

WILD 502

Lab 1 – Estimating Survival when Animal Fates are Known

Today’s lab will give you hands-on experience with estimating survival rates using logistic regression to estimate the parameters in a variety of competing models. The estimation will be done in Program MARK. Thus, you’ll gain experience with quite a variety of tasks today: inputting data into MARK, working with MARK, using logistic regression, working with transformed parameter estimates, evaluating competing models, and considering the key outputs that come from such an analysis. Given that the lab entails so many topics, many of which are likely new to you, you may need to spend a fair bit of time thinking about each aspect to get the most out of the lab.

1. The dataset

Radio telemetry was used to study survival of black ducks for 8 weeks during 1 winter. The dataset consists of the number of radiomarked birds that were studied each week and the number of birds that died and survived each week.

Week	# Females	# Died	# Survived
1	48	1	47
2	47	2	45
3	41	2	39
4	39	5	34
5	32	4	28
6	28	3	25
7	25	1	24
8	24	1	23
TOTAL			

- Using the data provided above, calculate the sampling variance and standard error for \hat{S} for each week’s data.
- Using the data provided, calculate the totals for the columns above along with the sampling variance and standard error for \hat{S} for data pooled over the entire study.

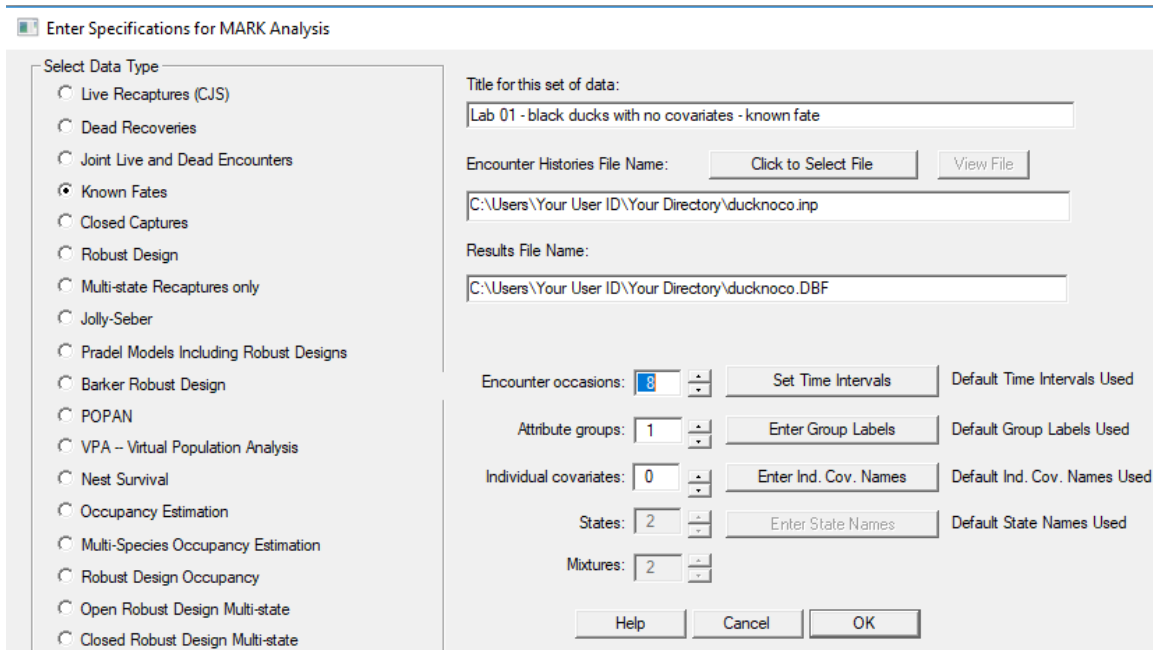
Note: You can accomplish the tasks in A & B in many ways (pencil, paper, and calculator; EXCEL; R; etc.). Working with a script in R is certainly a useful way to do it. If you’d like to use R, you could accomplish the tasks in A & B in many different ways. I find that the *dplyr* package is worth learning for data manipulation. You can learn about *dplyr* on the web, datacamp courses, and many recent R books. In case you’d like to try it out, the course schedule for today’s lab has a link to some starting code (*lab01-R-code.R*) that you can save, open in R, and adapt to work out answers for A & B above.

2. Inputting the data to Program MARK

First, for ease of communication with one another, we’ll label the beta’s we’ll be estimating this semester starting with “Beta0”. To do this, start Program MARK, choose

“File”, “Preferences” and click the 6th box down - now the 1st column of the design matrix will be labeled Beta0 instead of Beta1.

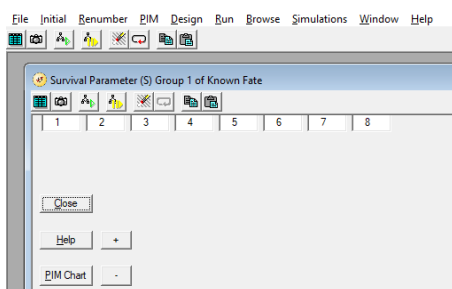
MARK can handle a wide variety of data types (something we’ll explore and take advantage of during the rest of the semester). Each data type has very specific data-formatting requirements. For this first exercise, we’ll be able to use a very simple format. In MARK, choose “File”, “New”, and fill in the window that comes up as seen below but using the correct directory information for your computer.



Once, you’ve done this, click on the “View File” button and examine the input file. Notice that the file has a comment field and that the data lines are separated by semicolons, which are important. Okay, the input file for this problem is simple enough.

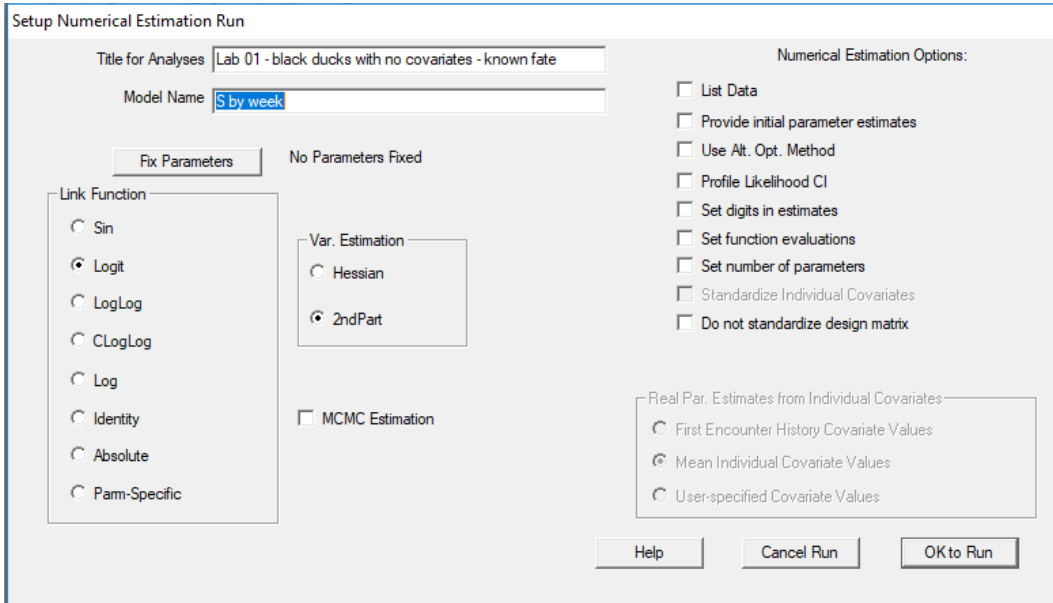
3. Build and estimate the parameters for one model of survival

- A. Close the Notepad window to go back to MARK and click on the “OK” button. This causes MARK to create a file and then put up the following window.



This window is the Parameter Index Matrix (PIM) for the survival parameters for group 1 (the only group in this analysis). MARK, by default, brought up a PIM that will estimate a different survival rate for each interval (week) in the study. The numbers in the PIM indicate how many parameters to estimate and that parameter 1 is \hat{S} for week 1, parameter 2 is \hat{S} for week 2, etc.

- B. Click on “Run,” “Current Model” to spawn the following window and to analyze the data using the “S by week” model dictated by the PIM. Fill in a name for the model. Your window should look like the following window. When you hit “OK to Run”, MARK asks if you’d like to use an identity matrix: click “Yes” and add the model to the Results Browser.



This created an 8 x 8 identity matrix that’s used as the model’s design matrix. Because the design matrix has 8 columns, the model estimates 8 $\hat{\beta}$ terms. Because it’s an identity matrix and because you chose the logit link function, each $\hat{\beta}$ will be multiplied by a 1 and provide an estimate of the log-odds of survival in the week corresponding to the row, e.g., the 3rd $\hat{\beta}$, (which is B2 because we started numbering with B0) represents the log-odds of survival in week 3.

B0	B1	B2	B3	Parm	B4	B5	B6	B7
1	0	0	0	1:S	0	0	0	0
0	1	0	0	2:S	0	0	0	0
0	0	1	0	3:S	0	0	0	0
0	0	0	1	4:S	0	0	0	0
0	0	0	0	5:S	1	0	0	0
0	0	0	0	6:S	0	1	0	0
0	0	0	0	7:S	0	0	1	0
0	0	0	0	8:S	0	0	0	1

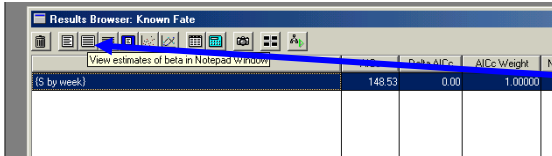
Note: The values in a row are the X-values that are used along with the estimated beta’s to estimate the Parameter modeled on that row. Also, here there’s no overall intercept term in the model. With this design matrix, each week’s survival rate is estimated using only the appropriate beta for that week.

Think about the logistic equation used in this model. Write it out as you think it looks. Be sure that you know that the equation reads (p_i is the probability of surviving a week):

$$L(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8 | x) = \frac{\exp(\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \beta_7 \cdot x_7)}{1 + \exp(\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \beta_7 \cdot x_7)}$$

Once you have the $\hat{\beta}$'s, how would you calculate the estimated survival rate for week i ? For week 4, you should know that you’d use $\exp(\beta_3 \cdot 1) / (1 + \exp(\beta_3 \cdot 1))$. Why? Because in the equation for week 4 (row 4 of the design matrix) there is a 1 in the cell corresponding to β_3 and 0’s corresponding to all the other betas such that only $\hat{\beta}_3$ is used to estimate S_4 .

- C. Let’s see if this works. Click the “OK to Run” button.



Click on the icon that yields “estimates of betas” – it’s the 2nd one to the right of the trash can.

This yields the Notepad window below, where values in the Beta column are the estimates for the specified logistic regression model. To get them into R, go to R and type or paste (but don’t yet execute) the following command:

```
beta <- read.table("clipboard", header = F,
  col.names = c("wk", "beta", "se", "lcl", "ucl"))
```

Next, highlight the output as shown below and click on Edit, Copy to copy the data to the Windows clipboard for input into R. Finally, go to R, and execute the R code above. You should now have the estimates in an R object named ‘beta’ that you can work with in R.

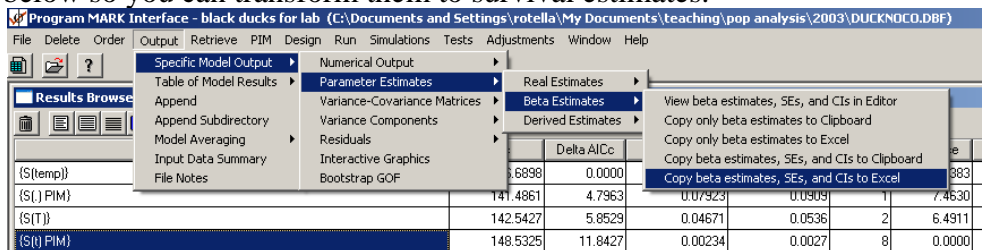
File Edit Format View Help

Lab 01 - black ducks with no covariates - known fate

LOGIT Link Function Parameters of {S by week}

Parameter	Beta	Standard Error	95% Confidence Interval	
			Lower	Upper
1:S	3.8501475	1.0105823	1.8694062	5.8308888
2:S	3.1135153	0.7226494	1.6971224	4.5299082
3:S	2.9704144	0.7250110	1.5493928	4.3914361
4:S	1.9169226	0.4789695	0.9781424	2.8557028
5:S	1.9459102	0.5345225	0.8982461	2.9935743
6:S	2.1202635	0.6110101	0.9226838	3.3178433
7:S	3.1780538	1.0206206	1.1776374	5.1784703
8:S	3.1354944	1.0215079	1.1333388	5.1376500

D. If you prefer, you can move the estimates to EXCEL instead. From MARK, copy the $\hat{\beta}$'s, $SE(\hat{\beta})$'s, and the 95% CIs into EXCEL as shown in the graphic below so you can transform them to survival estimates.



	A	B	C	D	E	F
1	Index	Label	Estimate	SE	LCI	UCI
2	1		3.850148	1.010582	1.869406	5.830889
3	2		3.113515	0.722649	1.697122	4.529908
4	3		2.970414	0.725011	1.549393	4.391436
5	4		1.916923	0.47897	0.978142	2.855703
6	5		1.94591	0.534523	0.898246	2.993574
7	6		2.120264	0.61101	0.922684	3.317843
8	7		3.178054	1.020621	1.177637	5.178471
9	8		3.135494	1.021508	1.133339	5.137649
10						

- E. Once you've got these values in R or EXCEL, create a column that back-transforms the beta values for each week to an estimate of \hat{S}_i .
- F. Create columns that back-transform the 95% LCI and UCI values in a way that's appropriate given that you used the logit link, i.e., create confidence intervals for \hat{S} using the CI limits for the estimated beta values. If you've done it correctly, the estimates you just created will match what you see in step G immediately below.
- G. In MARK, on the Results Browser window, click on the button that reveals the "estimates of real parameters" (3rd button to the right of the trash can).

Parameter	Estimate	Standard Error	95% Confidence Interval	
			Lower	Upper
1:S	0.9791667	0.0206152	0.8663896	0.9970731
2:S	0.9574468	0.0294425	0.8451585	0.9893333
3:S	0.9512195	0.0336412	0.8248260	0.9877685
4:S	0.8717949	0.0535337	0.7267395	0.9456127
5:S	0.8750000	0.0584634	0.7105889	0.9522830
6:S	0.8928571	0.0584512	0.7155886	0.9650359
7:S	0.9600000	0.0391918	0.7645227	0.9943950
8:S	0.9583333	0.0407894	0.7564545	0.9941628

- a. Do these numbers look like the numbers you calculated in R or EXCEL in step F? They should.
- b. Do the \hat{S}_i look like the numbers you calculated back at step 1.A.? They should. You might also check whether MARK's values for $SE(\hat{S}_i)$ match what you estimated in step 1.B.

*Note: the CI's for \hat{S}_i in MARK for this data type and many others are built by back-transforming the CI's of the $\text{logit}(\hat{S}_i)$ onto the real parameter scale. They are **not** built by working with the \hat{S}_i & $SE(\hat{S}_i)$'s to build CI's. The method MARK uses has better properties for coverage, etc. and leads to CI's on the real parameters that are **not** symmetric about \hat{S}_i . The CI's about the beta's are simply $\text{beta-hat} \pm 1.96 \cdot SE(\hat{\beta})$. Thus, the CI's about the beta's are symmetric around beta. The back-transformation leads to asymmetry.*

OK – what have you done so far?

- ✓ Input a dataset to MARK
- ✓ Considered how a design matrix relates to a model of interest
- ✓ Considered how a logistic regression model looks for a model of interest
- ✓ Analyzed a dataset using a logit link, i.e., estimated the model's parameters
- ✓ Back-transformed estimated beta's and CI's for betas to real parameters.

What's left to do?

- Consider some other models
- Consider the structure of each model's regression model and design matrix
- Estimate parameters for other models
- Evaluate the strength of evidence for each of the various models
- Consider the inferences that can be made given the dataset and model list

4. A model of constant survival

A. Perhaps survival doesn't vary among the 8 weeks. Without using Program MARK, how would you calculate an estimate of survival rate for a model that assumes that there is no time variation across the 8 weeks? If you're stuck go back to 1.A. Let's see how to run such a model in Program MARK.

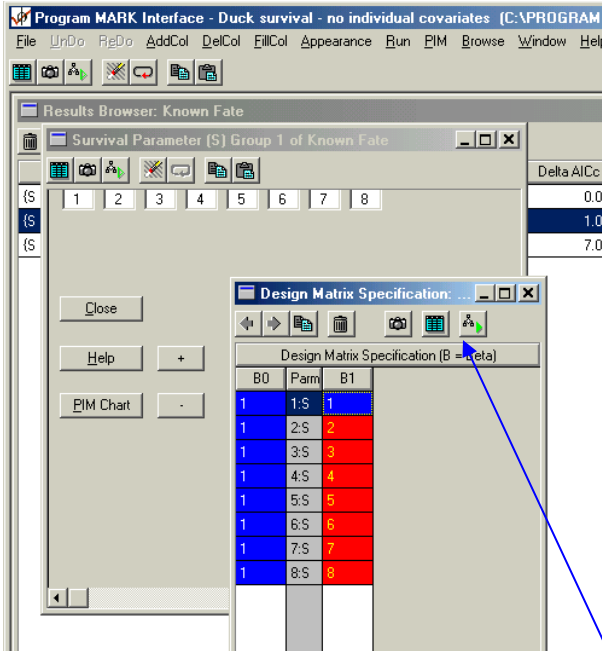
- a. You need to build an intercept-only model that holds survival rate constant for all 8 weeks. To do so, we can simply model the probability of surviving any week the same way using a logistic regression model that is an intercept-only model, i.e., $\Pr(S = 1 | \text{model with constant } S, \text{ data}) = \exp(\beta_0 \cdot 1) / (1 + \exp(\beta_0 \cdot 1))$ or we can simplify the right-hand side to be: $\exp(\beta_0) / (1 + \exp(\beta_0))$.
- b. One way to do this is to use a design matrix with 8 rows and 1 column and to fill that column with 1's, i.e., use an intercept-only model. To create such a matrix, click on the "Design" pull-down menu, choose "Reduced," and state that you want 1 column. Here, every week is *constrained by the model* to have the same survival estimate. Every $\hat{S}_i = \exp(\beta_0 \cdot 1) / (1 + \exp(\beta_0 \cdot 1))$.
- c. Analyze this model using the logit link, examine the $\hat{\beta}$, $SE(\hat{\beta})$, & the 95% CI, and make sure that you understand how to go from $\hat{\beta}$ to \hat{S}_i . Do the estimates of \hat{S}_i & $SE(\hat{S}_i)$ match what you got in step 1.A. for the **TOTAL** row? They should.

BO	Parm
1	1:S
1	2:S
1	3:S
1	4:S
1	5:S
1	6:S
1	7:S
1	8:S

5. A trend model

You might hypothesize that survival decreases as the season progresses and winter severity increases. For such a model, we need a covariate that follows a linear pattern, e.g., $x = 1$ in week 1, $x = 2$ in week 2, ..., $x = 8$ in week 8. There is nothing magic about using $x = 1-8$; other values could be used. But, remember that a slope represents the change in y per unit change in x , so it's convenient to have the x values go up by increments of 1. (*Test your understanding:* What do you think would happen to your slope estimates, if you made the x values 0.1, 0.2, ..., 0.8 or 10, 20, ..., 80? What do you think happens to your intercept value if you use $x = 0, 1, \dots, 7$ instead of 1, 2, ..., 8?) The keys are that we employ a model that constrains the \hat{S}_i such that they follow a linear trend over time on the logit scale, develop a covariate that follows a linear trend, and estimate a slope term for that covariate to evaluate a possible linear trend in \hat{S}_i over the 8 weeks. Work through the example below to improve your understanding and feel free to play with other versions of x values such as those mentioned a few lines earlier.

- A. Click on the “Design” pull-down menu, choose “Reduced,” and state that you want to use 2 covariate columns (1 for an intercept and 1 for our covariate).
- B. Fill in the 1st column with 1’s and the 2nd column with 1 through 8 so that your design matrix looks like the one below. You can fill in the numbers by hand or use the “FillCol” menu that appears when you open a Design Matrix.

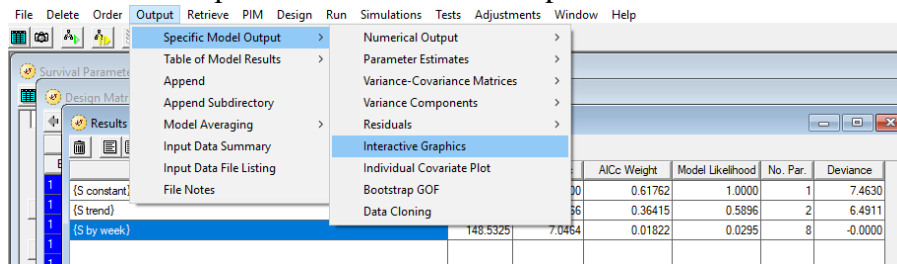


- C. Click on the “Run” menu or the “Run” icon. Once again, choose the logit link and run the model (call it something like ‘S(time trend)’). Think about what this logistic regression model looks like.

$$\Pr(S = 1 | \text{linear trend model for } S, \text{ data}) = \frac{\exp(\beta_0 + \beta_1 \cdot x_i)}{1 + \exp(\beta_0 + \beta_1 \cdot x_i)},$$

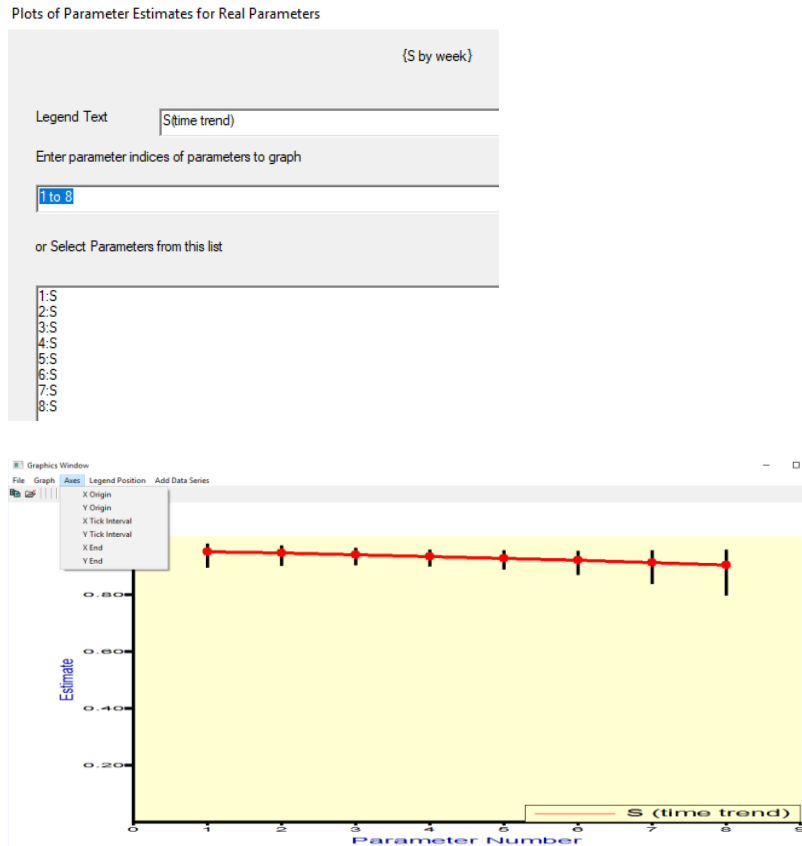
where x_i is the value in the 2nd column of the design matrix for a given week.

- D. Examine the beta estimates and consider the evidence for a linear trend in survival rate across the 8 weeks, what direction any trend might be, and how certain you are of the beta for trend.
- E. In MARK, look at the \hat{S}_i and check for overlap among CI’s. Is there much?
- F. In MARK’s Results Browser, highlight the S(time trend) model and find the “Interactive Graphics” item under the Output menu.



Use the Interactive Graphics to give your plot some brief legend text and plot parameters “1 to 8”. After you have your plot made, use the menu options on the graphics page to (a) adjust the y-axis to have an origin of 0, (b) set y-axis tick intervals of 0.20, and (c) place the legend text in the bottom right. The

images below will help you get those tasks done. Finally, save your image to a jpg file with a name of your choosing.



- G. Use R or EXCEL, the estimated values for the beta's, and x values 1, 2, ..., 8 to calculate (a) the log-odds of weekly survival rate or $(\hat{\beta}_0 + \hat{\beta}_1 \cdot x_i)$ and (b) the estimated survival rate for each week. Notice that:
- logit(S), which is $\ln(S/(1-S))$ is declining each week by a constant amount. What is that amount? What is $\hat{\beta}_1$? Hmm, does that make sense to you? Be sure that it does!
 - S-hat is declining each week but ... by a different amount each week and that amount increases by week. Does that make sense?
- H. Graph your values of x against logit(S) and against S-hat and check each for linearity?

6. A model considering weather

Imagine that you hypothesized that survival was lower in weeks with colder weather. Evaluate that hypothesis using a model that estimates survival as a function of the average temperature (Celsius) during the week. The 8 weeks had the following average temperatures: 11, 9, 8, 4, 6, 5, 10, 9. Build a model that estimates survival using this covariate. Once again, we need a design matrix with 2 columns: an intercept and a single x variable. But, this time you want to put the temperature values in as x_i values instead of the simple trend values you used last time.

- A. Get Program MARK's windows for PIM and Design Matrix to look like this.

B0:	Parm	B1:
1	1:S	11
1	2:S	9
1	3:S	8
1	4:S	4
1	5:S	6
1	6:S	5
1	7:S	10
1	8:S	9

- B. Run this model and give it a name like “S(temp)” – Take a look at the model’s estimates and see how they compare to those from the other models. Does it appear that temperature is related to survival? How so? How strongly?
- C. You could import the \hat{S}_i , $SE(\hat{S}_i)$, and associated CI’s into R or EXCEL. You could then type in the associated temperature values for each week. You could then plot \hat{S}_i (along with CI’s) against temperature. To do the work in R, enter the following code into R but don’t execute it yet.

```
s_temp <- read.table("clipboard",
                    header = F,
                    col.names = c("wk", "s_hat", "se", "lcl", "ucl"))
```

Next, go to Notepad, highlight the estimates of the real parameters, and copy them to the clipboard as shown below.

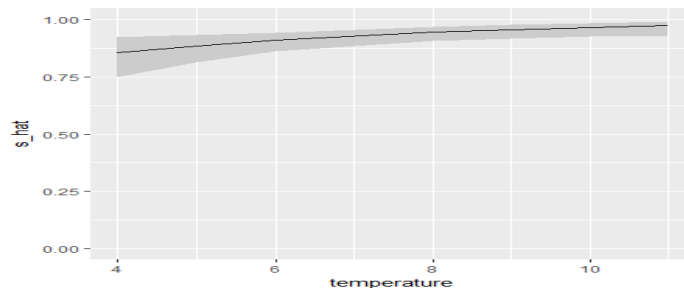
Lab 01 - black ducks with no covariates - known fate

Real Function Parameters of {S temp}

Parameter	Estimate	Standard Error	95% Confidence Interval	
			Lower	Upper
1:S	0.9740110	0.0130349	0.9317872	0.9903684
2:S	0.9569513	0.0142715	0.9185203	0.9776960
3:S	0.9448129	0.0146810	0.9079117	0.9674572
4:S	0.8576124	0.0437777	0.7489346	0.9240198
5:S	0.9103510	0.0202321	0.8620037	0.9428821
6:S	0.8866293	0.0291986	0.8156907	0.9325230
7:S	0.9665146	0.0138062	0.9259768	0.9852070
8:S	0.9569513	0.0142715	0.9185203	0.9776960

Next, return to R and execute the code above. Finally, to produce a useful graph of the estimated relationship between weekly survival rate and temperature, execute the following code, which adds the temperature values and makes a plot like the one at the top of the next page.

```
library(ggplot2)
s_temp$temperature <- c(11, 9, 8, 4, 6, 5, 10, 9)
ggplot(s_temp, aes(x = temperature, y = s_hat)) +
  geom_ribbon(aes(ymin = lcl, ymax = ucl), fill = "grey80") +
  geom_line() + ylim(0, 1)
```



Caution: in the exercises above, you copied and pasted estimates from various models among software platforms. It's important to be very careful when doing so, and to double-check that you pasted the correct values over. A safer way to proceed is to execute the MARK analyses from within R using the RMark package. I'll provide the RMark code each week for our labs, but I won't require you to do anything in RMark as that would be one more challenge in an already-challenging course.

Which model is best?

D. The information-theoretic approach is introduced in Chapter 4 of Williams et al. (2001). You're also learning about it in lecture and your readings of the Cooch and White on-line MARK book. To get some hands-on experience, consider the AICc values for the competing models we have here.

a. Look at the Results Browser and consider which model is best supported by the data.

The screenshot shows the 'Results Browser: Known Fate' window in the MARK software. It contains a table with the following data:

Model	AICc	Delta AICc	AICc Weight	No. Par.	Deviance
{S temp}	136.69	0.00	0.87171	2	0.6380
{S constant}	141.49	4.80	0.07924	1	7.4630
{S Trend}	142.54	5.85	0.04671	2	6.4910
{S by week}	148.53	11.84	0.00234	8	0.0000

E. The AICc Weight column is useful for putting the delta AICc values into perspective.

7. What inferences can be made from the model?

- A. Consider the estimated beta's and survival rates along with their associated measures of precision and uncertainty.
- B. Examine relevant plots of the output for models that contain covariates.
- C. To obtain estimates of weekly survival rate based on results for all of the models, use model-averaging on the model-specific estimates and the model weights. To do this, choose "Output", "Model Averaging", and "Real." Next, select each estimate, choose to have the estimates put into EXCEL, and click on "OK". The file that's opened in Notepad shows you a number of interesting details for each week regarding model-specific estimates and model weights along with model-averaged values and associated measures of uncertainty. The file that's opened in EXCEL provides the model-averaged estimate for each week along with the associated SE and confidence intervals.

We've covered many new topics. Take time to digest today's lab and be sure to understand today's concepts: they'll be used many times this semester. The formal assignment below is intended to help you do that.

Formal Assignment:

1. Describe the available data in a few short sentences.
2. Provide a table that contains the following columns for each of the candidate models evaluated in today's lab: (1) model name, (2) the logit of the logistic regression equation in symbols (*Note*: you don't have to typeset actual Greek letters, i.e., you can just type "Beta0" (*Hint*: logit for constant-survival model = β_0), and (3) the logit of the logistic regression equation with parameter values in place of symbols (*Hint*: logit for constant-survival model = 2.6352908).
3. Provide a table of model-selection results along with a few sentences describing in words what inferences could be made about which models were supported by the data, which ones weren't, and how much model-selection uncertainty was present.
4. Present key features of the best model (include beta estimates and SE's and survival rates and SE's) and any other model that was well-supported by the data. This presentation should include a graphic for the best model along with confidence intervals.
5. Provide model-averaged estimates of weekly survival along with their SE's and CI's.
6. What inferences can you draw from the modeling you did this week?
7. Please briefly list out any questions that you have about this week's lab that you'd like to have discussed at the start of next week's lab.