Section 13: Large sample hypothesis testing

ARE 210

November 28, 2017

1. Consider the linear regression model $Y_i|Z_i \sim N(\beta_0+\beta_1 Z_i, \sigma^2)$, where Z_i is a scalar, and consider the null hypothesis $\beta_1 = 0$.
   (a) What is θ_0? $Q_0(\theta)$? $Q_n(\theta)$? $\hat{\theta}_u$? $\hat{\theta}_c$? $a(\theta)$? $A(\theta)$? A_n? $\frac{\partial Q_0(\theta)}{\partial\theta}$? $\frac{\partial Q_n(\theta)}{\partial\theta}$? $\frac{\partial Q_n(\hat{\theta}_u)}{\partial\theta}$? $\frac{\partial Q_n(\hat{\theta}_c)}{\partial\theta}$? λ_n? Σ? Σ_n? Ψ? Ψ_n?

$$\theta_0^{\mathrm{T}} = (\beta_0, \beta_1, \sigma^2)$$
$$Q_0(\theta) = \mathbf{E}\left[-\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log\sigma^2 - \tfrac{(Y_i-\beta_0-\beta_1 Z_i)^2}{2\sigma^2}\right]$$
$$Q_n(\theta) = -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log\sigma^2 - \tfrac{\frac{1}{n}\sum_{i=1}^n (Y_i-\beta_0-\beta_1 Z_i)^2}{2\sigma^2}$$
$$\hat{\theta}_u^{\mathrm{T}} = (\overline{Y} - \hat{\beta}_1\overline{Z}, \tfrac{\hat{\mathrm{Cov}}(Y_i,Z_i)}{\hat{\mathrm{Var}}(Z_i)}, \tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Z_i)^2)$$
$$\hat{\theta}_c^{\mathrm{T}} = (\overline{Y}, 0, \tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i - \overline{Y}))$$
$$a(\theta) = \beta_1$$
$$A(\theta) = (0,1,0)$$
$$A_n = (0,1,0)$$
$$\frac{\partial Q_0(\theta)}{\partial\theta} = \begin{pmatrix} (\mathbf{E}[Y] - \beta_0 - \beta_1\mathbf{E}[Z])/\sigma^2 \\ \mathbf{E}[(Y-\beta_0-\beta_1 Z)Z]/\sigma^2 \\ -\frac{1}{2\sigma^2} + \frac{\mathbf{E}[(Y-\beta_0-\beta_1 Z)^2]}{2(\sigma^2)^2} \end{pmatrix}$$
$$\frac{\partial Q_n(\theta)}{\partial\theta} = \begin{pmatrix} (\overline{Y} - \beta_0 - \beta_1\overline{Z})/\sigma^2 \\ \frac{1}{n}\sum_{i=1}^n (Y_i-\beta_0-\beta_1 Z_i)Z_i/\sigma^2 \\ -\frac{1}{2\sigma^2} + \frac{\frac{1}{n}\sum_{i=1}^n (Y_i-\beta_0-\beta_1 Z_i)^2}{2(\sigma^2)^2} \end{pmatrix}$$
$$\frac{\partial Q_n(\hat{\theta}_u)}{\partial\theta}^{\mathrm{T}} = (0,0,0)$$
$$\frac{\partial Q_n(\hat{\theta}_c)}{\partial\theta}^{\mathrm{T}} = (0, \hat{\mathrm{Cov}}(Y_i,Z_i)/\sigma^2, 0)$$
$$\lambda_n = \hat{\mathrm{Cov}}(Y_i,Z_i)/\sigma^2$$
$$\Sigma = -\Psi = \begin{pmatrix} 1/\sigma^2 & \mathbf{E}[Z]/\sigma^2 & 0 \\ \mathbf{E}[Z]/\sigma^2 & \mathbf{E}[Z^2]/\sigma^2 & 0 \\ 0 & 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$
$$\Sigma_n = -\Psi_n = \begin{pmatrix} 1/\hat{\sigma}^2 & \overline{Z}/\hat{\sigma}^2 & 0 \\ \overline{Z}/\hat{\sigma}^2 & \overline{Z^2}/\hat{\sigma}^2 & 0 \\ 0 & 0 & \frac{1}{2\hat{\sigma}^4} \end{pmatrix}$$

(b) Derive the LR test and its asymptotic distribution.

$$\begin{aligned}\text{LR} &= 2n(Q_n(\hat{\theta}_u) - Q_n(\hat{\theta}_c)) \\ &= n \log \frac{\hat{\sigma}_c^2}{\hat{\sigma}_u^2}\end{aligned}$$

Here, we use the fact that $\hat{\sigma}_c^2 = \hat{V}[Y]$, and $\hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Z_i)^2 = \hat{V}[Y] - \hat{\beta}_1^2 \hat{V}[Z]$. Making these substitutions yields

$$\text{LR} = n \log \left(1 + \frac{\hat{\beta}_1^2 \hat{V}[Z]}{\hat{\sigma}_u^2} \right)$$

Asymptotically under the null (with an application of the delta method), we can approximate this LR $\approx n\frac{\hat{\beta}_1^2 \hat{V}[Z]}{\hat{\sigma}_u^2}$. From previous results on OLS, we know $\sqrt{n}\frac{\hat{\beta}_1 \hat{V}[Z]}{\hat{\sigma}_u^2} \xrightarrow{d} N(0,1)$, so LR $\xrightarrow{d} \chi_1^2$.

(c) Derive the Wald test and its asymptotic distribution.

$\text{WS} = na(\hat{\theta}_u)^{\text{T}} V_{\theta_0}[a(\hat{\theta}_u)]^{-1} a(\hat{\theta}_u)$. Here, we use $a(\hat{\theta}_u) = \hat{\beta}_1$, so $\hat{V}[a(\hat{\theta}_u)]^{-1} = \hat{V}[Z]/\hat{\sigma}_u^2$. Making this substitution yields WS $= n\frac{\hat{\beta}_1^2 \hat{V}[Z]}{\hat{\sigma}_u^2}$. As above, we can see WS $\xrightarrow{d} \chi_1^2$.

(d) Derive the score test and its asymptotic distribution.

$\text{ST} = n(\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta})^{\text{T}} \Sigma_n^{-1} (\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta})$ For this, we first use the fact that $(\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta})^{\text{T}} = (0, \hat{\text{Cov}}(Y,Z)/\hat{\sigma}_c^2, 0)$. Because of this, we only need to calculate $[\Sigma_n^{-1}]_{22} = \hat{\sigma}_c^2/\hat{V}[Z]$. Simplifying yields ST $= n\frac{\hat{\beta}_1^2 \hat{V}[Z]}{\hat{\sigma}_c^2}$. Again, we can see ST $\xrightarrow{d} \chi_1^2$.

(e) Compare the tests, do they differ in power? Size? How do their sizes compare to the t-test we discussed previously?

As a quick note, note that the LR test depends on $\hat{\theta}_u$ and $\hat{\theta}_c$, the Wald test depends on $\hat{\theta}_u$, and the score test depends on $\hat{\theta}_c$. As a consequence, which test is most appropriate may depend on the relative ease of estimating each of these quantities; while an LR test is often relatively easy with nested models estimated by maximum likelihood.

We can see that each test is identical except the LR. The LR has strictly lower power than the Wald test, since $\log(1+x)$ is always below x for positive x. The score test has strictly lower power than the Wald test, since $\hat{\sigma}_c$ is always larger than $\hat{\sigma}_u$ (however, if a degrees of freedom adjustment were used, this would no longer be true, and the two would have the same expectation under the null, although $\hat{\sigma_u}$ would have a higher variance). The Wald test has strictly higher power than the equivalent t-test. As a result, it will over reject in small samples relative to its asymptotic size. It is more difficult to compare the score test or the LR test to the equivalent t-test. However, it looks like the Wald is more aggressive than the score which is more aggressive than the LR test and the t-test. I'm not sure which of the t-test or the LR is more conservative.