R15 - Interactions in Logistic and Poisson Regression

HCI/PSYCH 522
Iowa State University

May 3, 2022

Interactions

Independent variables

- Categorical-continuous
- Categorical-categorical
- Continuous-continuous (not discussed)

Generalized linear models

- Poisson regression
- Logistic regression

Poisson regression

Poisson regression with two (continuous) independent variables:

- Data model

$$Y_i \overset{ind}{\sim} Po(\lambda_i)$$

- Main effects model

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2}$$

- Interaction model

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,1} X_{i,2}$$

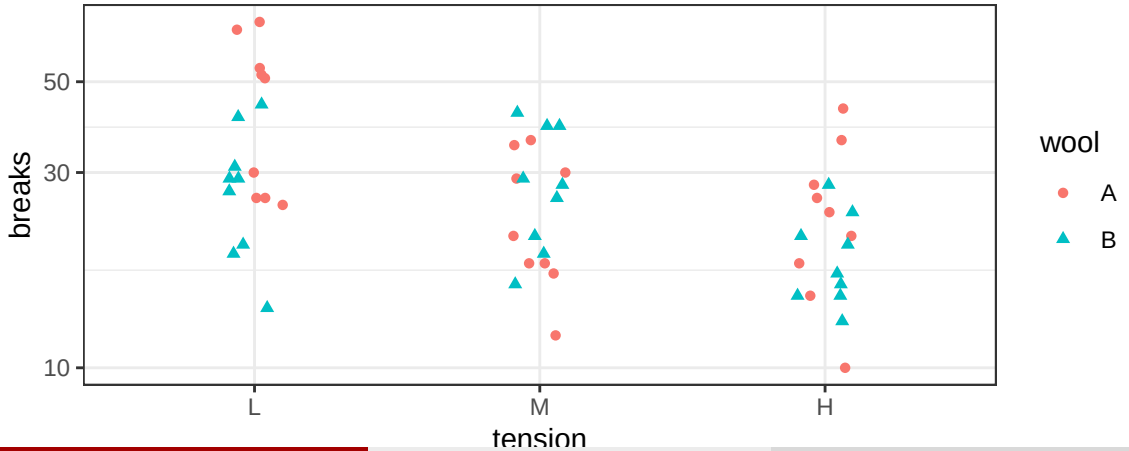
The interaction occurs on the log scale.

Warpbreaks - Descriptive Statistics

```
head(warpbreaks)

##   breaks wool tension
## 1     26    A       L
## 2     30    A       L
## 3     54    A       L
## 4     25    A       L
## 5     70    A       L
## 6     52    A       L

summary(warpbreaks)

##      breaks      wool   tension
##  Min.   :10.00   A:27   L:18
##  1st Qu.:18.25   B:27   M:18
##  Median :26.00          H:18
##  Mean   :28.15
##  3rd Qu.:34.00
##  Max.   :70.00
```

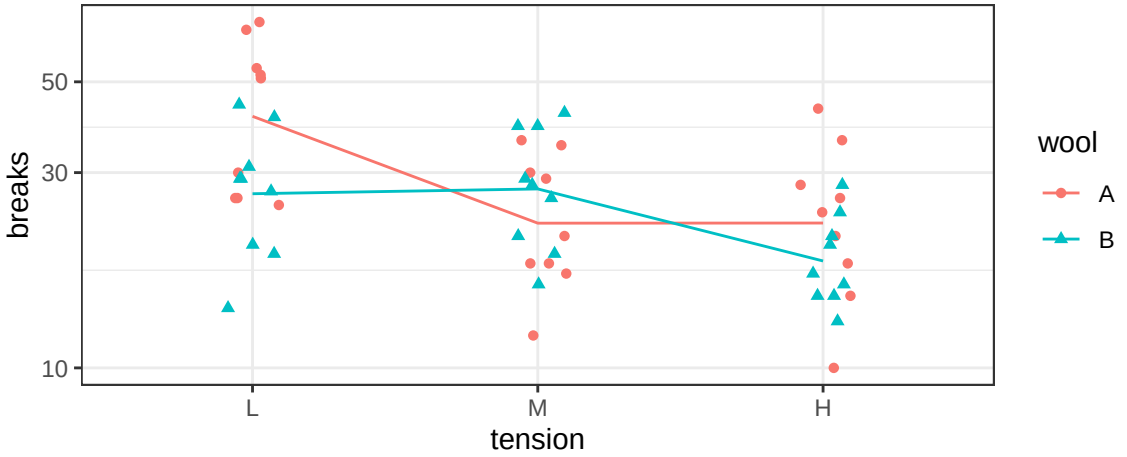
Warpbreaks - Graphical Statistics

```
g <- ggplot(warpbreaks, aes(x = tension, y = breaks,
                            color = wool, shape = wool, group = wool)) +
  geom_jitter(width=0.1, height=0) + scale_y_log10()
g
```



Warpbreaks - Graphical Statistics

```
g + stat_summary(fun = mean, geom = "line")
```

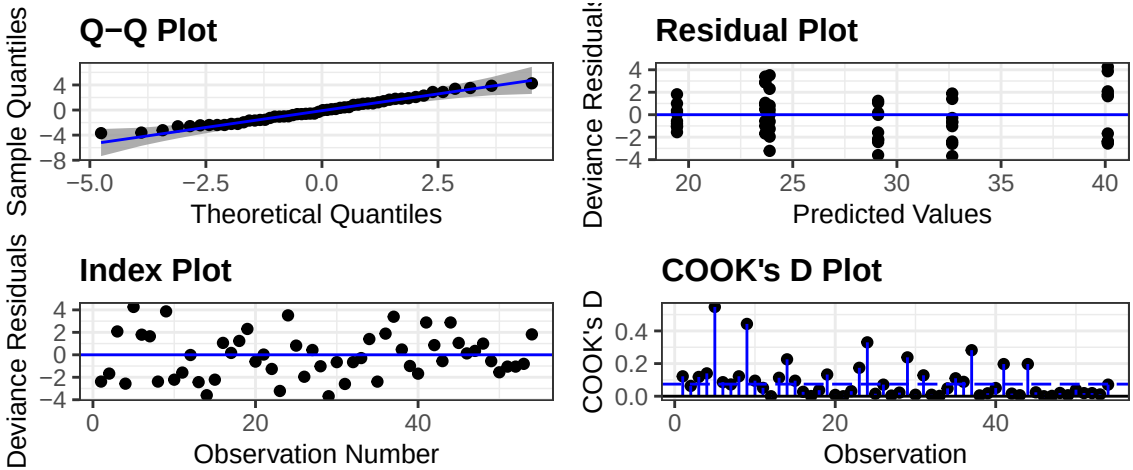


Warpbreaks - Poisson Regression Models

```
mM <- glm(breaks ~ wool + tension, data = warpbreaks, family = poisson) # Main effects model
mI <- glm(breaks ~ wool * tension, data = warpbreaks, family = poisson) # Interaction model

drop1(mI, test="Chi")

## Single term deletions
##
## Model:
## breaks ~ wool * tension
##              Df Deviance    AIC    LRT  Pr(>Chi)
## <none>            182.31 468.97
## wool:tension  2   210.39 493.06 28.087 7.962e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Warpbreaks - Diagnostics

```
resid_panel(mM, plots = c("qq", "resid", "index", "cookd"), qqbands = TRUE)
```



Warpbreaks - Main effects model

```
summary(mM)

##
## Call:
## glm(formula = breaks ~ wool + tension, family = poisson, data = warpbreaks)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
## woolB       -0.20599    0.05157  -3.994 6.49e-05 ***
## tensionM    -0.32132    0.06027  -5.332 9.73e-08 ***
## tensionH    -0.51849    0.06396  -8.107 5.21e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: 493.06
##
## Number of Fisher Scoring iterations: 4
```

Warpbreaks - Main effects model

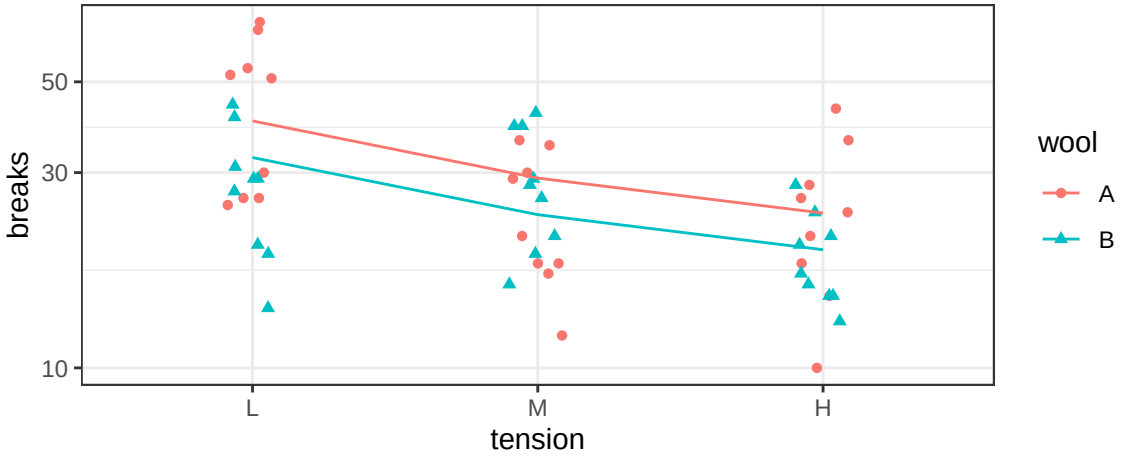
```
em <- emmeans(mM, pairwise ~ wool)
cm <- confint(em, type  = "unlink"); cm

## $emmeans
##  wool rate    SE  df asymp.LCL asymp.UCL
##  A    30.3 1.061 Inf      28.3      32.5
##  B    24.7 0.955 Inf      22.9      26.6
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B     1.23 0.0634 Inf      1.11      1.36
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

Warpbreaks - Main effects model

```
et <- emmeans(mM, pairwise ~ tension)
ct <- confint(et, type  = "unlink"); ct

## $emmeans
##  tension rate   SE  df asymp.LCL asymp.UCL
##  L       36.2 1.42 Inf      33.5      39.1
##  M       26.2 1.21 Inf      24.0      28.7
##  H       21.6 1.09 Inf      19.5      23.8
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M     1.38 0.0831 Inf      1.20      1.59
##  L / H     1.68 0.1074 Inf      1.45      1.95
##  M / H     1.22 0.0832 Inf      1.04      1.43
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## Intervals are back-transformed from the log scale
```

Warpbreaks - Main effects model

```
g + geom_line(mapping=aes(y=predict(mM, warpbreaks, type  = "response")))
```



Warpbreaks - Interaction model

```
summary(mI)

##
## Call:
## glm(formula = breaks ~ wool * tension, family = poisson, data = warpbreaks)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3383  -1.4844  -0.1291   1.1725   3.5153
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.79674    0.04994  76.030  < 2e-16 ***
## woolB          -0.45663    0.08019  -5.694 1.24e-08 ***
## tensionM       -0.61868    0.08440  -7.330 2.30e-13 ***
## tensionH       -0.59580    0.08378  -7.112 1.15e-12 ***
## woolB:tensionM  0.63818    0.12215   5.224 1.75e-07 ***
## woolB:tensionH  0.18836    0.12990   1.450    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 182.31  on 48  degrees of freedom
## AIC: 468.97
##
## Number of Fisher Scoring iterations: 4
```

Warpbreaks - Interaction model

```
em <- emmeans(mI, pairwise ~ wool | tension)
cm <- confint(em, type  = "unlink"); cm$emmeans

## tension = L:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    44.6 2.22 Inf      40.4      49.1
##  B    28.2 1.77 Inf      25.0      31.9
##
## tension = M:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    24.0 1.63 Inf      21.0      27.4
##  B    28.8 1.79 Inf      25.5      32.5
##
## tension = H:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    24.6 1.65 Inf      21.5      28.0
##  B    18.8 1.44 Inf      16.1      21.8
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

Warpbreaks - Interaction model

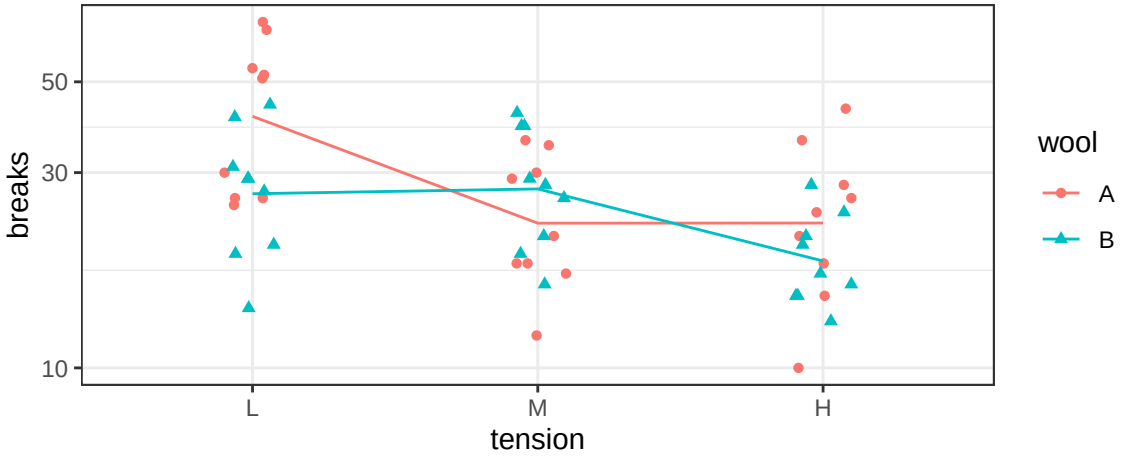
```
cm$contrasts

## tension = L:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B    1.579 0.1266 Inf     1.349     1.847
##
## tension = M:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B    0.834 0.0768 Inf     0.696     0.999
##
## tension = H:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B    1.308 0.1336 Inf     1.070     1.598
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

Warpbreaks - Interaction model

```
et <- emmeans(mI, pairwise ~ tension | wool)
ct <- confint(et, type  = "unlink"); ct$contrasts

## wool = A:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M    1.856 0.1567 Inf     1.523      2.26
##  L / H    1.814 0.1520 Inf     1.491      2.21
##  M / H    0.977 0.0935 Inf     0.781      1.22
##
## wool = B:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M    0.981 0.0866 Inf     0.797      1.21
##  L / H    1.503 0.1492 Inf     1.191      1.90
##  M / H    1.533 0.1515 Inf     1.216      1.93
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## Intervals are back-transformed from the log scale
```

Warpbreaks - Interaction model

```
g + stat_summary(fun = mean, geom = "line")
```



Logistic regression

Logistic regression with two (continuous) independent varibles.

- Data model

$$Y_g \overset{ind}{\sim} Bin(n_g, \theta_g)$$

- Main effects model

$$\log\left(\frac{\theta_g}{1-\theta_g}\right) = \beta_0 + \beta_1 X_{g,1} + \beta_2 X_{g,2}$$

- Interaction model

$$\log\left(\frac{\theta_g}{1-\theta_g}\right) = \beta_0 + \beta_1 X_{g,1} + \beta_2 X_{g,2} + \beta_3 X_{g,1} X_{g,2}$$

The interaction occurs on the log-odds scale.

Donner Party - Descriptive Statistics

```
head(case2001)
```

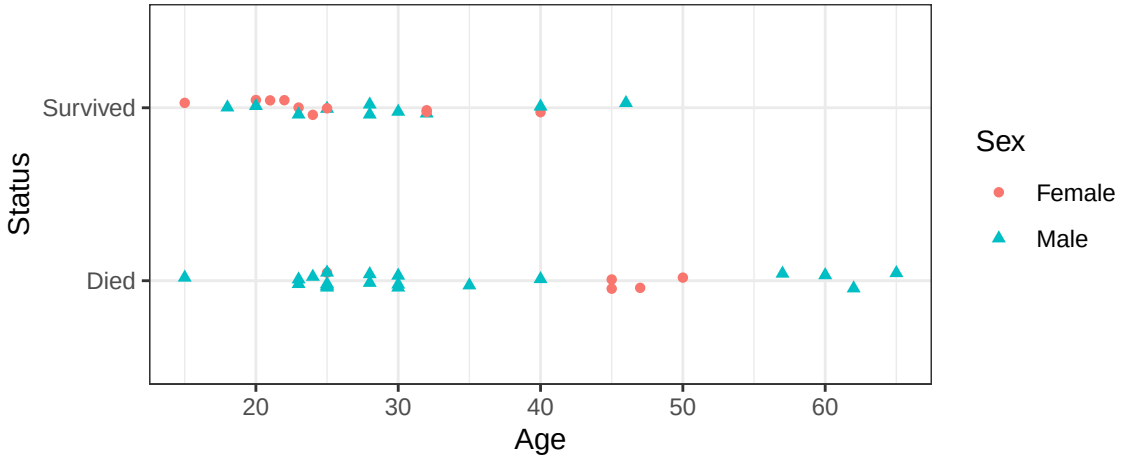
```
##   Age    Sex   Status
## 1  23   Male     Died
## 2  40 Female Survived
## 3  40   Male Survived
## 4  30   Male     Died
## 5  28   Male     Died
## 6  40   Male     Died
```

```
summary(case2001)
```

```
##       Age           Sex          Status
##  Min.   :15.0   Female:15   Died    :25
##  1st Qu.:24.0   Male  :30   Survived:20
##  Median :28.0
##  Mean   :31.8
##  3rd Qu.:40.0
##  Max.   :65.0
```

Donner Party - Graphical Statistics

```
g <- ggplot(case2001, aes(x = Age, y = Status, color = Sex, shape = Sex)) +
  geom_jitter(width=0, height = 0.05)
g
```

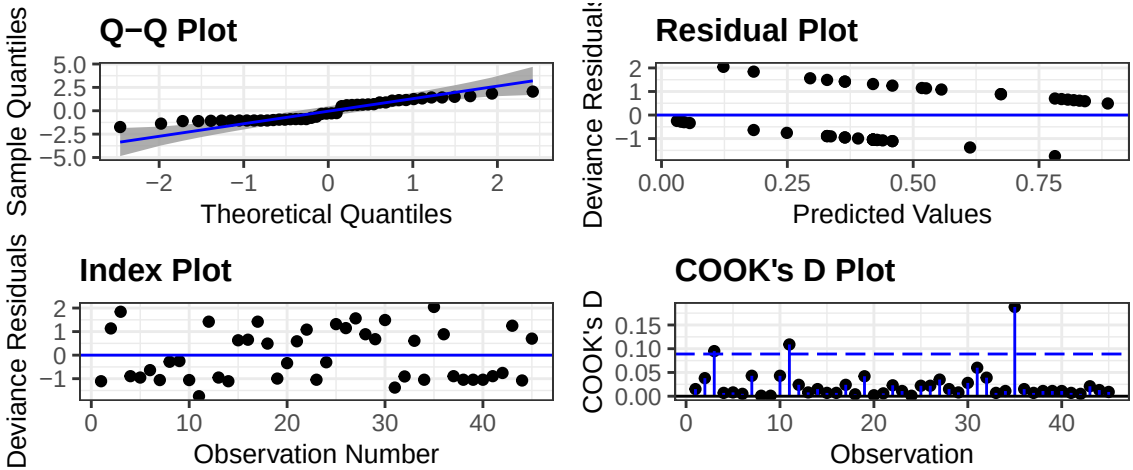


Donner Party - Logistic Regression Models

```
mM <- glm(Status == "Survived" ~ Sex + Age, data = case2001, family = binomial) # Main effects model
mI <- glm(Status == "Survived" ~ Sex * Age, data = case2001, family = binomial) # Interaction model

drop1(mI, test="Chi")

## Single term deletions
##
## Model:
## Status == "Survived" ~ Sex * Age
##         Df Deviance    AIC    LRT Pr(>Chi)
## <none>       47.346 55.346
## Sex:Age  1   51.256 57.256 3.9099    0.048 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Donner Party - Diagnostics

```
resid_panel(mM, plots = c("qq", "resid", "index", "cookd"), qqbands = TRUE)
```



Donner Party - Main effects model

```
summary(mM)

##
## Call:
## glm(formula = Status == "Survived" ~ Sex + Age, family = binomial,
##     data = case2001)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7445  -1.0441  -0.3029   0.8877   2.0472
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.23041    1.38686   2.329   0.0198 *
## SexMale     -1.59729    0.75547  -2.114   0.0345 *
## Age         -0.07820    0.03728  -2.097   0.0359 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
```

Donner Party - Main effects model

```
em <- emmeans(mM, pairwise ~ Sex | Age, at = list(Age = c(15,28,65)))
cm <- confint(em, type = "unlink"); cm$emmeans

## Age = 15:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female   0.8867 0.0918 Inf   0.56620     0.979
##  Male     0.6130 0.1484 Inf   0.31742     0.844
##
## Age = 28:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female   0.7390 0.1217 Inf   0.45129     0.907
##  Male     0.3643 0.0949 Inf   0.20429     0.561
##
## Age = 65:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female   0.1355 0.1539 Inf   0.01181     0.673
##  Male     0.0308 0.0433 Inf   0.00185     0.353
##
## Confidence level used: 0.95
## Intervals are back-transformed from the logit scale
```

Donner Party - Main effects model

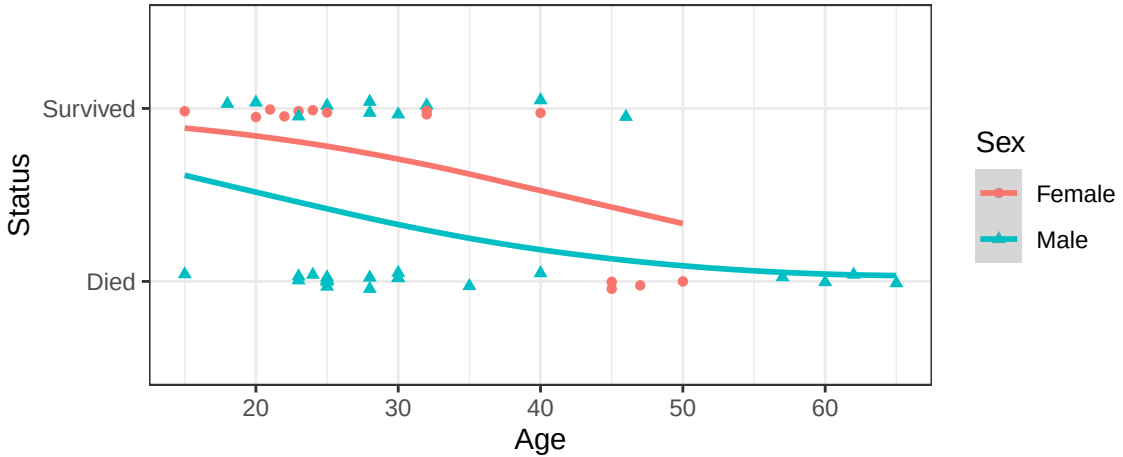
```
cm$contrasts

## Age = 15:
##  contrast      odds.ratio   SE  df asymp.LCL asymp.UCL
##  Female / Male       4.94 3.73 Inf      1.12      21.7
##
## Age = 28:
##  contrast      odds.ratio   SE  df asymp.LCL asymp.UCL
##  Female / Male       4.94 3.73 Inf      1.12      21.7
##
## Age = 65:
##  contrast      odds.ratio   SE  df asymp.LCL asymp.UCL
##  Female / Male       4.94 3.73 Inf      1.12      21.7
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log odds ratio scale
```

Donner Party - Main effects model

```
et <- emtrends(mM, pairwise ~ Sex, var = "Age")
ct <- confint(et); ct

## $emtrends
##  Sex    Age.trend     SE  df asymp.LCL asymp.UCL
##  Female   -0.0782 0.0373 Inf    -0.151  -0.00513
##  Male     -0.0782 0.0373 Inf    -0.151  -0.00513
##
## Results are given on the == (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##  contrast      estimate SE  df asymp.LCL asymp.UCL
##  Female - Male        0  0 Inf         0         0
##
## Note: contrasts are still on the == scale
## Results are given on the == (not the response) scale.
## Confidence level used: 0.95
```

Donner Party - Main effects model

```
g + geom_smooth(mapping=aes(y=predict(mM, case2001, type="response")+1))
```



Donner Party - Interaction model

```
summary(mI)

##
## Call:
## glm(formula = Status == "Survived" ~ Sex * Age, family = binomial,
##     data = case2001)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2279  -0.9388  -0.5550   0.7794   1.6998
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.24638    3.20517   2.261   0.0238 *
## SexMale     -6.92805    3.39887  -2.038   0.0415 *
## Age         -0.19407    0.08742  -2.220   0.0264 *
## SexMale:Age  0.16160    0.09426   1.714   0.0865 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 47.346  on 41  degrees of freedom
## AIC: 55.346
##
## Number of Fisher Scoring iterations: 5
```

Donner Party - Interaction model

```
em <- emmeans(mI, pairwise ~ Sex | Age, at = list(Age = c(15,28,65)))
cm <- confint(em, type = "unlink"); cm$emmeans

## Age = 15:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female  0.98707 0.0251 Inf  6.19e-01     1.000
##  Male    0.45789 0.1641 Inf  1.88e-01     0.755
##
## Age = 28:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female  0.85964 0.1258 Inf  4.42e-01     0.979
##  Male    0.35640 0.0917 Inf  2.02e-01     0.548
##
## Age = 65:
##  Sex    response     SE  df asymp.LCL asymp.UCL
##  Female  0.00464 0.0125 Inf  2.31e-05     0.484
##  Male    0.14274 0.1582 Inf  1.30e-02     0.677
##
## Confidence level used: 0.95
## Intervals are back-transformed from the logit scale
```

Donner Party - Interaction model

```
cm$contrasts

## Age = 15:
##  contrast      odds.ratio      SE  df asymp.LCL asymp.UCL
##  Female / Male     90.387 187.410 Inf  1.55e+00    5260.2
##
## Age = 28:
##  contrast      odds.ratio      SE  df asymp.LCL asymp.UCL
##  Female / Male     11.060  12.352 Inf  1.24e+00      98.7
##
## Age = 65:
##  contrast      odds.ratio      SE  df asymp.LCL asymp.UCL
##  Female / Male      0.028   0.084 Inf  7.83e-05      10.0
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log odds ratio scale
```

Donner Party - Interaction model

```
et <- emtrends(mI, pairwise ~ Sex, var = "Age")
ct <- confint(et); ct

## $emtrends
##  Sex    Age.trend     SE  df asymp.LCL asymp.UCL
##  Female   -0.1941 0.0874 Inf    -0.365   -0.0227
##  Male     -0.0325 0.0353 Inf    -0.102    0.0366
##
## Results are given on the == (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##  contrast      estimate     SE  df asymp.LCL asymp.UCL
##  Female - Male   -0.162 0.0943 Inf    -0.346    0.0232
##
## Note: contrasts are still on the == scale
## Results are given on the == (not the response) scale.
## Confidence level used: 0.95
```

Donner Party - Interaction model

```
g + geom_smooth(mapping=aes(y=predict(mI, case2001, type="response")+1))
```