Testing Hypothesis

Maura Mezzetti

Department of Economics and Finance
Università Tor Vergata

Hypothesis Testing

"*It is a mistake to confound strangeness with mystery*"

Sherlock Holmes
A Study in Scarlet

Outline

Large Sample Tests

Why do we need large sample hypothesis testing?

The advantage of a large sample for testing is that we do not need the distribution of the test statistics: less computations needed! ... but results hold asymptotically

Large Sample Tests

Suppose we wish to test $H_0 : \theta = \theta_0$ against $H_0 : \theta \neq \theta_0$. The likelihood-based give rise to several possible tests

Large Sample Tests

Let $logL(\theta)$ denote the loglikelihood and $\hat{\theta}_n$ the consistent root of the likelihood equation. Intuitively, the farther θ_0 is from $\hat{\theta}_n$, the stronger the evidence against the null hypothesis.

But how far is far enough?

Large Sample Tests

Let $logL(\theta)$ denote the loglikelihood and $\hat{\theta}_n$ the consistent root of the likelihood equation. Intuitively, the farther θ_0 is from $\hat{\theta}_n$, the stronger the evidence against the null hypothesis.

But how far is far enough?

If θ_0 is close to $\hat{\theta}_n$

- $logL(\theta_0)$ should also be close to $logL(\hat{\theta}_n)$
- $logL'(\theta_0)$ should also be close to $logL'(\hat{\theta}_n) = 0$.

Large Sample Tests

Wald test

If we base a test upon the value of $(\hat{\theta}_n - \theta_0)$

Large Sample Tests

Wald test

If we base a test upon the value of $(\hat{\theta}_n - \theta_0)$

Likelihood ratio test

If we base a test upon the value of $logL(\hat{\theta}_n) - logL(\theta_0)$

Large Sample Tests

Wald test

If we base a test upon the value of $(\hat{\theta}_n - \theta_0)$

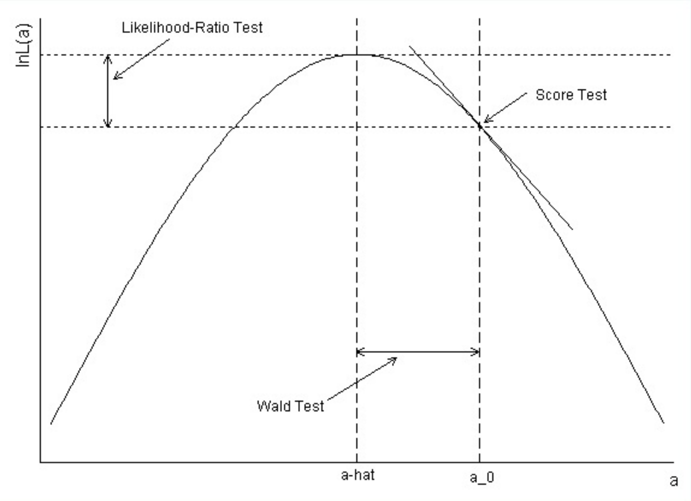
Likelihood ratio test

If we base a test upon the value of $logL(\hat{\theta}_n) - logL(\theta_0)$

(Rao) score test

If we base a test upon the value of $logL'(\theta_0)$

Three asymptotic tests



Large Sample Tests

We argued that under certain regularity conditions, the following facts are true under H_0:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow N\left(0, \frac{1}{I(\theta_0)}\right)$$

Large Sample Tests

We argued that under certain regularity conditions, the following facts are true under H_0:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow N\left(0, \frac{1}{I(\theta_0)}\right)$$

$$\frac{1}{\sqrt{n}}(logL'(\theta_0)) \longrightarrow N\left(0, I(\theta_0)\right)$$

Large Sample Tests

We argued that under certain regularity conditions, the following facts are true under H_0:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow N\left(0, \frac{1}{I(\theta_0)}\right)$$

$$\frac{1}{\sqrt{n}}(logL'(\theta_0)) \longrightarrow N\left(0, I(\theta_0)\right)$$

$$-\frac{1}{n}(logL''(\theta_0)) \longrightarrow I(\theta_0)$$

Wald Test

Under certain regularity conditions, the maximum likelihood estimator $\hat{\theta}$ has approximately in large samples a (multivariate) normal distribution with mean equal to the true parameter value and variance-covariance matrix given by the inverse of the information matrix, so that

$$\hat{\theta} \sim N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$$

Wald Test

Wald test statistics (parameter θ scalar)

- $$(\hat{\theta} - \theta_0) \times \sqrt{nI(\theta_0)} \sim N(0,1)$$
- A consistent estimator is used
$$(\hat{\theta} - \theta_0) \times \sqrt{nI(\hat{\theta})} \sim N(0,1)$$
- More generally, often it is written as:
$$\frac{\hat{\theta} - \theta_0}{\sqrt{Var(\hat{\theta})}} \sim N(0,1)$$

Wald Test Statistics

When θ is a parameter of dimension p. This result provides a basis for constructing tests of hypotheses and confidence regions. For example under the hypothesis $H_0 : \theta = \theta_0$ for a fixed value θ_0, the quadratic form

$$W = (\hat{\theta} - \theta_0)\prime I(\hat{\theta})(\hat{\theta} - \theta_0)$$

written also as:

$$W = (\hat{\theta} - \theta_0)\prime var^{-1}(\hat{\theta})(\hat{\theta} - \theta_0)$$

has approximately in large samples a chi-squared distribution with p degrees of freedom.

Comments on Wald Test Stastiscs

You may find also this formula:

$$W = (\hat{\theta} - \theta_0)\prime I(\theta_0)(\hat{\theta} - \theta_0)$$

and assumed an approximately chi-squared distribution with p degrees of freedom.

REMEMBER:

$$lim_{n \to \infty} I(\hat{\theta}) = I(\theta_0)$$

Score Test

Under some regularity conditions the score itself has an asymptotic normal distribution with mean 0 and variance equal to the information matrix, so that

$$u(\theta) \sim N(0, I_n(\theta))$$

Score Test Statistcs

$$u(\theta_0) \sim N(0, I_n(\theta_0))$$

$$\frac{\frac{\partial logL(\theta|x_1,\ldots,x_n)}{\partial\theta}}{\sqrt{nI(\theta_0)}} \sim N(0,1)$$

Score Test

When θ is a parameter of dimension p. Under some regularity conditions the score itself has an asymptotic normal distribution with mean 0 and variance-covariance matrix equal to the information matrix, so that

$$u(\theta) \sim N_p(0, I_n(\theta))$$

Score Test

This result provides another basis for constructing tests of hypotheses and confidence regions. For example under $H_0 : \theta = \theta_0$ for a fixed value θ_0, the quadratic form

$$Q = (u(\theta_0))\prime I^{-1}(\theta_0)u(\theta_0)$$

$$Q = (u(\theta_0))\prime I^{-1}(\theta_0)u(\theta_0)$$

has approximately in large samples a chi-squared distribution with p degrees of freedom.

Likelihood Ratio Tests

Under certain regularity conditions, minus twice the log of the likelihood ratio has approximately in large samples a chi-square distribution

$$-2 log \lambda = 2 log L(\hat{\theta}, x) - 2 log L(\theta_0, x)$$

In case θ is a scalar, $-2 log \lambda$ has a chi-square distribution with 1 degree of freedom.

Likelihood Ratio Tests

Under certain regularity conditions, minus twice the log of the likelihood ratio has approximately in large samples a chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models. Thus,

$$-2 log \lambda = 2 log L(\hat{\theta}_{\omega_2}, x) - 2 log L(\hat{\theta}_{\omega_1}, x)$$

where the degrees of freedom are $\nu = dim(\omega_2) - dim(\omega_1)$, the number of parameters in the larger model ω_2 minus the number of parameters in the smaller model ω_1.

Large Sample Tests

The LR, Wald, and Score tests (the "trinity" of test statistics) require different models to be estimated.

- LR test requires both the restricted and unrestricted models to be estimated
- Wald test requires only the unrestricted model to be estimated.
- Score test requires only the restricted model to be estimated.

Applicability of each test then depends on the nature of the hypotheses. For $H_0 : \theta = \theta_0$, the restricted model is trivial to estimate, and so LR or Score test might be preferred. For $H_0 : \theta > \theta_0$, restricted model is a constrained maximization problem, so Wald test might be preferred.

Large Sample Tests

The score test does not require $\hat{\theta}_n$ whereas the other two tests do.

Large Sample Tests

The score test does not require $\hat{\theta}_n$ whereas the other two tests do.

The Wald test is most easily interpretable and yields immediate confidence intervals.

Large Sample Tests

The score test does not require $\hat{\theta}_n$ whereas the other two tests do.

The Wald test is most easily interpretable and yields immediate confidence intervals.

The score test and likelihood ratio test are invariant under reparameterization, whereas the Wald test is not.

Large Sample Tests: $H_0 \, \theta = \theta_0$

Wald Test

Under the null, the normalised distance between $\hat{\theta}$ and θ_0 should be small.

Large Sample Tests: $H_0\,\theta = \theta_0$

Wald Test

Under the null, the normalised distance between $\hat{\theta}$ and θ_0 should be small.

Lagrange multiplier (or Score) test

Under the null, the score $u(\theta_0)$ evaluated at θ_0 should be close to zero.

Large Sample Tests: $H_0\,\theta = \theta_0$

Wald Test

Under the null, the normalised distance between $\hat{\theta}$ and θ_0 should be small.

Lagrange multiplier (or Score) test

Under the null, the score $u(\theta_0)$ evaluated at θ_0 should be close to zero.

Likelihood ratio test

The ratio between the likelihood $L(x, \hat{\theta})$ evaluated at at the estimate $\hat{\theta}$ and the likelihood $L(x, \theta_0)$ evaluated at θ_0 should be close to 1

Relationship between the three tests

From the Handbook Chapter 13 by Robert Engle

These three general principles have a certain symmetry which has revolutionized the teaching of hypothesis tests and the development of new procedures:

- Essentially, the Lagrance Multiplier approach starts at the null and asks whether movement toward the alternative would be an improvement
- The Wald approace starts at the alternative and considers movement toward the null
- The Likelihood ratio method compares the two hypothesis directly on an equal basis.

p-value

A common complaint about Hypothesis Testing is the he choice of the significance level α, that is essentially arbitrary. To make matters worse, the conclusions (Reject or DonÂt Reject H) depend on what value of α is used.

p-value

Definition: A p-value $p(x)$ is a test statistic satisfying $0 \leq p(x) \leq 1$ for every sample point x. Small values of $p(x)$ give evidence that H_1 is true.

Theorem: Let $W(X)$ be a test statistic such that large value of W give evidence that H_1 is true. For each sample point x, define

$$p(x) = \sup_{\theta \in \Theta_0} P_\theta(W(X) \geq W(x))$$

then $p(X)$ is a valid p-value.

Steps to find p-value

1. Let T be the test statistic.
2. Compute the value of T using the sample $x_1, \ldots, x_n$. Say it is a.
3. The p-value is given by

$$p - value = \begin{array}{cc} P(T < a|H_0), & \textit{if lower tail test} \\ P(T > a|H_0), & \textit{if upper tail test} \\ P(|T| > |a||H_0), & \textit{if two tail test} \end{array}$$

p-value

- The smallest level at which H_0 would be rejected.
- The the probability that you would obtain evidence against H_0 which is at least as strong as that actually observed, if H_0 were true.
- The smaller the $p - value$, the stronger the evidence that H_0 is false.

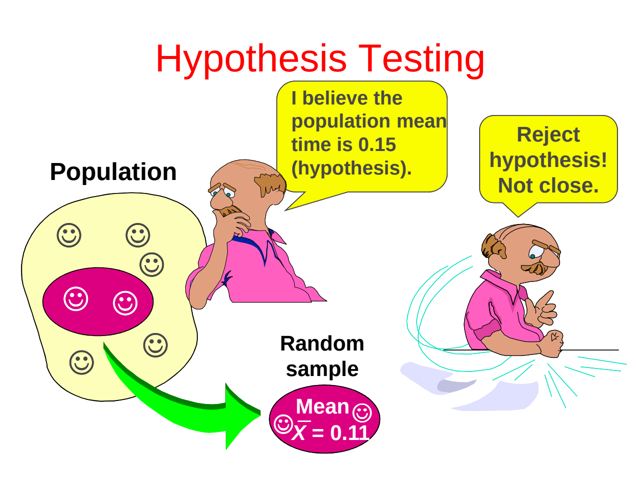
Example: psychiatrist problem

A psychiatrist wants to determine if a small amount of coke decreases the reaction time in adults.
The reaction time for a specified test is assumed normally distributed and has mean equal to $\mu = 0.15$ seconds and $\sigma = 0.2$.
A sample of 81 people were given a small amount of coke and their reaction time averaged 0.11 seconds.
Does this sample result support the psychiatrist's hypothesis?

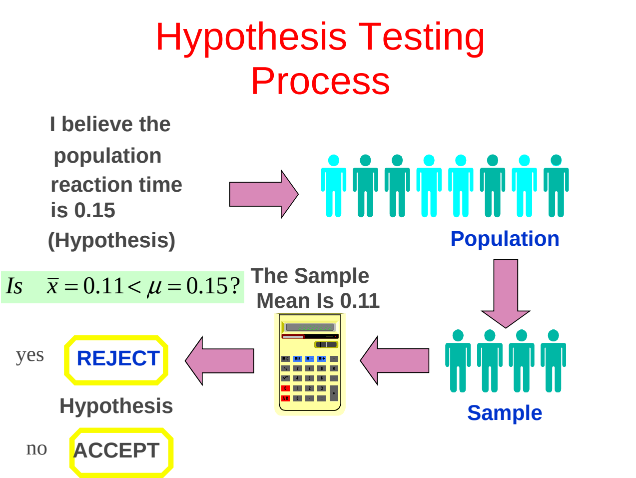
Example: psychiatrist problem

- The reaction time is expected to decrease
- What are null and alternative hypothesis?
- $H_0 : \mu = 0.15 \qquad H_1 : \mu < 0.15$
- If the null hypothesis were true, could I observe a sample mean of 0.11 from a sampling distribution centered in 0.15 and standard deviation $\sigma = 0.2$?
- $$P\left(\bar{X} < 0.11 | \mu = 0.15\right) = P\left(\frac{\bar{X} - 0.15}{\frac{0.2}{81}} < \frac{0.11 - 0.15}{\frac{0.2}{81}}\right) = 0.0359$$

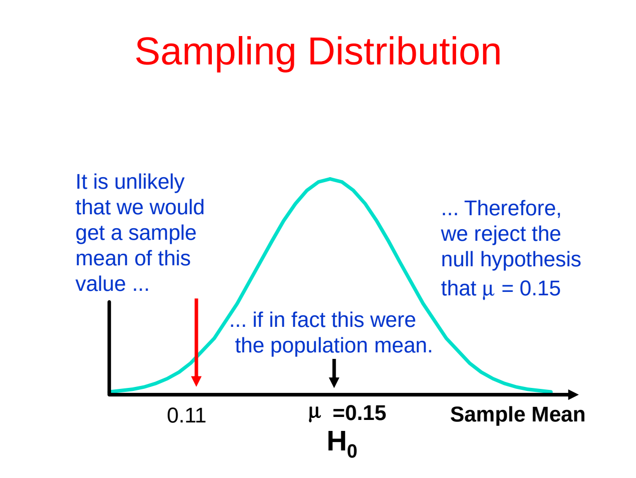
Example: psychiatrist problem



Example: psychiatrist problem



Example: psychiatrist problem



p-value: Example: psychiatrist problem

- 0.0359 represents the probability to observe a sample mean with a more extreme value than 0.11.
- Probability of obtaining a test statistic as extreme as, or more extreme, ($\leq$ *or*, $\geq$) than the actual sample value, given H_0 is true
- If we think the p-value is too low to believe the observed test statistic is obtained by chance only, then we would reject chance (reject the null hypothesis).
- Otherwise, we fail to reject chance and do not reject the null hypothesis.

Rejection Region: Example: psychiatrist problem

$$R = \left\{ (x_1, x_2, \ldots, x_n) : \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} < -z_{1-\alpha} \right\}$$

Rejection Region: Example: psychiatrist problem

- $\alpha = 0.05$

$$R = \left\{ (x_1, x_2, \ldots, x_n) : \frac{\bar{x} - 0.15}{\sqrt{\frac{0.2^2}{81}}} < -1.645 \right\}$$

$-1.8 < -1.645$ Reject null hypothesis

- $$R = \left\{ (x_1, x_2, \ldots, x_n) : \frac{\bar{x} - 0.15}{\sqrt{\frac{0.2^2}{81}}} < -2.32 \right\}$$

$-1.8 < -2.32$ Accept null hypothesis

Rejection Region: Example: psychiatrist problem

$$R = \left\{ (x_1, x_2, \ldots, x_n) : \bar{x} < \mu_0 - z_{1-\alpha} \sqrt{\frac{\sigma^2}{n}} \right\}$$

Rejection Region: Example: psychiatrist problem

- $\alpha = 0.05$

$$R = \{(x_1, x_2, \ldots, x_n) : \bar{x} < 0.1134\}$$

  Reject null hypothesis

- $$R = \{(x_1, x_2, \ldots, x_n) : \bar{x} < 0.0984\}$$

  Accept null hypothesis

Example: psychiatrist problem: Decision

- p-value$= 0.0359$ represents the probability to observe a sample mean with a more extreme value than 0.11
- If $\alpha = 0.05$, we reject null hypothesis, coke does have effect on time reaction (coke decreases reaction time)
- If $\alpha = 0.01$, we DO NOT reject null hypothesis, coke does NOT have effect on time reaction

Application to Poisson Distribution

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables distributed as a Poisson distribution with parameter λ

$$f(x; \lambda) = \frac{\lambda^x exp(-\lambda)}{x!}$$

We wish to test $H_0 : \lambda = 6$ versus $H_1 : \lambda \neq 6$
We observe a random sample $(X_1, \ldots, X_{100})$, such that $\sum_i x_i = 550$

Application to Poisson Distribution

Observe that the joint pdf of $X = (X_1, \ldots, X_n)$ is give by

$$
\begin{aligned}
f(x_1, \ldots, x : n; \lambda) &= \prod_i \frac{\lambda_i^x exp(-\lambda)}{x_i!} \\
L(\lambda) &= \frac{\lambda^{\sum_i x_i} exp(-n\lambda)}{\prod_i x_i!} \\
l(\lambda) &= \sum_i x_i log(\lambda) - n\lambda - log\left(\prod_i x_i!\right) \\
\frac{dl(\lambda)}{\lambda} &= \frac{\sum_i x_i}{\lambda} - n \\
\frac{d^2 l(\lambda)}{\lambda^2} &= -\frac{\sum_i x_i}{\lambda^2} \\
\hat{\lambda} &= \bar{x} = 5.5
\end{aligned}
$$

Application to Poisson Distribution

$$
\begin{aligned}
l(\lambda) &= \sum_i x_i log(\lambda) - n\lambda - log\left(\prod_i x_i!\right) \\
\frac{dl(\lambda)}{\lambda} &= \frac{\sum_i x_i}{\lambda} - n \\
\frac{d^2 l(\lambda)}{\lambda^2} &= -\frac{\sum_i x_i}{\lambda^2} \\
I_n(\lambda) &= -E\left(\frac{d^2 l(\lambda)}{\lambda^2}\right) = -\frac{n}{\lambda} \\
\hat{\lambda} &= \bar{x}
\end{aligned}
$$

Poisson- Large sample test

- WALD Test statistics

$$\frac{\bar{x}-\lambda_0}{\sqrt{\frac{\bar{x}}{n}}}$$

approximate distribution $N(0,1)$

-

$$\frac{\bar{x}-\lambda_0}{\sqrt{\frac{\bar{x}}{n}}} = \frac{5.5-6}{\sqrt{\frac{5.5}{100}}} = -2.13$$

p-value=0.0332

Poisson- Large sample test

- Score test statistics

$$\frac{\frac{\sum_i x_i}{\lambda_0} - n}{\sqrt{\frac{n}{\lambda_0}}}$$

  approximate distribution $N(0,1)$

- $$\frac{\frac{\sum_i x_i}{\lambda_0} - n}{\sqrt{\frac{n}{\lambda_0}}} = -2.0412$$

  p-value=0.04146

Poisson- Large sample test

- Likelihood Ratio Test Statistics

$$
\begin{aligned}
2 \times \left(l(\hat{\lambda}) - l(\lambda_0) \right) &= 2 \times \left(\sum_i x_i log \frac{\hat{\lambda}}{\lambda_0} - n(\hat{\lambda} - \lambda_0) \right) \\
\Lambda(x) &= 2 \times \left(\sum_i x_i log \frac{\bar{x}}{\lambda_0} - n(\bar{x} - \lambda_0) \right) \\
&= 4.2875 \\
p - value &= 0.0384
\end{aligned}
$$

approximate distribution χ_1