Multiparameter Testing for Gaussian Linear Models

Surya Tokdar

Effect of a categorical factor

- Recall chick weight data

  $$\text{weight}_i = \beta_1 + \beta_2\text{Diet}_2 + \beta_3\text{Diet}_3 + \beta_4\text{Diet}_4 + \beta_5\text{Time}_i + \epsilon_i$$

- Want to test Diet has no effect
- $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$.

Another look

- In a different parametrization:

  $$\text{weight}_i = \beta_1\text{Diet}_1 + \beta_2\text{Diet}_2 + \beta_3\text{Diet}_3 + \beta_4\text{Diet}_4 + \beta_5\text{Time}_i + \epsilon_i$$

- The same hypothesis is now represented by

  $$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

Multiple linear combinations

- These hypotheses relate to an important general class

  $$H_0 : a_1^T\beta = 0 \ \& \ a_2^T\beta = 0 \ \& \cdots \ \& \ a_r^T\beta = 0$$

- Or equivalently

  $$H_0 : A^T\beta = 0$$

  where A is the $p \times r$ matrix with columns $a_1, \cdots, a_r$.

ML calculations

- Let $\eta = A^T\beta$. Then the profile log-likelihood in η is

  $$\ell_y^*(\eta) = \text{const} - \frac{n}{2}\log\left[1 + \frac{(\eta - A^T\hat\beta_{\text{LS}})^T\{A^T(Z^TZ)^{-1}A\}^{-1}(\eta - A^T\hat\beta_{\text{LS}})}{(n-p)s_{y|z}^2}\right]$$

- With MLE $\hat\eta_{\text{MLE}}(y) = A^T\hat\beta_{\text{LS}}$. Call this $\hat\eta_{\text{LS}}$.

ML tests

- ML test rejects $H_0 : \eta = 0$ if

  $$\ell_y(0) < \ell_y(A^T\hat\beta_{\text{LS}}) - c^2/2$$

- Or equivalently, if

  $$F(y) = \frac{\hat\eta_{\text{LS}}^T\{A^T(Z^TZ)^{-1}A\}^{-1}\hat\eta_{\text{LS}}}{r \times s_{y|z}^2} > c_n$$

  with $c_n = \frac{n-p}{r}(e^{c^2/n} - 1)$

Distribution theory

- If (β, σ^2) is such that $A^T\beta = 0$ then
  1. $\hat{\eta}_{LS} = A^T\hat{\beta}_{LS} \sim N_r(0, \sigma^2 A^T(Z^TZ)^{-1}A)$
  2. So $\hat{\eta}_{LS}^T\{A^T(Z^TZ)^{-1}Z\}^{-1}\hat{\eta}_{LS}/\sigma^2 \sim \chi^2(r)$
  3. $(n-p)s^2_{y|z}/\sigma^2 \sim \chi^2(n-p)$
  4. These two quantities are independent
- So, $F(Y) \sim F(r, n-p)$: the F distribution with df r, $n-p$.

Size calculation

- Call $\delta_c(y)$ the ML test that rejects H_0 when $F(y) > c$.
- So power of δ_c equals $1 - F_{r,n-p}(c)$ over Θ_0
- And so size of δ_c is $1 - F_{r,n-p}(c)$.
- A size α test is found by taking $c = F^{-1}_{r,n-p}(1-\alpha)$.

The case of $r = 1$

- When $r = 1$, we can replace $A = a$, a p-dim vector
- $F(y) = \frac{(a^T\hat{\beta}_{LS})^2}{s^2_{y|z}/n_a}$
- $F(1, n-p) = t(n-p)^2$, i.e., $T \sim t(n-p)$ means $F = T^2 \sim F(1, n-p)$.
- So the size-α F-test for $H_0 : \eta = 0$ is same as

$$\text{reject } H_0 \text{ when } 0 \notin a^T\hat{\beta}_{LS} \mp z_{n-p}(\alpha)s_{y|z}/\sqrt{n_a},$$

  exactly as before.

Fixed level testing and P-value

- The works of R Fisher, J Neyman and E Pearson put classical hypothesis testing on a solid platform which led to the eventual acceptance and popularity of this quantitative technique in all scientific studies
- However Fisher and Neyman-Pearson differed in how to administer this

N-P's fixed level testing recipe

1. Choose a small positive fraction α, called *level of significance*, usually, 1%, 5% or 10%
2. Pick a test of size α with good power outside the null set (ML tests are great candidates for this, more on this later)
3. Reject or accept H_0 based on the outcome of this test
4. If rejected (or accepted) report H_0 rejected (accepted) at α level of significance

Fisher's critique

- Food data: $Y_i \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$, test $H_0 : \mu = 175$, level $= 5\%$
- 95% ML conf interval $\bar{y} \mp z_{n-1}(.05)s_y/\sqrt{n}$
- Observed: $\bar{y} = 143.64$, $s_y = 41.44$, $n = 22$
- $z_{21}(.05) = 2.08$
- Interval $= [125.26, 162.02]$: Reject H_0 at 5% level
- If instead: $\bar{y} = 156.54$ then interval $= [138.16, 174.92]$
- Same report reject H_0 at level 5%

No report of strength of evidence

- For either data we report the same
- But in the second we came close to taking the other decision
- Fail to convey strength of evidence against H_0

Fisher's recommendation: p-value

- Report the smallest α such that a size α (ML) test rejects H_0
- Call this number p-value
- The smaller the p-value the more evidence there is against H_0

Understanding p-value

Imagine an infinite number of testers, each using a different size α (ML) test. Together, they cover the whole range $\alpha \in (0, 1)$. The testers with smaller α are more conservative about H_0, they need to see more evidence against H_0 to reject it. Next you show your recorded data to all testers and each take a decision to reject/accept H_0. The most liberal testers, those with α very close to 1, would be quick to report "reject H_0" while the most conservative ones will stick to "accept H_0. In between, there's a point of switch, a value $\alpha_0(x)$ so that all testers with $\alpha \geq \alpha_0(x)$ have rejected H_0 and all testers with $\alpha < \alpha_0(x)$ have failed to reject H_0. This switch point is the p-value. The smaller the switch point, the more compelling the evidence against H_0 has been (converting more conservatives).

Operational details

- ML tests reject $H_0 : A^T\beta = 0$ if
$$F(y) = \frac{\hat{\eta}_{LS}^T\{A^T(Z^TZ)^{-1}A\}^{-1}\hat{\eta}_{LS}}{r \times s^2_{y|z}} > c$$
  with size $1 - F_{r,n-p}(c)$
- Calculate test statistic value $f = F(y)$
- p-value $= 1 - F_{r,n-p}(f) = P(F(Y) > f)$.

The $r = 1$ case: two-sided

- ML test reject $H_0 : a^T\beta = \eta_0$ if
$$T(y) = \frac{a^T\hat{\beta}_{LS} - \eta_0}{s_y/\sqrt{n}} < -c \text{ or } T(y) > c$$
  with size $2(1 - F_{n-1}(c))$.
- Calculate $t = T(y)$
- p-value $= 2(1 - F_{n-1}(|t|)) = P(|T(Y)| > |t|)$.

The $r = 1$ case: one-sided

- ML test reject $H_0 : a^T\beta \leq \eta_0$ if
$$T(y) > c$$
  with size $1 - F_{n-1}(c)$.
- Calculate $t = T(y)$
- If $t \leq 0$ then p-value is undefined (or you can take it to be 1)
- If $t > 0$ p-value $= 1 - F_{n-1}(t) = P(T(Y) > t)$.
- $H_0 : a^T\beta \geq \eta_0$ will be a mirror image of this