Домашнее задание по специфике медицинских данных

2023-11-12

#загрузка данных  
trauma <- read\_excel("trauma.xlsx")  
trauma <- trauma |>   
 select(-1)  
#head(trauma)  
#str(trauma)  
#summary(trauma)

# **Задание 1**

#Преобразуем необходимые переменные  
trauma <- trauma |>   
 mutate(Height = (as.numeric(str\_remove\_all(trauma$Height, "\"")) \* 2.54 / 100) |> round(2),  
 Weight = (Weight / 2.2) |> round(2),  
 Sex = ifelse(Sex == "Male", "Мужской", "Женский"),  
 Hb = if\_else(Hb == 0, NA, Hb))# заменим нулевые значения по уровню гемоглобина на NA

Выведем сводную статистику по переменным

trauma |>   
 select(!c(id, Name)) |>   
 tbl\_summary(  
 type = all\_continuous() ~ "continuous2",  
 statistic = all\_continuous() ~ c(  
 "{mean} ({sd})",  
 "{median} ({p25}, {p75})",  
 "{min}, {max}"  
 ),  
 missing = "no",  
 label = c("Sex" = "Пол",  
 "Age" = "Возраст, г",  
 "Height" = "Рост, м",  
 "Weight" = "Вес, кг",  
 "SBP" = "САД, мм.рт.ст.",  
 "DBP" = "ДАД, мм.рт.ст.",  
 "FOUR" = "Баллы по шкале комы FOUR",  
 "GSC" = "Баллы по шкале комы Глазго",  
 "Hb" = "Hb, г/дл",  
 "Death" = "Смерть"),  
 digits = c(all\_continuous() ~ c(1, 1),  
 all\_categorical() ~ c(0, 1))  
 ) |>   
 bold\_labels()

| **Characteristic** | **N = 1,024**1 |
| --- | --- |
| **Пол** |  |
| Женский | 384 (37.5%) |
| Мужской | 640 (62.5%) |
| **Возраст, г** |  |
| Mean (SD) | 40.9 (13.7) |
| Median (IQR) | 41.0 (29.0, 53.0) |
| Range | 18.0, 64.0 |
| **Рост, м** |  |
| Mean (SD) | 1.8 (0.1) |
| Median (IQR) | 1.9 (1.8, 1.9) |
| Range | 1.6, 2.0 |
| **Вес, кг** |  |
| Mean (SD) | 88.0 (8.3) |
| Median (IQR) | 88.1 (82.6, 93.4) |
| Range | 60.6, 115.1 |
| **САД, мм.рт.ст.** |  |
| Mean (SD) | 110.8 (7.5) |
| Median (IQR) | 110.0 (106.0, 116.0) |
| Range | 90.0, 134.0 |
| **ДАД, мм.рт.ст.** |  |
| Mean (SD) | 84.8 (8.4) |
| Median (IQR) | 84.0 (78.0, 90.0) |
| Range | 58.0, 110.0 |
| **Баллы по шкале комы FOUR** |  |
| Mean (SD) | 8.9 (3.0) |
| Median (IQR) | 9.0 (7.0, 11.0) |
| Range | 0.0, 16.0 |
| **Баллы по шкале комы Глазго** |  |
| Mean (SD) | 7.8 (2.7) |
| Median (IQR) | 8.0 (6.0, 10.0) |
| Range | 3.0, 14.0 |
| **Hb, г/дл** |  |
| Mean (SD) | 13.0 (1.3) |
| Median (IQR) | 13.1 (12.1, 14.0) |
| Range | 9.4, 16.2 |
| **Смерть** | 409 (39.9%) |
| 1n (%) | |

#Сниженный уровень гемоглобина  
  
trauma <- trauma |>   
 mutate( Hb\_code = (case\_when(  
 Hb < 12 & Sex == "Женский" ~ 1, # сниженный уровень для женщин  
 Hb < 13.5 & Sex == "Мужской" ~ 1, # сниженный уровень для мужчин  
 Hb >= 12 & Hb <= 14 & Sex == "Женский" ~ 2, # нормальный уровень для женщин  
 Hb >= 13.5 & Hb <= 16 & Sex == "Мужской" ~ 2, # нормальный уровень для мужчин  
 Hb > 14 & Sex == "Женский" ~ 3, # повышенный уровень для женщин  
 Hb > 16 & Sex == "Мужской" ~ 3 # повышенный уровень для мужчин  
 )) |> as.factor())  
  
trauma |>   
 filter(Hb\_code == 1) |>   
 summarise('Количество' = n()) |>   
 mutate('% случаев' = (Количество / nrow(trauma) \*100) |> round(1)) |>  
 flextable() |>   
 set\_table\_properties(width = 0.5, layout = "autofit") |>   
 add\_header\_row(values = 'Сниженный уровень гемоглобина в выборке', colwidths = 2)

| Сниженный уровень гемоглобина в выборке | |
| --- | --- |
| Количество | % случаев |
| 458 | 44.7 |

Сниженный уровень гемоглобина выявлен у 458 пациентов, что составило 44,7% от общего количества пациентов (n = 1024)

# **Задание 2**

# Создадим переменную ИМТ  
trauma <- trauma |>   
 mutate(BMI = Weight/(Height^2)|> round(1))

trauma |>   
 filter(BMI > 30) |>   
 summarise('Количество' = n()) |>   
 mutate('% случаев' = (Количество / nrow(trauma) \*100) |> round(1)) |>   
 flextable() |>   
 set\_table\_properties(width = 0.5, layout = "autofit") |>   
 add\_header\_row(values = 'Индекс массы тела > 30 кг/м2 в выборке', colwidths = 2)

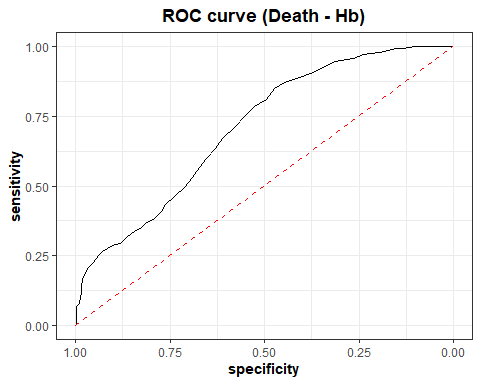
| Индекс массы тела > 30 кг/м2 в выборке | |
| --- | --- |
| Количество | % случаев |
| 74 | 7.2 |

Средний уровень ИМТ M (SD) у пациентов, включенных в исследование 26.1(2.6) кг/м2. Ожирение выявлено в 7.2% случаев от общего количества пациентов (n = 1024)

# **Задание 3**

Построим ROC-кривую для предсказания летального исхода в течение 24 часов по уровню гемоглобина.

roc\_hb <- roc(Death ~ Hb, data = trauma, ci = TRUE)  
  
roc\_hb |>   
 ggroc() +   
 geom\_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="red", linetype="dashed") +   
 theme\_bw() +  
 labs(title = "ROC curve (Death - Hb)") +  
 theme(axis.title = element\_text(face = "bold"),  
 plot.title = element\_text(face = "bold", hjust = 0.5))



Мы видим, что форма ROC-кривой приближается к диагональной линии, то есть уровень гемоглобина не является сильным предиктором летального исхода.

# **Задание 4**

Площадь под построенной ROC-кривой

roc\_hb$auc

## Area under the curve: 0.7078

95% двусторонний ДИ для площади под ROC-кривой

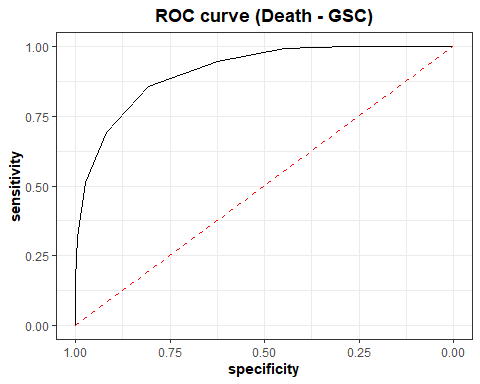
roc\_hb$ci

## 95% CI: 0.6763-0.7392 (DeLong)

# **Задание 5**

ROC-анализ для предсказания летального исхода в течение 24 часов по шкале комы Глазго.

roc\_gsc <- roc(Death ~ GSC, data = trauma, ci = TRUE)  
  
roc\_gsc |>   
 ggroc() +   
 geom\_segment(aes(x = 1, xend = 0, y = 0, yend = 1), color="red", linetype="dashed") +   
 theme\_bw() +  
 labs(title = "ROC curve (Death - GSC)") +  
 theme(axis.title = element\_text(face = "bold"),  
 plot.title = element\_text(face = "bold", hjust = 0.5))



Ниже в таблице представлены оптимальное пороговое значение для предсказания летального исхода по шкале комы Глазго и соответствующие значения чувствительности и специфичности

roc\_gsc\_best <- roc\_gsc |> coords(x = "best", best.method = "closest.topleft")   
  
roc\_gsc\_best |>   
 mutate(specificity = round(specificity, 2),  
 sensitivity = round(sensitivity, 2)) |>   
 rename(`Пороговое значение` = threshold,  
 Чувствительность = sensitivity,  
 Специфичность = specificity) |>   
 flextable() |>   
 set\_table\_properties(width = 0.5, layout = "autofit")

| Пороговое значение | Специфичность | Чувствительность |
| --- | --- | --- |
| 7.5 | 0.81 | 0.86 |

Для предсказания летального исхода в течение 24 часов оптимальным является пороговое значение 7,5 баллов по шкале комы Глазго с чувствительностью 86% и специфичностью 81%.

# **Задание 6**

ROC-анализ для всех количественных переменных в датасете

trauma |>   
 select(Death, where(is.numeric) & !id) |>   
 pivot\_longer(cols = !Death) |>   
 rename("Переменная" = name) |>   
 group\_by(Переменная) |>   
 summarise(AUC = roc(Death, value, ci = TRUE)$ci[2] |> round(3),  
 "Нижняя граница 95% ДИ" = roc(Death, value, ci = T)$ci[1] %>% round(3),  
 "Верхняя граница 95% ДИ" = roc(Death, value, ci = T)$ci[3] %>% round(3)) |>   
 arrange(desc(AUC)) |>   
 flextable() |>   
 set\_table\_properties(width = 0.5, layout = "autofit")

| Переменная | AUC | Нижняя граница 95% ДИ | Верхняя граница 95% ДИ |
| --- | --- | --- | --- |
| FOUR | 0.934 | 0.920 | 0.948 |
| GSC | 0.912 | 0.896 | 0.929 |
| SBP | 0.786 | 0.759 | 0.814 |
| DBP | 0.744 | 0.714 | 0.774 |
| Hb | 0.708 | 0.676 | 0.739 |
| Age | 0.527 | 0.491 | 0.563 |
| Weight | 0.518 | 0.482 | 0.554 |
| BMI | 0.501 | 0.464 | 0.537 |
| Height | 0.481 | 0.445 | 0.517 |

Наибольшей площадью под ROC-кривой обладает оценка по шкале комы FOUR. Это означает, что такая шкала является наиболее важным предиктором для определения вероятности смерти пациента. Это обусловлено тем, что она и была разработана для анализа тяжести церебрального повреждения и прогноза исхода.

Минимальную площадь под ROC-кривой имеет рост пациента. Такой показатель является неспецифическим для прогноза летальности.