HW1 (Due Sept 26, in D2L)

Gustavo

9/20/2021

Homework 1

Using the Gout data set:

1) Fit a linear model of the form `su~race+sex+age`, report your results, and summarize (in no more than three sentences) your conclusions.

2) Consider now expanding the model to inclue race-by-sex interactions.

- Explain with words what an interaction term different than zero means in this model.
- Fit the model with the interaction term, report your results and conclusions.

3) Consider now testing the hypothesis that sex has **any** effect on su, that is, whether sex has an effect that is the same for white and black people, or an effect that is different in black and white people.

- Describe the null and the alternative hypothesis,
- Test the null using `anova()`, and
- Summarize your findings.

4) Reproducing the results of the F-test:

- Review the F-statistic in the class notes and
- Develop a function that takes as input two `lm` objects and return a table identical to the one produced by `anova()`.
- Test your function using the H0 and Ha you used in Q3.

5) Wald's test

Like the F-test, Wald's test can also be used for tests involving 1 or more than 1 df. The test can be used with any null that can be expressed in linear form. The general form of the test is as follows:

- **Ha**: **y=Xb+e** (for this case use your Ha of Q3). Here, **y** is a nx1 vector (the *response*), **X** is an nxp incidence matrix for the pxy vector of effects **b**, and **e** is an nx1 error vector.
- **H0**: **Tb=a**, where **T** is a contrast matrix of dimensions qxp, and **a** is a qx1 vector (often **a=0**).

The covariance matrix of the contrast ($\hat{\mathbf{d}} = \mathbf{T}\hat{\mathbf{b}}$) is $Cov(\hat{\mathbf{d}}) = \mathbf{T}\mathbf{Cov}(\hat{\mathbf{b}})\mathbf{T}' = \mathbf{S}$, where $Cov(\hat{b})$ is the (co)variance matrix of estimates (Hint: use vcov(fm) to obtain it, here fm is the fitted alternative hypothesis).

Because of the CLT, in large samples, $\hat{\mathbf{d}} = \mathbf{T}\hat{\mathbf{b}}$ follows a multivariate normal distribution with (co)variance matrix **S**. Therefore, under the null, $(\hat{\mathbf{d}} - \mathbf{a})'\mathbf{S}^{-1}(\hat{\mathbf{d}} - \mathbf{a})$ follows a chi-square distribution with df equal to the rank of **T**.

- Create a function that Implement Wald's test (your function should take a fitted model, representing Ha, and a matrix of contrasts (T). The function should return the test-statistic, test DF, and the p-value.
- Test youf function for the test in 1.3, compare your p-value with that of the F-test.