Figures

```
library(tidyverse)
library(marginaleffects)
library(lubridate)
library(MASS)
library(survival)
library(survminer)
library(conflicted)
conflict_prefer("select", "dplyr")
```

```
## [conflicted] Removing existing preference.
## [conflicted] Will prefer dplyr::select over any other package.
```

Figure 1: Packages

```
my_url <- "http://ritsokiguess.site/datafiles/spector.csv"
spector <- read_csv(my_url)
```

```
## Rows: 32 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): grade_improved, psi
## dbl (2): tuce, gpa
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
spector
```

```
## # A tibble: 32 x 4
##    grade_improved psi    tuce   gpa
##    <chr>          <chr> <dbl> <dbl>
##  1 no             no       20  2.66
##  2 no             no       22  2.89
##  3 no             no       24  3.28
##  4 no             no       12  2.92
##  5 yes            no       21  4
##  6 no             no       17  2.86
##  7 no             no       17  2.76
##  8 no             no       21  2.87
##  9 no             no       25  3.03
## 10 yes            no       29  3.92
## # ... with 22 more rows
```

Figure 2: Teaching method study data (some)

```
spector.1 <- glm(factor(grade_improved) ~ psi + tuce + gpa, data = spector, family = "binomial")
summary(spector.1)
```

```
##
## Call:
## glm(formula = factor(grade_improved) ~ psi + tuce + gpa, family = "binomial",
##     data = spector)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9551  -0.6453  -0.2570   0.5888   2.0966
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -13.02135    4.93127  -2.641  0.00828 **
## psiyes        2.37869    1.06456   2.234  0.02545 *
## tuce          0.09516    0.14155   0.672  0.50143
## gpa           2.82611    1.26293   2.238  0.02524 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 41.183  on 31  degrees of freedom
## Residual deviance: 25.779  on 28  degrees of freedom
## AIC: 33.779
##
## Number of Fisher Scoring iterations: 5
```

Figure 3: Teaching method study analysis part 1

```
spector.2 <- update(spector.1, .~. - tuce)
summary(spector.2)
```

```
##
## Call:
## glm(formula = factor(grade_improved) ~ psi + gpa, family = "binomial",
##     data = spector)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8396  -0.6282  -0.3045   0.5629   2.0378
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -11.602      4.213  -2.754  0.00589 **
## psiyes         2.338      1.041   2.246  0.02470 *
## gpa            3.063      1.223   2.505  0.01224 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 41.183  on 31  degrees of freedom
## Residual deviance: 26.253  on 29  degrees of freedom
## AIC: 32.253
##
## Number of Fisher Scoring iterations: 5
```

Figure 4: Teaching method study analysis part 2

```
psis <- c("no", "yes")
gpas <- c(2.8, 3.1, 3.5)
new <- datagrid(psi = psis, gpa = gpas, model = spector.2)
new %>% select(psi, gpa)
```

```
##   psi gpa
## 1  no 2.8
## 2  no 3.1
## 3  no 3.5
## 4 yes 2.8
## 5 yes 3.1
## 6 yes 3.5
```

```
cbind(predictions(spector.2, newdata = new)) %>%
  select(estimate, conf.low, conf.high, psi, gpa)
```

```
##    estimate    conf.low conf.high psi gpa
## 1 0.0463473 0.006413699 0.2678833  no 2.8
## 2 0.1085996 0.024086475 0.3755384  no 3.1
## 3 0.2932235 0.088128350 0.6404104  no 3.5
## 4 0.3348431 0.100939593 0.6929838 yes 2.8
## 5 0.5579014 0.257191196 0.8214082 yes 3.1
## 6 0.8112260 0.468529084 0.9544377 yes 3.5
```

Figure 5: Teaching method study predictions

```
troutegg
```

```
## # A tibble: 20 x 4
##    survive total location period
##      <dbl> <dbl> <chr>     <dbl>
##  1      89    94 A             4
##  2     106   108 B             4
##  3     119   123 C             4
##  4     104   104 D             4
##  5      49    93 E             4
##  6      94    98 A             7
##  7      91   106 B             7
##  8     100   130 C             7
##  9      80    97 D             7
## 10      11   113 E             7
## 11      77    86 A             8
## 12      87    96 B             8
## 13      88   119 C             8
## 14      67    99 D             8
## 15      18    88 E             8
## 16     141   155 A            11
## 17     104   122 B            11
## 18      91   125 C            11
## 19     111   132 D            11
## 20       0   138 E            11
```

Figure 6: Trout egg data

```
troutegg %>%
  mutate(dead = total - survive) %>%
  select(survive, dead) %>%
  as.matrix() -> response
response
```

```
##       survive dead
##  [1,]      89    5
##  [2,]     106    2
##  [3,]     119    4
##  [4,]     104    0
##  [5,]      49   44
##  [6,]      94    4
##  [7,]      91   15
##  [8,]     100   30
##  [9,]      80   17
## [10,]      11  102
## [11,]      77    9
## [12,]      87    9
## [13,]      88   31
## [14,]      67   32
## [15,]      18   70
## [16,]     141   14
## [17,]     104   18
## [18,]      91   34
## [19,]     111   21
## [20,]       0  138
```

Figure 7: Trout egg code

```r
troutegg.1 <- glm(response ~ period + location + I(period^2),
                  family = "binomial", data = troutegg)
summary(troutegg.1)
```

```
##
## Call:
## glm(formula = response ~ period + location + I(period^2), family = "binomial",
##     data = troutegg)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.8628  -0.5488  -0.0178   0.7256   3.5177
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  9.69222    0.79999  12.115  < 2e-16 ***
## period      -1.60962    0.19677  -8.180 2.83e-16 ***
## locationB   -0.41505    0.24618  -1.686   0.0918 .
## locationC   -1.24068    0.21952  -5.652 1.59e-08 ***
## locationD   -0.94708    0.22882  -4.139 3.49e-05 ***
## locationE   -4.60742    0.24955 -18.463  < 2e-16 ***
## I(period^2)  0.08436    0.01200   7.031 2.05e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1021.469  on 19  degrees of freedom
## Residual deviance:   65.687  on 13  degrees of freedom
## AIC: 156.22
##
## Number of Fisher Scoring iterations: 5
```

Figure 8: Trout egg logistic regression

```
wg93
```

```
## # A tibble: 871 x 3
##        D sex      edu
##    <dbl> <chr>  <dbl>
##  1     3 male       3
##  2     3 female     4
##  3     4 male       2
##  4     2 female     3
##  5     3 female     2
##  6     5 female     2
##  7     4 male       2
##  8     2 female     3
##  9     1 female     2
## 10     2 female     2
## # ... with 861 more rows
```

Figure 9: German science survey data

```
D.1 <- polr(factor(D) ~ sex + edu, data = wg93)
drop1(D.1, test = "Chisq")
```

```
## Single term deletions
##
## Model:
## factor(D) ~ sex + edu
##        Df    AIC    LRT Pr(>Chi)
## <none>    2664.9
## sex     1 2666.4 3.4622 0.062785 .
## edu     1 2671.7 8.8181 0.002983 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10: German science survey model

```
new <- datagrid(sex = c("female", "male"), edu = c(1, 6), model = D.1)
cbind(predictions(D.1, newdata = new)) %>%
  select(group, estimate, sex, edu) %>%
  pivot_wider(names_from = "group", values_from = "estimate")
```

```
##
## Re-fitting to get Hessian
##
##
## Re-fitting to get Hessian
## # A tibble: 4 x 7
##   sex      edu    `1`   `2`   `3`   `4`   `5`
##   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 female     1 0.0969 0.328 0.235 0.218 0.123
## 2 female     6 0.0499 0.215 0.221 0.291 0.222
## 3 male       1 0.0788 0.292 0.237 0.244 0.149
## 4 male       6 0.0402 0.183 0.207 0.305 0.264
```

Figure 11: German science survey predictions

```
receivables
```

```
## # A tibble: 2,466 x 4
##    invoice_date due_date   invoice_amount settled_date
##    <chr>        <chr>               <dbl> <chr>
##  1 1/2/2013     2/1/2013             55.9 1/15/2013
##  2 1/26/2013    2/25/2013            61.7 3/3/2013
##  3 7/3/2013     8/2/2013             65.9 7/8/2013
##  4 2/10/2013    3/12/2013           106.  3/17/2013
##  5 10/25/2012   11/24/2012           72.3 11/28/2012
##  6 1/27/2012    2/26/2012            94   2/22/2012
##  7 8/13/2013    9/12/2013            74.7 9/9/2013
##  8 12/16/2012   1/15/2013            75.1 1/12/2013
##  9 5/14/2012    6/13/2012            80.1 7/1/2012
## 10 7/1/2013     7/31/2013            48.3 7/26/2013
## # ... with 2,456 more rows
```

Figure 12: Accounts receivable data

```
psych
```

```
##    sex age time death
## 1    2  51    1     1
## 2    2  58    1     1
## 3    2  55    2     1
## 4    2  28   22     1
## 5    1  21   30     0
## 6    1  19   28     1
## 7    2  25   32     1
## 8    2  48   11     1
## 9    2  47   14     1
## 10   2  25   36     0
## 11   2  31   31     0
## 12   1  24   33     0
## 13   1  25   33     0
## 14   2  30   37     0
## 15   2  33   35     0
## 16   1  36   25     1
## 17   1  30   31     0
## 18   1  41   22     1
## 19   2  43   26     1
## 20   2  45   24     1
## 21   2  35   35     0
## 22   1  29   34     0
## 23   1  35   30     0
## 24   1  32   35     1
## 25   2  36   40     1
## 26   1  32   39     0
```

Figure 13: Psychiatric patients data

```
psych %>%
  mutate(y = Surv(time, death == 1)) -> psych
head(psych)
```

```
##   sex age time death   y
## 1   2  51    1     1   1
## 2   2  58    1     1   1
## 3   2  55    2     1   2
## 4   2  28   22     1  22
## 5   1  21   30     0 30+
## 6   1  19   28     1  28
```

Figure 14: Psychiatric patients: some code and its output. Note that "head" displays the first six lines of its input.

```
psych.1 <- coxph(y ~ age + sex, data = psych)
summary(psych.1)
```

```
## Call:
## coxph(formula = y ~ age + sex, data = psych)
##
##   n= 26, number of events= 14
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## age  0.20753   1.23063  0.05828  3.561  0.00037 ***
## sex -0.52374   0.59230  0.73753 -0.710  0.47762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## age    1.2306     0.8126    1.0978     1.380
## sex    0.5923     1.6883    0.1396     2.514
##
## Concordance= 0.816  (se = 0.081 )
## Likelihood ratio test= 20.91  on 2 df,   p=3e-05
## Wald test            = 14.3  on 2 df,   p=8e-04
## Score (logrank) test = 21.27  on 2 df,   p=2e-05
```

Figure 15: Psychiatric patients: model and output

```
new <- datagrid(age = c(28, 35, 42), sex = c(1, 2), model = psych.1)
new
```

```
##   age sex
## 1  28   1
## 2  28   2
## 3  35   1
## 4  35   2
## 5  42   1
## 6  42   2
```

Figure 16: Psychiatric patients: values to predict

```
s <- survfit(psych.1, newdata = new, data = psych)
ggsurvplot(s, conf.int = FALSE)
```

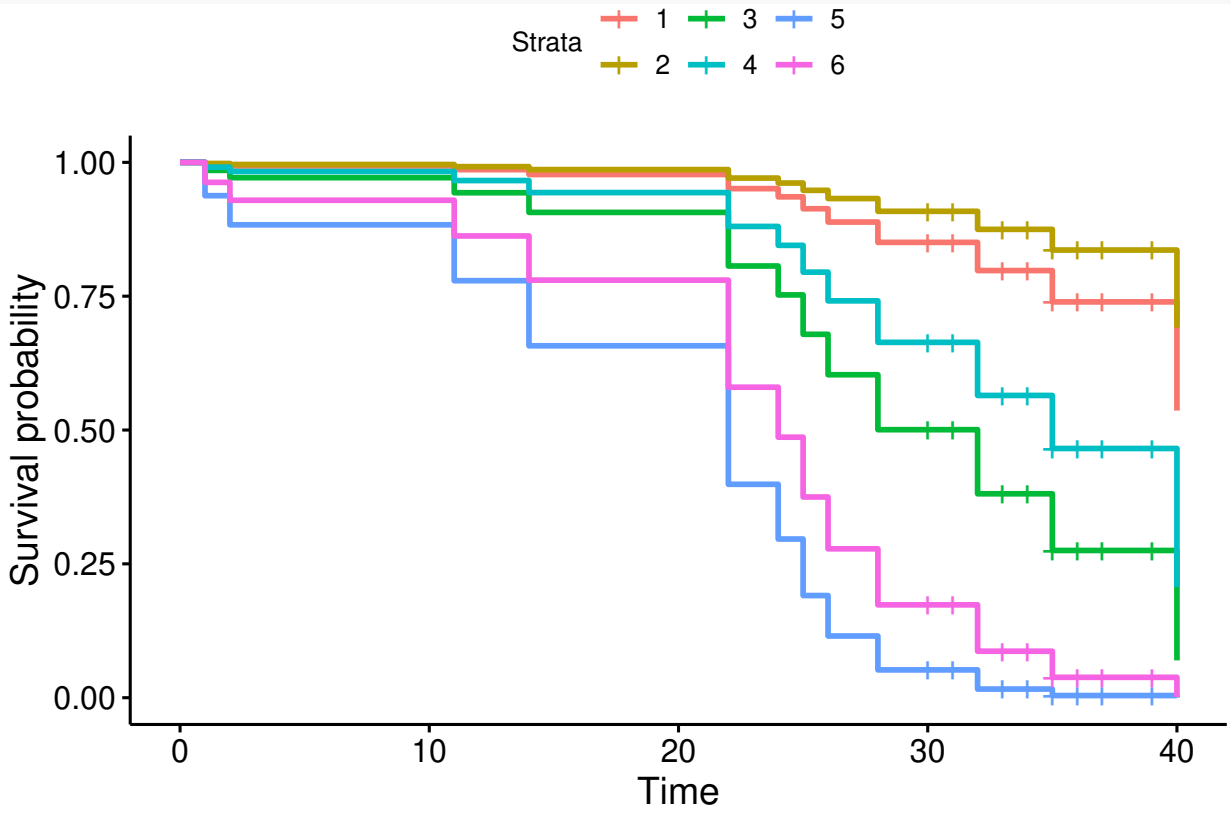

Figure 17: Psychiatric patients: predictions and plot of estimated survival curves