Here is the output of the R code,

|  |
| --- |
| > #Multiple Logistic Regression in R  >  > library(faraway)  > library(StepReg)  >  > logistic <- glm(test ~ age + bmi + diastolic + diabetes + glucose + insulin + pregnant + triceps, family=binomial(logit),data=pima)  > summary(logistic)  Call:  glm(formula = test ~ age + bmi + diastolic + diabetes + glucose +  insulin + pregnant + triceps, family = binomial(logit), data = pima)  Deviance Residuals:  Min 1Q Median 3Q Max  -2.5566 -0.7274 -0.4159 0.7267 2.9297  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -8.4046964 0.7166359 -11.728 < 2e-16 \*\*\*  age 0.0148690 0.0093348 1.593 0.111192  bmi 0.0897010 0.0150876 5.945 2.76e-09 \*\*\*  diastolic -0.0132955 0.0052336 -2.540 0.011072 \*  diabetes 0.9451797 0.2991475 3.160 0.001580 \*\*  glucose 0.0351637 0.0037087 9.481 < 2e-16 \*\*\*  insulin -0.0011917 0.0009012 -1.322 0.186065  pregnant 0.1231823 0.0320776 3.840 0.000123 \*\*\*  triceps 0.0006190 0.0068994 0.090 0.928515  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 993.48 on 767 degrees of freedom  Residual deviance: 723.45 on 759 degrees of freedom  AIC: 741.45  Number of Fisher Scoring iterations: 5  > analysis(logistic)  Error in analysis(logistic) : could not find function "analysis"  >  > y <- "test"  > stepwiselogit(data=pima,y, exclude = NULL, include = NULL, selection = "forward",  + select = "AIC", sle = 0.15, sls = 0.15)  $SummaryOfSelection  Step EnteredEffect RemovedEffect DF NumberIn AIC  1 1 glucose 1 1 812.7196  2 2 bmi 1 2 777.403  3 3 pregnant 1 3 752.1249  4 4 diabetes 1 4 744.3059  5 5 diastolic 1 5 740.5596  6 6 age 1 6 739.4617  7 7 insulin 1 7 739.4534  $AnalysisOfMaximumLikelihoodEstimate  Parameter Estimate Std. Error z value Pr(>|z|)  (Intercept) (Intercept) -8.4051 0.7167 -11.7275 0  glucose glucose 0.0351 0.0037 9.587 0  bmi bmi 0.0901 0.0145 6.2294 0  pregnant pregnant 0.1232 0.0321 3.8409 1e-04  diabetes diabetes 0.9476 0.298 3.1798 0.0015  diastolic diastolic -0.0132 0.0052 -2.5639 0.0103  age age 0.0148 0.0093 1.592 0.1114  insulin insulin -0.0012 8e-04 -1.4211 0.1553  >  > stepwiselogit(data=pima,y, exclude = NULL, include = NULL, selection = "backward",  + select = "AIC", sle = 0.15, sls = 0.15)  $SummaryOfSelection  Step EnteredEffect RemovedEffect DF NumberIn AIC  1 1 triceps 1 7 739.4534  $AnalysisOfMaximumLikelihoodEstimate  Parameter Estimate Std. Error z value Pr(>|z|)  (Intercept) (Intercept) -8.4051 0.7167 -11.7275 0  pregnant pregnant 0.1232 0.0321 3.8409 1e-04  glucose glucose 0.0351 0.0037 9.587 0  diastolic diastolic -0.0132 0.0052 -2.5639 0.0103  insulin insulin -0.0012 8e-04 -1.4211 0.1553  bmi bmi 0.0901 0.0145 6.2294 0  diabetes diabetes 0.9476 0.298 3.1798 0.0015  age age 0.0148 0.0093 1.592 0.1114  >  > stepwiselogit(data=pima,y, exclude = NULL, include = NULL, selection = "bidirection",  + select = "AIC", sle = 0.15, sls = 0.15)  $SummaryOfSelection  Step EnteredEffect RemovedEffect DF NumberIn AIC  1 1 glucose 1 1 812.7196  2 2 bmi 1 2 777.403  3 3 pregnant 1 3 752.1249  4 4 diabetes 1 4 744.3059  5 5 diastolic 1 5 740.5596  6 6 age 1 6 739.4617  7 7 insulin 1 7 739.4534  $AnalysisOfMaximumLikelihoodEstimate  Parameter Estimate Std. Error z value Pr(>|z|)  (Intercept) (Intercept) -8.4051 0.7167 -11.7275 0  glucose glucose 0.0351 0.0037 9.587 0  bmi bmi 0.0901 0.0145 6.2294 0  pregnant pregnant 0.1232 0.0321 3.8409 1e-04  diabetes diabetes 0.9476 0.298 3.1798 0.0015  diastolic diastolic -0.0132 0.0052 -2.5639 0.0103  age age 0.0148 0.0093 1.592 0.1114  insulin insulin -0.0012 8e-04 -1.4211 0.1553 |
|  |
| |  | | --- | | > | |

The forward regression data says that the variable glucose is the most needed as it is the first one that came out in the AIC regression. The variables bmi and pregnant are also important, while insulin is the least important

Looking at the backward elimination regression, we can see that the variable triceps was removed, that means, the variable triceps is the least needed.

Looking at the bidirectional regression model, we don’t see any variables getting removed as any variable gets added. The results are quite similar as forward regression results.