Problem 4.9 - Sleep Equation Analysis

Problem 4.9

In Problem 3.3, we estimated the equation

$$\widehat{sleep} = 3{,}638.25 - 0.148\,totwrk - 11.13\,educ + 2.20\,age$$
$$(112.28) \quad\quad (0.017) \quad\quad\quad (5.88) \quad\quad (1.45)$$
$$n = 706,\ R^2 = 0.113,$$

where we now report standard errors along with the estimates.

(i) Is either *educ* or *age* individually significant at the 5% level against a two-sided alternative? Show your work.

Answer

For each of these, we do an individual test:

$$\begin{cases} H_0 : \beta_{educ} = 0 \\ H_1 : \beta_{educ} \neq 0 \end{cases} \implies t_{educ} = \frac{-11.13 - 0}{5.88} \approx -1.89.$$

Our critical values are $t_{0.025,702} \approx -1.96$ and $t_{0.975,702} \approx +1.96$. The statistic is not in the rejection region, so we fail to reject H_0 at the 5% level. For *age*, the statistic is ≈ 1.52, which leads to the same conclusion.

(ii) Dropping *educ* and *age* from the equation gives

$$\widehat{sleep} = 3{,}586.38 - 0.151\,totwrk$$
$$(38.91) \quad (0.017)$$
$$n = 706,\ R^2 = 0.103.$$

Are *educ* and *age* jointly significant in the original equation at the 5% level? Justify your answer.

Answer

We are requested to test:

$$\begin{cases} H_0 : \beta_{educ} = \beta_{age} = 0 \\ H_1 : H_0 \text{ is false} \end{cases}$$

We can do an F-test for this. The statistic is:

$$F = \frac{(R^2_{full} - R^2_{restricted})/q}{(1 - R^2_{full})/(n - k - 1)} = \frac{(0.113 - 0.103)/2}{(1 - 0.113)/(706 - 4)} \approx 3.97.$$

The critical value for $F_{2,702}$ at the 5% level is approximately 3.08, so we reject the null hypothesis that *educ* and *age* are jointly insignificant.

Note that the F-test statistic is equivalent to the alternative one

$$\begin{aligned}
F &= \frac{(R^2_{full} - R^2_{restricted})/q}{(1 - R^2_{full})/(n - k - 1)} \\
&= \frac{\left(\frac{SCE_{full}}{SCT} - \frac{SCE_{restricted}}{SCT}\right)/q}{(1 - \frac{SCE_{full}}{SCT})/(n - k - 1)} \\
&= \frac{(SCE_{full} - SCE_{restricted})/q}{(SCT - SCE_{full})/(n - k - 1)} \\
&= \frac{(SCE_{full} - SCE_{restricted} + SCT - SCT)/q}{(SCT - SCE_{full})/(n - k - 1)} \\
&= \frac{(SCR_{restricted} - SCR_{full})/q}{(SCR_{full})/(n - k - 1)}
\end{aligned}$$

Where we have used the equality $SCT = SCE_{full} + SCR_{full}$ and $SCT = SCE_{restricted} + SCR_{restricted}$.

(iii) Does including *educ* and *age* in the model greatly affect the estimated tradeoff between sleeping and working?

Answer

Not significantly. We have discussed endogeneity issues with great detail in the past so I will ommit that. For now, focus on the precision.

We could think of these from two effects outweighting each other. Including more variables in the models requires us to estimate more parameters. This typically leads to an increase in the variance of our estimates (less precision). On the same note, we are including variables, which we expect to have some degree of collinearity with the variable we are interested in (*totwrk*).

2

However, if those variables explain the outcome in a significant proportion, then the variance of the residual decreases, leading to more precise estimates.

(iv) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (i) and (ii)?

Answer

Our previous tests rely on estimates of the standard errors, which are based on the assumption of homoskedasticity. If there is heteroskedasticity, then these standard errors are generally biased, leading to incorrect conclusions in our hypothesis tests. In essence, even though the estimators of the parameters of interest (e.g., $\hat{\beta}_{totwrk}$) remain unbiased, the tests become invalid.