Part a:

i.

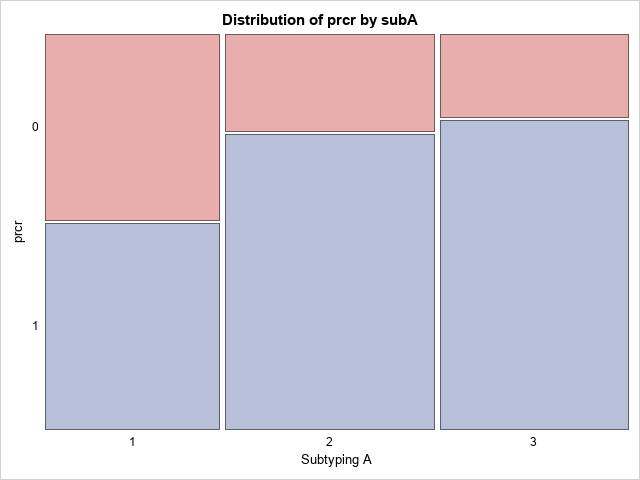
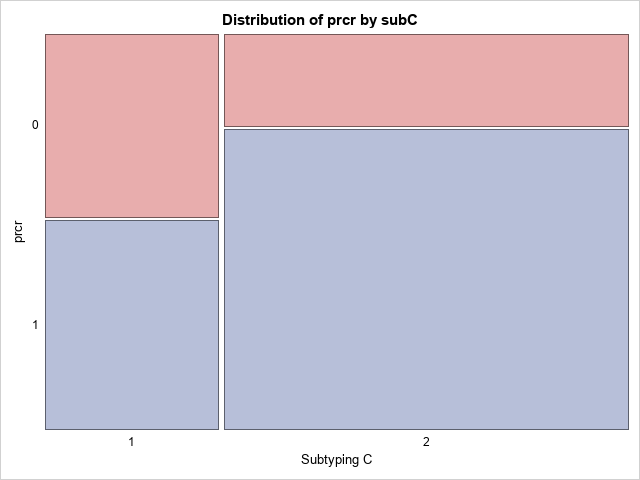
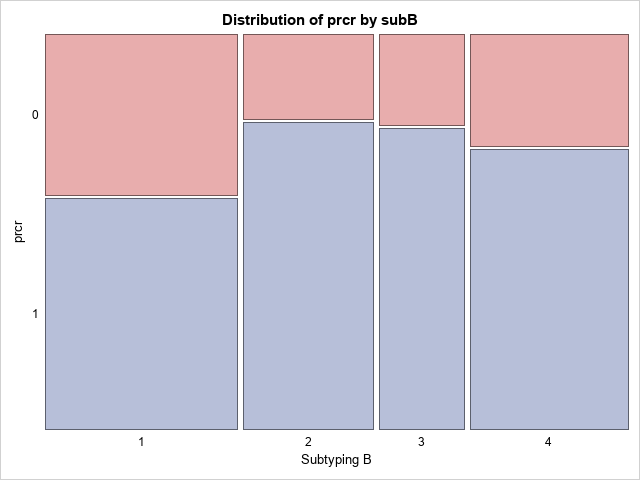
Tables 1-3 show the percentage of subjects within each response category by subtype for each of the three subtyping strategies. For strategy A, more than 50% of subjects with subtypes A2 and A3 fall into the complete response category. This is a similar trend for subtypes B2-4 and C2. Subtype A1, B1 and C1 shows a more even distribution of response categories, with higher percentages trending towards progressive disease.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Response** | **Subtyping A** | | | |
| **1** | **2** | **3** | **Total** |
| **1** | 31.15 | 6.85 | 10.61 |  |
| **2** | 16.39 | 17.81 | 10.61 |  |
| **3** | 32.79 | 23.29 | 19.7 |  |
| **4** | 19.67 | 52.05 | 59.09 |  |
| **Total** | 61 | 73 | 66 | 200 |

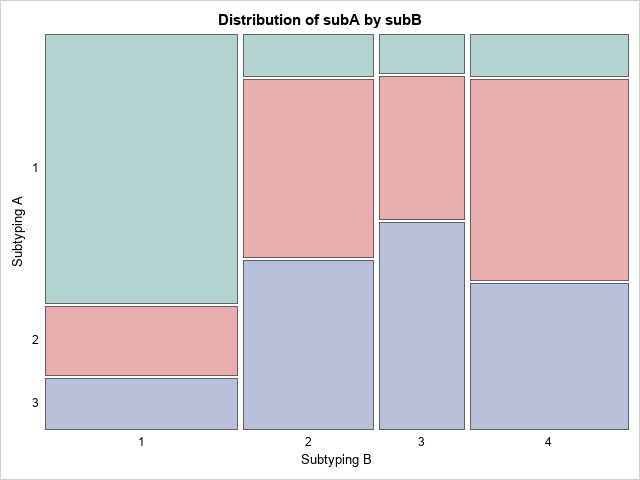
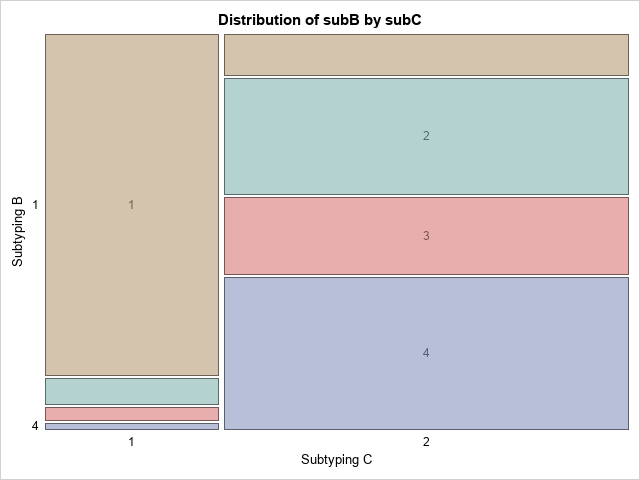
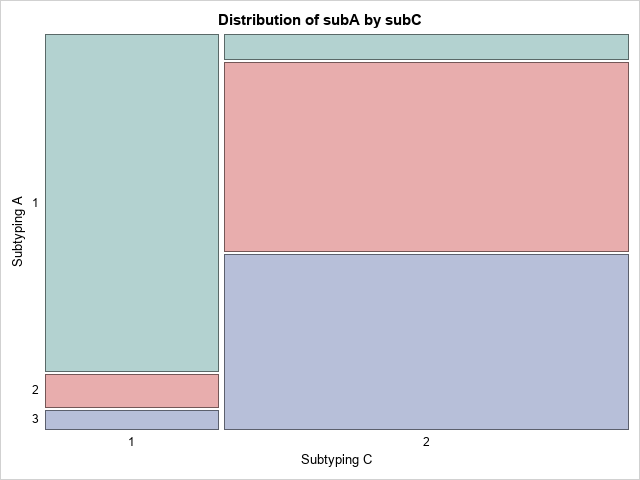
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Response** | **Subtyping B** | | | | |
| **1** | **2** | **3** | **4** | **Total** |
| **1** | 27.94 | 13.04 | 0 | 10.71 |  |
| **2** | 13.24 | 8.7 | 23.33 | 17.86 |  |
| **3** | 33.82 | 23.91 | 20 | 17.86 |  |
| **4** | 25 | 54.35 | 56.67 | 53.57 |  |
| **Total** | 68 | 46 | 30 | 56 | 200 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Response** | **Subtyping C** | | |
| **1** | **2** | **Total** |
| **1** | 30 | 9.29 |  |
| **2** | 16.67 | 14.29 |  |
| **3** | 33.33 | 21.43 |  |
| **4** | 20 | 55 |  |
| **Total** | 60 | 140 | 200 |

ii. Figure 1 shows the proportion of participants with partial or complete response (1, purple), vs those with stable or progressive disease (0, pink) for each subtyping approach, pooling across study arms.

iii. Figure 2 shows the overlap between subtyping approaches A vs B, B vs C, and A vs C. There appears to be an extremely high correlation between subtypes A1, B1 and C1. Subtypes B2, B3 and B4 appear to have similar overlap across subtypes A2 and A3. The majority of people with subtype C2 also have subtypes B2-B4 but few have subtype B1. Of those with subtype C2, most have subtype A2 and A3, but few have subtype A1.

iv.

Subtypes A2 and A3 have similar profiles, subtypes B2-4 have similar profiles, and subtype C2 appears to be similar to the aforementioned subtypes. Subtypes A1, B1 and C1 appear to have similar proportions of response category. It appears that a more complicated strategy such as B is not warranted, and a dichotomous subtyping approach such as C appears to capture most of the variability in response between subtypes.

Part b:

i.

Fit a proportional odds model for each subtyping strategy with subtype as the only predictor. Check the model assumptions using appropriate diagnostics.

Model: logit (theta\_hij) = alpha\_k + x’\_hi \* beta\_k

Theta = probability of worse disease state (lower vs higher number of response category, where 1=progressive disease and 4=complete response)

Alpha1 = log odds of progressive disease vs all other categories, alpha2 = log odds of progressive disease or stable disease vs other categories, alpha3 = log odds of progressive disease, stable disease or partial response vs complete response

Beta1 = increment to log odds of more favorable outcome for subtype 2 vs 1, Beta2 = increment to log odds of more favorable outcome for subtype 3 vs 1, Beta3 = increment to log odds of more favorable outcome for subtype 4 vs 1

[Gamma1 = increment to log odds of more favorable outcome for arm 2 vs arm 1 in part D]

Proportional odds assumption states that each of these betas are equal, so the differences in log odds amongst subtypes can be assumed to be equal for each cumulative logit. For all 3 subtyping strategies the p value for the proportional odds assumption is greater than 0.05, indicating that the probabilities of each response category vary by cumulative logit. The model for subtyping strategy B has an associated p value of 0.0835, indicating some degree of homogeneity across cumulative logits. This could potentially be explained by the graph in part a ii, which shows that subtype B1 has a different response profile from B2, B3 and B4 which are similar.

ii.

Perform test of null hypothesis that subtype is not associated with response category.

Subtype A: p <0.0001, chi squared 2 df, Wald test

Subtype B: p=0.0012, chi squared 3 df

Subtype C: p<0.0001, chi squared 1 df

For all three subtyping approaches, subtype is significantly associated with response category.

Part c:

The approach with the smallest p value is not necessarily the best option for categorizing patients by predicted clinical outcome. It is possible that a subtyping approach with more subtypes than other options may simply be overfitting the data. Similarly, an approach such as C with only two categories may not be capturing the subtle nuances of certain subtypes since there are 4 response categories.

Part d: (describe stat means to compare each subtype approach, given models in B)

Ideas: Goal is to correctly classify disease response class of patients based on subtyping approach. In order to pick the best subtyping approach it would be important to examine the proportional odds assumption and how strongly associated each subtype is associated with outcome for each approach. The test of the proportional odds assumption casts a wide net and makes a very general assumption about all variables and all logits. If two subtypes for one subtype approach equally could predict the same disease outcome it would make sense to collapse those subtypes into one subtype.

Part e: Repeat analysis in B and D using arm as a covariate in proportional odds model.

Perform test of null hypothesis that subtype is not associated with response category.

Subtype A: p <0.0001, chi squared 2 df, Wald test

Subtype B: p=0.0012, chi squared 3 df

Subtype C: p<0.0001, chi squared 1 df

i.

For subtyping approach C, the proportional odds model indicates that the log odds of better response for someone with subtype C1 in arm 2 compared to someone with subtype C1 in arm 1 is 0.822. Therefore, for all cumulative logits, an individual in arm 2 compared to an individual in arm 1, is 2.275 times are likely to have a better outcome. The test of the null hypothesis of no treatment effect follows a chi square distribution with 1 degree of freedom. The test statistic has a value of 9.00 and an associated p value of 0.0027. Thus, we reject the null hypothesis of no treatment effect when applying subtyping approach C. The confidence interval for this odds ratio is 1.33 to 3.89.

An interaction term was included between arm and subtype to see if treatment effects depend on subtype. The p value for the interaction is 0.0079, indicating that treatment has different effects across subtypes.

ii. (Which drug would you recommend for each subtype based on these results?)

I would recommend Gemcitabine + Abraxane (arm 2) for subtype C1 and Folfirinox (arm 1) for subtype C2.

For subtyping approach A, Folfirinox (arm 1) appears to have a higher proportion of complete response for all subtypes.

For subtyping approach B, Folfirinox (arm 1) appears to have a higher proportion of complete response for subtypes B2-B4. Both arms appear to have the same proportion of complete response for patients with subtype B1, but I would recommend Gemcitabine + Abraxane for those with subtype B1 because there is a higher proportion of either complete response or partial response as compared with Folfirinox.