**Odds & ORs, Part 2:**

**when the explanatory variable is NUMERICAL**

**Instructions:**

**Step 1:** Work alone or with the person next to you to answer all the parts of Question #1.

**Step 2:** Come up to the front of the room and check your answers with me.

…Then continue with Question #2. After completing Question #2, come up and check your answers with me.

1. **Med School Acceptance: MCAT scores**   
   Open the **MedGPA** data set (from the Stat2Data package) in R and use ?MedGPA to learn about the data.

The variable “MCAT” contains the total MCAT scores for these 55 students. There is a lot of variability in MCAT scores, and there are very few students who have the same value. Instead of looking at the odds of success (acceptance) for each individual score value, we’ll group them, following the procedure described on p. 472- 474 (Example 9.12).

**Note**: We are doing this for two reasons:

1. to practice *interpreting* ORs of different numerical “groups”
2. to *visualize* the proportion of success vs. MCAT and the log(odds) vs MCAT…which we need to do to check the linearity condition

Importantly, we do NOT have to do this when actually *fitting* a logistic regression model, because R knows how to deal with numerical predictors. However, it is important that you know how to do this since it’s how we check the linearity condition.

* 1. First, we divide the range of MCAT scores into intervals with roughly equal numbers of cases. I have made (what I think is) a judicious division below into 5 intervals (however, there is no “right” number of intervals in these cases!).

Fill in the rest of the table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group # | # Cases | Range of MCAT | Midpoint of range | Admitted | | Proportion admitted | Odds of admission | Odds Ratios |
| Yes | No |
| 1 |  | 18 – 32 |  |  |  |  |  |  |
|  |
| 2 |  | 33 – 34 |  |  |  |  |  |
|  |
| 3 |  | 35 – 36 |  |  |  |  |  |
|  |
| 4 |  | 37 – 39 |  |  |  |  |  |
|  |
| 5 |  | 40 – 48 |  |  |  |  |  |
|  |

* 1. Interpret the odds of admission for Group 2.
  2. Interpret (in context!) the odds ratio for an increase in MCAT score from Group 3 to Group 4.
  3. Create three vectors from the table above: “midpoints”, that contains the range midpoints for each group; “props”, that contains the proportion of admission for each group; and “odds”, that contains the odds from each group. You can create these vectors easily using (for example)

midpoints <- c(25, 33.5, …)

props <- c(2/11, 4/9, …)  
odds <- c(2/9, 4/5, …)

* 1. Make three scatterplots:

(1) proportion of success vs. MCAT score (“midpoints”);

(2) odds vs. MCAT score;

(3) log(odds) vs. MCAT score.

Look carefully at the difference between these three plots, and make sure you understand what each is plotting. What form do we expect in plot (1)? What form do we expect in plot (3)?

1. **ICU Survival: Age**Open the **ICU** data set (from the Stat2Data package) in R and use ?ICU to learn about the data.  
   1. Make a boxplot of survival status by age. Does there appear to be a relationship between these two variables?
   2. I have made (what I think is) a judicious division of Age into intervals (however, there is no “right” number of intervals in these cases!).

Fill in the rest of the table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group # | # Cases | Range of Age | Midpoint of range | Survived | | Proportion survived | Odds of survival | Odds Ratios |
| Yes | No |
| 1 |  | 15 – 29 |  |  |  |  |  |  |
|  |
| 2 |  | 30 – 49 |  |  |  |  |  |
|  |
| 3 |  | 50 – 59 |  |  |  |  |  |
|  |
| 4 |  | 60 – 64 |  |  |  |  |  |
|  |
| 5 |  | 65 – 69 |  |  |  |  |  |
|  |
| 6 |  | 70 – 74 |  |  |  |  |  |
|  |
| 7 |  | 75 – 79 |  |  |  |  |  |
|  |
| 8 |  | 80 – 92 |  |  |  |  |  |
|  |

**Remember**: We will not have to create a table like this when actually fitting a model, but we *will* need to create one to check our linearity condition.

* 1. Interpret the odds of survival for Group 2.
  2. Interpret (in context!) the odds ratio for Group 2 compared to Group 3.
  3. Interpret (in context!) the odds ratio for Group 6 compared to Group 2.
  4. Create three vectors from the table above: “midpoints”, that contains the range midpoints for each group; “props”, that contains the proportion of survival for each group; and “odds”, that contains the odds from each group.   
     Then make three scatterplots:

(1) proportion of survival vs. age (“midpoints”);

(2) odds vs. age;

(3) log(odds) vs. age.

Look carefully at the difference between these three plots, and make sure you understand what each is plotting. What form do we expect in plot (1)? What form do we expect in plot (3)?