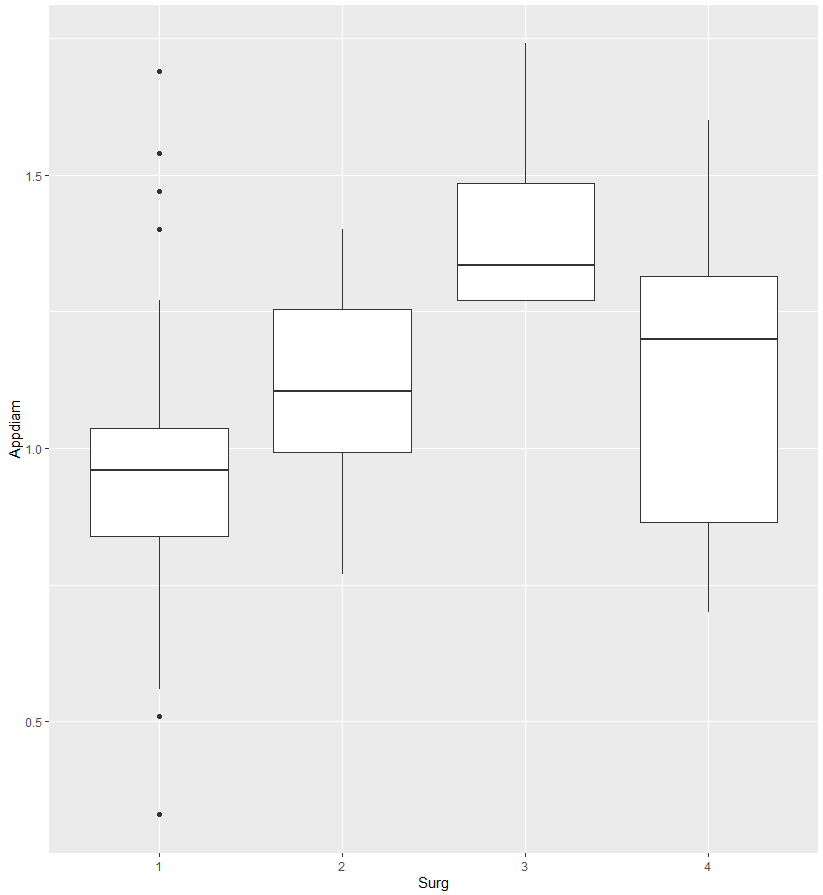
**We removed patients without surgical result and the specific characteristic in each part.**

**Largest diameter of the appendix (cm)**

****

**Anova:** F = 0.221 p-value = 0.882

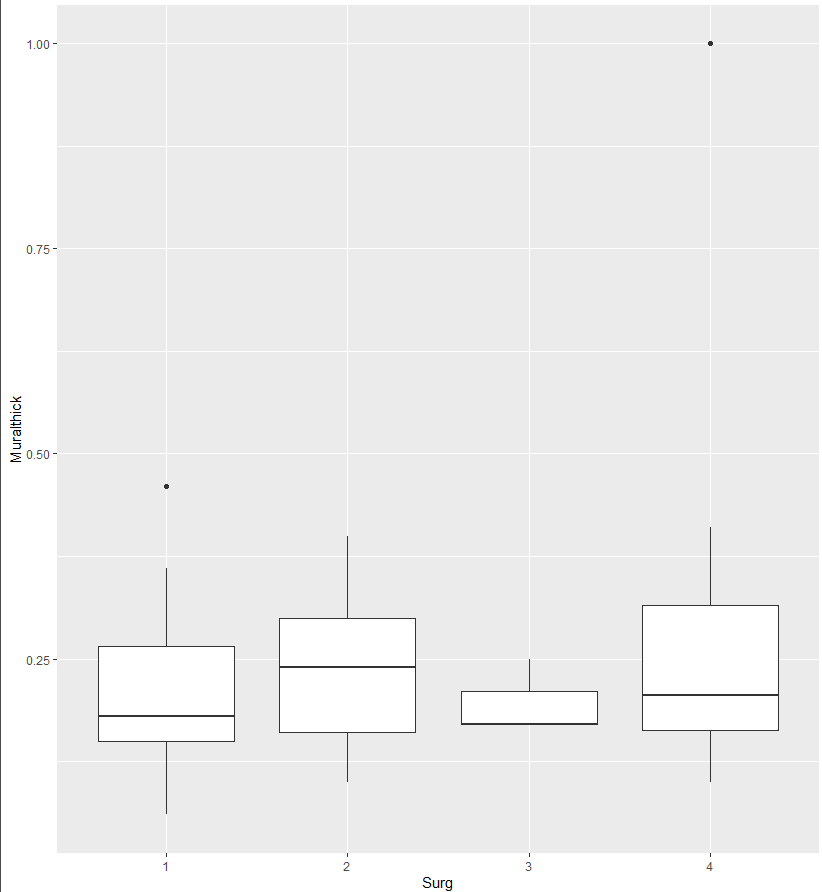
This means we fail to reject the means of appendix diameter are different in different surgical stages.

Also, I checked the pairwise comparison by Tukey’s method. The table below show the pairwise difference, confidence interval and the p-value of it.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | diff | lower | upper | *p-value* |
| 2-1 | 0.020 | -0.596 | 0.637 | *0.999* |
| 3-1 | 0.323 | -0.732 | 1.377 | *0.852* |
| 4-1 | 0.425 | -0.635 | 0.721 | *0.998* |
| 3-2 | 0.302 | -0.845 | 1.450 | *0.899* |
| 4-2 | 0.022 | -0.973 | 0.838 | *0.999* |
| 4-3 | -0.280 | -1.462 | 0.902 | *0.924* |

The result shows that we fail to reject the means of appendix diameter are different in each pair of surgical stages.

**Maximal mural diameter (cm)**



**Anova:** F = 1.185 p-value = 0.323

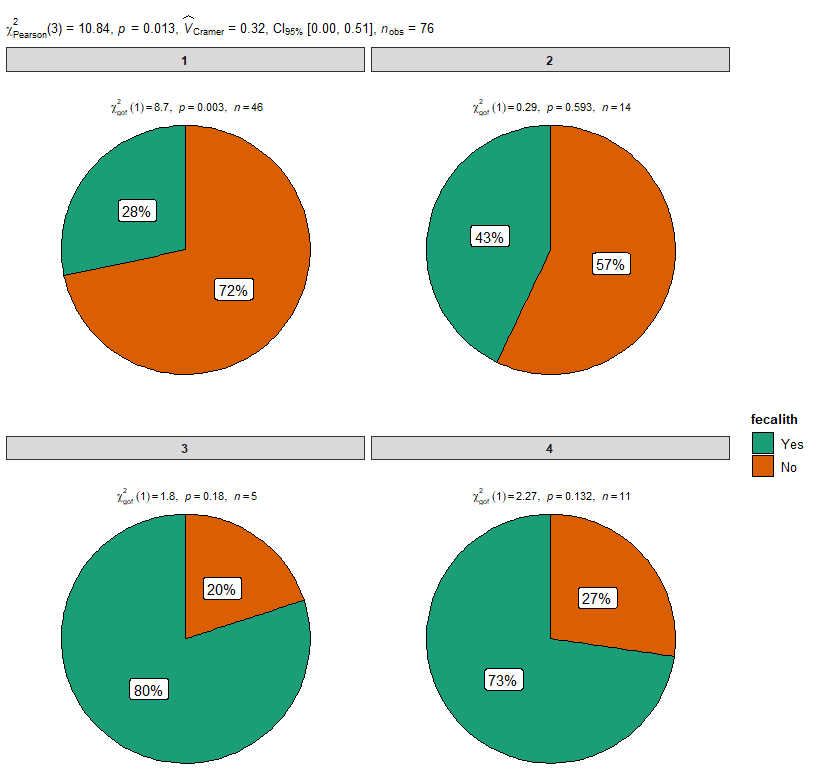
This means we fail to reject the means of mural diameter are different in different surgical stages.

Also, I checked the pairwise comparison by Tukey’s method. The table below show the pairwise difference, confidence interval and the p-value of it.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | diff | lower | upper | p-value |
| 2-1 | 0.020 | -0.095 | 0.134 | 0.968 |
| 3-1 | -0.011 | -0.223 | 0.200 | 0.998 |
| 4-1 | 0.088 | -0.038 | 0.214 | 0.263 |
| 3-2 | -0.031 | -0.256 | 0.194 | 0.8983 |
| 4-2 | 0.068 | -0.079 | 0.216 | 0.616 |
| 4-3 | 0.099 | -0.132 | 0.331 | 0.670 |

The result shows that we fail to reject the means of mural diameter are different in each pair of surgical stages.

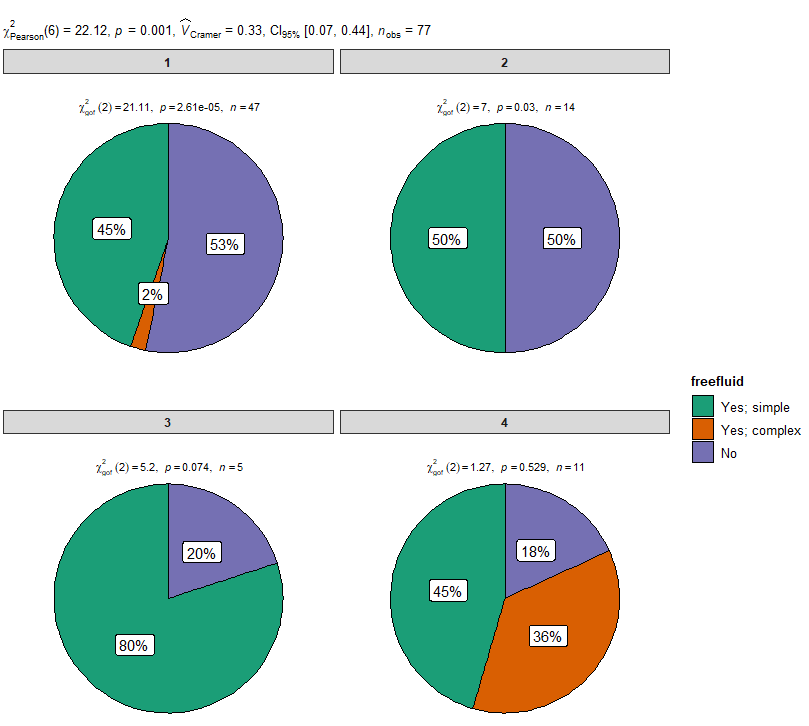
**Fecalith**

****

By Pearson chi-square test, the p-value is 0.013. Therefore, we can conclude there is association between fecalith and surgical stage.

Due to the limited sample size, I also checked the association by Fisher’s exact test. P-value is 0.0095 which indicates that there is association between between fecalith and surgical stage.

**Fluid**

****

By Pearson chi-square test, the p-value is 0.013. Therefore, we can conclude there is association between fluid and surgical stage.

Due to the limited sample size, I also checked the association by Fisher’s exact test. P-value is 0.0136 which indicates that there is association between between fluid and surgical stage.

**Regression (Based on 74 patients after remove NA)**

From above visualization, I add fecalith and fluid as covariates in the prediction regression model for surgical stage.

All following variables are binary variable.

surg, it means surgical stage which combine combine 1&2, 3&4

app, it is indicator of diameter of appendix with cutoff point at 1mm.

lossSM, it means loss of normal echogenic submucosal layer. In the dataset, loss of SM layer is captured by the ultrasound stage (combine 1&2, 3&4)

fecalith: presence of fecalith

freefluid: Presence of surrounding free fluid

I have tried following 4 logistic regression models:

Model 1: surg ~ app\_ + lossSM + fecalith + freefluid

**Model 2: surg ~** **app\_ + lossSM + fecalith**

Model 3: surg ~ app\_ + lossSM + freefluid

Model 4: surg ~ lossSM + fecalith + freefluid

Model 5: surg ~ lossSM + app\_

Model 6: surg ~ lossSM + fecalith

Model 7: surg ~ lossSM + freefluid

Model 8: surg ~ lossSM (The model picked by stepwise model selection by AIC)

I checked their performance by confusion matrix and related statistics. (cut at probability 0.5)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | 1 | **2** | 3 | 4 | **5** | 6 | 7 | 8 |
| Accuracy | 0.838 | **0.865** | 0.824 | 0.797 | **0.811** | 0.811 | 0.797 | 0.730 |
| Kappa | 0.410 | **0.560** | 0.443 | 0.388 | **0.491** | 0.491 |  | 0.442 |
| Sensitivity | 0.400 | **0.600** | 0.533 | 0.533 | **0.733** | 0.733 |  | 1.000 |
| Specificity | 0.949 | **0.932** | 0.8983 | 0.864 | **0.831** | 0.831 |  | 0.661 |
| AUC | 0.906 | **0.894** | 0.885 | 0.870 | **0.870** | 0.870 |  |  |

Graphical user interface

Description automatically generated with low confidence

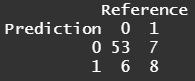
Text

Description automatically generated with low confidenceText

Description automatically generated with medium confidenceText

Description automatically generated with low confidenceText

Description automatically generated with low confidenceText

Description automatically generated with medium confidenceA picture containing text

Description automatically generated

Model 2 have the highest accuracy and the best kappa. But considering the balance between sensitivity and specificity, I would like to choose Model 5.

After that I checked the correlation between lossSM and all other variables by fisher exact test. The result shows that all the other variables are correlated with lossSM, which might be the reason the model selection remove other variables.