3. Using the data in RDCHEM, the following equation was obtained by OLS:

$$\widehat{rdintens} = 2.613 + .00030\ sales - .0000000070\ sales^2$$
$$(.429)\quad (.00014)\quad\quad (.0000000037)$$
$$n = 32, R^2 = .1484.$$

i. At what point does the marginal effect of *sales* on *rdintens* become negative?

ii. Would you keep the quadratic term in the model? Explain.

iii. Define *salesbil* as sales measured in billions of dollars: $salesbil = sales/1{,}000$. Rewrite the estimated equation with *salesbil* and $salesbil^2$ as the independent variables. Be sure to report standard errors and the *R*-squared. [*Hint*: Note that $salesbil^2 = sales^2/(1{,}000)^2$.]

iv. For the purpose of reporting the results, which equation do you prefer?

i) $\dfrac{\partial \widehat{rdintens}}{\partial sales} = 0.00030 - 0.000000014\, sales < 0$

$sales > 21428.57143$

→ when sale more than 21428.57143 → marginal effect on rdintens will be negative.

ii) $t = \dfrac{\hat{\beta}sales^2 - \beta\, sales^2}{std.\ error}$

$= \dfrac{-0.00000007 - 0}{0.0000000037}$

$= -18.92 \rightarrow H_0 : \beta sales^2 < 0$

iii) Define salesbil as sales measured in billions of dollars.

$rdintens = \hat{\beta}_0 + \hat{\beta}_1\, sales + \hat{\beta}_2\, sales^2$

$= \hat{\beta}_0 + \hat{\beta}_1 (1000 \times sales) + \hat{\beta}_2 (1000^2 \cdot salesbil)$

$= 2.613 + 0.003 (1000 \times sales) + 0.00000007 (1000^2 \cdot salesbil)^2$

$= 2.613 + 3\, sales + 0.07 \cdot salesbil^2$

$(0.429)\quad (0.14)\quad\quad (0.0037)$

$n = 32,\ R^2 = 0.1484$

$salesbil = \dfrac{sales}{1000}$

$sales = 1{,}000 \times salesbil$

iv) for the

1. Using the data in SLEEP75 (see also Problem 3 in Chapter 3), we obtain the estimated equation

$$\begin{aligned}\widehat{sleep} &= \underset{(235.11)}{3{,}840.83} - \underset{(.018)}{.163}\, totwrk - \underset{(5.86)}{11.71}\, educ - \underset{(11.21)}{8.70}\, age \\ &\quad + \underset{(.134)}{.128}\, age^2 + \underset{(34.33)}{87.75}\, male \\ n &= 706,\ R^2 = .123,\ \overline{R}^2 = .117.\end{aligned}$$

The variable *sleep* is total minutes per week spent sleeping at night, *totwrk* is total weekly minutes spent working, *educ* and *age* are measured in years, and *male* is a gender dummy.

i. All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

ii. Is there a statistically significant tradeoff between working and sleeping? What is the estimated tradeoff?

iii. What other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping?

i. All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

$H_0 = 0$

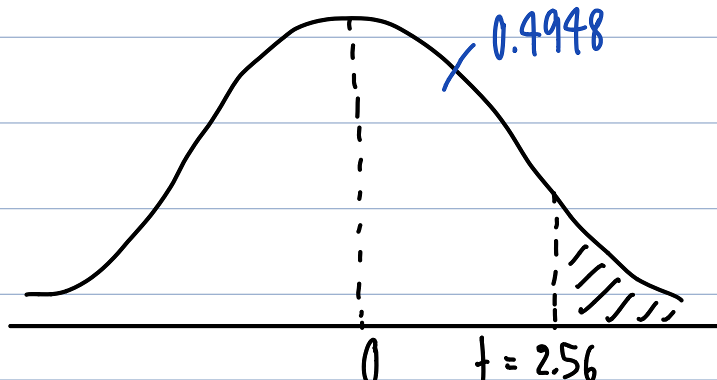
$H_a \neq 0$

to test whether X_j has a statiscally significant impact on Y.

$$t_{d.f} = \frac{\hat{\beta}_j - a_j}{s.e\ a_j} = \frac{\text{estimate} - \text{hypothesized value}}{s.e.}$$

$$t_{male} = \frac{\hat{\beta}_{male} - \beta_{male}}{se.(\hat{\beta}_{male})}$$

$$= \frac{87.75 - 0}{34.33} = 2.556$$



Since, $n-k-1 > 90 \rightarrow$ use z table.

P-value $= 0.5 - 0.4948 = 0.0052$

Given 1% significant level $\rightarrow \frac{0.01}{2} = 0.005$

So, P-value > significant level that is $0.0052 > 0.005$
→ men sleep more than women at about 1% level of significant.

ii. Is there a statistically significant tradeoff between working and sleeping? What is the estimated tradeoff?

$$t_{d.f} = \frac{\hat{\beta}_j - a_j}{s.e\ a_j}$$

$$\rightarrow t_{totwrk} = \frac{\hat{\beta}_{totwrk} - \beta_{totwrk}}{s.e.\ totwrk}$$

$$= \frac{-0.163 - 0}{0.018} = -9.055$$ statistically significant

To gain an additional hour of work, it would cause $0.163(60) = 9.78$ minutes of sleep time.

iii. What other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping?

→ we to test multiple regression by F-test, $H_0: \hat{\beta}_3 = 0, \hat{\beta}_4 = 0$, H_a: age has effect on sleeping: $\hat{\beta}_3 = 0$ and $\hat{\beta}_4 \neq 0$ or $\hat{\beta}_3 \neq 0$ and $\hat{\beta}_4 = 0$ or $\hat{\beta}_3 \neq 0$ and $\hat{\beta}_4 \neq 0$.

8. Suppose you collect data from a survey on wages, education, experience, and gender. In addition, you ask for information about marijuana usage. The original question is: "On how many separate occasions last month did you smoke marijuana?"

i. Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, "Smoking marijuana five more times per month is estimated to change wage by *x*%."

i) $\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u$

$100\beta_1$ is the percentage change in wage when marijuana usage rise by 1 time.

ii. Write a model that would allow you to test whether drug usage has different effects on wages for men and women. How would you test that there are no differences in the effects of drug usage for men and women?

$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female \cdot usage + u.$

$H_0 : \beta_6 = 0$

$H_a : \beta_6 \neq 0$

iii. Suppose you think it is better to measure marijuana usage by putting people into one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and heavy user (more

(δ_1) = 1; nonuser (1 to 5 times per month)
0 ; if not a nonuser.

(δ_2) 1 ; moderate (6 to 10 times per month)
0 ; if not a moderate users.

(δ_3) 1; heavy user (more than 10 times per month)
0; if not a heavy users.

$\log(wage) = \beta_0 + \delta_1 nonuser + \delta_2 moderate + \delta_3 heavy + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$

iv. Using the model in part (iii), explain in detail how to test the null hypothesis that marijuana usage has no effect on wage. Be very specific and include a careful listing of degrees of freedom.

$$\left.\begin{array}{l} H_0 : \delta_1 = \delta_2 = \delta_3 = 0 \\ q = 3 \text{ restrictions} \\ n = \text{Sample size} \\ \text{d.f.} = \text{Unrestricted models} \end{array}\right\} \text{we gain } F_{q,n-8} \text{ distribution}$$

v. What are some potential problems with drawing causal inference using the survey data that you collected?

→ Factors : we need to hold other fixed in order to get precise outcome.