Name: CHAN King Yeung

SID: 1155119394

STAT3008 Assignment 3

Question 2

1. After executed the R code, it returns

The estimated annual salary for Mary is USD 16606.5

1. After executed the R code, it returns

1. After executed the R code, it returns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Regression** | 4 | 642448811 | 160612203 | 27.146 | 1.708e-11 |
| **Residual** | 45 | 266244659 | 5916548 |  |  |
| **Total** | 49 | 908693470 |  |  |  |

1. Since the p-value , we reject at

We have sufficient evidence that Rank is important to explain the Salary, which is more appropriate model comparing to

1. After executed the R code, it returns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Regression** | 3 | 10748075 | 3582692 | 0.6055 | 0.6148 |
| **Residual** | 45 | 266244659 | 5916548 |  |  |
| **Total** | 48 | 276992734 |  |  |  |

1. Since the p-value , we do not reject at

We do not have sufficient evidence that Sex is playing an important role to describe the salary in between male and female, that is, , and are not differ from 0

Question 3

1. Forward selection

|  |  |  |
| --- | --- | --- |
| Start: AIC=253.58  y ~ 1  Df Sum of Sq RSS AIC  + x4 1 1661.66 884.73 181.57  + x2 1 1120.45 1425.95 214.99  + x1 1 504.47 2041.92 240.12  + x3 1 462.22 2084.18 241.55  + x6 1 360.22 2186.17 244.90  + x5 1 281.47 2264.92 247.38  <none> 2546.39 253.58 | Step: AIC=181.58  y ~ x4  Df Sum of Sq RSS AIC  + x3 1 141.523 743.21 171.37  + x5 1 90.016 794.72 176.06  + x6 1 49.592 835.14 179.54  + x2 1 25.046 859.69 181.56  <none> 884.73 181.57  + x1 1 8.291 876.44 182.92 | Step: AIC=171.37  y ~ x4 + x3  Df Sum of Sq RSS AIC  + x6 1 45.431 697.78 168.96  + x2 1 21.519 721.69 171.32  <none> 743.21 171.37  + x1 1 8.535 734.68 172.56  + x5 1 1.637 741.57 173.22 |
| Step: AIC=168.96  y ~ x4 + x3 + x6  Df Sum of Sq RSS AIC  <none> 697.78 168.96  + x1 1 15.1423 682.64 169.42  + x2 1 13.2307 684.55 169.62  + x5 1 4.2635 693.52 170.53 |  |  |

The parsimonious model under forward selection is

Backward selection

|  |  |  |
| --- | --- | --- |
| Start: AIC=172.67  y ~ x1 + x2 + x3 + x4 + x5 + x6  Df Sum of Sq RSS AIC  - x2 1 2.93 678.29 170.98  - x5 1 2.95 678.31 170.98  - x1 1 7.01 682.38 171.40  <none> 675.36 172.67  - x6 1 44.82 720.18 175.17  - x3 1 59.18 734.54 176.55  - x4 1 646.12 1321.48 217.66 | Step: AIC=170.98  y ~ x1 + x3 + x4 + x5 + x6  Df Sum of Sq RSS AIC  - x5 1 4.34 682.64 169.42  - x1 1 15.22 693.52 170.53  <none> 678.29 170.98  - x6 1 54.83 733.12 174.42  - x3 1 74.01 752.30 176.22  - x4 1 1190.09 1868.39 239.90 | Step: AIC=169.42  y ~ x1 + x3 + x4 + x6  Df Sum of Sq RSS AIC  - x1 1 15.14 697.78 168.96  <none> 682.64 169.42  - x6 1 52.04 734.68 172.56  - x3 1 148.70 831.34 181.22  - x4 1 1222.35 1904.99 239.26 |
| Step: AIC=168.96  y ~ x3 + x4 + x6  Df Sum of Sq RSS AIC  <none> 697.78 168.96  - x6 1 45.43 743.21 171.37  - x3 1 137.36 835.14 179.54  - x4 1 1278.97 1976.75 239.85 |  |  |

The parsimonious model under backward selection is

The parsimonious models under both forward selection and backward selection are the same. Thus, the parsimonious based on AIC is

1. Forward selection

|  |  |  |
| --- | --- | --- |
| Start: AIC=255.82  y ~ 1  Df Sum of Sq RSS AIC  + x4 1 1661.66 884.73 186.07  + x2 1 1120.45 1425.95 219.48  + x1 1 504.47 2041.92 244.62  + x3 1 462.22 2084.18 246.05  + x6 1 360.22 2186.17 249.40  + x5 1 281.47 2264.92 251.87  <none> 2546.39 255.82 | Step: AIC=186.07  y ~ x4  Df Sum of Sq RSS AIC  + x3 1 141.523 743.21 178.12  + x5 1 90.016 794.72 182.81  <none> 884.73 186.07  + x6 1 49.592 835.14 186.28  + x2 1 25.046 859.69 188.31  + x1 1 8.291 876.44 189.66 | Step: AIC=178.12  y ~ x4 + x3  Df Sum of Sq RSS AIC  + x6 1 45.431 697.78 177.95  <none> 743.21 178.12  + x2 1 21.519 721.69 180.31  + x1 1 8.535 734.68 181.56  + x5 1 1.637 741.57 182.21 |
| Step: AIC=177.95  y ~ x4 + x3 + x6  Df Sum of Sq RSS AIC  <none> 697.78 177.95  + x1 1 15.1423 682.64 180.66  + x2 1 13.2307 684.55 180.86  + x5 1 4.2635 693.52 181.77 |  |  |

The parsimonious model under forward selection is

Backward selection

|  |  |  |
| --- | --- | --- |
| Start: AIC=188.41  y ~ x1 + x2 + x3 + x4 + x5 + x6  Df Sum of Sq RSS AIC  - x2 1 2.93 678.29 184.47  - x5 1 2.95 678.31 184.47  - x1 1 7.01 682.38 184.89  <none> 675.36 188.41  - x6 1 44.82 720.18 188.66  - x3 1 59.18 734.54 190.04  - x4 1 646.12 1321.48 231.15 | Step: AIC=184.47  y ~ x1 + x3 + x4 + x5 + x6  Df Sum of Sq RSS AIC  - x5 1 4.34 682.64 180.66  - x1 1 15.22 693.52 181.77  <none> 678.29 184.47  - x6 1 54.83 733.12 185.66  - x3 1 74.01 752.30 187.47  - x4 1 1190.09 1868.39 251.15 | Step: AIC=180.67  y ~ x1 + x3 + x4 + x6  Df Sum of Sq RSS AIC  - x1 1 15.14 697.78 177.95  <none> 682.64 180.66  - x6 1 52.04 734.68 181.56  - x3 1 148.70 831.34 190.21  - x4 1 1222.35 1904.99 248.26 |
| Step: AIC=177.95  y ~ x3 + x4 + x6  Df Sum of Sq RSS AIC  <none> 697.78 177.95  - x6 1 45.43 743.21 178.12  - x3 1 137.36 835.14 186.28  - x4 1 1278.97 1976.75 246.59 |  |  |

The parsimonious model under backward selection is

The parsimonious models under both forward selection and backward selection are the same. Thus, the parsimonious based on BIC is

1. After executed the R code, it returns 6.034562. That is