerful** (ump) size α test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \Theta_1$.

Likelihood ratio test

- the likelihood ratio test is optimal for testing a simple versus a simple hypothesis

- we will develop a generalisation of this test for use in situations in which the hypothesis are not simple, such tests are generally not optimal, but they are typically nonoptimal in situations for which no optimal test exists, and they usually perform reasonably well

- suppose a random sample is selected from a distribution and that the likelihood function $L(x_1, \ldots, x_n|\theta_1, \ldots, \theta_k)$ is a function of k parameters, that is, $\Theta = (\theta_1, \ldots, \theta_k)$

- we can write the likelihood function as $L(\Theta)$

Likelihood ratio test

- suppose that the null hypothesis specifies that Θ lies in a particular set of possible values, say, Ω_0, and that the alternative hypothesis specifies that Θ lies in another set of possible values Ω_a, which does not overlap Ω_0

- let $L(\hat{\Omega}_0)$ denote the maximum of the likelihood function for all $\Theta \in \Omega_0$

- similar, $L(\hat{\Omega})$ represents the best explanation for the observed data for all $\Theta \in \Omega = \Omega_0 \cup \Omega_a$

Likelihood ratio test

- if $L(\hat{\Omega}_0) = L(\hat{\Omega})$, then a best explanation for the observed data can be found inside Ω_0 and we should not reject the null hypothesis $H_0 : \Theta \in \Omega_0$

- however, if $L(\hat{\Omega}_0) < L(\hat{\Omega})$, then the best explanation for the observed data can be found inside Ω_a, and we should consider rejecting H_0 in favour of H_a

- a **likelihood ratio test** is based on the ratio λ defined by

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{\underset{\Theta \in \Omega_0}{max}\, L(\Theta)}{\underset{\Theta \in \Omega}{max} L(\Theta)}$$

for $H_0 : \Theta \in \Omega_0$ versus $H_a : \Theta \in \Omega_a$, where λ is the test statistic, and the rejection region is determined by $\lambda \leq k$

Likelihood ratio test

- when the sample size is large, we can obtain an approximation to the distribution of λ,

$$-2\log\lambda \approx \chi_v^2$$

where v denotes the degree of freedom, which equals to the difference between the number of free parameters specified by H_0 and the number of free parameters specified by H_1.

Likelihood ratio test

- **Example 5:** Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a manufacturing plant. One hundred independent observations on the number of complaints gave means $\bar{x} = 20$ for shift 1 and $\bar{y} = 22$ for shift 2. Assume that the number of complaints per week on the ith shift has a Poisson distribution with mean θ_i, for $i = 1, 2$. Use the likelihood ratio method to test $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$ with $\alpha \approx 0.01$.

  Sol: In Lecture.