Problem 1

(a.) First, define the difference between the effects of experience and tenure by:

$$\theta = \beta_{\text{Exper}} - \beta_{\text{Tenure}}$$

Then, our null and alternative hypotheses are:

$$H_0 : \theta = 0$$
$$H_A : \theta \neq 0$$

Next, solve for the coefficient on experience and substitute into the regression equation.

$$\log(wage) = \beta_0 + \beta_{\text{Educ}} \text{Educ} + (\beta_{\text{Tenure}} + \theta) \text{Exper} + \beta_{\text{Tenure}} \text{Tenure} + \beta_{\text{IQ}} \text{IQ} + u$$

Running this regression:

```r
> summary(lm(log(wage)~educ+exper+I(exper+tenure)+iq,data=x))
```

Call:
`lm(formula = log(wage) ~ educ + exper + I(exper + tenure) + iq, data = x)`

Residuals:
```
Min     1Q    Median     3Q    Max
-1.89718 -0.23345  0.01131  0.24372  1.29658
```

Coefficients:

```
                       Estimate  Std. Error t value     Pr(>|t|)
(Intercept)            5.2098581  0.1201247  43.370 < 2e-16 ***
educ                  0.0555855  0.0072675   7.648 5.07e-14 ***
exper                 0.0030545  0.0046714   0.654    0.513
I(exper + tenure)    0.0123705  0.0025519   4.847 1.46e-06 ***
ig                    0.0054564  0.0009704   5.623 2.378e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.3815 on 930 degrees of freedom
Multiple R-squared: 0.1829,  Adjusted R-squared: 0.1794
F-statistic: 52.04 on 4 and 930 DF,  p-value: < 2.2e-16

The coefficient on exper is the estimate for theta, which is the difference between the effects of experience and tenure. The t-statistic is 0.654, which is very small
relative to the t-critical value 1.96. Hence, we fail to reject the hypothesis that the effects of experience and tenure are the same.

(b.) Here, the model in b is a nested model within a. So we use an F test. Hence,

```R
> regUR<-lm(log(wage)~educ+exper+tenure+iq,data=x)
> summary(regUR)

Call:
  lm(formula = log(wage) ~ educ + exper + tenure + iq, data = x)

Residuals:
          Min     1Q Median     3Q    Max
-1.89718 -0.23345  0.01131  0.24372  1.29658

Coefficients:                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       5.2098581  0.1201247  43.370  < 2e-16 ***
educ              0.0555855  0.0072675   7.648 5.07e-14 ***
exper             0.0154250  0.0033155   4.652 3.76e-06 ***
tenure            0.0123705  0.0025519   4.847 1.46e-06 ***
          iq    0.0054564  0.0009704   5.623 2.48e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3815 on 930 degrees of freedom
Multiple R-squared: 0.1829,  Adjusted R-squared: 0.1794
F-statistic: 52.04 on 4 and 930 DF,  p-value: < 2.2e-16

> SSR_UR<-sum(regUR$residuals^2,na.rm=TRUE)
> SSR_UR
[1] 135.3594

> regR<-lm(log(wage)~educ+iq,data=x)
> summary(regR)

Call:
  lm(formula = log(wage) ~ educ + iq, data = x)

Residuals:
          Min     1Q Median     3Q    Max
-2.01601 -0.24367  0.03359  0.27960  1.23783

Coefficients:                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       5.6582876  0.0962408  58.793  < 2e-16 ***
educ              0.0391199  0.0068382   5.721 1.43e-08 ***
          iq    0.0058631  0.0009979   5.875 5.87e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3933 on 932 degrees of freedom
Multiple R-squared: 0.1297,  Adjusted R-squared: 0.1278
F-statistic: 69.42 on 2 and 932 DF,  p-value: < 2.2e-16

> SSR_R <- sum(regR$residuals^2, na.rm=TRUE)
> SSR_R
[1] 144.1783

> Fstat <- ((SSR_R - SSR_UR) / 2) / (SSR_UR / regUR$df)
> Fstat
[1] 30.29567

Since the Fstat is very large, we reject the restrictions imposed from going to ‘b’ to ‘a’.

(c.) Here, the model in ‘b’ is a non-nested model relative to ‘c’. Hence, we use adjusted R^2. In ‘b’, this value is 0.1278. In ‘c’, this value is obtained by running the regression in ‘c’.

> summary(lm(log(wage) ~ educ + exper + tenure + sibs, data=x))

Call:
  lm(formula = log(wage) ~ educ + exper + tenure + sibs, data = x)

Residuals:
     Min      1Q  Median      3Q     Max
-1.80947 -0.23746  0.01715  0.25714  1.32598

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.583363   0.117298  47.600  < 2e-16 ***
educ          0.071484   0.006683  10.697  < 2e-16 ***
exper         0.015032   0.003366   4.466  8.93e-06 ***
tenure        0.013165   0.002584   5.095  4.22e-07 ***
sibs          -0.012311   0.005666  -2.173   0.0300 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.387 on 930 degrees of freedom
Multiple R-squared: 0.1594,  Adjusted R-squared: 0.1558
F-statistic: 44.08 on 4 and 930 DF,  p-value: < 2.2e-16

The adjusted R^2 in ‘c’ is 0.1558, which is larger than ‘b’. Hence, absent the nested models required for an F-test, the values of adjusted R^2 suggest that we choose the model in ‘c’.
Problem 2

> summary(lm(wage~urban+south+I(urban*south),x))

Call:
  lm(formula = wage ~ urban + south + I(urban * south), data = x)

Residuals:
     Min      1Q  Median      3Q     Max
-932.95  -261.30   -54.24  187.47  2158.76

Coefficients:                         Estimate Std. Error t value Pr(>|t|)
(Intercept)        871.07      31.87  27.332  < 2e-16 ***
  urban              176.88      36.72   4.817  1.7e-06 ***
    south          -96.54      48.93  -1.973   0.0488 *
  I(urban * south)  -32.17      58.93  -0.546   0.5853
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 392.9 on 931 degrees of freedom
Multiple R-squared: 0.05883, Adjusted R-squared: 0.05579
F-statistic: 19.4 on 3 and 931 DF,  p-value: 3.327e-12

(a.) Use the equation for the confidence interval:

\[
\hat{\beta}_{urban} - t_{crit} \cdot se(\hat{\beta}_{urban}) < \beta_{urban} < \hat{\beta}_{urban} + t_{crit} \cdot se(\hat{\beta}_{urban})
\]

\[
176.88 - 1.64 \cdot 36.72 < \beta_{urban} < 176.88 + 1.64 \cdot 36.72
\]

\[
116.66 < \beta_{urban} < 237.10
\]

With 90% confidence, an urban resident in the north earns between 166.66 and 237.10 more than a rural resident in the north.

(b.) It helps to take a “derivative” with respect to urban:

\[
\Delta wage = (\hat{\beta}_{urban} + \hat{\beta}_{urban \cdot south \cdot South}) \Delta Urban
\]

The returns to being an urban resident are 32.17 dollars per month less if you live in the south. However, this is not a significant difference. The tstat is -0.546, and at the 95% level the t critical values is 1.94. You can also answer this question by stating the returns to living the south being 32.17 dollars less if you live in an urban area.