HW0310

3.1 There were 64 countries in 1992 that competed in the Olympics and won at least one medal. Let *MEDALS* be the total number of medals won, and let *GDPB* be GDP (billions of 1995 dollars). A linear regression model explaining the number of medals won is $MEDALS = \beta_1 + \beta_2 GDPB + e$. The estimated relationship is

$$\widehat{MEDALS} = b_1 + b_2 GDPB = 7.61733 + 0.01309 GDPB$$
$$\text{(se)} \qquad\qquad (2.38994) \quad (0.00215) \qquad\qquad \text{(XR3.1)}$$

a. We wish to test the hypothesis that there is no relationship between the number of medals won and *GDP* against the alternative there is a positive relationship. State the null and alternative hypotheses in terms of the model parameters.

b. What is the test statistic for part (a) and what is its distribution if the null hypothesis is true?

c. What happens to the distribution of the test statistic for part (a) if the alternative hypothesis is true? Is the distribution shifted to the left or right, relative to the usual *t*-distribution? [*Hint*: What is the expected value of b_2 if the null hypothesis is true, and what is it if the alternative is true?]

d. For a test at the 1% level of significance, for what values of the *t*-statistic will we reject the null hypothesis in part (a)? For what values will we fail to reject the null hypothesis?

e. Carry out the *t*-test for the null hypothesis in part (a) at the 1% level of significance. What is your economic conclusion? What does 1% level of significance mean in this example?

a. H0 : b2 = 0 (獎牌數與 GDP 無關)
   H1 : b2 > 0 (獎牌數與 GDP 呈現正相關)

b. t-value = b2/SE(b2) = 0.01309 / 0.00215 = 6.0884，若 H0 為真，檢定統計量符合自由度為 n-2 = 62 的 t 分配。

c. 若 H1 為真，則檢定統計量的期望值將大於零，分配將呈現右長尾。

d. $t_{0.01,62}$ = 2.39，因此若檢定統計量大於 2.39，則有證據拒絕虛無假設，反之將無法拒絕虛無假設。

e. 由於 t-value = 6.0884 > 2.39，因此有證據拒絕虛無假設，即獎牌數與 GDP 呈現正相關；1%的顯著水準表示有 1%的機率會錯誤地拒絕虛無假設。

3.7 We have 2008 data on *INCOME* = income per capita (in thousands of dollars) and *BACHELOR* = percentage of the population with a bachelor's degree or more for the 50 U.S. States plus the District of Columbia, a total of $N = 51$ observations. The results from a simple linear regression of *INCOME* on *BACHELOR* are

$$\begin{array}{rcc} \widehat{INCOME} = & (a) & + \ 1.029BACHELOR \\ \text{se} & (2.672) & (c) \\ t & (4.31) & (10.75) \end{array}$$

a. Using the information provided calculate the estimated intercept. Show your work.
b. Sketch the estimated relationship. Is it increasing or decreasing? Is it a positive or inverse relationship? Is it increasing or decreasing at a constant rate or is it increasing or decreasing at an increasing rate?
c. Using the information provided calculate the standard error of the slope coefficient. Show your work.
d. What is the value of the *t*-statistic for the null hypothesis that the intercept parameter equals 10?
e. The *p*-value for a two-tail test that the intercept parameter equals 10, from part (d), is 0.572. Show the *p*-value in a sketch. On the sketch, show the rejection region if $\alpha = 0.05$.
f. Construct a 99% interval estimate of the slope. Interpret the interval estimate.
g. Test the null hypothesis that the slope coefficient is one against the alternative that it is not one at the 5% level of significance. State the economic result of the test, in the context of this problem.

a. t-value = intercept / SE(intercept)；4.31 = intercept / 2.672，intercept = 11.5163。
b. 由斜率 =1.029 可知收入與大學學歷之關係為線性正相關。
c. SE(slope) = slope / t-value = 1.029 / 10.75 = 0.0957。
d. H0 : intercept = 10，t-value = (11.5163-10)/2.672 = 0.5675。
e. p-value = 0.572(two-tail)，設定 alpha = 0.05，則拒絕域為 p-value < 0.05，此例中 p-value 遠高於 0.05，因此無法拒絕虛無假設。
f. 斜率在 99%信心水準之下的的區間估計為 1.029+-$t_{0.005, 49}$ * 0.0957，即 [0.7725, 1.2855]。
g. H0 : slope = 1；H1 : slope != 1(two-tail)，alpha = 0.05
檢定統計量 t = (1.029 – 1)/ 0.0957 = 0.3030，在大多數情況下均無法拒絕虛無假設，即無法拒絕大學學歷對收入的邊際影響為 1。

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.

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

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

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. H0 : slope = 1.8；H1 : slope > 1.8，alpha = 0.05
檢定統計量 t = 2.46 – 1.8 / 0.16 = 4.125 > $t_{0.05, 986}$ ~= 1.646，有證據拒絕虛無假設，即斜率顯著大於 1.8。
b. Expected Wage = -4.88 + 1.8 * 16 = 23.92，95% interval estimation = [23.92 – 1.96 * 0.833, 23.92 + 1.96 * 0.833] = [22.2873, 25.5527]
c. Expected wage = -10.76 + 2.46 * 16 = 28.6，se(WAGE) = $\sqrt{}$(2.27^2 + 16^2 * 0.16^2 – 2 * 16 * 0.345) = 0.8164，95% interval = [28.6 – 1.96 * 0.8164, 28.6 + 1.96 * 0.8164] = [26.9999, 30.2001]，在都市的估計區間要小於鄉村的估計區間，我認為是合理的，因為鄉村的貧富差距通常較大。
d. H0 : b1 >= 4；H1 : b1 < 4，alpha = 0.01，檢定統計量 t = (-4.88 – 4) / 3.29 = -2.6991 < $t_{0.01, 214}$ ~=2.33，沒有證據支持截距項大於 4。

3.19 The owners of a motel discovered that a defective product was used during construction. It took 7 months to correct the defects during which approximately 14 rooms in the 100-unit motel were taken out of service for 1 month at a time. The data are in the file *motel*.

a. Plot $MOTEL\_PCT$ and $COMP\_PCT$ versus $TIME$ on the same graph. What can you say about the occupancy rates over time? Do they tend to move together? Which seems to have the higher occupancy rates? Estimate the regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$. Construct a 95% interval estimate for the parameter β_2. Have we estimated the association between $MOTEL\_PCT$ and $COMP\_PCT$ relatively precisely, or not? Explain your reasoning.

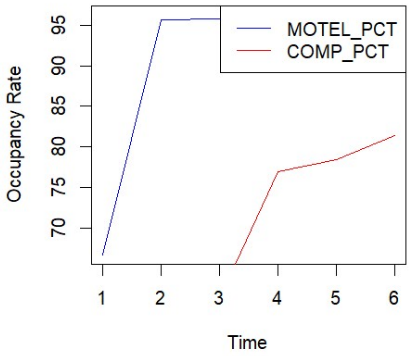
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, $MOTEL\_PCT$, given that $COMP\_PCT = 70$.

c. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0: \beta_2 \leq 0$ against the alternative hypothesis $H_0: \beta_2 > 0$ at the $\alpha = 0.01$ level of significance. Discuss your conclusion. Clearly define the test statistic used and the rejection region.

d. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0: \beta_2 = 1$ against the alternative hypothesis $H_0: \beta_2 \neq 1$ at the $\alpha = 0.01$ level of significance. If the null hypothesis were true, what would that imply about the motel's occupancy rate versus their competitor's occupancy rate? Discuss your conclusion. Clearly define the test statistic used and the rejection region.

e. Calculate the least squares residuals from the regression of $MOTEL\_PCT$ on $COMP\_PCT$ and plot them against $TIME$. Are there any unusual features to the plot? What is the predominant sign of the residuals during time periods 17–23 (July, 2004 to January, 2005)?

a. 由下圖可知旅館的入住率跟對手的入住率有正相關，並且自己有較高的入住率。




```
Residuals:
1       2       3       4      5       6
-19.254  10.407   8.456   3.693    1.496  -4.797

Coefficients:
Estimate Std. Error
(Intercept)  67.0567    31.6396
COMP_PCT      0.3310     0.4574
t value Pr(>|t|)
(Intercept)   2.119    0.101
COMP_PCT      0.724    0.509
```

```
Residual standard error: 12.14 on 4 degrees of freedom
Multiple R-squared:  0.1157,       Adjusted R-squared:  -0.1053
F-statistic: 0.5235 on 1 and 4 DF,  p-value: 0.5094

                  2.5 %     97.5 %
(Intercept) -20.7888893 154.902310
COMP_PCT     -0.9390301   1.600946
```

b.
```
     fit      lwr      upr
90.22378 79.53225 100.9153
```

c. 由先前的敘述統計圖表可知，斜率的 p-value 為 0.509，在 alpha = 0.01 下無法拒絕斜率小魚等魚 0 的虛無假設。

d.
```
  Res.Df    RSS Df Sum of Sq      F
1      5 904.66
2      4 589.41  1    315.24 2.1394
  Pr(>F)
1
2 0.2174
```

上述之 p-value 在 alpha = 0.01 之下無法拒絕虛無假設，即競爭對手的入住率增加 1%，汽車旅觀的入住率也增加 1%。

e.