Probability, log-odds, and odds

WILD 502- Jay Rotella

To better understand the connections between the log-odds of an outcome, the odds of an outcome, and the probability of an outcome, it is helpful to work with a range of values on one scale and convert it to the others. It's also helpful to visualize the relationships with some plots. Recall that if the probability of an event is 0.2, that

1. the odds of the event occurring are

$$odds = \frac{0.2}{0.8} = 0.25$$

2. the log-odds of the event occurring are

$$ln\left(\frac{0.2}{0.8}\right) = -1.3863$$

or

$$ln(0.25) = -1.3863$$

3. the probability can be reconstructed as

$$\frac{odds}{1 + odds} = \frac{0.25}{1.25} = 0.2$$

4. the probability can also be reconstructed as

$$\frac{exp(ln(odds))}{1 + exp(ln(odds))} = \frac{exp(-1.3683)}{1 + exp(-1.3683)} = \frac{0.25}{1.25} = 0.2$$

In R, you can

1. obtain the odds for a given probability by dividing the probability by 1 minus the probability, e.g., `odds = 0.2/(1-0.2) = 0.25`

2. obtain the log-odds for a given probability by taking the natural logarithm of the odds, e.g., `log(0.25)` = -1.3862944 or using the `qlogis` function on the probability value, e.g., `qlogis(0.2)` = -1.3862944.

3. obtain the probability from the log-odds using $\frac{exp(x)}{1+exp(x)}$, where x represents the log-odds value either by writing the expression out, e.g., `exp(-1.3862944)/(1 + exp(-1.3862944))`, or by using the `plogis` function, e.g., `plogis(-1.3862944)` = 0.2.

4. obtain the probability from the odds by using $\frac{odds}{1+odds}$, e.g., `0.25/1.25` = 0.2.

Probability values range from 0 to 1. It turns out that for $\frac{exp(x)}{1+exp(x)}$, values of x ranging from -5 to +5 create probabilities that range from just above 0 to very close to 1. Values of x ranging from -1 to +1 create probabilities that range from about 0.25 to 0.75. The material below will let you explore the relationships for yourself.

```
library(ggplot2)
log_odds = seq(from = -5, to = 5, by = 0.25)
odds = exp(log_odds)
# use 'plogis' function to calculate exp(x)/(1 + exp(x))
p = plogis(log_odds)
# use odds/(1+odds) to calculate p a different way
p2 = odds/(1 + odds)
# store probability of failure (1-p)
q = 1 - p
# store log_odds and y in data frame for use with ggplot
```

```
d = data.frame(log_odds, odds, p, p2, q)
head(d, 4)
```

```
##   log_odds        odds           p          p2         q
## 1    -5.00 0.006737947 0.006692851 0.006692851 0.9933071
## 2    -4.75 0.008651695 0.008577485 0.008577485 0.9914225
## 3    -4.50 0.011108997 0.010986943 0.010986943 0.9890131
## 4    -4.25 0.014264234 0.014063627 0.014063627 0.9859364
```

```
d[19:23, ]
```

```
##    log_odds      odds         p        p2         q
## 19    -0.50 0.6065307 0.3775407 0.3775407 0.6224593
## 20    -0.25 0.7788008 0.4378235 0.4378235 0.5621765
## 21     0.00 1.0000000 0.5000000 0.5000000 0.5000000
## 22     0.25 1.2840254 0.5621765 0.5621765 0.4378235
## 23     0.50 1.6487213 0.6224593 0.6224593 0.3775407
```

```
tail(d, 4)
```

```
##    log_odds       odds         p        p2          q
## 38     4.25   70.10541 0.9859364 0.9859364 0.014063627
## 39     4.50   90.01713 0.9890131 0.9890131 0.010986943
## 40     4.75  115.58428 0.9914225 0.9914225 0.008577485
## 41     5.00  148.41316 0.9933071 0.9933071 0.006692851
```

Below, we plot the relationship, so you can see the pattern among the values for log-odds and associated probabilities. You might wonder what happens if you get log-odds values that are very very small (e.g., -24, -147, or -2421) or very big (e.g.,14, 250, or 1250). You should use the `plogis` function on such values (no commas in your numbers, e.g., `plogis(-2421)`) to find out for yourself.

```
ggplot(d, aes(x = log_odds, y = odds)) +
  geom_line() +
  scale_x_continuous(breaks = seq(-5, 5, by = 1)) +
  labs(title = "odds versus log-odds")
```
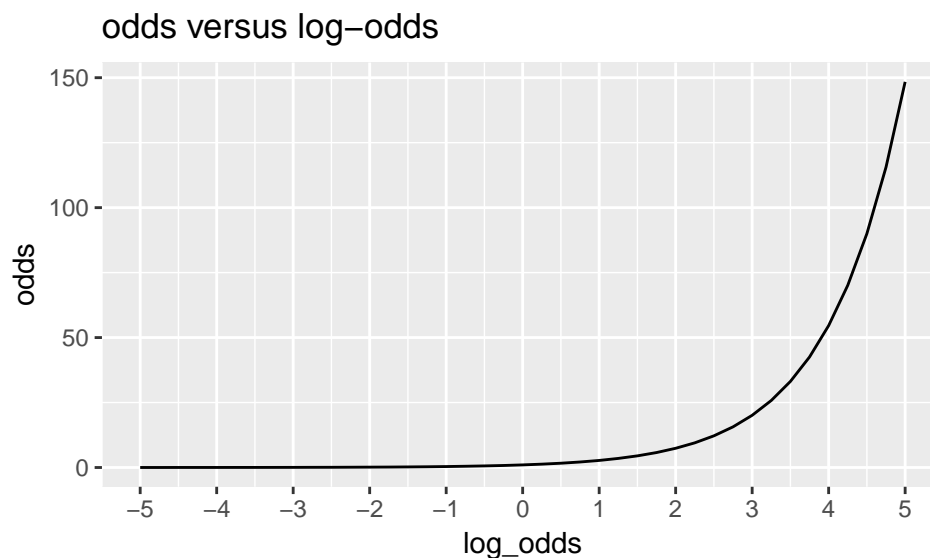


```
ggplot(d, aes(x = odds, y = p)) +
  geom_line() +
```

```
labs(title = "probability versus odds")
```

probability versus odds
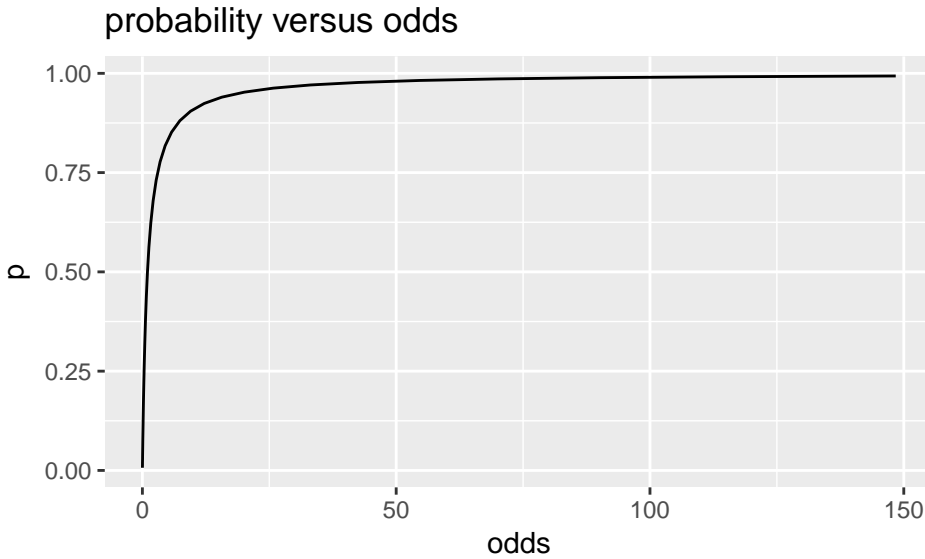
Finally, this is the plot that I think you'll find most useful because in logistic regression your regression equation, e.g., $\hat{\beta}_0 + \hat{\beta}_0 \cdot x_1$ yields the log-odds, and you're interested in how that relates to the probability of survival (or later in the course, the probability of detection or some other probability of interest).

```
ggplot(d, aes(x = log_odds, y = p)) +
  geom_line() +
  geom_hline(aes(yintercept = 0.5),
             colour = "gray",
             linetype = "dashed") +
             geom_vline(aes(xintercept = 0.0),
             colour = "gray",
             linetype = "dashed") +
  scale_x_continuous(breaks = seq(-5, 5, by = 1)) +
  labs(title = "probability versus log-odds")
```

probability versus log−odds