übung2

Lilli Schuckert

27 4 2021

## Aufgabe 1

## # A tibble: 79 x 15  
## Species Status Length Mass Range Migr Insect Diet Clutch Broods Wood  
## <chr> <int> <int> <dbl> <dbl> <int> <int> <int> <dbl> <int> <int>  
## 1 Cyg\_ol~ 1 1520 9600 1.21 1 12 2 6 1 0  
## 2 Cyg\_at~ 1 1250 5000 0.56 1 0 1 6 1 0  
## 3 Cer\_no~ 1 870 3360 0.07 1 0 1 4 1 0  
## 4 Ans\_ca~ 0 720 2517 1.1 3 12 2 3.8 1 0  
## 5 Ans\_an~ 0 820 3170 3.45 3 0 1 5.9 1 0  
## 6 Bra\_ca~ 1 770 4390 2.96 2 0 1 5.9 1 0  
## 7 Bra\_sa~ 0 50 1930 0.01 1 0 1 4 2 0  
## 8 Alo\_ae~ 0 680 2040 2.71 1 NA 2 8.5 1 0  
## 9 Ana\_pl~ 1 570 1020 9.01 2 6 2 12.6 1 0  
## 10 Ana\_ac~ 0 580 910 7.9 3 6 2 8.3 1 0  
## # ... with 69 more rows, and 4 more variables: Upland <int>, Water <int>,  
## # Release <int>, Indiv <int>

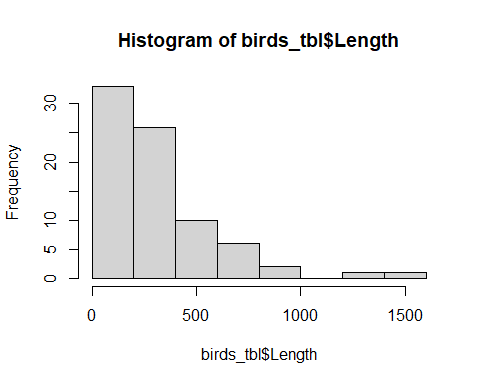
### 1.Univariate Übersichtstabelle mit tableone

### 2. Fehlende Werte und nonnormal Variablen

## 'data.frame': 79 obs. of 15 variables:  
## $ Species: chr "Cyg\_olor" "Cyg\_atra" "Cer\_nova" "Ans\_caer" ...  
## $ Status : int 1 1 1 0 0 1 0 0 1 0 ...  
## $ Length : int 1520 1250 870 720 820 770 50 680 570 580 ...  
## $ Mass : num 9600 5000 3360 2517 3170 ...  
## $ Range : num 1.21 0.56 0.07 1.1 3.45 2.96 0.01 2.71 9.01 7.9 ...  
## $ Migr : int 1 1 1 3 3 2 1 1 2 3 ...  
## $ Insect : int 12 0 0 12 0 0 0 NA 6 6 ...  
## $ Diet : int 2 1 1 2 1 1 1 2 2 2 ...  
## $ Clutch : num 6 6 4 3.8 5.9 5.9 4 8.5 12.6 8.3 ...  
## $ Broods : int 1 1 1 1 1 1 2 1 1 1 ...  
## $ Wood : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Upland : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Water : int 1 1 1 1 1 1 0 1 1 1 ...  
## $ Release: int 6 10 3 1 2 10 1 1 17 3 ...  
## $ Indiv : int 29 85 8 10 7 60 2 8 1539 102 ...

##   
## level Overall   
## n 79   
## Status (%) 0 52 (65.8)   
## 1 27 (34.2)   
## Length (mean (SD)) 329.38 (262.41)   
## Mass (mean (SD)) 663.89 (1434.93)  
## Range (mean (SD)) 2.34 (2.18)   
## Migr (%) 1 40 (50.6)   
## 2 24 (30.4)   
## 3 15 (19.0)   
## Insect (mean (SD)) 7.29 (4.78)   
## Diet (%) 1 28 (35.4)   
## 2 43 (54.4)   
## 3 8 (10.1)   
## Clutch (mean (SD)) 6.00 (3.36)   
## Broods (%) 1 37 (50.0)   
## 2 23 (31.1)   
## 3 12 (16.2)   
## 5 1 ( 1.4)   
## 7 1 ( 1.4)   
## Wood (%) 0 35 (44.3)   
## 1 44 (55.7)   
## Upland (%) 0 71 (89.9)   
## 1 8 (10.1)   
## Water (%) 0 59 (74.7)   
## 1 20 (25.3)   
## Release (mean (SD)) 6.29 (5.91)   
## Indiv (mean (SD)) 185.20 (314.20)

##   
## ### Summary of continuous variables ###  
##   
## strata: Overall  
## n miss p.miss mean sd median p25 p75 min max skew kurt  
## Length 79 0 0 329 262 230 149.5 418 50.00 1520 2.2 6.0  
## Mass 79 5 6 664 1435 166 23.5 577 11.50 9600 4.2 21.6  
## Range 79 0 0 2 2 2 0.6 3 0.01 9 1.2 1.2  
## Insect 79 2 3 7 5 7 3.0 12 0.00 12 -0.3 -1.5  
## Clutch 79 0 0 6 3 5 3.9 8 1.90 16 1.3 0.7  
## Release 79 0 0 6 6 4 2.0 10 1.00 27 1.3 1.3  
## Indiv 79 3 4 185 314 33 8.0 227 2.00 1539 2.5 7.1  
##   
## =======================================================================================  
##   
## ### Summary of categorical variables ###   
##   
## strata: Overall  
## var n miss p.miss level freq percent cum.percent  
## Status 79 0 0.0 0 52 65.8 65.8  
## 1 27 34.2 100.0  
##   
## Migr 79 0 0.0 1 40 50.6 50.6  
## 2 24 30.4 81.0  
## 3 15 19.0 100.0  
##   
## Diet 79 0 0.0 1 28 35.4 35.4  
## 2 43 54.4 89.9  
## 3 8 10.1 100.0  
##   
## Broods 79 5 6.3 1 37 50.0 50.0  
## 2 23 31.1 81.1  
## 3 12 16.2 97.3  
## 5 1 1.4 98.6  
## 7 1 1.4 100.0  
##   
## Wood 79 0 0.0 0 35 44.3 44.3  
## 1 44 55.7 100.0  
##   
## Upland 79 0 0.0 0 71 89.9 89.9  
## 1 8 10.1 100.0  
##   
## Water 79 0 0.0 0 59 74.7 74.7  
## 1 20 25.3 100.0  
##



##   
## Overall   
## n 79   
## Status = 1 (%) 27 (34.2)   
## Length (median [IQR]) 230.00 [149.50, 417.50]  
## Mass (median [IQR]) 166.00 [23.52, 577.25]   
## Range (mean (SD)) 2.34 (2.18)   
## Migr (%)   
## 1 40 (50.6)   
## 2 24 (30.4)   
## 3 15 (19.0)   
## Insect (mean (SD)) 7.29 (4.78)   
## Diet (%)   
## 1 28 (35.4)   
## 2 43 (54.4)   
## 3 8 (10.1)   
## Clutch (mean (SD)) 6.00 (3.36)   
## Broods (%)   
## 1 37 (50.0)   
## 2 23 (31.1)   
## 3 12 (16.2)   
## 5 1 ( 1.4)   
## 7 1 ( 1.4)   
## Wood = 1 (%) 44 (55.7)   
## Upland = 1 (%) 8 (10.1)   
## Water = 1 (%) 20 (25.3)   
## Release (mean (SD)) 6.29 (5.91)   
## Indiv (median [IQR]) 33.00 [8.00, 226.75]

### 3. Stratifizierung nach einer beliebigen Variable

## Stratified by Diet  
## 1 2   
## n 28 43   
## Status = 1 (%) 8 ( 28.6) 16 ( 37.2)   
## Length (median [IQR]) 265.50 [120.00, 436.25] 222.00 [155.00, 417.50]  
## Mass (median [IQR]) 227.50 [18.62, 803.75] 79.80 [23.60, 539.00]   
## Range (mean (SD)) 1.90 (1.85) 2.55 (2.21)   
## Migr (%)   
## 1 19 ( 67.9) 16 ( 37.2)   
## 2 6 ( 21.4) 16 ( 37.2)   
## 3 3 ( 10.7) 11 ( 25.6)   
## Insect (mean (SD)) 2.14 (2.03) 10.17 (2.89)   
## Diet (%)   
## 1 28 (100.0) 0 ( 0.0)   
## 2 0 ( 0.0) 43 (100.0)   
## 3 0 ( 0.0) 0 ( 0.0)   
## Clutch (mean (SD)) 7.07 (3.92) 5.77 (2.96)   
## Broods (%)   
## 1 13 ( 50.0) 19 ( 47.5)   
## 2 6 ( 23.1) 15 ( 37.5)   
## 3 6 ( 23.1) 6 ( 15.0)   
## 5 0 ( 0.0) 0 ( 0.0)   
## 7 1 ( 3.8) 0 ( 0.0)   
## Wood = 1 (%) 10 ( 35.7) 28 ( 65.1)   
## Upland = 1 (%) 6 ( 21.4) 2 ( 4.7)   
## Water = 1 (%) 7 ( 25.0) 11 ( 25.6)   
## Release (mean (SD)) 6.43 (6.12) 6.40 (6.04)   
## Indiv (median [IQR]) 54.00 [10.00, 285.50] 29.00 [7.25, 228.50]   
## Stratified by Diet  
## 3 p test   
## n 8   
## Status = 1 (%) 3 ( 37.5) 0.739   
## Length (median [IQR]) 292.50 [261.25, 355.00] 0.541 nonnorm  
## Mass (median [IQR]) 298.00 [201.00, 349.00] 0.601 nonnorm  
## Range (mean (SD)) 2.67 (3.05) 0.428   
## Migr (%) 0.134   
## 1 5 ( 62.5)   
## 2 2 ( 25.0)   
## 3 1 ( 12.5)   
## Insect (mean (SD)) 10.50 (4.24) <0.001   
## Diet (%) <0.001   
## 1 0 ( 0.0)   
## 2 0 ( 0.0)   
## 3 8 (100.0)   
## Clutch (mean (SD)) 3.50 (1.22) 0.022   
## Broods (%) 0.091   
## 1 5 ( 62.5)   
## 2 2 ( 25.0)   
## 3 0 ( 0.0)   
## 5 1 ( 12.5)   
## 7 0 ( 0.0)   
## Wood = 1 (%) 6 ( 75.0) 0.026   
## Upland = 1 (%) 0 ( 0.0) 0.044   
## Water = 1 (%) 2 ( 25.0) 0.998   
## Release (mean (SD)) 5.25 (4.98) 0.874   
## Indiv (median [IQR]) 21.00 [5.00, 172.50] 0.611 nonnorm

## Aufgabe 2

## Warning in table1.formula(~Water + Status | Diet, data = birds\_tbl, overall =  
## "Total"): Terms to the right of '|' in formula 'x' define table columns and are  
## expected to be factors with meaningful labels.

## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th class='firstrow lastrow'><span class='stratlabel'>1<br><span class='stratn'>(N=28)</span></span></th>\n<th class='firstrow lastrow'><span class='stratlabel'>2<br><span class='stratn'>(N=43)</span></span></th>\n<th class='firstrow lastrow'><span class='stratlabel'>3<br><span class='stratn'>(N=8)</span></span></th>\n<th class='firstrow lastrow'><span class='stratlabel'>Total<br><span class='stratn'>(N=79)</span></span></th>\n</tr>\n</thead>\n<tbody>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Water</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Mean (SD)</td>\n<td>0.250 (0.441)</td>\n<td>0.256 (0.441)</td>\n<td>0.250 (0.463)</td>\n<td>0.253 (0.438)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Median [Min, Max]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n</tr>\n<tr>\n<td class='rowlabel firstrow'><span class='varlabel'>Status</span></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n<td class='firstrow'></td>\n</tr>\n<tr>\n<td class='rowlabel'>Mean (SD)</td>\n<td>0.286 (0.460)</td>\n<td>0.372 (0.489)</td>\n<td>0.375 (0.518)</td>\n<td>0.342 (0.477)</td>\n</tr>\n<tr>\n<td class='rowlabel lastrow'>Median [Min, Max]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n<td class='lastrow'>0 [0, 1.00]</td>\n</tr>\n</tbody>\n</table>\n"

## Aufgabe 3: Multiple gaussian linear regression

##   
## Call:  
## lm(formula = Status ~ Length + Wood + Diet, data = birds\_tbl)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5757 -0.3515 -0.2384 0.5815 0.7922   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0941239 0.1774960 0.530 0.597   
## Length 0.0003533 0.0002115 1.671 0.099 .  
## Wood 0.1341073 0.1159735 1.156 0.251   
## Diet 0.0323941 0.0893010 0.363 0.718   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4747 on 75 degrees of freedom  
## Multiple R-squared: 0.04898, Adjusted R-squared: 0.01094   
## F-statistic: 1.288 on 3 and 75 DF, p-value: 0.2849

## Aufgabe 4: Multiple logistic linear regression

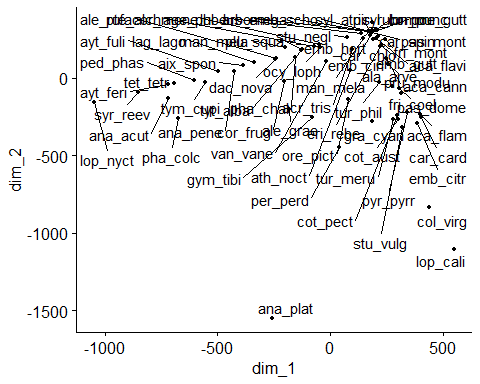
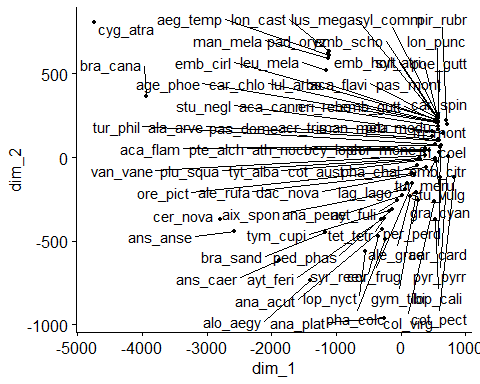
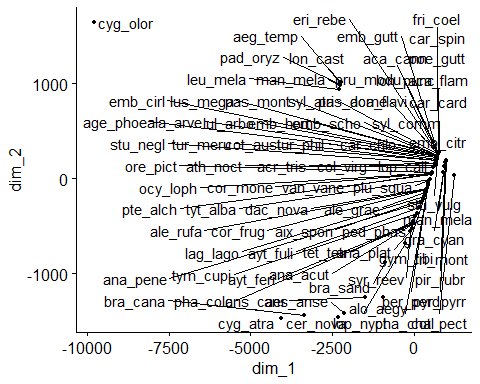
##   
## -- Column specification --------------------------------------------------------  
## cols(  
## Species = col\_character(),  
## Status = col\_double(),  
## Length = col\_double(),  
## Mass = col\_double(),  
## Range = col\_double(),  
## Migr = col\_double(),  
## Insect = col\_double(),  
## Diet = col\_double(),  
## Clutch = col\_double(),  
## Broods = col\_double(),  
## Wood = col\_double(),  
## Upland = col\_double(),  
## Water = col\_double(),  
## Release = col\_double(),  
## Indiv = col\_double()  
## )

## # A tibble: 79 x 15  
## species status length mass range migr insect diet clutch broods wood  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 cyg\_ol~ 1 1520 9600 1.21 1 12 2 6 1 0  
## 2 cyg\_at~ 1 1250 5000 0.56 1 0 1 6 1 0  
## 3 cer\_no~ 1 870 3360 0.07 1 0 1 4 1 0  
## 4 ans\_ca~ 0 720 2517 1.1 3 12 2 3.8 1 0  
## 5 ans\_an~ 0 820 3170 3.45 3 0 1 5.9 1 0  
## 6 bra\_ca~ 1 770 4390 2.96 2 0 1 5.9 1 0  
## 7 bra\_sa~ 0 50 1930 0.01 1 0 1 4 2 0  
## 8 alo\_ae~ 0 680 2040 2.71 1 NA 2 8.5 1 0  
## 9 ana\_pl~ 1 570 1020 9.01 2 6 2 12.6 1 0  
## 10 ana\_ac~ 0 580 910 7.9 3 6 2 8.3 1 0  
## # ... with 69 more rows, and 4 more variables: upland <dbl>, water <dbl>,  
## # release <dbl>, indiv <dbl>

## # A tibble: 79 x 14  
## status length mass range migr insect diet clutch broods wood upland water  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 1520 9600 1.21 1 12 2 6 1 0 0 1  
## 2 1 1250 5000 0.56 1 0 1 6 1 0 0 1  
## 3 1 870 3360 0.07 1 0 1 4 1 0 0 1  
## 4 0 720 2517 1.1 3 12 2 3.8 1 0 0 1  
## 5 0 820 3170 3.45 3 0 1 5.9 1 0 0 1  
## 6 1 770 4390 2.96 2 0 1 5.9 1 0 0 1  
## 7 0 50 1930 0.01 1 0 1 4 2 0 0 0  
## 8 0 680 2040 2.71 1 NA 2 8.5 1 0 0 1  
## 9 1 570 1020 9.01 2 6 2 12.6 1 0 0 1  
## 10 0 580 910 7.9 3 6 2 8.3 1 0 0 1  
## # ... with 69 more rows, and 2 more variables: release <dbl>, indiv <dbl>

##   
## ### Summary of continuous variables ###  
##   
## strata: Overall  
## n miss p.miss mean sd median p25 p75 min max skew kurt  
## status 79 0 0 0.3 5e-01 0 0.0 1.0 0.00 1 0.7 -1.6  
## length 79 0 0 329.4 3e+02 230 149.5 417.5 50.00 1520 2.2 6.0  
## mass 79 5 6 663.9 1e+03 166 23.5 577.2 11.50 9600 4.2 21.6  
## range 79 0 0 2.3 2e+00 2 0.6 3.4 0.01 9 1.2 1.2  
## migr 79 0 0 1.7 8e-01 1 1.0 2.0 1.00 3 0.6 -1.1  
## insect 79 2 3 7.3 5e+00 7 3.0 12.0 0.00 12 -0.3 -1.5  
## diet 79 0 0 1.7 6e-01 2 1.0 2.0 1.00 3 0.3 -0.6  
## clutch 79 0 0 6.0 3e+00 5 3.9 7.5 1.90 16 1.3 0.7  
## broods 79 5 6 1.8 1e+00 2 1.0 2.0 1.00 7 2.3 8.2  
## wood 79 0 0 0.6 5e-01 1 0.0 1.0 0.00 1 -0.2 -2.0  
## upland 79 0 0 0.1 3e-01 0 0.0 0.0 0.00 1 2.7 5.4  
## water 79 0 0 0.3 4e-01 0 0.0 0.5 0.00 1 1.2 -0.7  
## release 79 0 0 6.3 6e+00 4 2.0 10.0 1.00 27 1.3 1.3  
## indiv 79 3 4 185.2 3e+02 33 8.0 226.8 2.00 1539 2.5 7.1

## Warning: The `x` argument of `as\_tibble.matrix()` must have unique column names if `.name\_repair` is omitted as of tibble 2.0.0.  
## Using compatibility `.name\_repair`.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_warnings()` to see where this warning was generated.



##   
## 0 1   
## 37 23

##   
## Call:  
## glm(formula = status ~ 1, family = binomial(link = "logit"),   
## data = birds\_clean\_tbl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9833 -0.9833 -0.9833 1.3848 1.3848   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.4754 0.2655 -1.79 0.0734 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 79.881 on 59 degrees of freedom  
## Residual deviance: 79.881 on 59 degrees of freedom  
## AIC: 81.881  
##   
## Number of Fisher Scoring iterations: 4

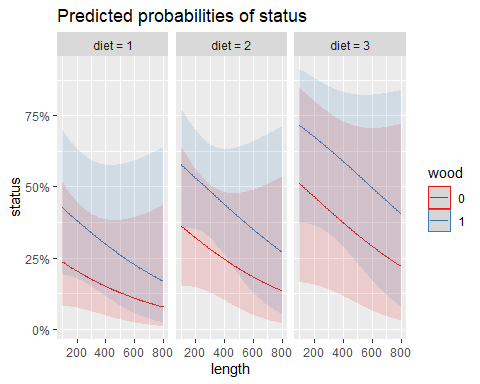
##   
## Call:  
## glm(formula = status ~ length + wood + diet, family = binomial(link = "logit"),   
## data = birds\_clean\_tbl)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3839 -0.9167 -0.7209 1.0939 1.8904   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.597596 0.997444 -1.602 0.109  
## length -0.001865 0.001747 -1.068 0.286  
## wood 0.876356 0.593305 1.477 0.140  
## diet 0.609586 0.467651 1.304 0.192  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 79.881 on 59 degrees of freedom  
## Residual deviance: 74.182 on 56 degrees of freedom  
## AIC: 82.182  
##   
## Number of Fisher Scoring iterations: 4

## # A tibble: 4 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.202 0.997 -1.60 0.109  
## 2 length 0.998 0.00175 -1.07 0.286  
## 3 wood 2.40 0.593 1.48 0.140  
## 4 diet 1.84 0.468 1.30 0.192

## Model Fit Statistics   
## ---------------------------------------------------------------------------------  
## Log-Lik Intercept Only: -39.940 Log-Lik Full Model: -37.091   
## Deviance(56): 74.182 LR(3): 5.699   
## Prob > LR: 0.127   
## MCFadden's R2 0.071 McFadden's Adj R2: -0.029   
## ML (Cox-Snell) R2: 0.091 Cragg-Uhler(Nagelkerke) R2: 0.123   
## McKelvey & Zavoina's R2: 0.121 Efron's R2: 0.099   
## Count R2: 0.717 Adj Count R2: 0.261   
## BIC: 90.559 AIC: 82.182   
## ---------------------------------------------------------------------------------

## Confusion Matrix and Statistics   
##   
## Reference  
## Prediction 0 1  
## 0 30 10  
## 1 7 13  
##   
##   
## Accuracy : 0.7167   
## No Information Rate : 0.6167   
##   
## Kappa : 0.3855   
##   
## McNemars's Test P-Value : 0.6276   
##   
## Sensitivity : 0.5652   
## Specificity : 0.8108   
## Pos Pred Value : 0.6500   
## Neg Pred Value : 0.7500   
## Prevalence : 0.3833   
## Detection Rate : 0.2167   
## Detection Prevalence : 0.3333   
## Balanced Accuracy : 0.6880   
## Precision : 0.6500   
## Recall : 0.5652   
##   
## 'Positive' Class : 1

## Data were 'prettified'. Consider using `terms="length [all]"` to get smooth plots.



## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties

## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties  
  
## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties  
  
## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties  
  
## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties  
  
## Warning in cor.test.default(as.numeric(x), as.numeric(y), method = method):  
## Cannot compute exact p-value with ties

