Confidence Intervals on Transformed Quantities

WILD 502 - Jay Rotella

Expressing Uncertainty on Estimated Probabilities

After you look at $\hat{\beta}'s$ and their associated $\widehat{SE}(\hat{\beta})'s$ results of a logistic regression , you are typically also interested in generating estimates of survival. To do so, you have learned to back-transform the regression from the logit scale to the probability scale using $\frac{exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 ...)}{1 + exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 ...)}$. You also need to have a way of obtaining estimates of the SE's for $\hat{S}$. There are a variety of ways of doing so, but for now, we'll discuss the *delta method*, which is dicussed in detail in Appendix B of the Cooch and White on-line book. The key idea is that you work with (1) the $\hat{\beta}'s$ and the variance-covariance matrix for the $\hat{\beta}'s$, (2) the transformation being used to convert the $\hat{\beta}'s$ to $\hat{S}$, and (3) a set of partial derivatives that indicate how much $\hat{S}$ changes as each of the $\hat{\beta}'s$ changes, i.e., (how sensitive is the outcome $\hat{S}$ to the uncertainties about the $\hat{\beta}'s$).

A Simple Example

For the *S(constant)* model on the fawns data, you estimated $\hat{\beta}_0$ as -0.3545 and $\widehat{SE}(\hat{\beta}_0)$ as 0.1903. The transformation is $\frac{exp(\hat{\beta}_0)}{1 + exp(\hat{\beta}_0)}$. You can read about the derivatives in Appendix B of the Cooch and White book, which is beyond the level of detail that we'll get into at this point in the course. Fortunately, we can use the *deltavar* function of the *emdbook* package to do the calculations quite easily as it calculates the derivatives and does the required matrix math for us.

```
library(knitr)   # for printing nice tables
library(emdbook) # for deltavar function
library(ggplot2) # for plotting
b0 <- -0.354545
Beta.hats <- c(b0 = b0)
se_b0 <- 0.1902681

# Calculate ln odds of Survival rate
ln_odds_S <- b0

# Estimate the 95% confidence limits for ln odds of S
se_ln_odds_S <- se_b0 # simple for this problem with only 1 beta_hat
lcl_ln_odds_S <- ln_odds_S - 1.96 * se_ln_odds_S
ucl_ln_odds_S <- ln_odds_S + 1.96 * se_ln_odds_S

# backtransform log-odds to obtain estimate of S
S <- plogis(ln_odds_S)

# Estimate SE for S with delta method
# create var-cov matrix (this one's simply a 1 x 1 matrix)
var_Beta = matrix(se_b0^2, nrow = 1, ncol = 1)
se_S <- sqrt(deltavar(exp(b0)/(1 + exp(b0)),
                meanval = Beta.hats,
                Sigma = var_Beta))
cbind(S, se_S)

##              S        se_S
## [1,] 0.4122807 0.04610297
```

You can obtain 95% confidence limits for $\hat{S}$ as shown below. Notice that here you don't use the $\widehat{SE}(\hat{S})$ to obtain the confidence limits. Rather, you work with confidence limits on the logit scale and back-transform those values, which creates confidence limits for $\hat{S}$ with better properties, e.g., they'll stay within the 0 to 1 bounds.

```r
# obtain confidence limits for S_hat by back-transforming
# from the log-odds and confidence limits for log-odds
lcl_S <- plogis(lcl_ln_odds_S)
ucl_S <- plogis(ucl_ln_odds_S)
cbind(S, se_S, lcl_S, ucl_S)
```

```
##              S       se_S     lcl_S     ucl_S
## [1,] 0.4122807 0.04610297 0.3257501 0.504595
```

A Slightly More Complex Example

You also estimated the parameters of the *S(length)* model that contains an intercept, $\hat{\beta}_0$, and a slope, $\hat{\beta}_1$, which is multiplied by the length covariate. For this model, you need to first use the *delta method* to obtain estimates of the log-odds and associated standard errors for the log-odds of survival for animals of different lengths. Once you have those, you can obtain the confidence bounds on the estimated log-odds.

```r
# Store beta_hats
b0 <- -10.23404 # intercept
b1 <-   0.07999 # beta for length

# Store variance-covariance matrix for beta_hats
sigma <- matrix(c(
23.02372737, -0.185786837,
-0.185786837,  0.001501638), nrow = 2, ncol = 2)

# Provide length values over range of data
Length <- seq(from = 108, to = 135.5, by = 0.5)

# Calculate ln odds of Survival rate for any length
ln_odds_S <- b0 + b1 * Length

# Estimate the Standard Errors for ln odds of S
se_ln_odds_S <- sqrt(deltavar(b0 + b1 * Length,
                     meanval = c(b0 = -10.635836, b1 = 0.0831520),
                     Sigma = sigma))

# Estimate the 95% confidence limits for ln odds of S
lcl_ln_odds_S <- ln_odds_S - 1.96 * se_ln_odds_S
ucl_ln_odds_S <- ln_odds_S + 1.96 * se_ln_odds_S

# Store all as dataframe
log_odds <- data.frame(Length, ln_odds_S,
                     se_ln_odds_S,
                     lcl_ln_odds_S,
                     ucl_ln_odds_S)
# take a look
kable(head(log_odds, 4), digits = 4)
```

Length	ln_odds_S	se_ln_odds_S	lcl_ln_odds_S	ucl_ln_odds_S
108.0	-1.5951	0.6394	-2.8484	-0.3418
108.5	-1.5551	0.6210	-2.7723	-0.3380
109.0	-1.5151	0.6026	-2.6963	-0.3340
109.5	-1.4751	0.5843	-2.6204	-0.3299

```
# Plot estimated relationship between the log-odds and body length.
ggplot(log_odds, aes(x = Length, y = ln_odds_S)) +
  geom_line(size = 1.5) +
  geom_ribbon(aes(ymin = lcl_ln_odds_S, ymax = ucl_ln_odds_S), alpha = 0.2) +
    xlab("Body Length (cm)") +
    ylab("Estimated Log-Odds of Survival")
```
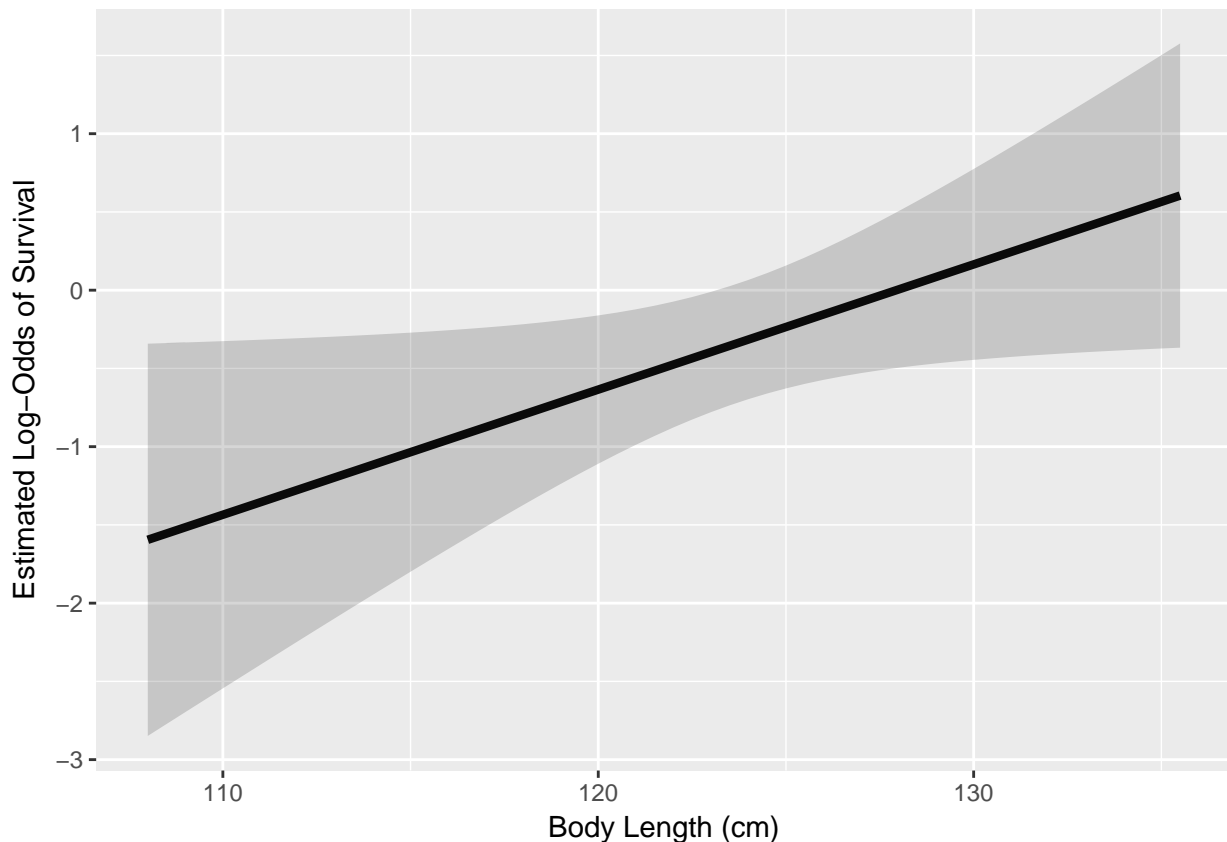

Finally, to obtain $\hat{S}$, $\widehat{SE}(\hat{S})$, and confidence limits for $\hat{S}$, you do the following work. Notice that as in the simpler example above that you don't use the $\widehat{SE}(\hat{S})$ to obtain the confidence limits. Rather, you work with confidence limits on the logit scale and back-transform those values, which creates confidence limits for $\hat{S}$ with better properties, e.g., they'll stay within the 0 to 1 bounds.

```
# back transform the log-odds to obtain estimates of S
S <- plogis(ln_odds_S)
lcl_S <- plogis(lcl_ln_odds_S)
ucl_S <- plogis(ucl_ln_odds_S)

# Estimate the Standard Errors for S
se_S <- sqrt(deltavar(
  exp(b0 + b1 * Length) / (1 + exp(b0 + b1 * Length)),
```

```
  meanval = c(b0 = b0, b1 = b1),
  Sigma = sigma))

# Store all as a data.frame
surv_ests <- data.frame(Length, S, se_S, lcl_S, ucl_S)
# look at a few rows
kable(head(surv_ests, 4), digits = 4)
```

Length	S	se_S	lcl_S	ucl_S
108.0	0.1687	0.0897	0.0548	0.4154
108.5	0.1743	0.0894	0.0588	0.4163
109.0	0.1802	0.0890	0.0632	0.4173
109.5	0.1862	0.0885	0.0678	0.4183

```
# Plot estimated relationship between survival rate and body length.
ggplot(surv_ests, aes(x = Length, y = S)) +
  geom_line(size = 1.5) +
  geom_ribbon(aes(ymin = lcl_S, ymax = ucl_S), alpha = 0.2) +
    xlab("Body Length (cm)") +
    ylab("Estimated Survival Rate")
```