3.17 Consider the regression model $WAGE = \beta_1 + \beta_2 EDUC + e$. Where *WAGE* is hourly wage rate in US 2013 dollars. *EDUC* is years of schooling. The model is estimated twice, once using individuals from an urban area, and again for individuals in a rural area.

$$\text{Urban} \qquad \begin{aligned} \widehat{WAGE} &= -10.76 + 2.46EDUC, \quad N = 986 \\ \text{(se)} \quad & \quad (2.27) \quad (0.16) \end{aligned}$$

$$\text{Rural} \qquad \begin{aligned} \widehat{WAGE} &= -4.88 + 1.80EDUC, \quad N = 214 \\ \text{(se)} \quad & \quad (3.29) \quad (0.24) \end{aligned}$$

a. Using the urban regression, test the null hypothesis that the regression slope equals 1.80 against the alternative that it is greater than 1.80. Use the $\alpha = 0.05$ level of significance. Show all steps, including a graph of the critical region and state your conclusion.

b. Using the rural regression, compute a 95% interval estimate for expected *WAGE* if $EDUC = 16$. The required standard error is 0.833. Show how it is calculated using the fact that the estimated covariance between the intercept and slope coefficients is -0.761.

c. Using the urban regression, compute a 95% interval estimate for expected *WAGE* if $EDUC = 16$. The estimated covariance between the intercept and slope coefficients is -0.345. Is the interval estimate for the urban regression wider or narrower than that for the rural regression in (b). Do you find this plausible? Explain.

d. Using the rural regression, test the hypothesis that the intercept parameter β_1 equals four, or more, against the alternative that it is less than four, at the 1% level of significance.

a. $H_0: \beta_2 = 1.80$, $H_1: \beta_2 > 1.80$

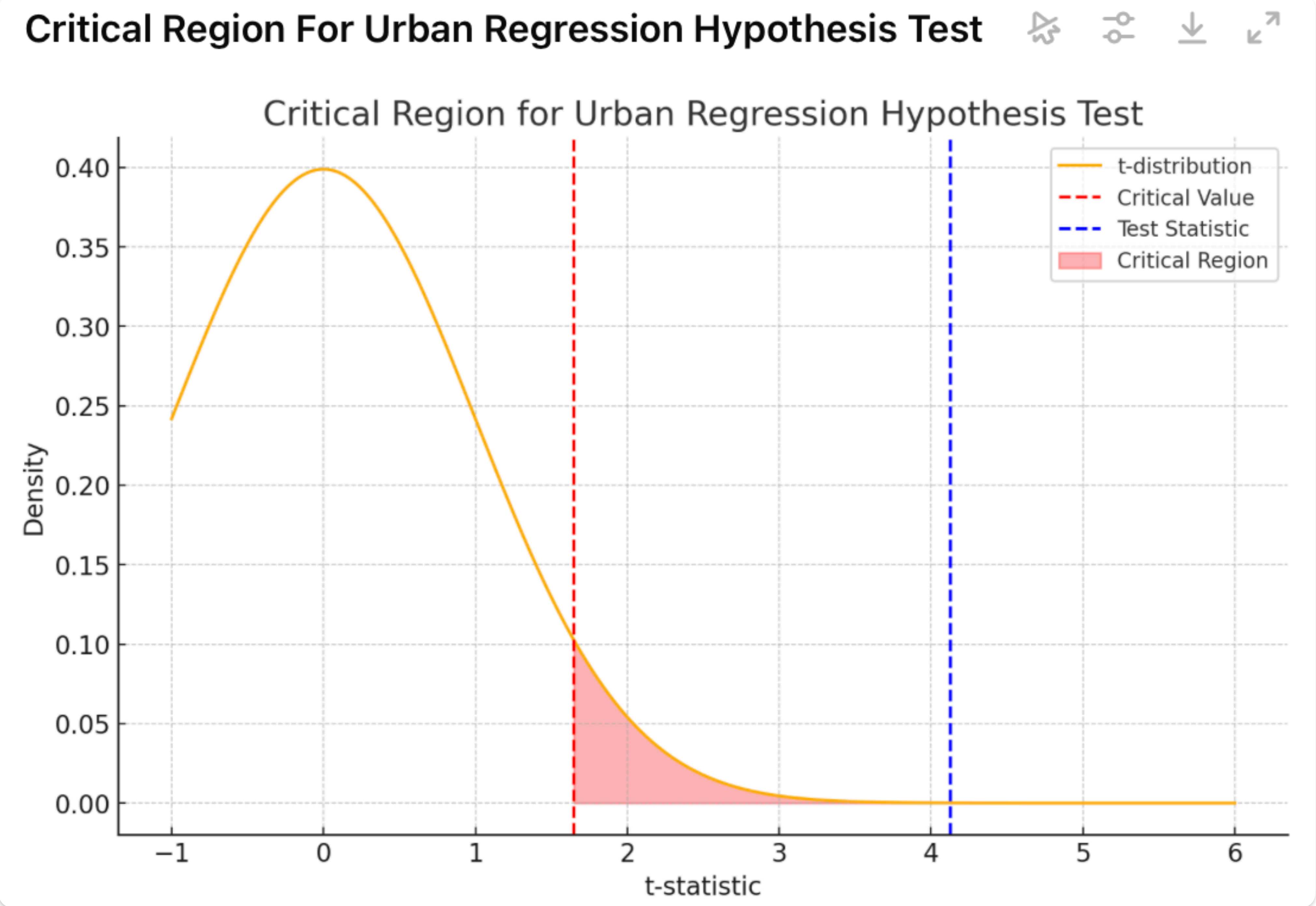
$\alpha = 0.05$, $df = N-2 = 986-2 = 984$

```
> qt(1 - 0.05, 984)
[1] 1.646404
```

$t = \frac{2.46-1.80}{0.16} = 4.125 > t_{0.95,984} = 1.646$

∴ we reject H_0 that the regression slope equals 1.80.

This means the slope in the urban model is significantly greater than 1.80.



b. $\widehat{wage} = -4.88 + 1.80 \times 16 = 23.92$

$\alpha = 0.05$, $df = N-2 = 214-2 = 212$

```
> qt(1 - 0.05 / 2, 212)
[1] 1.971217
```

$t_{0.975,212} = 1.9712$

$C.I. = \widehat{wage} \pm 1.9712 \times SE(\widehat{wage}) = 23.92 \pm 1.9712 \times 0.833$
$= [22.2780, 25.5620]$

$SE(\widehat{wage}) = \sqrt{SE(b_1)^2 + SE(b_2)^2 \times EDUC^2 + 2 \times cov(b_1, b_2) \times EDUC}$
$= \sqrt{3.29^2 + 0.24^2 \times 16^2 + 2 \times (-0.761) \times 16} = 1.1035$

c. $\widehat{wage} = -10.76 + 2.46 \times 16 = 28.6$

$SE(\widehat{wage}) = \sqrt{SE(b_1)^2 + SE(b_2)^2 \times EDUC^2 + 2 \times cov(b_1, b_2) \times EDUC}$
$= \sqrt{2.27^2 + 0.16^2 \times 16^2 + 2 \times (-0.345) \times 16} = 0.8164$

$\alpha = 0.05$, $df = N-2 = 986-2 = 984$

```
> qt(1 - 0.05 / 2, 984)
[1] 1.962378
```

$t_{0.975,984} = 1.9624$

$C.I. = \widehat{wage} \pm 1.9624 \times SE(\widehat{wage}) = 28.6 \pm 1.9624 \times 0.8164$
$= [26.9979, 30.2021]$

Comparison of Interval Widths:

Rural: $25.5620 - 22.2780 = 3.284$

Urban: $30.2021 - 26.9979 = 3.2042$

The confidence interval width for the urban area is norrower than for the rural area.

d. $H_0: \beta_1 \geq 4$, $H_1: \beta_1 < 4$

$\alpha = 0.01$, $df = N-2 = 214-2 = 212$

```
> qt(0.01, 212)
[1] -2.344066
```

$t = \frac{-4.88-4}{3.29} = -2.6991 < t_{0.01,212} = -2.3441$

∴ we reject H_0 that the intercept ≥ 4.

This means the intercept in the rural model is significantly less than 4.