STAT2183

Min Yang

Assignment 5

Assignment 5 Report

**Introduction**

The researcher wants to study what factor influence the infection risk. The sample is collected randomly from 113 hospitals. The researcher considered nine variables as possible factors:

1. length of stay: Average length of stay in hospital (in days)
2. age: average age of patients (in years)
3. routing culturing ratio: average estimated probability of acquiring infection in hospital (in percent)
4. routine chest X-ray ratio: Ratio of number of X-rays performed to number of patients without signs or symptoms of pneumonia, times 100
5. number of beds: Average number of beds in hospital
6. medical school affiliation: 0 = Yes, 1 = No
7. average daily census: Average number of patients in hospital per day
8. number of nurses: Average number of full-time licensed practical nurses
9. available facilities and services: Percent of 35 potential facilities and services that are provided by the hospital

Among all these factors, medical school affiliation is a qualitative variable.

The dependent variable infection risk is recorded as average estimated probability of acquiring infection in hospital (in percent).

**Method**

First, we will determine which factor individually effect the infection risk using simple linear regression model for each factor.

Then, we use forward selection, backward elimination, and stepwise regression to fit the data in a best model, we use C(p) to determine the best model, as the model with lowest C(p) value.

Then, we perform a residual analysis on the chosen model, and test the conditions for regression model. There are three assumptions: 1. The residuals are independent. 2. The variance of the residuals is constant. 3. The residuals are normally distributed.

Lastly, we use the best model to predict that last five observation to see if the model is doing well.

**Test for individual factor effect on infection risk**

H0: = 0

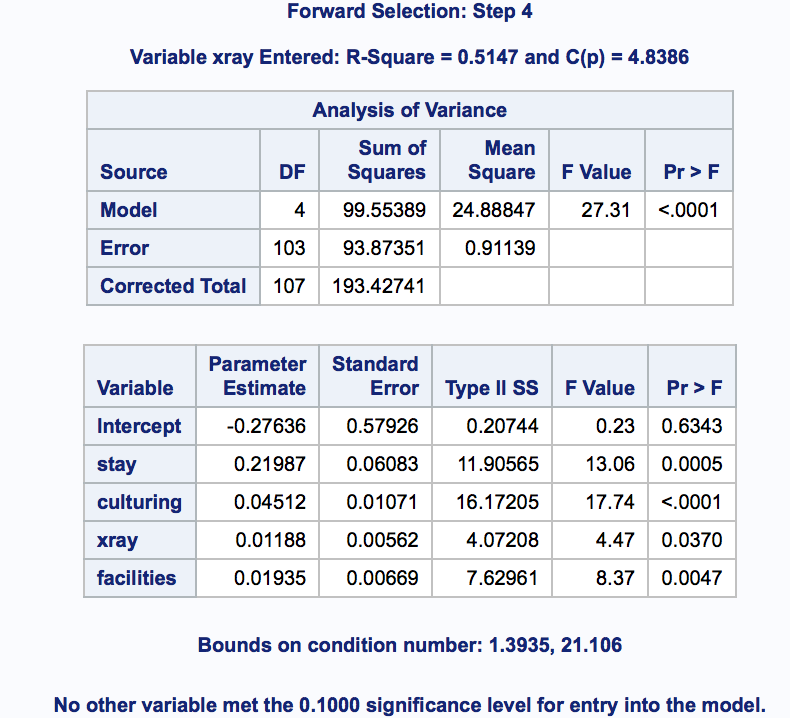
Ha:

|  |  |
| --- | --- |
| Stay |  |
| Age |  |
| Culturing |  |
| Xray |  |
| beds |  |
| school |  |
| census |  |
| nurses |  |
| facilities |  |

At a=0.05 significance level, stay, culturing, x-ray, beds, school, census, nurses, and facilities individually effect to the infection risk as they all have a p-value smaller than 0.05, while we fail to reject the null hypothesis for age as it has a p-value = 0.89>0.05, so we cannot conclude age has an effect the infection risk.

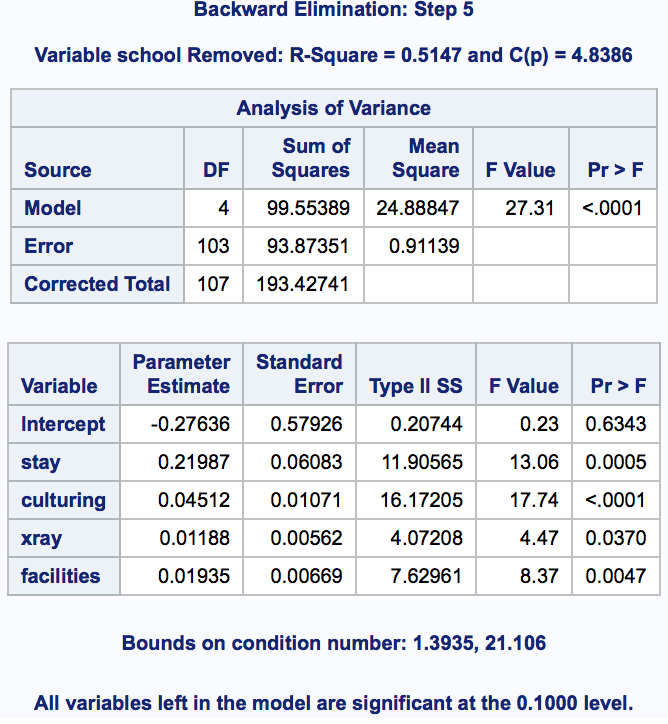
**Model Selection**

1. Forward Selection



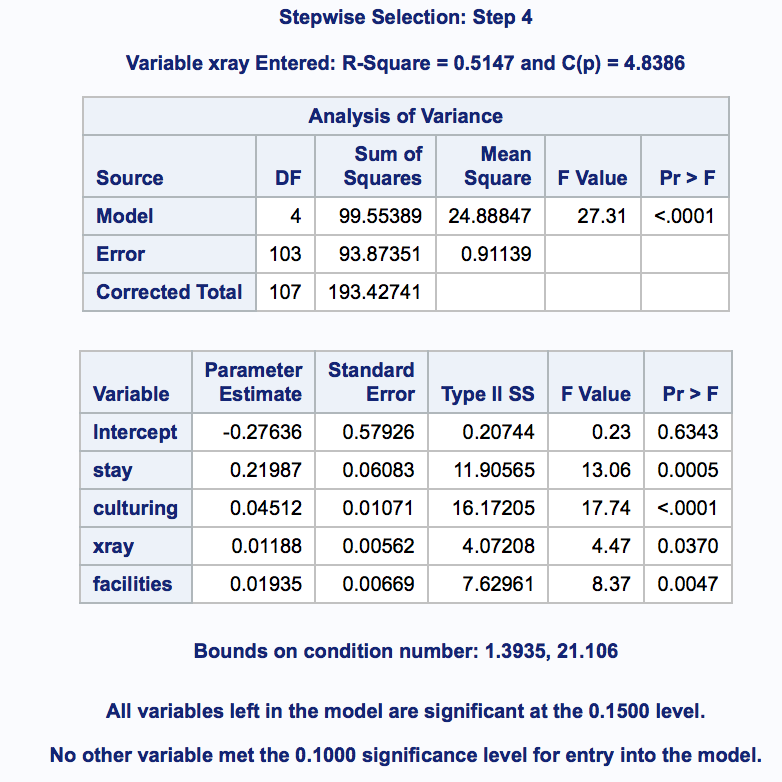
From forward selection, the best model is:

1. Backward Elimination



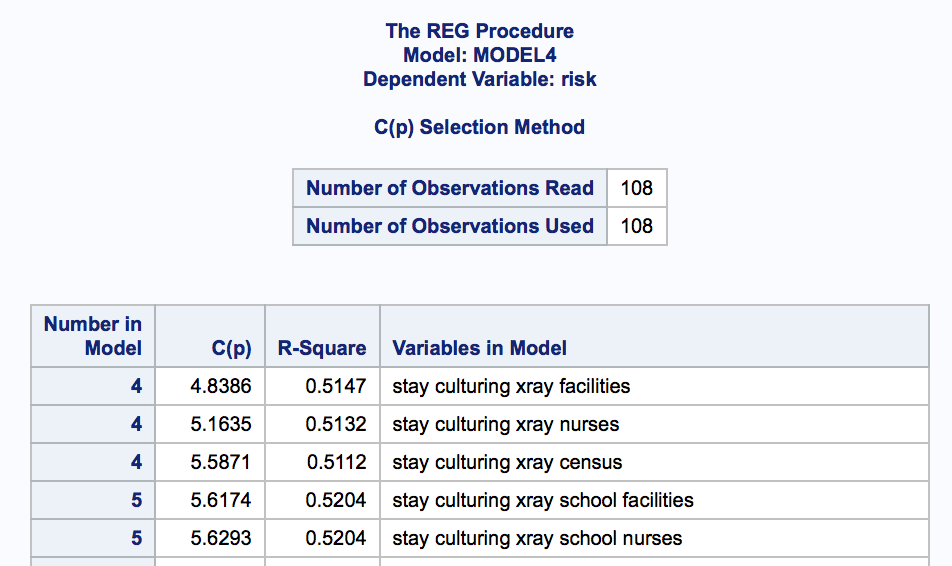
From backward elimination, the best model is:

1. Stepwise Regression



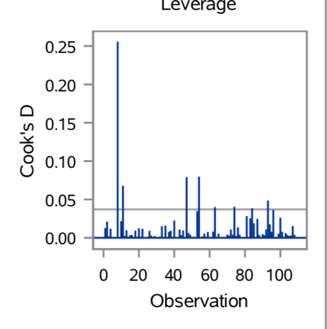
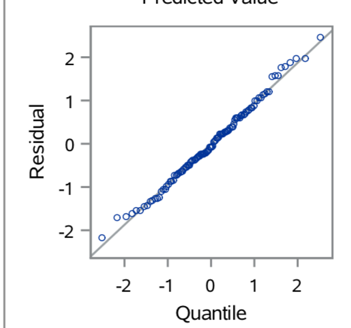
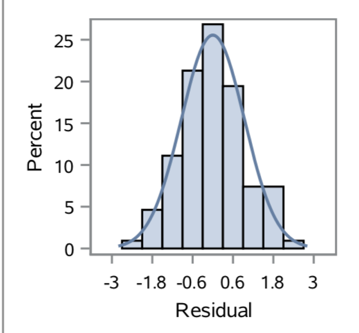
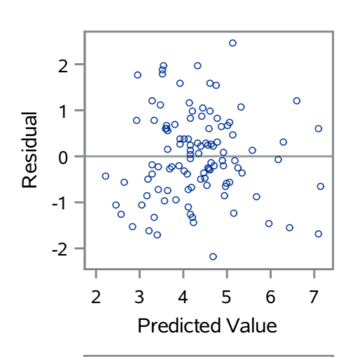
From stepwise regression, the best model is:

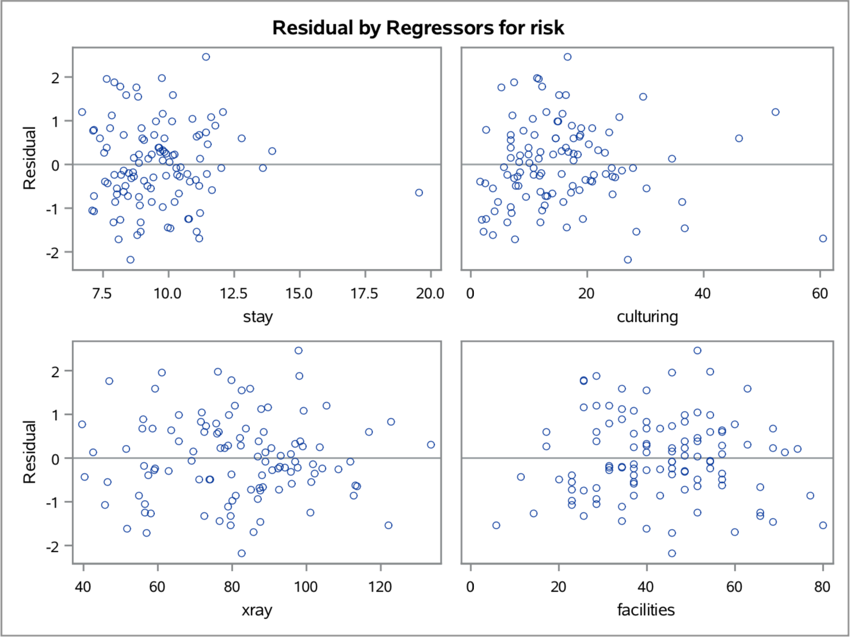
**C(p) rank lowest five models**



The model that has the lowest C(p) is model with variables stay, culturing, x-ray, and facilities, with a C(p) of 4.8386. And forward selection, backward elimination, and stepwise regression all agreed on this model to be the best model.

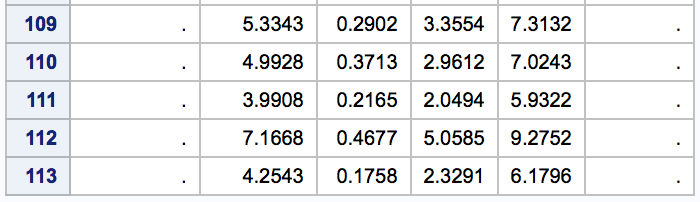
**Test for residual conditions**





From the residual-predicted value plot, there is no pattern or trend show in the plot, from the cock’s distance, there is no outstanding outliers, and the rest two plots also shows that the residuals are independent and normally distributed. The four individual plots all show no pattern, so the variance of residuals is constant. Thus, all assumptions for regression conditions are satisfied.

**Prediction for last five values**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Order | Predict Value | Actual Value | 95% CL Predict LB | 95% CL Predict UB |
| 109 | 5.3343 | 5.7 | 3.355 | 7.3132 |
| 110 | 4.9928 | 5.8 | 2.9612 | 7.0243 |
| 111 | 3.9908 | 4.4 | 2.0494 | 5.9322 |
| 112 | 7.1668 | 5.9 | 5.0585 | 9.2752 |
| 113 | 4.2543 | 3.1 | 2.3291 | 6.1796 |

Since all five actual value are within the 95% confidence interval predict value, so this model is doing pretty well.

**Conclusion**

At a=0.05 significance level, stay, culturing, x-ray, beds, school, census, nurses, and facilities individually effect the infection risk as they all have a p-value smaller than 0.05, while we fail to reject the null hypothesis for age as it has a p-value = 0.89>0.05, so we cannot conclude age has an effect to the infection risk. From forward selection, backward elimination, and stepwise regression, the best model is .

From the residual-predicted value plot, there is no pattern or trend show in the plot, from the cock’s distance, there is no outstanding outliers, and the rest two plots also shows that the residuals are independent and normally distributed. The four individual plots all show no pattern, so the variance of residuals is constant. Thus, all assumptions for regression conditions are satisfied.

The prediction for last five observations of the data are 5.3343, 4.9928, 3.9908, 7.1668, 4.2543 respectively. All of them lie in between the 95% predicted confidence interval, so the model is doing well.

Appendix

data data;

input risk id stay age culturing xray beds school census nurses facilities;

cards;

4.1 1 7.13 55.7 9 39.6 279 1 207 241 60

1.6 2 8.82 58.2 3.8 51.7 80 1 51 52 40

2.7 3 8.34 56.9 8.1 74 107 1 82 54 20

5.6 4 8.95 53.7 18.9 122.8 147 1 53 148 40

5.7 5 11.2 56.5 34.5 88.9 180 1 134 151 40

5.1 6 9.76 50.9 21.9 97 150 1 147 106 40

4.6 7 9.68 57.8 16.7 79 186 1 151 129 40

5.4 8 11.18 45.7 60.5 85.8 640 0 399 360 60

4.3 9 8.67 48.2 24.4 90.8 182 1 130 118 40

6.3 10 8.84 56.3 29.6 82.6 85 1 59 66 40

4.9 11 11.07 53.2 28.5 122 768 0 591 656 80

4.3 12 8.3 57.2 6.8 83.8 167 1 105 59 40

7.7 13 12.78 56.8 46 116.9 322 0 252 349 57.1

3.7 14 7.58 56.7 20.8 88 97 1 59 79 37.1

4.2 15 9 56.3 14.6 76.4 72 1 61 38 17.1

5.5 16 11.08 50.2 18.6 63.6 387 1 326 405 57.1

4.5 17 8.28 48.1 26 101.8 108 1 84 73 37.1

6.4 18 11.62 53.9 25.5 99.2 133 1 113 101 37.1

4.2 19 9.06 52.8 6.9 75.9 134 1 103 125 37.1

4.1 20 9.35 53.8 15.9 80.9 833 1 547 519 77.1

4.2 21 7.53 42 23.1 98.9 95 1 47 49 17.1

4.8 22 10.24 49 36.3 112.6 195 1 163 170 37.1

5 23 9.78 52.3 17.6 95.9 270 0 240 198 57.1

4.8 24 9.84 62.2 12 82.3 600 1 468 497 57.1

4 25 9.2 52.2 17.5 71.1 298 0 244 236 57.1

3.9 26 8.28 49.5 12 113.1 546 0 413 436 57.1

4.5 27 9.31 47.2 30.2 101.3 170 1 124 173 37.1

3.2 28 8.19 52.1 10.8 59.2 176 1 156 88 37.1

4.4 29 11.65 54.5 18.6 96.1 248 1 217 189 37.1

4.9 30 9.89 50.5 17.7 103.6 167 1 113 106 37.1

5 31 11.03 49.9 19.7 102.1 318 1 270 335 57.1

5.2 32 9.84 53 17.7 72.6 210 1 200 239 54.3

5.3 33 11.77 54.1 17.3 56 196 1 164 165 34.3

6.1 34 13.59 54 24.2 111.7 312 1 258 169 54.3

6.3 35 9.74 54.4 11.4 76.1 221 1 170 172 54.3

5 36 10.33 55.8 21.2 104.3 266 1 181 149 54.3

2.8 37 9.97 58.2 16.5 76.5 90 1 69 42 34.3

4.6 38 7.84 49.1 7.1 87.9 60 1 50 45 34.3

4.1 39 10.47 53.2 5.7 69.1 196 1 168 153 54.3

1.3 40 8.16 60.9 1.9 58 73 1 49 21 14.3

3.7 41 8.48 51.1 12.1 92.8 166 1 145 118 34.3

4.7 42 10.72 53.8 23.2 94.1 113 1 90 107 34.3

3 43 11.2 45 7 78.9 130 1 95 56 34.3

5.6 44 10.12 51.7 14.9 79.1 362 0 313 264 54.3

5.5 45 8.37 50.7 15.1 84.8 115 1 96 88 34.3

4.6 46 10.16 54.2 8.4 51.5 831 0 581 629 74.3

6.5 47 19.56 59.9 17.2 113.7 306 1 273 172 51.4

5.5 48 10.9 57.2 10.6 71.9 593 1 446 211 51.4

1.8 49 7.67 51.7 2.5 40.4 106 1 93 35 11.4

4.2 50 8.88 51.5 10.1 86.9 305 1 238 197 51.4

5.6 51 11.48 57.6 20.3 82 252 1 207 251 51.4

4.3 52 9.23 51.6 11.6 42.6 620 1 413 420 71.4

7.6 53 11.41 61.1 16.6 97.9 535 1 330 273 51.4

7.8 54 12.07 43.7 52.4 105.3 157 1 115 76 31.4

3.1 55 8.63 54 8.4 56.2 76 1 39 44 31.4

3.9 56 11.15 56.5 7.7 73.9 281 1 217 199 51.4

3.7 57 7.14 59 2.6 75.8 70 1 37 35 31.4

4.3 58 7.65 47.1 16.4 65.7 318 1 265 314 51.4

3.9 59 10.73 50.6 19.3 101 445 0 374 345 51.4

4.5 60 11.46 56.9 15.6 97.7 191 1 153 132 31.4

3.4 61 10.42 58 8 59 119 1 67 64 31.4

5.7 62 11.18 51 18.8 55.9 595 0 546 392 68.6

5.4 63 7.93 64.1 7.5 98.1 68 1 42 49 28.6

4.4 64 9.66 52.1 9.9 98.3 83 1 66 95 28.6

5 65 7.78 45.5 20.9 71.6 489 1 391 329 48.6

4.3 66 9.42 50.6 24.8 62.8 508 1 421 528 48.6

4.4 67 10.02 49.5 8.3 93 265 1 191 202 48.6

3.7 68 8.58 55 7.4 95.9 304 1 248 218 48.6

4.5 69 9.61 52.4 6.9 87.2 487 1 404 220 48.6

3.5 70 8.03 54.2 24.3 87.3 97 1 65 55 28.6

4.2 71 7.39 51 14.6 88.4 72 1 38 67 28.6

2 72 7.08 52 12.3 56.4 87 1 52 57 28.6

5.2 73 9.53 51.5 15 65.7 298 1 241 193 48.6

4.5 74 10.05 52 36.7 87.5 184 0 144 151 68.6

3.4 75 8.45 38.8 12.9 85 235 1 143 124 48.6

4.5 76 6.7 48.6 13 80.8 76 1 51 79 28.6

2.9 77 8.9 49.7 12.7 86.9 52 1 37 35 28.6

4.9 78 10.23 53.2 9.9 77.9 752 0 595 446 68.6

4.4 79 8.88 55.8 14.1 76.8 237 1 165 182 48.6

5.1 80 10.3 59.6 27.8 88.9 175 1 113 73 45.7

2.9 81 10.79 44.2 2.6 56.6 461 0 320 196 65.7

3.5 82 7.94 49.5 6.2 92.3 195 1 139 116 45.7

5.5 83 7.63 52.1 11.6 61.1 197 1 109 110 45.7

4.7 84 8.77 54.5 5.2 47 143 1 85 87 25.7

1.7 85 8.09 56.9 7.6 56.9 92 1 61 61 45.7

4.1 86 9.05 51.2 20.5 79.8 195 1 127 112 45.7

2.9 87 7.91 52.8 11.9 79.5 477 1 349 188 65.7

4.3 88 10.39 54.6 14 88.3 353 1 223 200 65.7

4.8 89 9.36 54.1 18.3 90.6 165 1 127 158 45.7

5.8 90 11.41 50.4 23.8 73 424 0 359 335 45.7

2.9 91 8.86 51.3 9.5 87.5 100 1 65 53 25.7

2 92 8.93 56 6.2 72.5 95 1 59 56 25.7

1.3 93 8.92 53.9 2.2 79.5 56 1 40 14 5.7

5.3 94 8.15 54.9 12.3 79.8 99 1 55 71 25.7

5.3 95 9.77 50.2 15.7 89.7 154 1 123 148 25.7

2.5 96 8.54 56.1 27 82.5 98 1 57 75 45.7

3.8 97 8.66 52.8 6.8 69.5 246 1 178 177 45.7

4.8 98 12.01 52.8 10.8 96.9 298 1 237 115 45.7

2.3 99 7.95 51.8 4.6 54.9 163 1 128 93 42.9

6.2 100 10.15 51.9 16.4 59.2 568 0 452 371 62.9

2.6 101 9.76 53.2 6.9 80.1 64 1 47 55 22.9

4.3 102 9.89 45.2 11.8 108.7 190 1 141 112 42.9

2.7 103 7.14 57.6 13.1 92.6 92 1 40 50 22.9

6.6 104 13.95 65.9 15.6 133.5 356 1 308 182 62.9

4.5 105 9.44 52.5 10.9 58.5 297 1 230 263 42.9

2.9 106 10.8 63.9 1.6 57.4 130 1 69 62 22.9

1.4 107 7.14 51.7 4.1 45.7 115 1 90 19 22.9

2.1 108 8.02 55 3.8 46.5 91 1 44 32 22.9

. 109 11.8 53.8 9.1 116.9 571 0 441 469 62.9

. 110 9.5 49.3 42 70.9 98 1 68 46 22.9

. 111 7.7 56.9 12.2 67.9 129 1 85 136 62.9

. 112 17.94 56.2 26.4 91.8 835 0 791 407 62.9

. 113 9.41 59.5 20.6 91.7 29 1 20 22 22.9

;

run;

proc reg data=data;

model risk=stay;

model risk=age;

model risk=culturing;

model risk=xray;

model risk=beds;

model risk=school;

model risk=census;

model risk=nurses;

model risk=facilities;

run;

proc reg data=data;

model risk=stay age culturing xray beds school census nurses facilities / selection=f slentry=0.1;

model risk=stay age culturing xray beds school census nurses facilities / selection=b slentry=0.1;

model risk=stay age culturing xray beds school census nurses facilities / selection=stepwise slentry=0.1;

model risk=stay age culturing xray beds school census nurses facilities / selection=cp;

run;

proc reg data=data;

model risk=stay culturing xray facilities/ cli;

run;