EXERCISE 1

1) The GENDER variable is categorical with 2 levels
It is encoded with 1 dummy variable, i.e.

$$GENDER1_i = \begin{cases} 1 & \text{if individual } i \text{ is female } (GENDER_i = 1) \\ 0 & \text{otherwise} \end{cases}$$

The RACE variable is categorical with 3 levels.
It is encoded with 2 (= 3 - 1, to avoid collinearity) dummy variables.
Specifically, from the output we see that we have parameters associated with the levels 2 and 3 (RACE2, RACE3), hence RACE = 1 is the baseline.
Thus

$$RACE2_i = \begin{cases} 1 & \text{if } RACE_i = 2 \text{ (individual } i \text{ is hispanic)} \\ 0 & \text{otherwise} \end{cases}$$

$$RACE3_i = \begin{cases} 1 & \text{if } RACE_i = 3 \text{ (individual } i \text{ is white)} \\ 0 & \text{otherwise} \end{cases}$$

2) Gaussian linear model
denoting with y_i the wage of individual i

$$Y_i = \beta_1 + \beta_2 EDU_i + \beta_3 SOUTH1_i + \beta_4 GENDER1_i + \beta_5 EXPER_i + \beta_6 UNION1_i + \beta_7 AGE_i + \beta_8 RACE2_i + \beta_9 RACE3_i + \beta_{10} MARR1_i + \varepsilon_i$$

$\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

The model's assumptions are:
(i) normality, homoscedasticity, independence $\Rightarrow$ $\varepsilon_i \sim N(0,\sigma^2)$ iid for $i = 1, \dots, 534$
(ii) linearity w.r.t. $\beta_1, \dots, B_{10}$
(iii) the covariates are linearly independent

3) t-value of β_6 (UNION1)

$$t^{obs} = \frac{\hat{\beta}_6 - 0}{\sqrt{\hat{var}(\hat{\beta}_6)}} = \frac{\hat{\beta}_6}{\hat{se}(\hat{\beta}_6)} = \frac{1.4336}{0.5087} = 2.818$$

p-value of the test $\begin{cases} H_0: \beta_6 = 0 \\ H_1: \beta_6 \neq 0 \end{cases}$

in the table

$$\text{p-value} = \mathbb{P}_{H_0}(|T| > |t^{obs}|) = 2 \cdot \mathbb{P}_{H_0}(T > |t^{obs}|) = 2 \cdot \mathbb{P}_{H_0}(T > 2.818) = 2 \cdot (1 - 0.9975) = 0.005$$

where $T = \dfrac{\hat{\beta}_6(\underline{Y})}{\hat{se}(\hat{\beta}_6(\underline{Y}))}$ $\quad T \overset{H_0}{\sim} t_{n-p} = t_{524}$

Estimate of β_9 (RACE3)

$$t^{obs} = \frac{\hat{\beta}_3}{\hat{se}(\hat{\beta}_3)} \Rightarrow \hat{\beta}_3 = t^{obs} \cdot \hat{se}(\hat{\beta}_3) = 1.66 \cdot 0.5860 = 0.9727$$

4) EDUCATION is numeric
Hence β_2 represents the (additive) change in the expected wage for an additional year of education, keeping the other covariates fixed
In other words, for every additional year of education, the mean wage increases of 1.26 \$, with all other variables held constant.

RACE is categorical
I consider two individuals j and k such that
$RACE_j = 1$ and $RACE_k = 2$
while all other covariates are equal (i.e., $EDU_j = EDU_k$, $SOUTH_j = SOUTH_k, \dots$)

$$\mu_j = \beta_1 + \beta_2 EDU_j + \dots + \beta_7 AGE_j + \beta_8 \underset{=0}{RACE2_j} + \beta_9 \underset{=0}{RACE3_j} + \beta_{10} MARR1_j$$

$$\mu_k = \beta_1 + \beta_2 EDU_k + \dots + \beta_7 AGE_k + \beta_8 \underset{=1}{RACE2_k} + \beta_9 \underset{=0}{RACE3_k} + \beta_{10} MARR1_k$$

I consider $\mu_k - \mu_j = \mathbb{E}[Y_k] - \mathbb{E}[Y_j]$

$$\mu_k - \mu_j = \cancel{\beta_1 - \beta_1} + \beta_2 \overset{0}{(\cancel{EDU_k - EDU_j})} + \dots + \beta_7 \overset{0}{(\cancel{AGE_k - AGE_j})} + \beta_8 \overbrace{(RACE2_k - RACE2_j)}^{1} + \beta_9 \underset{0}{(\cancel{RACE3_k - RACE3_j})} + \beta_{10} \underset{0}{(\cancel{MARR1_k - MARR1_j})} = \beta_8$$

Hence β_8 represents the additive change in the mean hourly wage if I consider an individual in the hispanic population compared to an individual in the "other" population (keeping other covariates equal).

parameter associated with RACE2

Following a similar reasoning, β_9 (associated with RACE3) represents the additive change in the mean hourly wage if I consider an individual in the white population compared to an individual in the "other" population (keeping other covariates equal).

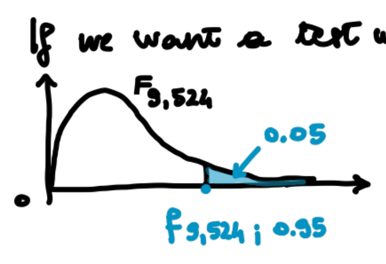
MARR1 is binary (categorical with 2 categories)
If I consider two individuals, identical for all covariates but the marital status, the married one has a mean hourly wage of 0.4563 \$ higher than the unmarried one.

5) $\begin{cases} H_0: \beta_2 = \beta_3 = \dots = \beta_{10} = 0 \\ H_1: \text{at least one } \beta_j \text{ is } \neq 0 \quad (j = 2, \dots, 10) \end{cases}$

I use the test statistic $F = \dfrac{R^2}{1-R^2} \cdot \dfrac{n-p}{p-1} = \dfrac{R^2}{1-R^2} \cdot \dfrac{524}{9}$ such that $F \overset{H_0}{\sim} F_{9,524}$

The observed value of the test statistic is $f^{obs} = \dfrac{0.2753}{1 - 0.2753} \cdot \dfrac{524}{9} = 22.117$

If we want a test with a 5% significance level



the reject region is $R = (f_{9,524;0.95}; +\infty) = (1.8977, +\infty)$
Since $f^{obs} \in R$, I reject H_0

6) The model is

$$Y_i = \gamma_1 + \gamma_2 EXPER_i + \gamma_3 GENDER1_i + \varepsilon_i$$

$\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

This model is nested to model A, hence I can compare them through a test

$$\begin{cases} H_0: \beta_2 = \beta_3 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0 \\ H_1: \text{at least one is } \neq 0 \end{cases}$$

The test is based on the statistic

$$F = \frac{SSE_B - SSE_A}{SSE_A} \cdot \frac{n - p_A}{p_A - p_B} = \frac{SSE_B - SSE_A}{SSE_A} \cdot \frac{534 - 10}{10 - 3} \qquad F \overset{H_0}{\sim} F_{7,524}$$

To compute the observed value, we first need to obtain SSE_B and SSE_A.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SSE_A = (\text{residual s.e.})^2 \cdot (n - 10) = 4.412^2 \cdot 524 = 10\,200.05$$

$$SSE_B = (\text{residual s.e.})^2 \cdot (n - 3) = 5.011^2 \cdot 531 = 13\,333.47$$

$$f^{obs} = \frac{13\,333 - 10\,200}{10\,200} \cdot \frac{524}{7} = 22.99$$

I reject H_0 for all usual significance level. I prefer model A.

7) No, because the R^2 always increase (or stays the same) when I add covariates.
I should use the adjusted R^2.

8) The inclusion of the interaction allows studying if the effect on the mean wage of an additional year of experience is different for men and women.

The model is

$$Y_i = \xi_1 + \xi_2 EXPER_i + \xi_3 GENDER1_i + \xi_4 EXPER{:}GENDER1_i + \varepsilon_i$$

$\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.
where

$$EXPER{:}GENDER1_i = \begin{cases} EXPER_i & \text{if } GENDER_i = 1 \\ 0 & \text{if } GENDER_i = 0 \end{cases}$$

If I consider a man, the expected wage is

$$\mathbb{E}[Y_i] = \xi_1 + \xi_2 EXPER_i$$

If I consider a woman, the expected wage is

$$\mathbb{E}[Y_i] = \xi_1 + \xi_2 EXPER_i + \xi_3 + \xi_4 EXPER_i = (\xi_1 + \xi_3) + (\xi_2 + \xi_4) EXPER_i$$

Hence ξ_4 is the change in the effect of an additional year of experience on the mean wage due to being a woman (compared to being a man)
In other terms: an additional year of experience leads to an increase of 0.809 \$ in the mean wage for a man, while it leads to an increase of (0.0809 - 0.0798) = 0.011 \$ for a woman.