**Recap:** Generalized Linear Models (GLMs) model a random component as a function of systematic components through a link function.

- Multiple regression is a type of GLM with the identity link and normal distribution.
- Logistic regression is a type of GLM with the logit link function and binomial random component.
  - \( Y \) is Bernoulli random variable so \( \pi(x) = P(Y = 1) = E(Y) = \mu \).
  - \( g(\mu) = \log(\mu/(1-\mu)) \) is the “logit” link.
  - \( \log \text{ odds} = \ln(\pi(x)/(1-\pi(x))) = \alpha + \beta x \).
  - \( \pi(x) = \exp(\alpha + \beta x)/(1+ \exp(\alpha + \beta x)) \).

When you give JMP Fit Y by X with a categorical response variable and a quantitative X, Factor, JMP fits a logistic curve.

![Graph showing log odds of died/survived versus age](image)

At each age, the dot is randomly placed at a height above or below the blue curve depending on whether the observation was a success or failure.

JMP trick: Right click on a column and choose Label. Then when you click on a point on the graph, the label will display.

(a) Based on this graph, can you identify any unusual observations (“outliers”)? Can you determine whether the model appears to fit the data?

(b) Interpret the estimated intercept coefficient.

(c) What is the odds ratio of survival comparing a 25 year old to a 45 year old? Comparing a 15 year old to a 65 year old?
(d) Verify your calculations by using the Logistic pull-down menu and selecting **Odds Ratios**. JMP reports a “unit odds ratio” and an “odds ratio” for the range of the data. Why might the latter be useful?

(e) What is the predicted probability of survival for a 20-year-old?

(f) Verify your calculation by selecting **Save Probability Formula** and finding a 20 year old in the data set. Then enter an age of 21 in row 46 of the age column and press Return. A value should appear in the Prob(died) column.

(g) At what age do we estimate the probability of death to be .5? How does this relate to the parameter values in general?

You can also find the *incremental rate of change in the fitted probability at a given x*: \( \hat{\beta} \hat{\pi} (1-\hat{\pi}) \)

So we would say that at \( x \), the probability of “success” increases at a rate of \( \hat{\beta} \hat{\pi} (1-\hat{\pi}) \) per 1 unit increase in \( x \). You can construct a confidence interval for this by substituting the endpoints of the confidence interval for \( \beta \).

(h) Report this value for the \( x \) values used in (f), comparing a 20-year-old to a 21-year-old. How does this compare to the incremental change for a 59-year-old? Why does it differ?

**Practice Problem – to submit in PolyLearn**

(i) How will these incremental rates of change compare for males and females at \( x = 40 \)? How does the median \( x \) value differ between males and females?

<table>
<thead>
<tr>
<th>Females:</th>
<th>Males:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
<td><strong>Estimate</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.2463844</td>
</tr>
<tr>
<td>age</td>
<td>0.19407412</td>
</tr>
</tbody>
</table>