

# Binary Logistic Regression

Conway

```
## Warning: package 'knitr' was built under R version 3.5.3
```

## Prompt

The data are based on a mock jury study conducted by Shari Diamond and Jonathan Casper. Subjects (N = 100) watched a videotaped sentencing phase trial in which the defendant had already been found guilty. The issue for the jurors to decide was whether the defendant deserved the death penalty. These data were collected “pre-deliberation” (i.e., each juror was asked to provide his/her vote on the death penalty verdict, then the jurors met as a group to decide the overall jury verdict). The initial individual verdicts are given in this data set. Verdict is dummy coded: 0 = life sentence, 1 = death penalty.

## Load packages and import data

```
# Load packages
library(pacman)
```

```
## Warning: package 'pacman' was built under R version 3.5.3
```

```
p_load(psych, jmv, aod, QuantPsyc, ggeffects, ggplot2)
```

```
# Import data
```

```
BL <- read.csv("https://www.dropbox.com/s/hd43va8a7hjfj27/Diamond.csv?dl=1")
```

## Descriptive statistics

```
# Descriptive statistics
```

```
# Including the binary outcome so we can see the frequencies (prints below Descriptives table)
```

```
# baseline classification success is equal to the reference frequency for Retention (No = 2)
```

```
desc <- descriptives(data = BL,
  vars = c('verdict', 'danger', 'rehab', 'punish', 'gendet', 'specdet', 'incap'),
  sd = TRUE,
  skew = TRUE,
  kurt = TRUE,
  freq = TRUE,
  hist = TRUE)
```

```
desc
```

```
##
```

```
## DESCRIPTIVES
```

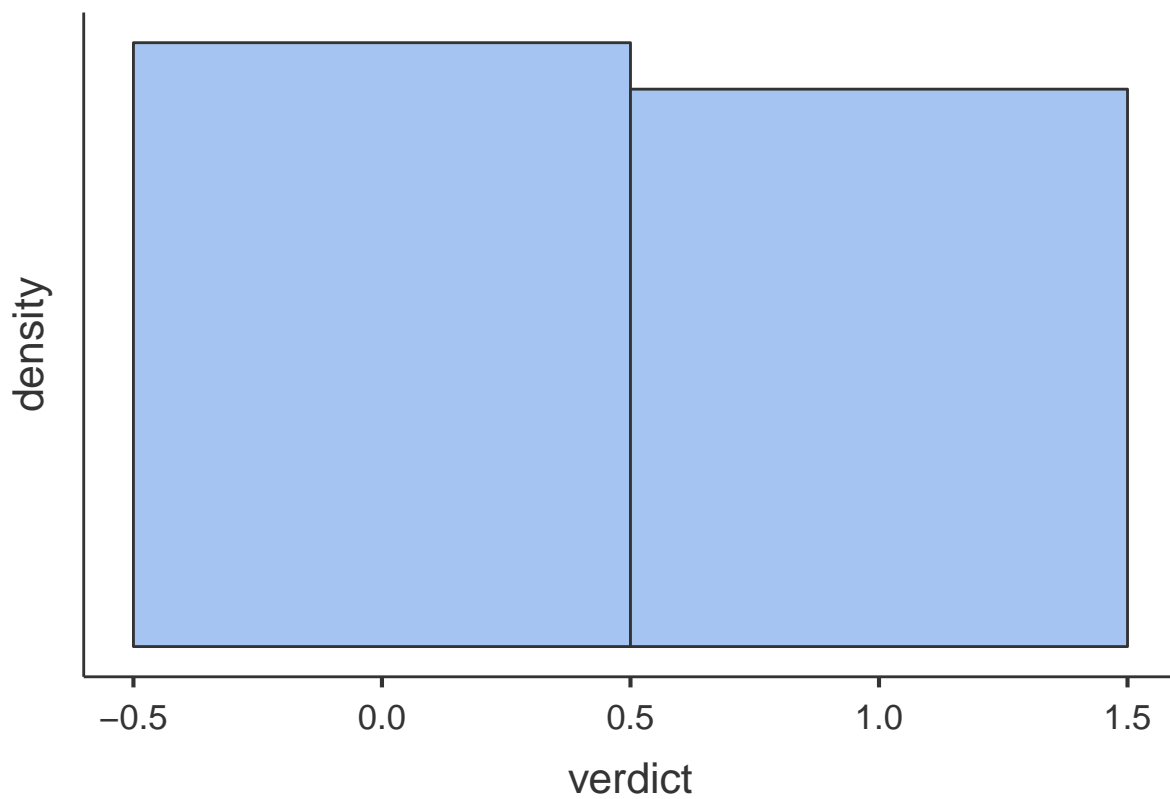
```
##
```

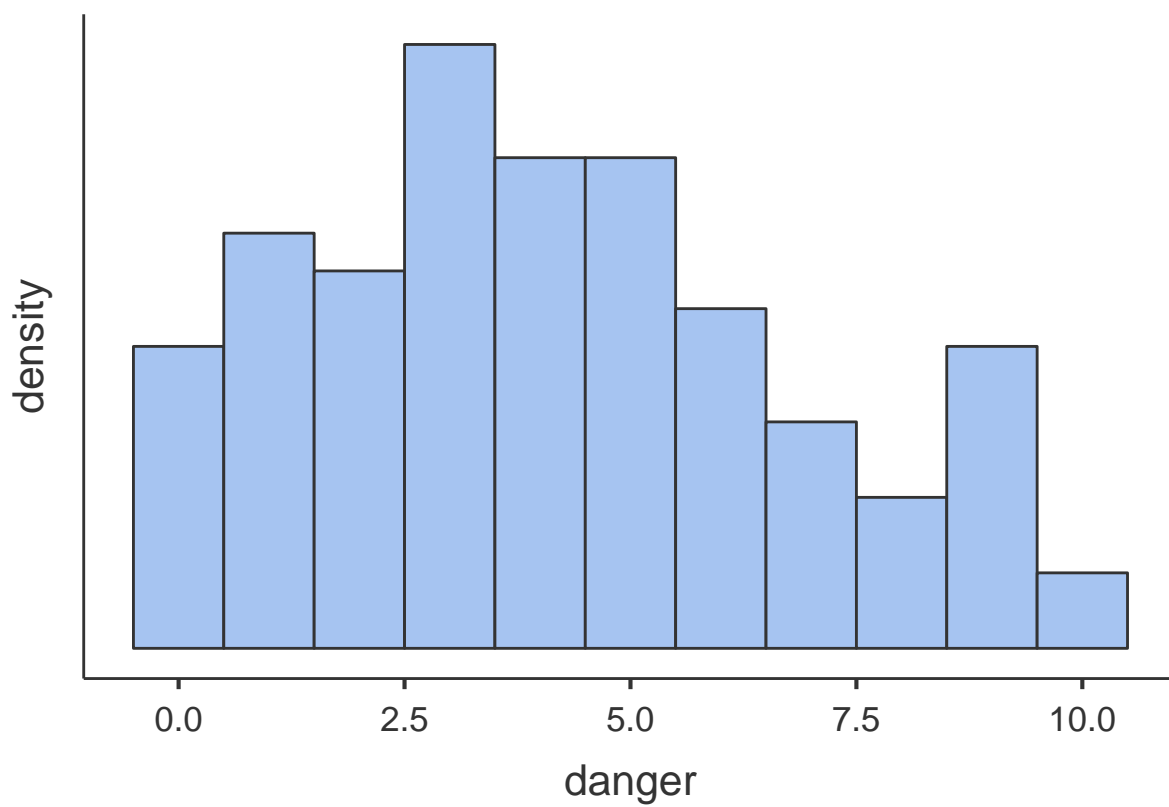
```
## Descriptives
```

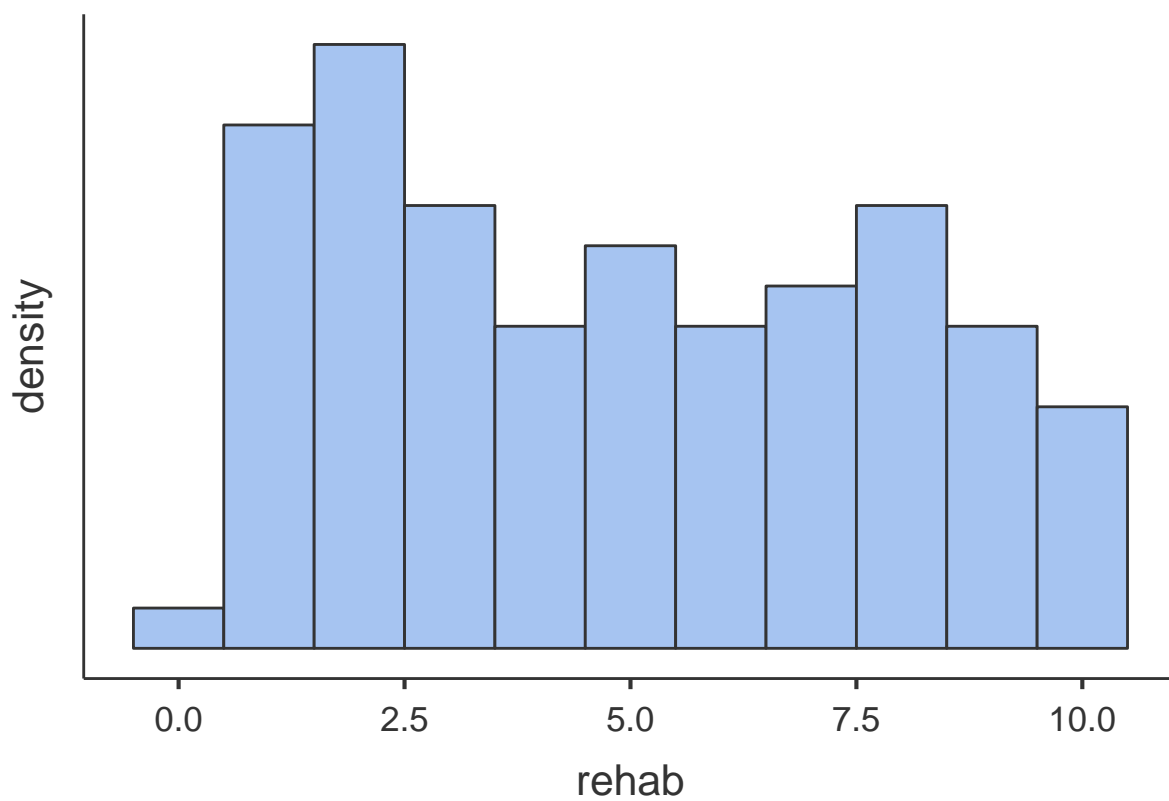
```
##
```

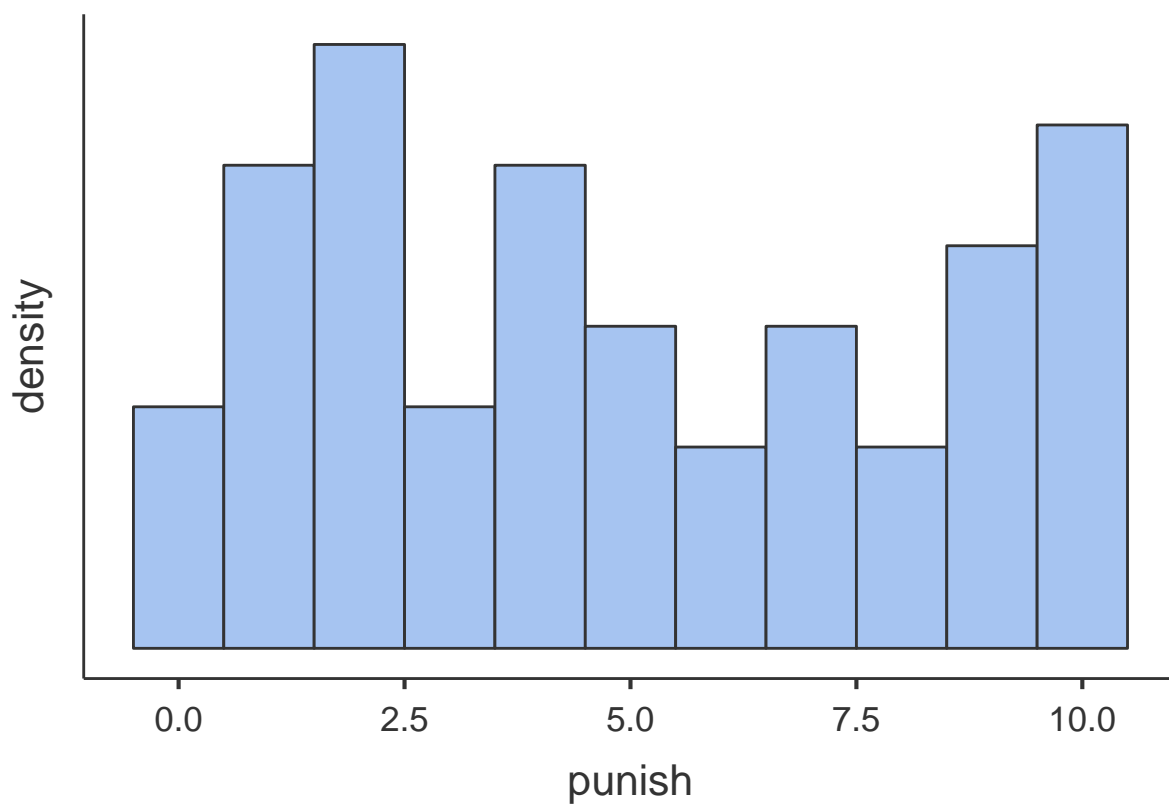
	verdict	danger	rehab	punish	gendet	specdet	incap
N	100	100	100	100	100	100	100
Missing	0	0	0	0	0	0	0
Mean	0.480	4.16	4.89	4.94	5.17	4.76	4.92
Median	0.00	4.00	5.00	4.00	5.00	4.00	5.00
Standard deviation	0.502	2.70	2.91	3.30	3.10	3.01	3.13
Minimum	0	0	0	0	0	0	0

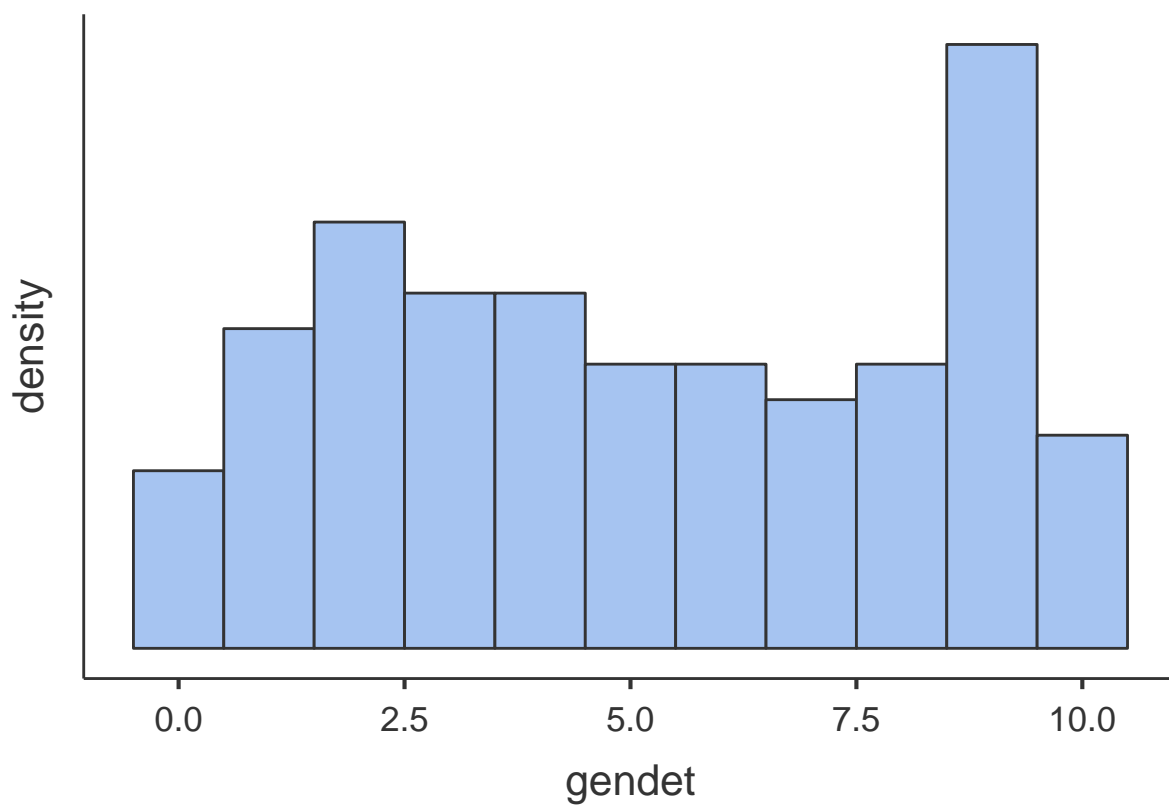
```
##      Maximum      1      10      10      10      10      10      10
##      Skewness    0.0813  0.360  0.191  0.195  0.00548  0.113  0.0856
##      Std. error skewness  0.241  0.241  0.241  0.241  0.241  0.241  0.241
##      Kurtosis    -2.03  -0.690 -1.25  -1.33  -1.33  -1.19  -1.03
##      Std. error kurtosis  0.478  0.478  0.478  0.478  0.478  0.478  0.478
## -----
##
##
## FREQUENCIES
##
## Frequencies of verdict
## -----
##      Levels      Counts      % of Total      Cumulative %
## -----
##      0           52         52.0          52.0
##      1           48         48.0          100.0
## -----
```

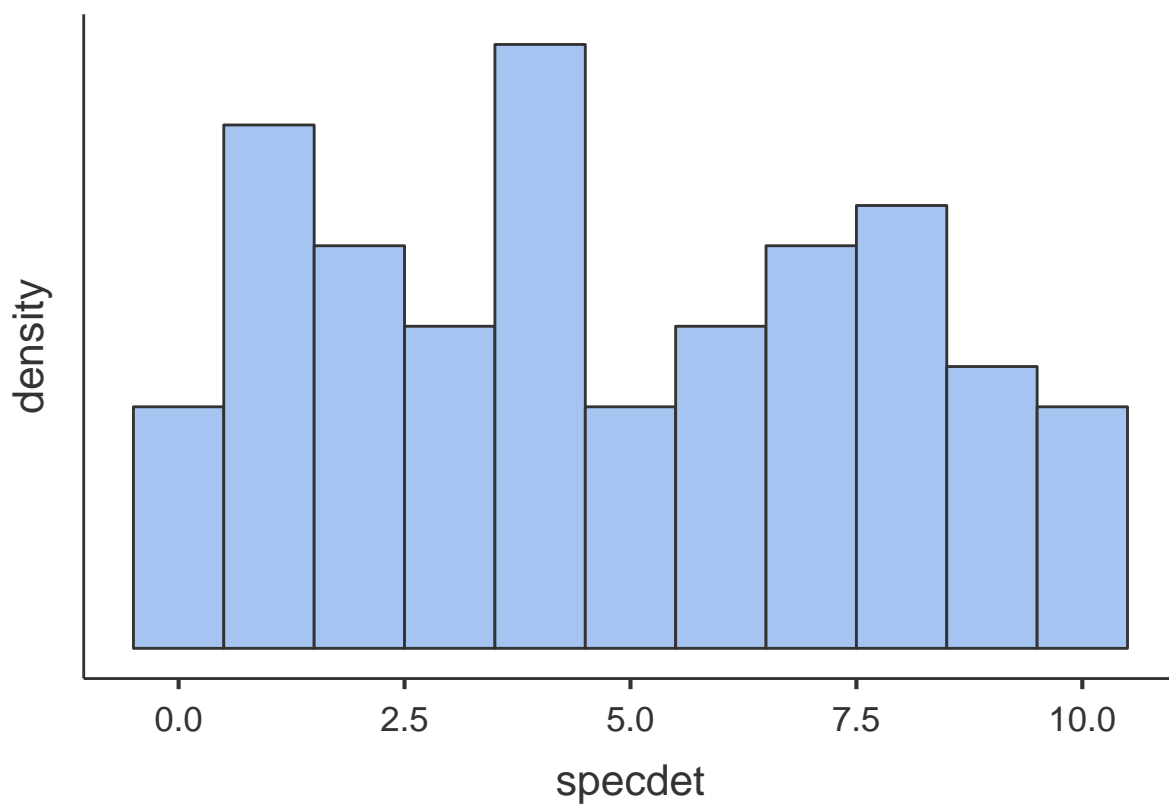


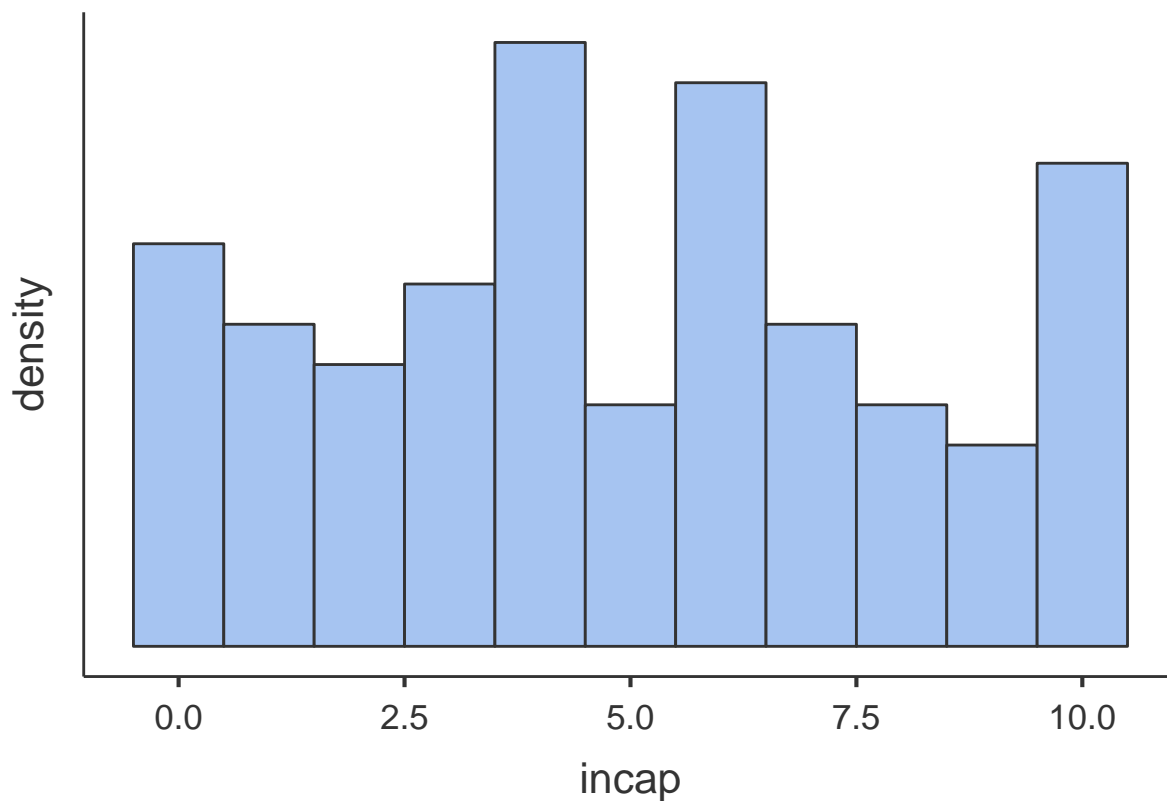












Assumptions 1. Independence of Observations 2. Predictor Variables Normally Distributed 3. Multicollinearity

### Correlations

```
# Correlation
cortable <- corrMatrix(data = BL,
  vars = c('verdict', 'danger', 'rehab', 'punish', 'gendet', 'specdet', 'incap'),
  flag = TRUE)
cortable
```

```
##
## CORRELATION MATRIX
##
## Correlation Matrix
## -----
```

		verdict	danger	rehab	punish	gendet	specdet	incap
-----								
verdict	Pearson's r		0.346	-0.226	0.042	0.232	0.057	0.070
	p-value		< .001	0.024	0.679	0.020	0.574	0.491
-----								
danger	Pearson's r			-0.087	-0.101	0.038	0.083	0.124
	p-value			0.392	0.317	0.709	0.411	0.221
-----								
rehab	Pearson's r				0.073	0.091	0.002	-0.014
	p-value				0.472	0.365	0.988	0.888
-----								
punish	Pearson's r					0.023	0.016	-0.107
	p-value							

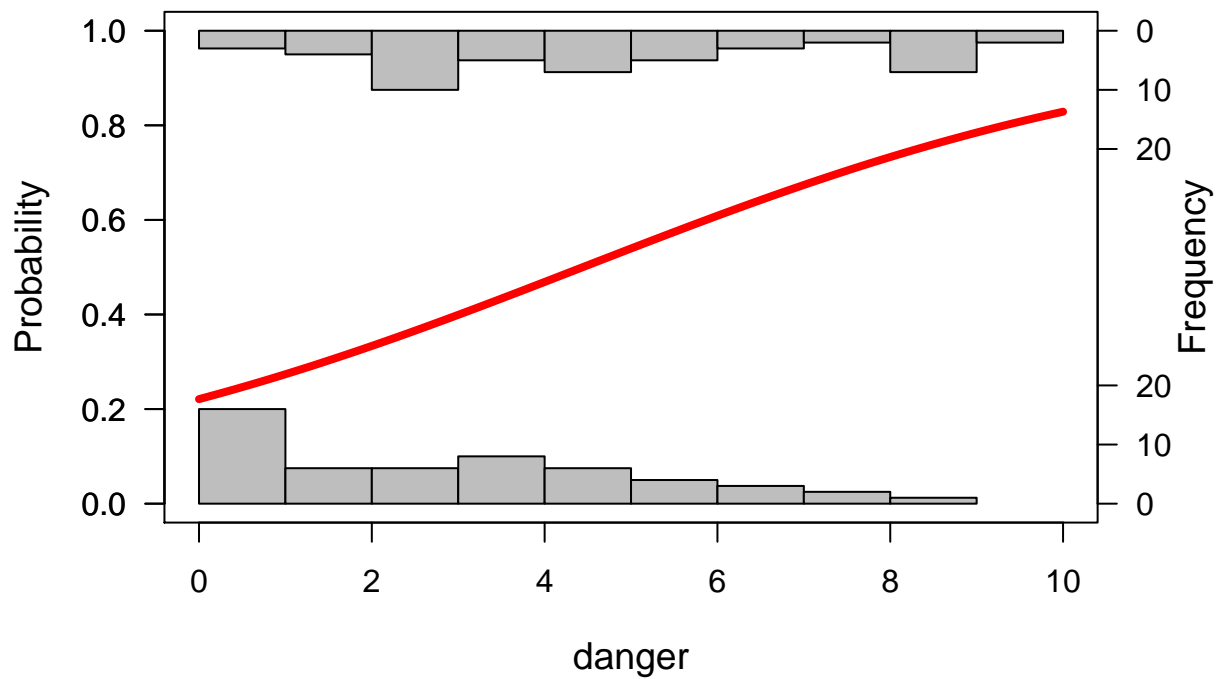
```
##
##
```



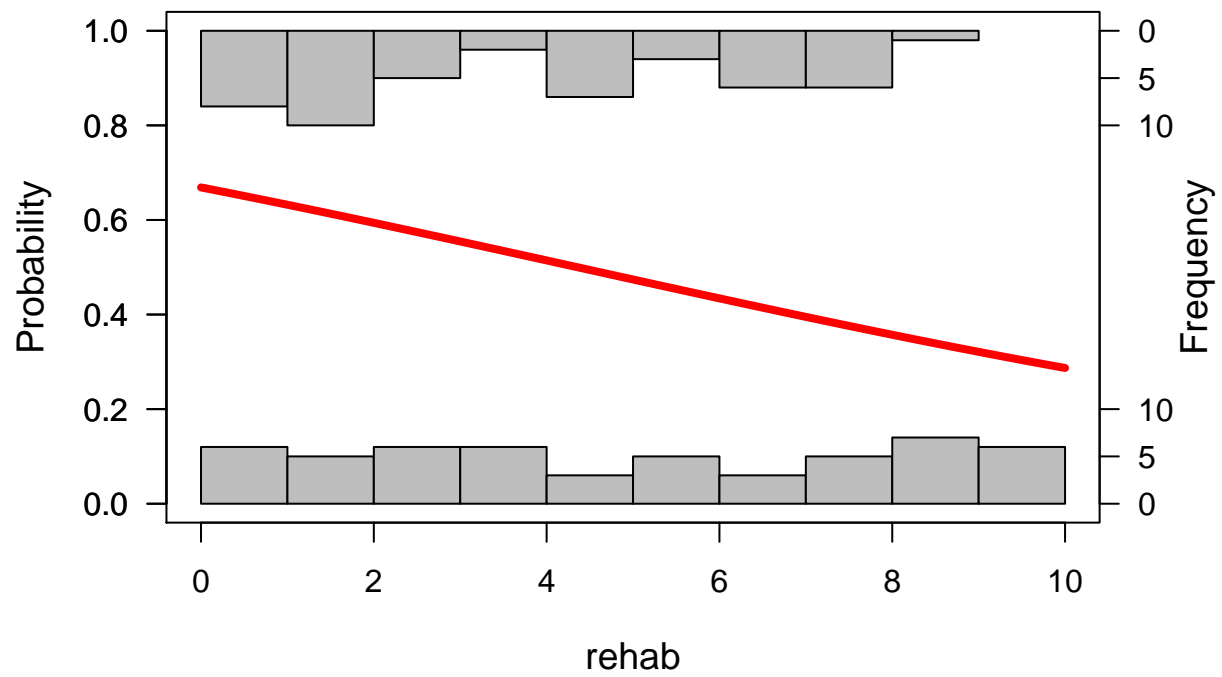
```
##          p-value          0.823      0.876      0.290
##
##   gendet   Pearson's r          0.094      0.112
##          p-value          0.352      0.269
##
##   specdet   Pearson's r          -0.064
##          p-value          0.526
##
##   incap     Pearson's r
##          p-value
## -----
##   Note. * p < .05, ** p < .01, *** p < .001
```

### Logistic Plots

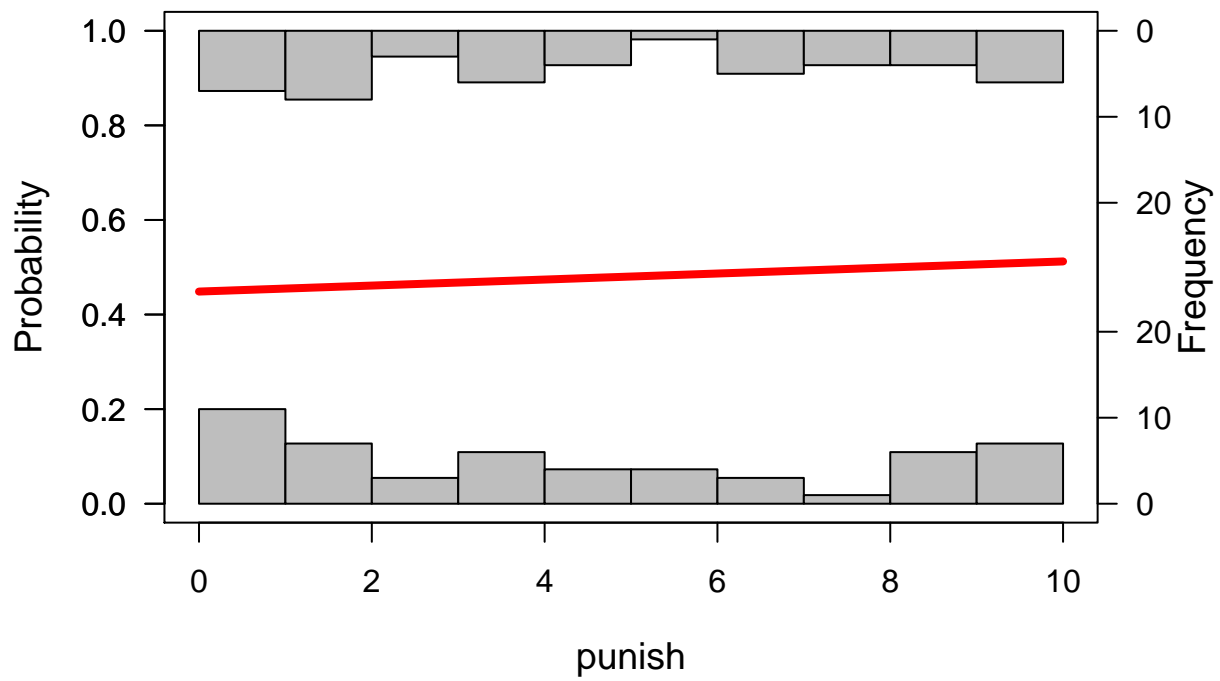
```
p_load(popbio)
logi.hist.plot(BL$danger, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "danger")
```



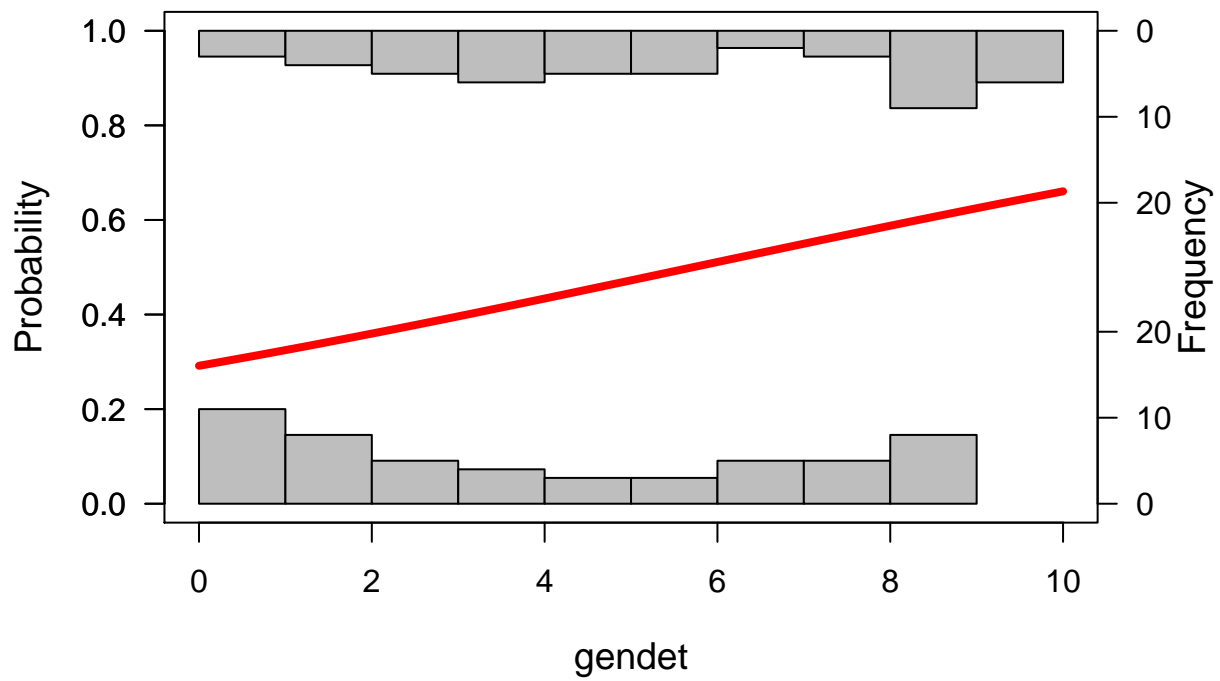
```
logi.hist.plot(BL$rehab, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "rehab")
```



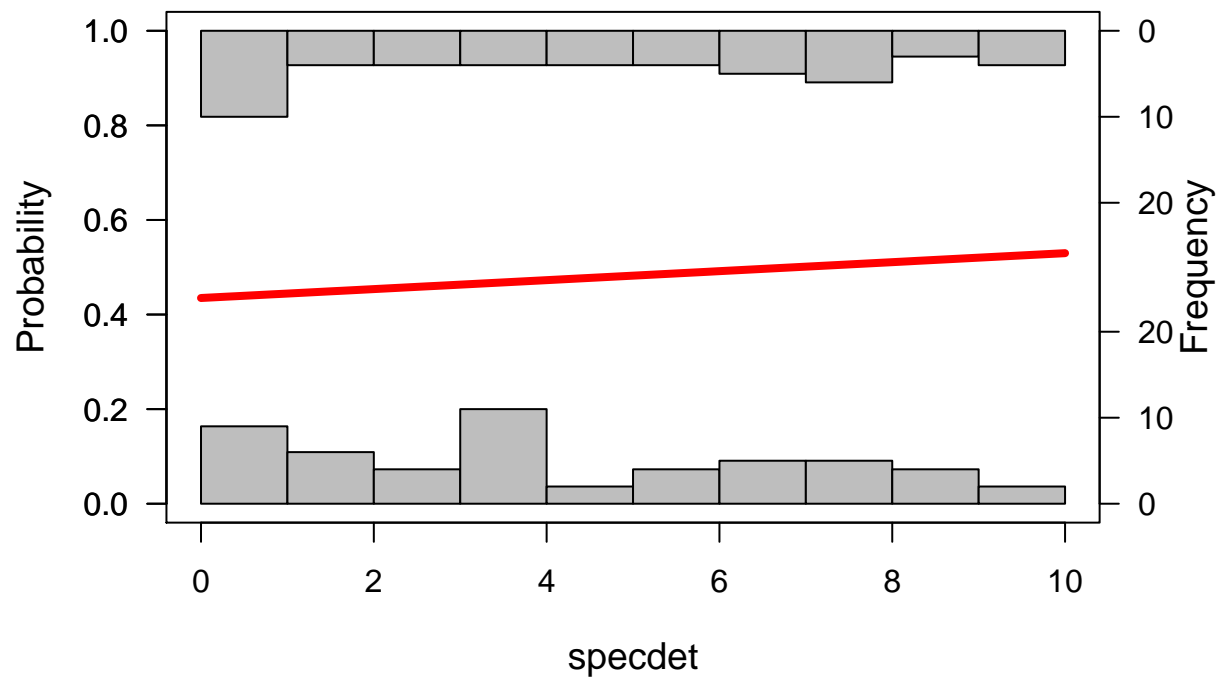
```
logi.hist.plot(BL$punish, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "punish")
```



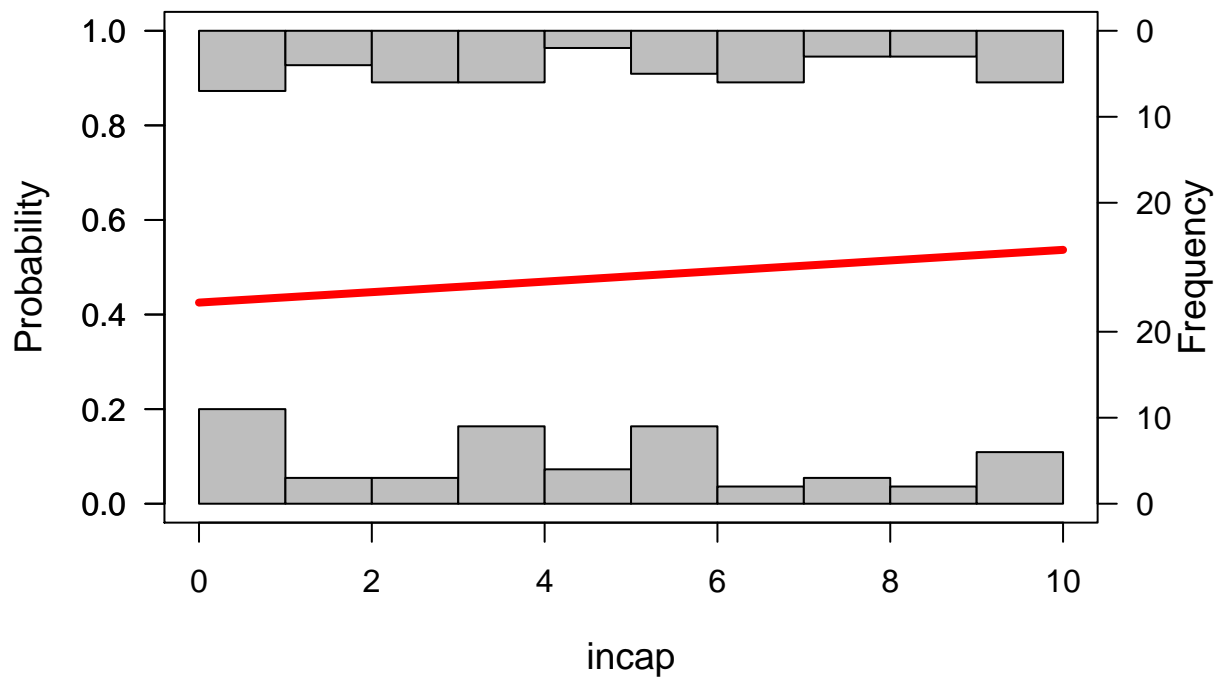
```
logi.hist.plot(BL$gendet, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "gendet")
```



```
logi.hist.plot(BL$specdet, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "specdet")
```



```
logi.hist.plot(BL$incap, BL$verdict, boxp=FALSE, type="hist", col="gray", xlabel = "incap")
```



### BiLoRe Models

```
# Null model
# Null deviance = Chi squared for the model
# df = N - (# of parameters)
model0 <- glm(BL$verdict ~ 1, family = binomial)
summary(model0)

##
## Call:
## glm(formula = BL$verdict ~ 1, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.144  -1.144  -1.144   1.212   1.212
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.08004    0.20016   -0.4    0.689
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 138.47  on 99  degrees of freedom
## Residual deviance: 138.47  on 99  degrees of freedom
## AIC: 140.47
##
## Number of Fisher Scoring iterations: 3
```

```

print("Logit")

## [1] "Logit"
coef(model0)

## (Intercept)
## -0.08004271
model0.odds <- exp(coef(model0)) #converts coefficient to odds [P(outcome)/(1-P(outcome))]
print("Odds")

## [1] "Odds"
model0.odds

## (Intercept)
## 0.9230769
model0.probs <- model0.odds / (1 + model0.odds) #
print("Probabilities")

## [1] "Probabilities"
model0.probs

## (Intercept)
## 0.48
print("Columns = Observed, Rows = Predicted")

## [1] "Columns = Observed, Rows = Predicted"
print("Null model")

## [1] "Null model"
ClassLog(model0, BL$verdict) # classification success under the null model (baseline)

## $rawtab
##      resp
##      0 1
## FALSE 52 48
##
## $classtab
##      resp
##      0 1
## FALSE 1 1
##
## $overall
## [1] 0.52
##
## $mcFadden
## [1] 2.220446e-16

# Model with predictors specific to the defendent.
# when odds ratio < 1 just flip (invert) the result(in relation to "no" instead of in relation to "yes".
# AIC is used to compare non-nested models for fit (lower means better fit)
# top chi-squared indicates the change of chi-squared vs the null model (Deviance + chi squared)

```

```

model1.jmv <- jmv::logRegBin(
  data = BL,
  dep = verdict,
  covs = vars(danger, specdet, incap, rehab, punish, gendet),
  blocks = list(
    list(
      'danger',
      'specdet',
      'incap')),
  refLevels = list(
    list(
      var = 'verdict',
      ref = '0')),
  modelTest = TRUE,
  OR = TRUE,
  class = TRUE,
  acc = TRUE,
  collin = TRUE)

```

model1.jmv

```

##
##  BINOMIAL LOGISTIC REGRESSION
##
##  Model Fit Measures
##  -----
##    Model    Deviance    AIC    R2-McF    <U+03C7>2    df    p
##  -----
##         1         126     134     0.0916     12.7     3    0.005
##  -----
##
##
##  MODEL SPECIFIC RESULTS
##
##  MODEL 1
##
##  Model Coefficients
##  -----
##    Predictor    Estimate    SE        Z        p        Odds ratio
##  -----
##    Intercept    -1.4592     0.6289    -2.320    0.020        0.232
##    danger        0.2786     0.0868     3.208    0.001        1.321
##    specdet       0.0228     0.0719     0.316    0.752        1.023
##    incap        0.0220     0.0696     0.316    0.752        1.022
##  -----
##    Note. Estimates represent the log odds of "verdict = 1" vs.
##    "verdict = 0"
##
##
##  ASSUMPTION CHECKS
##
##  Collinearity Statistics
##  -----
##              VIF      Tolerance

```



```
## -----
##    danger      1.01      0.986
##    specdet     1.01      0.988
##    incap       1.02      0.984
## -----
##
##
## PREDICTION
##
## Classification Table  verdict
## -----
##    Observed    0    1    % Correct
## -----
##           0    34    18        65.4
##           1    21    27        56.2
## -----
##    Note. The cut-off value is set
##    to 0.5
##
##
## Predictive Measures
## -----
##    Accuracy
## -----
##    0.610
## -----
##    Note. The
##    cut-off value
##    is set to 0.5
```

```
# Model with predictors about the criminal justice system.
```

```
#Multicollinearity
```

```
#Tolerance = 1 - R squared --> for our purpose < .4 is bad
```

```
#VIF = 1/Tolerance
```

```
# Small VIF values indicates low correlation among variables under ideal conditions
```

```
#Multicollinearity occurs when two or more predictors in the model are correlated and provide redundant
```

```
model2.jmv <- jmv::logRegBin(
  data = BL,
  dep = verdict,
  covs = vars(danger, specdet, incap, rehab, punish, gendet),
  blocks = list(
    list(
      'rehab',
      'punish',
      'gendet')),
  refLevels = list(
    list(
      var = 'verdict',
      ref = '0')),
  modelTest = TRUE,
  OR = TRUE,
```

```
class = TRUE,
acc = TRUE,
collin = TRUE)
```

model2.jmv

```
##
## BINOMIAL LOGISTIC REGRESSION
##
## Model Fit Measures
## -----
##      Model      Deviance      AIC      R2-McF      <U+03C7>2      df      p
## -----
##           1           126       134       0.0901       12.5       3       0.006
## -----
##
##
## MODEL SPECIFIC RESULTS
##
## MODEL 1
##
## Model Coefficients
## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----
##      Intercept      -0.2619      0.5875     -0.446     0.656       0.770
##      rehab          -0.1946      0.0770     -2.527     0.011       0.823
##      punish          0.0373      0.0650      0.575     0.566       1.038
##      gendet          0.1830      0.0720      2.540     0.011       1.201
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
## ASSUMPTION CHECKS
##
## Collinearity Statistics
## -----
##              VIF      Tolerance
## -----
##      rehab      1.06      0.944
##      punish      1.01      0.991
##      gendet      1.05      0.952
## -----
##
##
## PREDICTION
##
## Classification Table      verdict
## -----
##      Observed      0      1      % Correct
## -----
##              0      39      13      75.0
##              1      22      26      54.2
```

```
## -----
## Note. The cut-off value is set
## to 0.5
##
## Predictive Measures
## -----
## Accuracy
## -----
## 0.650
## -----
## Note. The
## cut-off value
## is set to 0.5
```

*# Model with all predictors*

```
model3.jmv <- jmv::logRegBin(
  data = BL,
  dep = verdict,
  covs = vars(danger, specdet, incap, rehab, punish, gendet),
  blocks = list(
    list(
      'danger',
      'specdet',
      'incap',
      'rehab',
      'punish',
      'gendet')),
  refLevels = list(
    list(
      var = 'verdict',
      ref = '0')),
  modelTest = TRUE,
  OR = TRUE,
  class = TRUE,
  acc = TRUE,
  collin = TRUE)
```

model3.jmv

```
##
## BINOMIAL LOGISTIC REGRESSION
##
## Model Fit Measures
## -----
## Model    Deviance    AIC    R2-McF    <U+03C7>2    df    p
## -----
## 1         114        128    0.176    24.4    6    < .001
## -----
##
##
## MODEL SPECIFIC RESULTS
##
## MODEL 1
```

```

##
## Model Coefficients
## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----
##      Intercept    -1.74758     0.9173   -1.9052   0.057     0.174
##      danger        0.29339     0.0929    3.1575   0.002     1.341
##      specdet       0.00590     0.0786    0.0751   0.940     1.006
##      incap         0.00353     0.0759    0.0465   0.963     1.004
##      rehab        -0.18784     0.0814   -2.3077   0.021     0.829
##      punish        0.07012     0.0711    0.9861   0.324     1.073
##      gendet        0.18574     0.0773    2.4019   0.016     1.204
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
## ASSUMPTION CHECKS
##
## Collinearity Statistics
## -----
##              VIF      Tolerance
## -----
##      danger      1.07      0.933
##      specdet     1.02      0.977
##      incap       1.05      0.955
##      rehab       1.04      0.957
##      punish      1.06      0.941
##      gendet      1.06      0.948
## -----
##
##
## PREDICTION
##
## Classification Table  verdict
## -----
##      Observed    0    1    % Correct
## -----
##              0    39    13      75.0
##              1    16    32      66.7
## -----
##      Note. The cut-off value is set
##      to 0.5
##
##
## Predictive Measures
## -----
##      Accuracy
## -----
##      0.710
## -----
##      Note. The
##      cut-off value
##      is set to 0.5

```

## Model Comparison

*# Model1 vs. Model3*

*# This is a similar set up to multiple regression in terms of the code, so if you know you will be running*

```
modelcomp.jmv <- jmv::logRegBin(  
  data = BL,  
  dep = verdict,  
  covs = vars(danger, specdet, incap, rehab, punish, gendet),  
  blocks = list(  
    list(  
      'danger',  
      'specdet',  
      'incap'),  
    list(  
      'rehab',  
      'punish',  
      'gendet')),  
  refLevels = list(  
    list(  
      var = 'verdict',  
      ref = '0')),  
  modelTest = TRUE,  
  OR = TRUE,  
  class = TRUE,  
  acc = TRUE,  
  collin = TRUE)
```

modelcomp.jmv

```
##  
##  BINOMIAL LOGISTIC REGRESSION  
##  
##  Model Fit Measures  
##  -----  
##    Model      Deviance    AIC    R2-McF    <U+03C7>2    df    p  
##  -----  
##        1          126     134    0.0916     12.7     3    0.005  
##        2          114     128    0.1763     24.4     6    < .001  
##  -----  
##  
##  
##  Model Comparisons  
##  -----  
##    Model      Model    <U+03C7>2    df    p  
##  -----  
##        1    -         2     11.7     3    0.008  
##  -----  
##  
##  
##  MODEL SPECIFIC RESULTS  
##  
##  MODEL 1  
##  
##  Model Coefficients
```

```

## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----
##      Intercept      -1.4592      0.6289     -2.320    0.020      0.232
##      danger          0.2786      0.0868      3.208    0.001      1.321
##      specdet         0.0228      0.0719      0.316    0.752      1.023
##      incap           0.0220      0.0696      0.316    0.752      1.022
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
## ASSUMPTION CHECKS
##
## Collinearity Statistics
## -----
##              VIF      Tolerance
## -----
##      danger      1.01      0.986
##      specdet     1.01      0.988
##      incap       1.02      0.984
## -----
##
##
## PREDICTION
##
## Classification Table  verdict
## -----
##      Observed      0      1      % Correct
## -----
##              0      34      18      65.4
##              1      21      27      56.2
## -----
##
##      Note. The cut-off value is set
##      to 0.5
##
##
## Predictive Measures
## -----
##      Accuracy
## -----
##      0.610
## -----
##
##      Note. The
##      cut-off value
##      is set to 0.5
##
##
## MODEL 2
##
## Model Coefficients
## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----

```

```
##      Intercept      -1.74758      0.9173      -1.9052      0.057      0.174
##      danger         0.29339      0.0929      3.1575      0.002      1.341
##      specdet        0.00590      0.0786      0.0751      0.940      1.006
##      incap          0.00353      0.0759      0.0465      0.963      1.004
##      rehab          -0.18784      0.0814      -2.3077      0.021      0.829
##      punish         0.07012      0.0711      0.9861      0.324      1.073
##      gendet         0.18574      0.0773      2.4019      0.016      1.204
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
##      ASSUMPTION CHECKS
##
##      Collinearity Statistics
##      -----
##              VIF      Tolerance
##      -----
##      danger      1.07      0.933
##      specdet     1.02      0.977
##      incap       1.05      0.955
##      rehab       1.04      0.957
##      punish      1.06      0.941
##      gendet      1.06      0.948
##      -----
##
##
##      PREDICTION
##
##      Classification Table  verdict
##      -----
##      Observed    0      1      % Correct
##      -----
##              0     39     13         75.0
##              1     16     32         66.7
##      -----
##      Note. The cut-off value is set
##      to 0.5
##
##
##      Predictive Measures
##      -----
##      Accuracy
##      -----
##      0.710
##      -----
##      Note. The
##      cut-off value
##      is set to 0.5
```

### Center Predictors

```
BL$dangerC <- BL$danger - mean(BL$danger)
BL$rehabC <- BL$rehab - mean(BL$rehab)
BL$punishC <- BL$punish - mean(BL$punish)
```

```
BL$gendetC <- BL$gendet - mean(BL$gendet)
BL$specdetC <- BL$specdet - mean(BL$specdet)
BL$incapC <- BL$incap - mean(BL$incap)
```

Re-run model with all predictors and run parsimonious model.

```
#remove not significant predictors

finalmodel.jmv <- logRegBin(
  data = BL,
  dep = verdict,
  covs = vars(dangerC, specdetC, incapC, rehabC, punishC, gendetC),
  blocks = list(
    list(
      'dangerC', # significant predictors only
      'rehabC',
      'gendetC'),
    list(
      'specdetC', # full model
      'incapC',
      'punishC')),
  refLevels = list(
    list(
      var = 'verdict',
      ref = '0')),
  modelTest = TRUE,
  OR = TRUE,
  class = TRUE,
  acc = TRUE,
  collin = TRUE)

finalmodel.jmv
```

```
##
##  BINOMIAL LOGISTIC REGRESSION
##
##  Model Fit Measures
##  -----
##    Model    Deviance    AIC    R2-McF    <U+03C7>2    df    p
##  -----
##      1         115      123    0.169     23.4     3    < .001
##      2         114      128    0.176     24.4     6    < .001
##  -----
##
##
##  Model Comparisons
##  -----
##    Model      Model    <U+03C7>2    df    p
##  -----
##      1    -      2    1.00     3    0.801
##  -----
##
##
##  MODEL SPECIFIC RESULTS
```



```

##
## MODEL 1
##
## Model Coefficients
## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----
##      Intercept      -0.0955      0.2263     -0.422    0.673      0.909
##      dangerC         0.2797      0.0888      3.151    0.002      1.323
##      rehabC          -0.1807      0.0806     -2.242    0.025      0.835
##      gendetC         0.1881      0.0765      2.460    0.014      1.207
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
## ASSUMPTION CHECKS
##
## Collinearity Statistics
## -----
##              VIF      Tolerance
## -----
##      dangerC      1.01      0.992
##      rehabC       1.03      0.972
##      gendetC      1.04      0.964
## -----
##
##
## PREDICTION
##
## Classification Table  verdict
## -----
##      Observed      0      1      % Correct
## -----
##              0      37      15      71.2
##              1      16      32      66.7
## -----
##      Note. The cut-off value is set
##      to 0.5
##
##
## Predictive Measures
## -----
##      Accuracy
## -----
##      0.690
## -----
##      Note. The
##      cut-off value
##      is set to 0.5
##
##
## MODEL 2
##

```

```

## Model Coefficients
## -----
##      Predictor      Estimate      SE      Z      p      Odds ratio
## -----
##      Intercept    -0.09352    0.2277    -0.4108    0.681      0.911
##      dangerC       0.29339    0.0929     3.1575    0.002      1.341
##      rehabC       -0.18784    0.0814    -2.3077    0.021      0.829
##      gendetC       0.18574    0.0773     2.4019    0.016      1.204
##      specdetC      0.00590    0.0786     0.0751    0.940      1.006
##      incapC       0.00353    0.0759     0.0465    0.963      1.004
##      punishC      0.07012    0.0711     0.9861    0.324      1.073
## -----
##      Note. Estimates represent the log odds of "verdict = 1" vs.
##      "verdict = 0"
##
##
## ASSUMPTION CHECKS
##
## Collinearity Statistics
## -----
##              VIF      Tolerance
## -----
##      dangerC      1.07      0.933
##      rehabC       1.04      0.957
##      gendetC      1.06      0.948
##      specdetC     1.02      0.977
##      incapC       1.05      0.955
##      punishC      1.06      0.941
## -----
##
##
## PREDICTION
##
## Classification Table  verdict
## -----
##      Observed      0      1      % Correct
## -----
##              0      39      13      75.0
##              1      16      32      66.7
## -----
##      Note. The cut-off value is set
##      to 0.5
##
##
## Predictive Measures
## -----
##      Accuracy
## -----
##      0.710
## -----
##      Note. The
##      cut-off value
##      is set to 0.5

```

Use regression equation to calculate predicted logit, odds, and probability

```
# Let D = danger, R = rehab, G = gendet
D = 10
R = 0
G = 10

predlogit <- -.096 + (.278*D) + (-.181*R) + (.188*G)
predodds <- exp(predlogit)
predprob <- predodds / (1 + predodds)

predlogit

## [1] 4.564
predodds

## [1] 95.96658
predprob

## [1] 0.9896872
```