WS #11 - Forward selection

Math 150, Jo Hardin

Monday, March 10, 2025

Your Name: _____

Names of people you worked with: _____

Which is better: spring forward or fall back?

Task: Consider the final forward model from the bird nest data.

```
glm(`Closed?` ~ Length + Location + No.eggs + Nestling,
    data = nests, family = "binomial") |> tidy()
```

```
# A tibble: 10 x 5
   term           estimate std.error statistic p.value
   <chr>             <dbl>     <dbl>     <dbl>   <dbl>
 1 (Intercept)      11.1   3328.      0.00334  0.997
 2 Length           -0.219    0.0754 -2.91     0.00364
 3 Locationconif   -19.3   3328.     -0.00580  0.995
 4 Locationdecid   -16.9   3328.     -0.00507  0.996
 5 Locationground  -20.5   3328.     -0.00617  0.995
 6 Locationshrub   -18.6   3328.     -0.00560  0.996
 7 Locationsnag      0.695 4313.      0.000161 1.00
 8 Locationwall    -18.3   3328.     -0.00550  0.996
 9 No.eggs           0.795    0.262   3.04     0.00238
10 Nestling          0.398    0.144   2.76     0.00577
```

```
nests |> select(Location) |> table()
```

```
Location
  bank  conif  decid ground  shrub   snag   wall
     3     14     25     19     17      4      4
```

a. Find the odds ratio to compare the odds of a nest being `Closed` for birds on the ground with 3 eggs versus birds in coniferous trees with 6 eggs, holding `Length` and `Nestling` constant.
b. Using a single odds ratio, interpret the coefficient on `Locationsnag`, $b_6 = 0.69$.

Solution:

To find the ORs, it is often a good idea to calculate the two different odds values first!

a. The first odds ratio is found by dividing the two separate odds values.

$$\widehat{\text{odds}}_1 = e^{11.1-0.22\cdot\text{Length}-20.52\cdot 1+0.79\cdot 3+0.40\cdot\text{Nestling}}$$

$$\widehat{\text{odds}}_2 = e^{11.1-0.22\cdot\text{Length}-19.29\cdot 1+0.79\cdot 6+0.40\cdot\text{Nestling}}$$

$$\widehat{\text{OR}} = e^{-20.52+19.29+0.79\cdot(3-6)} = 0.027$$

For birds on the ground with 3 eggs, the odds of having a closed nest are 0.027 times the odds of having a closed nest for birds in coniferous trees with 6 eggs, holding `Length` and `Nestling` constant.

b. The second OR is given by comparing the level of interest to the baseline value

$$\widehat{\text{OR}} = e^{0.69} = 1.99$$

For birds on a snag, the odds of having a closed nest are 1.99 times the odds of having a closed nest **for birds on a bank**, holding all the other variables constant. Notice that the baseline level of `Location` is bank, there is no dummy variable for bank!