Hypothesis Testing

$H_0 : \{P_\theta : \theta \in \Theta_0\}$ Null

$H_1 : \{P_\theta : \theta \in \Theta_1\}$ Alternative

Distinguish between two sets of probability measures

Test: $\varphi : \{Z_i\}_{i=1}^N \to \{0, 1\}$

(or $\{Z_i\}_{i=1}^N$ and other measurable random variables)

Usually constructed by using a test statistic

$T\left(\{Z_i\}_{i=1}^N\right)$ such that under H_0

it has approximately known distribution

with sets $\underset{\substack{\uparrow \\ \text{acceptance} \\ \text{region}}}{A}$ and $\underset{\substack{\uparrow \\ \text{rejection} \\ \text{region}}}{A^c}$

We define a test as $\varphi = \mathbb{1}\{T \in A^c\}$

We want a test ideally such that

$\varphi = 1$ when $P \in H_1$ and 0 when $P \in H_0$

with high probabilities

- Size of a test is $\sup_{P \in H_0} \underbrace{P(\varphi = 1)}_{\text{Type I error}}$

we often call this α Type I err

want it to be small

~ Power of a test at $P \in H_1$

is $1 - \underbrace{P(\varphi = 0)}_{\text{Type II error}}$

we want power to be large

Scalar tests: will apply to individual coefficient in a linear regression

Suppose under H_0: we have T such that

$T \xrightarrow{d} Z \sim F$ some, known continuous CDF

α quantile of Z is value q_α

such that $F(q_\alpha) = \alpha$

Then $\Pr\left(q_{\alpha/2}(Z) \leq T \leq q_{1-\alpha/2}(Z)\right) \to 1-\alpha$

So $A = \left\{ \left[q_{\alpha/2}(Z),\, q_{1-\alpha/2}(Z)\right] \ni T \right\}$

is acceptance region

This guarantees that size of the test is asymptotically α. We also say such a test has (asymptotic) significance level α

Any region with probability $1-\alpha$ would do this including any pair of quantiles

eg: - $[q_\alpha(Z), \infty)$ left tail

- $(-\infty, q_{1-\alpha}(Z)]$ right tail

"one tailed tests"

T-test for OLS coefficients

$H_0: \beta_j = \beta_j^*$ (and all linear model assumptions)

$H_1: \beta_j \neq \beta_j^*$ (and all linear model assumptions)

Under H_0: $\sqrt{N}(\hat{\beta}_j - \beta_j^*) \xrightarrow{d} \tilde{Z} \sim N(0, V_{jj})$

where V is $\Sigma_{XX}^{-1} E[X_i \varepsilon_i \varepsilon_i X_i'] \Sigma_{XX}^{-1}$

[$E\varepsilon_i X_i' = 0$, $\{\varepsilon_i, X_i\}_{i=1}^N$ iid, moments exist all assumed]

V_{jj} is j,j^{th} element of V matrix

- Because V is unknown, standardize by an estimate

$\hat{V}_{jj}$, eg $\hat{V}_{jj}^{robust} = \left[\left(\frac{1}{N}\sum x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum x_i x_i' \hat{\varepsilon}_i^2\right)\left(\frac{1}{N}\sum x_i x_i'\right)^{-1}\right]_{jj}$

Then

$$t = \frac{\sqrt{N}(\hat{\beta}_j - \beta_j^*)}{\sqrt{\hat{V}_{jj}}} \xrightarrow{d} \frac{\tilde{Z}}{\sqrt{V_{jj}}} = Z \sim N(0,1)$$

by Slutsky, continuous mapping

"t-statistic"

- Reject H_0 if $t \notin [q_{\alpha/2}(Z), q_{1-\alpha/2}(Z)]$

- Under H_1: $\beta_j = \beta_j^a \neq \beta_j^*$ for some β_j^a

$$t = \underbrace{\underbrace{\frac{\sqrt{N}(\hat{\beta}_j - \beta_j^a)}{\sqrt{\hat{V}_{jj}}}}_{\xrightarrow{d} N(0,1)} + \underbrace{\left\{\frac{\sqrt{N}(\beta_j^a - \beta_j^*)}{\sqrt{\hat{V}_{jj}}}\right\}}_{\xrightarrow{p} \text{constant} \neq 0 \quad \to \infty \text{ or } -\infty}}_{\to \infty \text{ or } -\infty}$$

does not converge in distribution to anything finite.

So if $\beta_j \neq \beta_j^*$

$$P\left(t \notin \left[q_{\alpha/2}(z),\ q_{1-\alpha/2}(z)\right]\right) \to 1$$

So power is asymptotically 1 for any finite critical region.

That analysis is not very accurate for finite N

Failure to reject H_0 when it is false does happen in finite samples especially when β_j is close to β_j^*

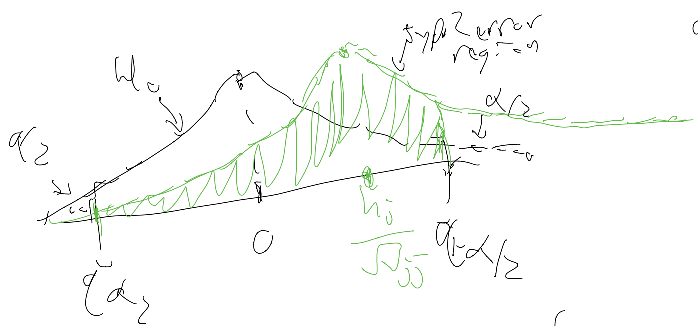
Consider power under "local alternatives"

Under H_1: $\beta = \beta^a = \beta^* + \frac{h}{\sqrt{N}}$

$$\sqrt{N}(\hat{\beta} - \beta^*) = \underbrace{h + \underbrace{\left(\frac{1}{N}\sum x_i x_i'\right)^{-1} \frac{1}{\sqrt{N}} \sum x_i \varepsilon_i}_{\xrightarrow{d} N(0,V)}}_{\xrightarrow{d} N(h,V)}$$

For t-test, under H_1, $t \xrightarrow{d} N\left(\frac{h_j}{\sqrt{v_{jj}}}, 1\right)$

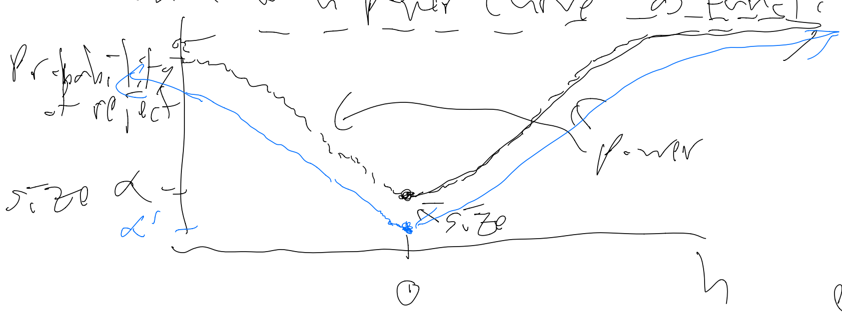
Power equals 1 − Pr (this distribution hits acceptance region)



$$\text{Power is } 1 - \Phi\left(q_{1-\alpha/2} - \frac{h_j}{\sqrt{v_{jj}}}\right) - \Phi\left(q_{\alpha/2} - \frac{h_j}{\sqrt{v_{jj}}}\right)$$

where Φ is CDF of standard Normal

Can draw a "power curve" as function of h



Trade-off of size and power: small α reduces type I errors but also shifts power curve down

In theory, choose α to accord with value of type I and type II errors

In practice RA Fisher said "use $\alpha = 0.05$" and it stuck

---

Confidence Interval

A region $[a, b]$ such that

$$\Pr([a,b] \ni \beta_j) \approx 1-\alpha$$

is called an (approximate) $1-\alpha$ confidence interval

$[a, b]$ are random: this is a statistic.

Use that $\Pr\left(q_{\alpha/2}(Z) \leq \frac{(\hat{\beta}_j - \beta_j)}{\sqrt{\hat{V}_{jj}}/\sqrt{N}} \leq q_{1-\alpha/2}(Z)\right) \to 1-\alpha$

equivalent to

$$\Pr\Big(\hat{\beta}_j - q_{1-\alpha/2}(Z) \cdot \underbrace{\frac{\sqrt{\hat{V}_{jj}}}{\sqrt{N}}}_{\text{"standard error"}} \leq \beta_j \leq \hat{\beta}_j - q_{\alpha/2}(Z) \frac{\sqrt{\hat{V}_{jj}}}{\sqrt{N}}\Big)$$

$\to 1-\alpha$ for any distribution satisfying OLS assumption

So interval $\left[\hat{\beta}_j - q_{1-\alpha/2}\widehat{SE},\ \hat{\beta}_j - q_{\alpha/2}\widehat{SE}\right]$

is an approximate $1-\alpha$ confidence interval

If $\alpha = 0.05$ and we use $N(0,1)$ distribution

Then $q_{1-\alpha/2} = 1.96 \approx 2$

$q_{\alpha/2} = -1.96 \approx -2$

So this is $\hat{\beta}_j \pm 1.96\, \widehat{SE}(\hat{\beta}_j)$

In general, can construct $1-\alpha$ confidence interval by choosing parameters not rejected by a level α test.

Multivariate Hypotheses

Covariant linear combination of parameters.

still use linear model $Y = X\beta + \varepsilon$

$$H_0: R\beta^* - r = 0$$

$$H_1: R\beta^* - r \neq 0$$

where R is an $h \times K$ matrix and β^* is the true parameter vector, and r is $h \times 1$ vector

Each row of R is a linear restriction

e.g. to test $\beta_1 = 1$, and $\beta_2 - \beta_3 = 2$ when $k=4$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix} \qquad r = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Applications: - single coefficient tests: same hypothesis as a t-test.

- Single equation, eg $H_0: \beta_1 + \beta_2 = 1$ (constant returns to scale, for example)

- "Joint significance for group of variables"

$$H_0: \beta_{j+1} = 0, \text{ and } \beta_{j+2} = 0 \ \dots \text{ and } \beta_{j+r} = 0$$

$$H_1: \beta_{j+1} \neq 0 \ \underline{\text{or}} \ \beta_{j+2} \neq 0, \ \dots \ \text{or } \beta_{j+r} \neq 0$$

At least one coefficient in group is significant.

often done for all coefficients, or all except constant term
asks if regression equation has any predictive power.

or $X\beta$ contains $\underbrace{w\beta_1 + w^2\beta_2 + w^3\beta_3} + \dots$

test if $\beta_1, \beta_2, \beta_3 = 0$ jointly
tests if w is related to y

Preliminaries

$X \sim N(\mu, \Omega)$ is multivariate normal in $\mathbb{R}^N$

properties:

- $X_i \sim N(\mu_i, \Omega_{ii})$ for each $i = 1 \dots N$
  pdf is $\frac{1}{\sqrt{2\pi\Omega_{ii}}} e^{-\frac{(x-\mu_i)^2}{2\Omega_{ii}}}$
- $\mathrm{cov}(X_i, X_\ell) = \Omega_{i\ell}$
- iff $z'X \sim N(z'\mu, z'\Omega z)$
  for any $z \in \mathbb{R}^N$

Can produce confidence ellipsis for (X_i, X_ℓ) in $\mathbb{R}^2$
generalization of an interval:

$\{z : \sum a_i z_i^2 \le C\}$, for some a_i's, C

Multivariate standard normal: $Z \sim N(0, I_N)$

each Z_i is standard normal $N(0,1)$
and independent of others.

For a multivariate normal, if z_i, z_j multivariate normal and $\mathrm{Cov}(z_i, z_j) = 0$, they are independent.

If $A \in \mathbb{R}^{k \times N}$ then $Ax \sim N(A\mu, A\Omega A')$

multivariate normality is preserved by linear and affine transformations

Chi-squared distribution

Let $z \sim N(0, I)$ be m-dimensional standard normal

$y \sim \chi^2_m$ "is distributed chi-squared with m degrees of freedom"

if $y = \sum_{i=1}^m z_i^2$ sum of squared independent standard normals
$= z'z$

Facts

If $x \sim N(0, \Omega) \in \mathbb{R}^m$

Then $x'\Omega^{-1}x \sim \chi^2_m$

Proof sketch: $\Omega^{-1} = A'A \quad A = \Omega^{-1/2}$

$$\Omega^{-\frac{1}{2}} x \sim N(0, \underbrace{\Omega^{-\frac{1}{2}} \Omega \Omega^{-\frac{1}{2}'}}_{\Omega^{-\frac{1}{2}} \Omega^{\frac{1}{2}} \Omega^{\frac{1}{2}} \Omega^{-\frac{1}{2}}})$$

$$\underbrace{\Omega^{-\frac{1}{2}} \Omega^{\frac{1}{2}}}_{I}$$

If $z \sim N(0, I_m)$ and P is an orthogonal projection matrix with rank r, then

$$z' P z \sim \chi^2_r$$

Proof involves eigendecomposition of P
(or prove)

$$P = V D_r V'$$

r ones, $m-r$ zeros

where $D_r = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$

$(m-r) \times (m-r)$

and $V'V = I_m$ (orthogonality)

so $V'z \sim N(0, V'V)$
$\sim N(0, I_m)$

so $z'Pz$ is sum of r squared independent normals

so projecting reduces degrees of freedom.

Noncentral χ^2 distribution

- let $z \sim N(\mu, I)$ be m-dimensional normal

Then $z'z \sim \chi^2_m(\lambda)$ "noncentral χ^2 with m degrees of freedom and noncentrality parameter $\lambda = \mu'\mu$

Facts: If $x \sim N(\mu, \Omega)$ in $\mathbb{R}^m$
then $x'\Omega^{-1}x \sim \chi^2_m(\mu'\Omega^{-1}\mu)$