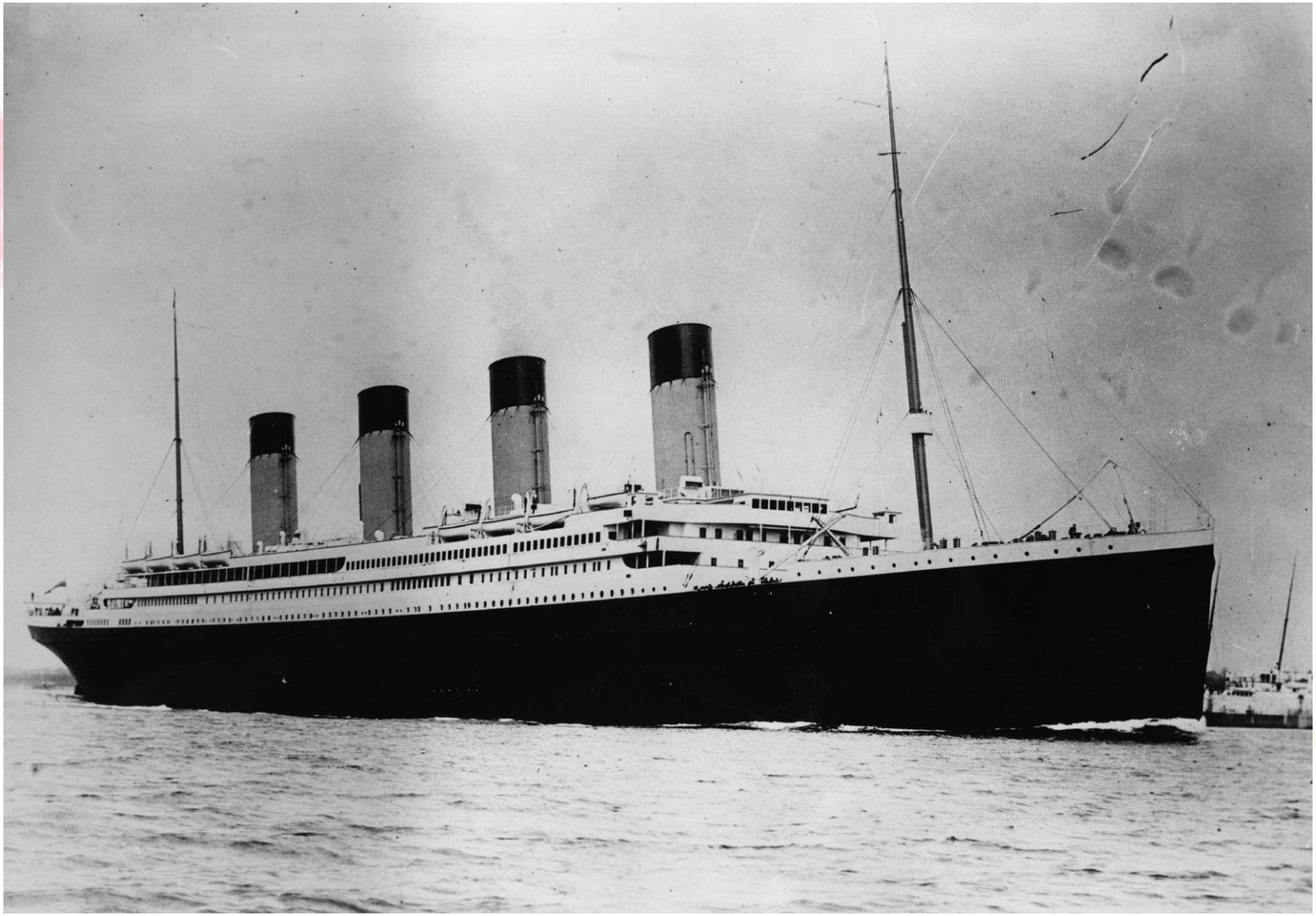
statgraphics®
centurion

# Would you have survived the sinking of the Titanic?

**Using the Statgraphics mosaic charts, multiple correspondence analysis and logistic regression procedures.**

Presented by Dr. Neil W. Polhemus

# RMS Titanic

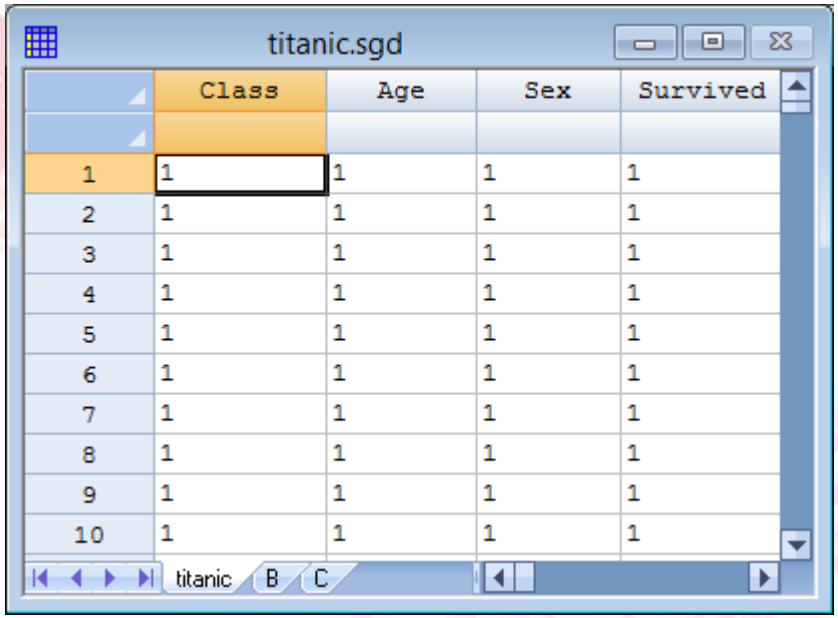


statgraphics®
statgraphics.com centurion

# Sinking of the Titanic

- Struck an iceberg on April 14, 1912 and sank the next morning.
- Carried over 2000 passengers and crew, many of whom did not survive.
- Controversies have centered on the lack of sufficient lifeboats and the differential treatment of individuals based on class of accommodation.

# Sample Data File #1

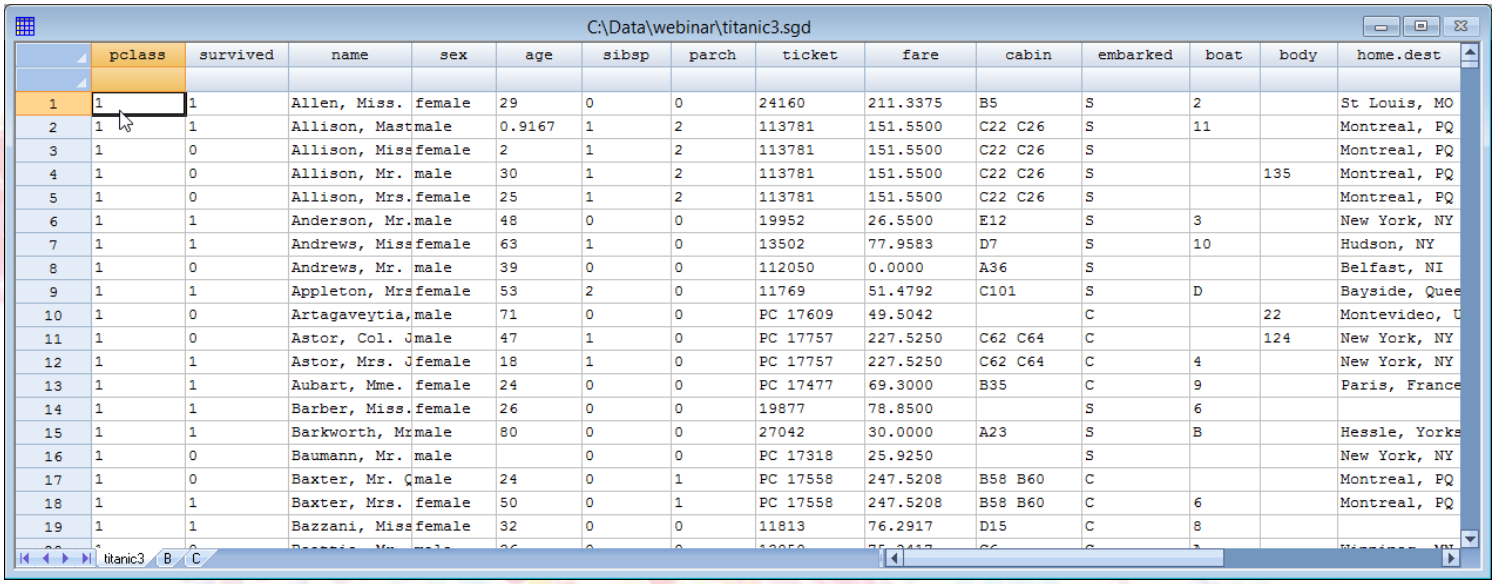



VARIABLE DESCRIPTIONS:
Class (1 = first, 2 = second, 3 = third, 0 = crew)
Age (1 = adult, 0 = child)
Sex (1 = male, 0 = female)
Survived (1 = yes, 0 = no)

n=2,201 observations (all people at risk)

Source: Robert J. MacG. Dawson, Saint Mary's University
http://www.amstat.org/publications/jse/datasets/titanic.txt

# Sample Data File #2



C:\Data\webinar\titanic3.sgd

| | pclass | survived | name | sex | age | sibsp | parch | ticket | fare | cabin | embarked | boat | body | home.dest |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | Allen, Miss. | female | 29 | 0 | 0 | 24160 | 211.3375 | B5 | S | 2 | | St Louis, MO |
| 2 | 1 | 1 | Allison, Mast | male | 0.9167 | 1 | 2 | 113781 | 151.5500 | C22 C26 | S | 11 | | Montreal, PQ |
| 3 | 1 | 0 | Allison, Miss | female | 2 | 1 | 2 | 113781 | 151.5500 | C22 C26 | S | | | Montreal, PQ |
| 4 | 1 | 0 | Allison, Mr. | male | 30 | 1 | 2 | 113781 | 151.5500 | C22 C26 | S | | 135 | Montreal, PQ |
| 5 | 1 | 0 | Allison, Mrs. | female | 25 | 1 | 2 | 113781 | 151.5500 | C22 C26 | S | | | Montreal, PQ |
| 6 | 1 | 1 | Anderson, Mr. | male | 48 | 0 | 0 | 19952 | 26.5500 | E12 | S | 3 | | New York, NY |
| 7 | 1 | 1 | Andrews, Miss | female | 63 | 1 | 0 | 13502 | 77.9583 | D7 | S | 10 | | Hudson, NY |
| 8 | 1 | 0 | Andrews, Mr. | male | 39 | 0 | 0 | 112050 | 0.0000 | A36 | S | | | Belfast, NI |
| 9 | 1 | 1 | Appleton, Mrs | female | 53 | 2 | 0 | 11769 | 51.4792 | C101 | S | D | | Bayside, Quee |
| 10 | 1 | 0 | Artagaveytia, | male | 71 | 0 | 0 | PC 17609 | 49.5042 | | C | | 22 | Montevideo, U |
| 11 | 1 | 0 | Astor, Col. J | male | 47 | 1 | 0 | PC 17757 | 227.5250 | C62 C64 | C | | 124 | New York, NY |
| 12 | 1 | 1 | Astor, Mrs. J | female | 18 | 1 | 0 | PC 17757 | 227.5250 | C62 C64 | C | 4 | | New York, NY |
| 13 | 1 | 1 | Aubart, Mme. | female | 24 | 0 | 0 | PC 17477 | 69.3000 | B35 | C | 9 | | Paris, France |
| 14 | 1 | 1 | Barber, Miss. | female | 26 | 0 | 0 | 19877 | 78.8500 | | S | 6 | | |
| 15 | 1 | 1 | Barkworth, Mr | male | 80 | 0 | 0 | 27042 | 30.0000 | A23 | S | B | | Hessle, Yorks |
| 16 | 1 | 0 | Baumann, Mr. | male | | 0 | 0 | PC 17318 | 25.9250 | | S | | | New York, NY |
| 17 | 1 | 0 | Baxter, Mr. Q | male | 24 | 0 | 1 | PC 17558 | 247.5208 | B58 B60 | C | | | Montreal, PQ |
| 18 | 1 | 1 | Baxter, Mrs. | female | 50 | 0 | 1 | PC 17558 | 247.5208 | B58 B60 | C | 6 | | Montreal, PQ |
| 19 | 1 | 1 | Bazzani, Miss | female | 32 | 0 | 0 | 11813 | 76.2917 | D15 | C | 8 | | |

titanic3 B C

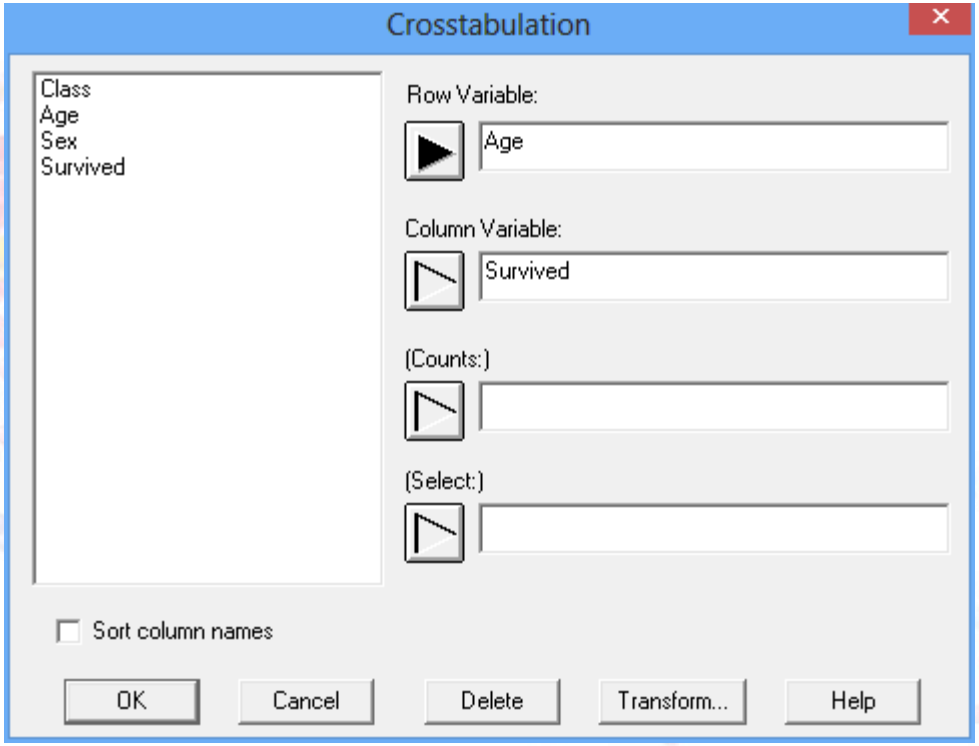
n=1,309 observations (passengers only)

Note that age is numeric.

Source: Frank Harrell and Thomas Cason, University of Virginia
http://biostat.mc.vanderbilt.edu/twiki/pub/Main/DataSets/titanic.html

statgraphics® centurion
statgraphics.com

# Crosstabulation



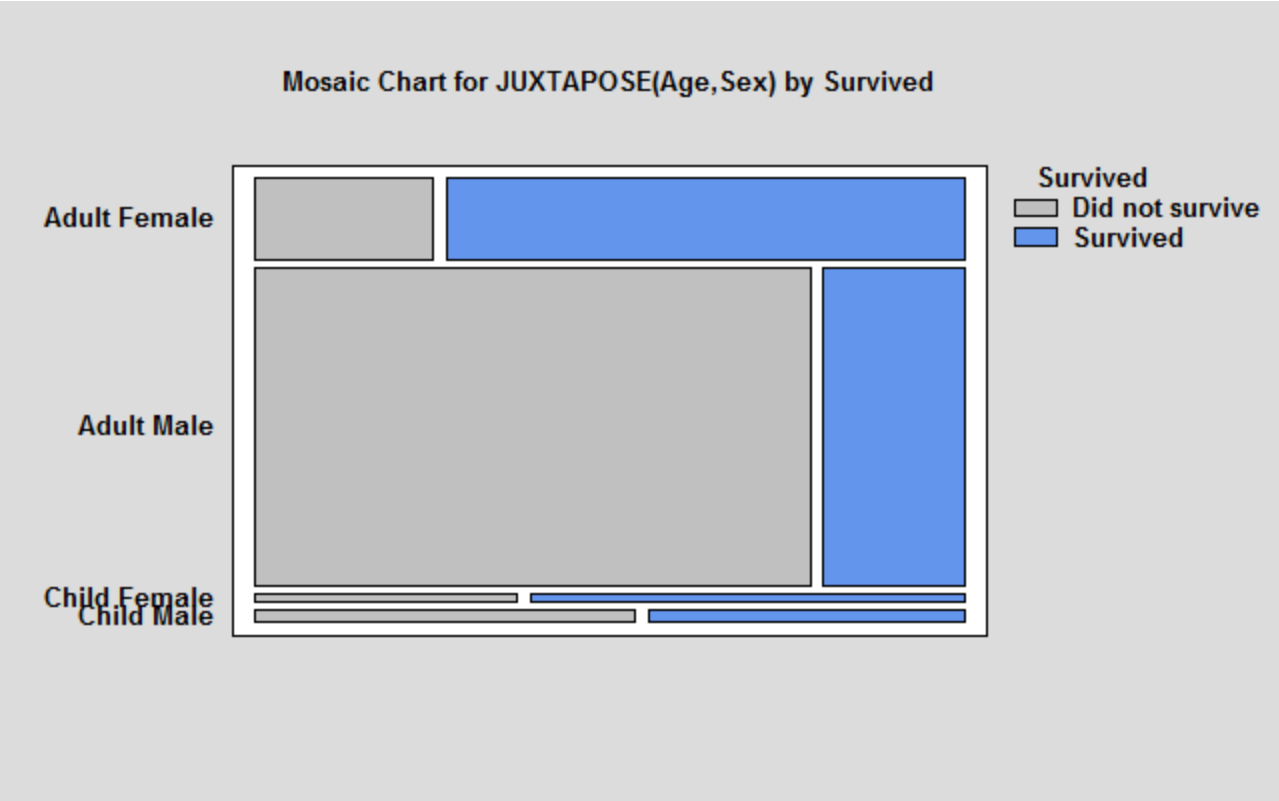
# Mosaic Charts

# Mosaic Charts


statgraphics®
statgraphics.com centurion

# Mosaic Charts

# Mosaic Charts



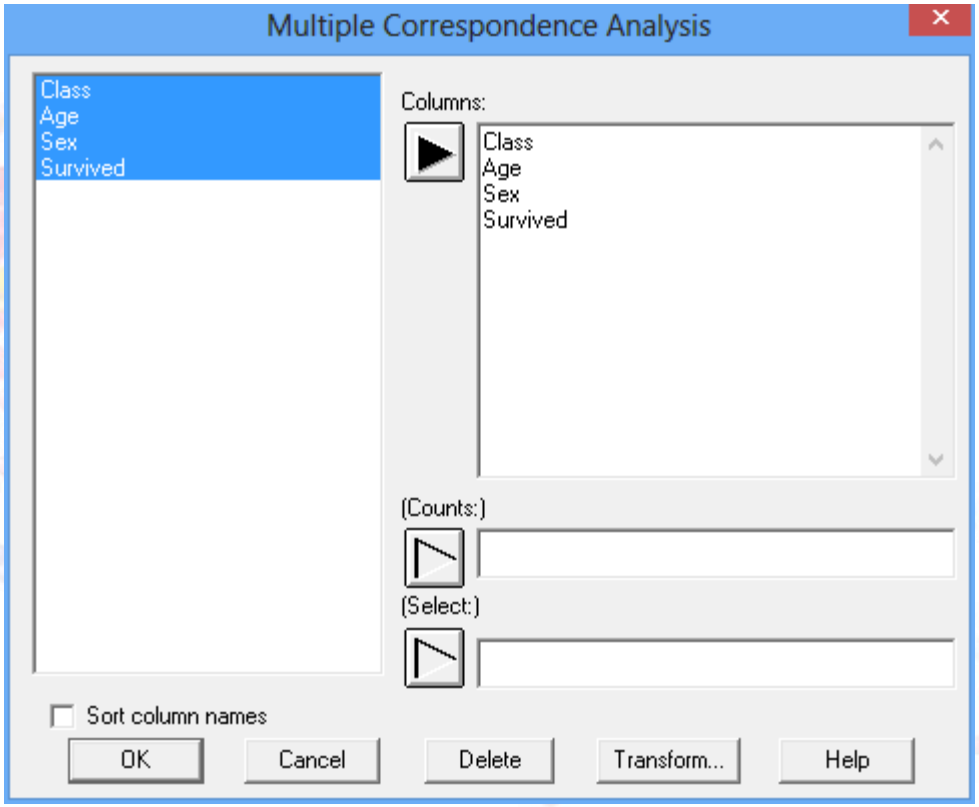
# Multiple Correspondence Analysis

- Creates a map of the associations amongst the categories of 2 or more variables.

- Seeks a small number of dimensions that describe most of the variability or “inertia” amongst the categories.

- Very useful in demonstrating the interrelationships amongst the variables.

# Data Input




statgraphics®
statgraphics.com
centurion

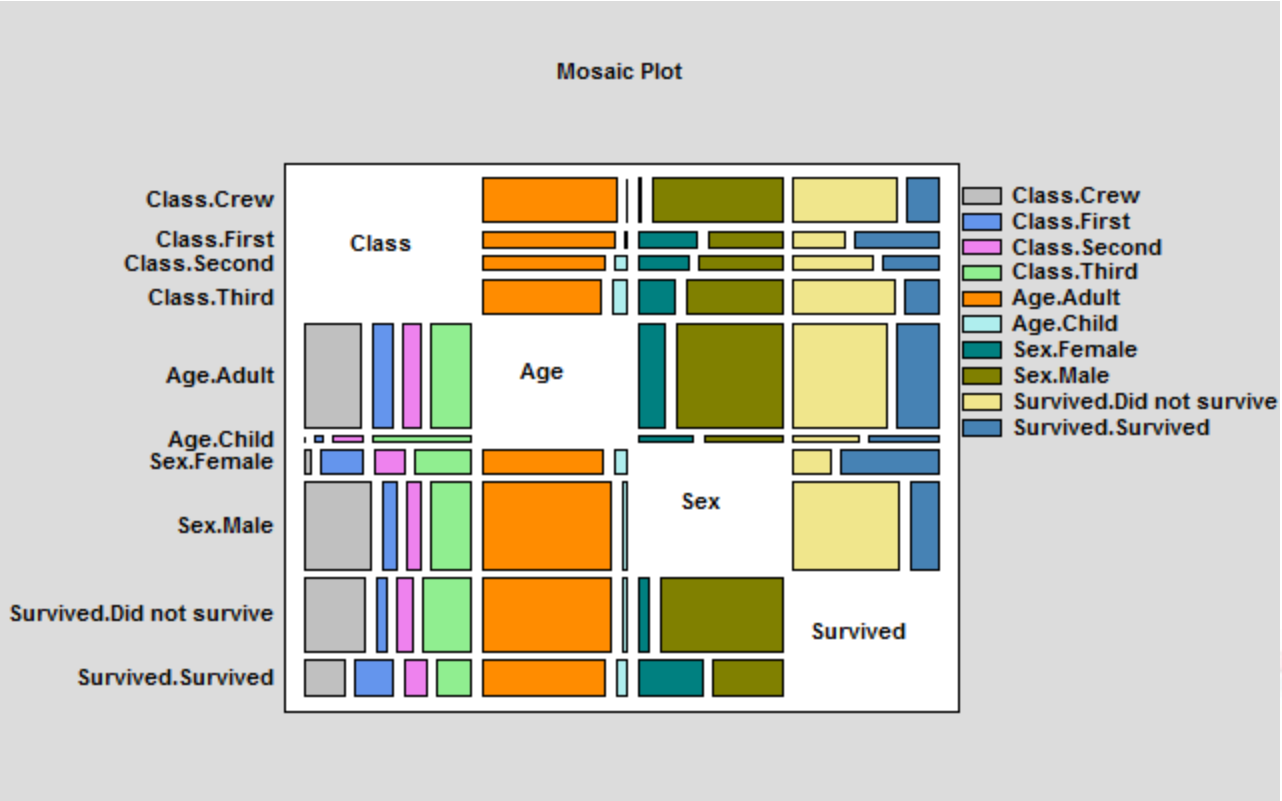
# Burt Table

Burt Table

| | *Class.Crew* | *Class.First* | *Class.Second* | *Class.Third* | *Age.Adult* | *Age.Child* |
|---|---|---|---|---|---|---|
| Class.Crew | 885 | 0 | 0 | 0 | 885 | 0 |
| Class.First | 0 | 325 | 0 | 0 | 319 | 6 |
| Class.Second | 0 | 0 | 285 | 0 | 261 | 24 |
| Class.Third | 0 | 0 | 0 | 706 | 627 | 79 |
| Age.Adult | 885 | 319 | 261 | 627 | 2092 | 0 |
| Age.Child | 0 | 6 | 24 | 79 | 0 | 109 |
| Sex.Female | 23 | 145 | 106 | 196 | 425 | 45 |
| Sex.Male | 862 | 180 | 179 | 510 | 1667 | 64 |
| Survived.Did not survive | 673 | 122 | 167 | 528 | 1438 | 52 |
| Survived.Survived | 212 | 203 | 118 | 178 | 654 | 57 |

| | *Sex.Female* | *Sex.Male* | *Survived.Did not survive* | *Survived.Survived* |
|---|---|---|---|---|
| Class.Crew | 23 | 862 | 673 | 212 |
| Class.First | 145 | 180 | 122 | 203 |
| Class.Second | 106 | 179 | 167 | 118 |
| Class.Third | 196 | 510 | 528 | 178 |
| Age.Adult | 425 | 1667 | 1438 | 654 |
| Age.Child | 45 | 64 | 52 | 57 |
| Sex.Female | 470 | 0 | 126 | 344 |
| Sex.Male | 0 | 1731 | 1364 | 367 |
| Survived.Did not survive | 126 | 1364 | 1490 | 0 |
| Survived.Survived | 344 | 367 | 0 | 711 |

# Multiple Mosaic Plot

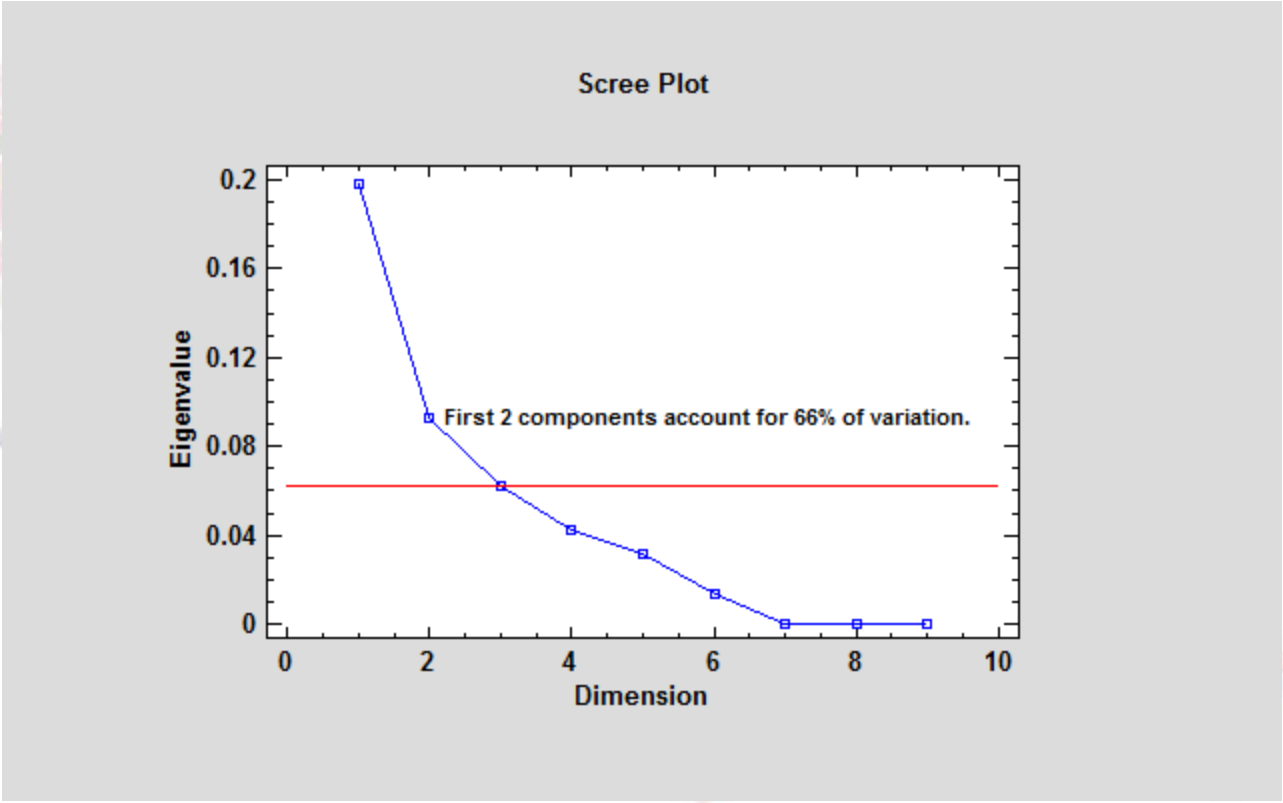


# Inertia and Chi-Square Decomposition

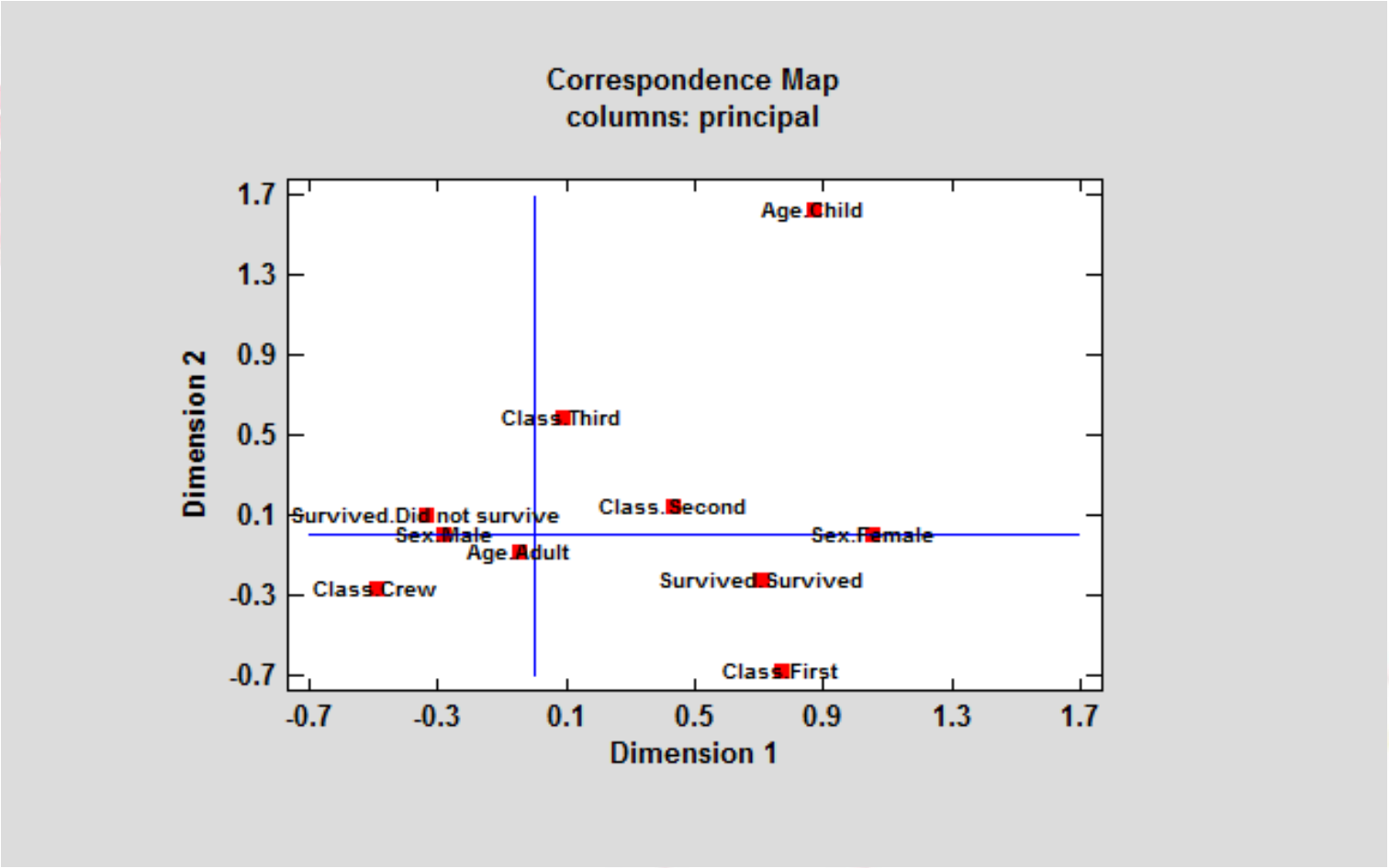
**Inertia and Chi-Square Decomposition**

| | *Singular* | | *Chi-* | | *Cumulative* | |
|---|---|---|---|---|---|---|
| *Dimension* | *Value* | *Inertia* | *Square* | *Percentage* | *Percentage* | *Histogram* |
| 1 | 0.4451 | 0.1981 | 6976.1395 | 44.9107 | 44.9107 | ********** |
| 2 | 0.3050 | 0.0931 | 3276.9079 | 21.0959 | 66.0066 | ******* |
| 3 | 0.2500 | 0.0625 | 2201.1057 | 14.1702 | 80.1768 | ***** |
| 4 | 0.2050 | 0.0420 | 1480.4911 | 9.5310 | 89.7078 | *** |
| 5 | 0.1785 | 0.0319 | 1122.2516 | 7.2248 | 96.9326 | *** |
| 6 | 0.1163 | 0.0135 | 476.4708 | 3.0674 | 100.0000 | * |
| TOTAL | | 0.4411 | 15533.366 | | | |

# Scree Plot



# Correspondence Map




statgraphics®
statgraphics.com centurion

# Logistic Regression

- Binary logistic regression is designed to model situations in which there are only 2 possible outcomes, such as “survived” or “did not survive”.

- It estimates a model that predicts the probability of those outcomes as a function of 1 or more predictor variables.

- The predictor variables may be either quantitative or categorical.

# Bernoulli Events

- A Bernoulli event is an event in which 2 outcomes are possible.
- Let $Y_i = 1$ if i$^{th}$ person survived, 0 otherwise.
- The pmf for a Bernoulli event is

$$Prob(y) = p^y(1-p)^{1-y} \text{ where } 0 \leq p \leq 1$$

# Predictor Variables

- We now want to let *p* be a function of one or more predictor variables *X*. In particular:
  - $X_1$ = class
  - $X_2$ = age
  - $X_3$ = sex

Then:

$$p = f(X_1, X_2, X_3)$$

# Logistic Model

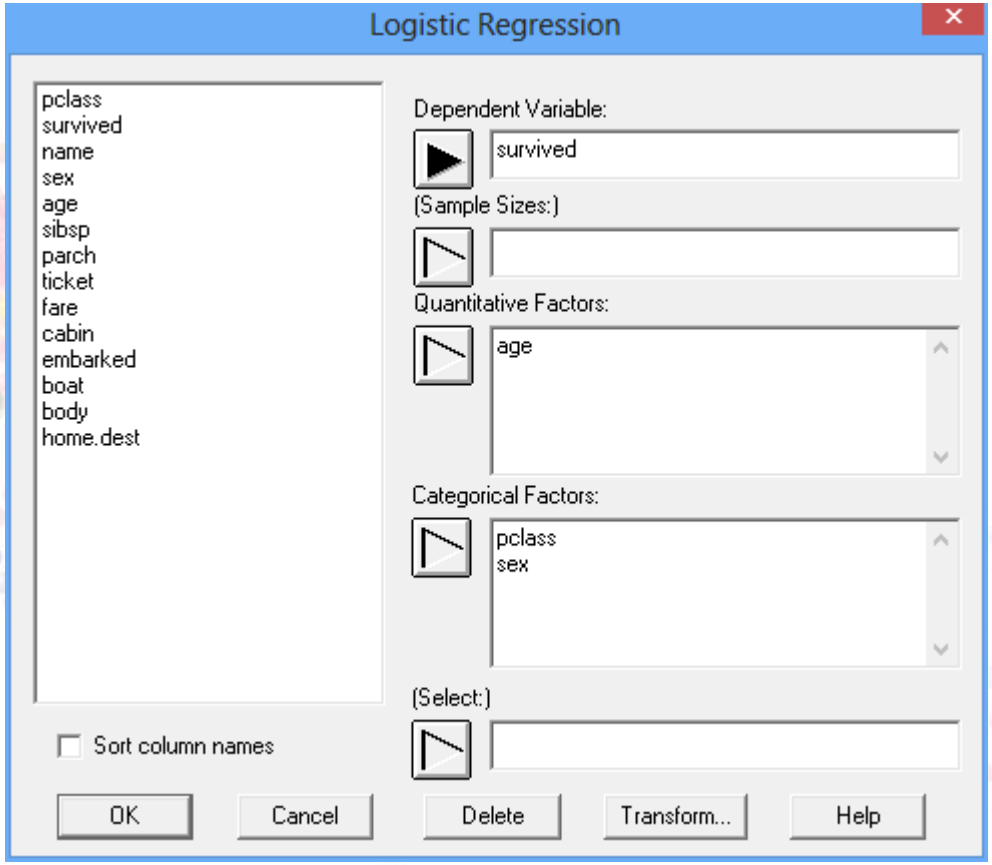
- When modeling *p,* we need a model that remains between 0 and 1.

- Most common choice is the logistic model:

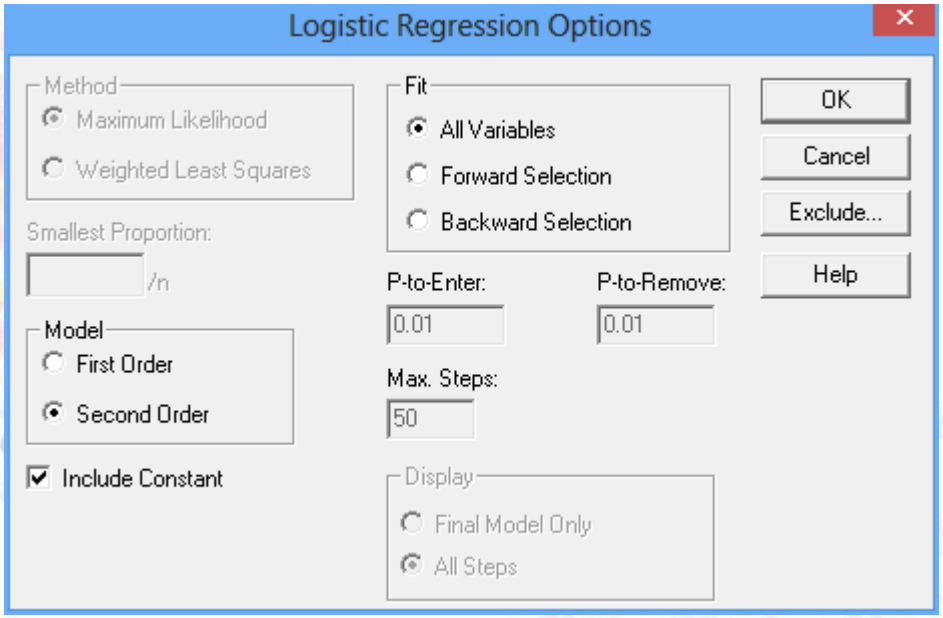
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$$

- It is a model for the log odds. If less than 0, odds are against the event. If greater than 0, odds are in favor of the event.

# Data Input



# Analysis Options




statgraphics®
centurion
statgraphics.com

# Likelihood Ratio Test

- Full second order model

**Likelihood Ratio Tests**

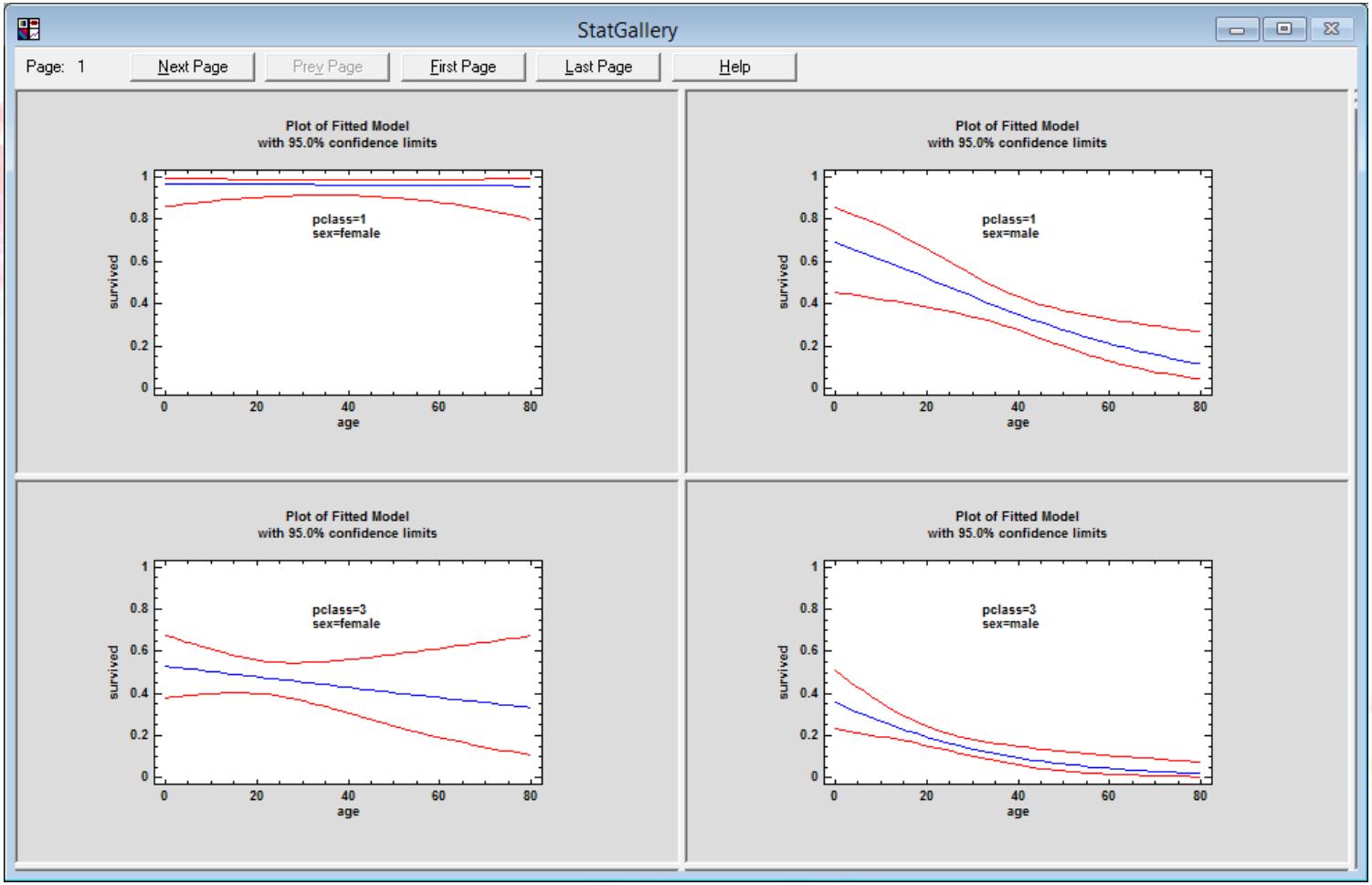
| *Factor* | *Chi-Square* | *Df* | *P-Value* |
|---|---|---|---|
| age | 9.7946 | 1 | 0.0017 |
| pclass | 11.6284 | 2 | 0.0030 |
| sex | 3.36405 | 1 | 0.0666 |
| age^2 | 1.91558 | 1 | 0.1663 |
| age*pclass | 9.37229 | 2 | 0.0092 |
| age*sex | 3.89 | 1 | 0.0486 |
| pclass*sex | 38.6254 | 2 | 0.0000 |

- Simplified model

**Likelihood Ratio Tests**

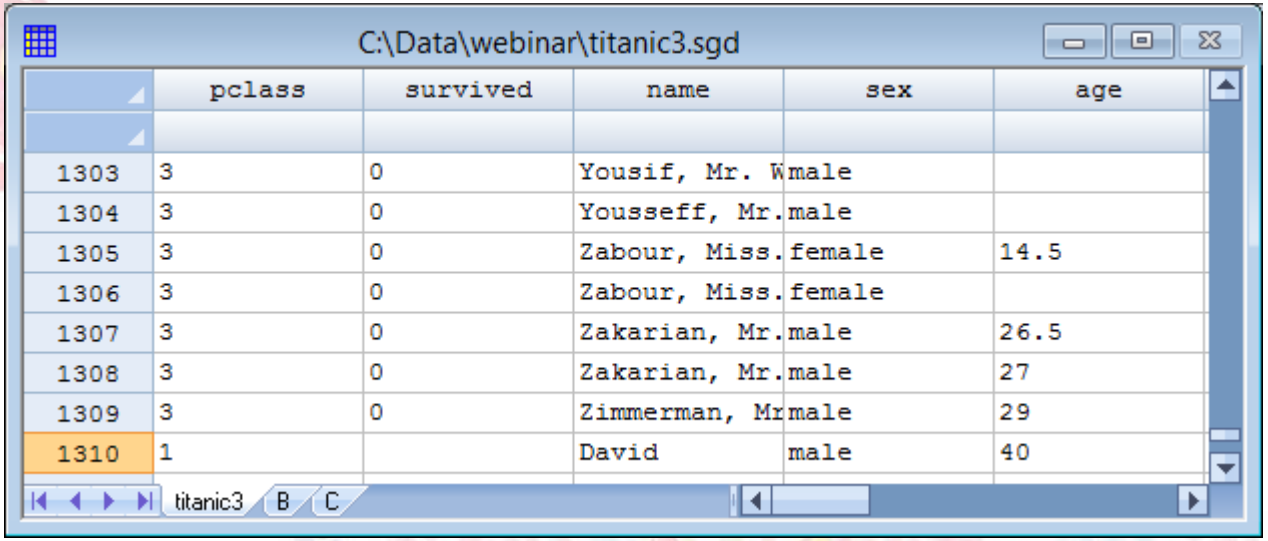
| *Factor* | *Chi-Square* | *Df* | *P-Value* |
|---|---|---|---|
| age | 12.208 | 1 | 0.0005 |
| pclass | 9.71294 | 2 | 0.0078 |
| sex | 3.10055 | 1 | 0.0783 |
| age*pclass | 10.0606 | 2 | 0.0065 |
| age*sex | 4.33079 | 1 | 0.0374 |
| pclass*sex | 38.1605 | 2 | 0.0000 |

# Plot of Fitted Model




statgraphics® centurion
statgraphics.com

# Predictions




| | pclass | survived | name | sex | age |
|---|---|---|---|---|---|
| 1303 | 3 | 0 | Yousif, Mr. W | male | |
| 1304 | 3 | 0 | Yousseff, Mr. | male | |
| 1305 | 3 | 0 | Zabour, Miss. | female | 14.5 |
| 1306 | 3 | 0 | Zabour, Miss. | female | |
| 1307 | 3 | 0 | Zakarian, Mr. | male | 26.5 |
| 1308 | 3 | 0 | Zakarian, Mr. | male | 27 |
| 1309 | 3 | 0 | Zimmerman, Mr | male | 29 |
| 1310 | 1 | | David | male | 40 |

Predictions for survived

| | *name* | *Observed* | *Fitted* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|
| *Row* | | *Value* | *Value* | *Conf. Limit* | *Conf. Limit* |
| 1310 | David | | 0.350919 | 0.276982 | 0.432779 |

statgraphics® centurion

statgraphics.com

# Recommended Reading

- Correspondence Analysis in Practice (second edition) by Michael Greenacre (Chapman and Hall, 2007)

- Applied Logistic Regression (third edition) by David Hosmer and Stanley Lemeshow (Wiley, 2013)

# Recorded Webinar

- You may find the recorded webinar, PowerPoint slides and sample data at:

  www.statgraphics.com

- Look for “Instructional Videos”.

statgraphics® centurion

statgraphics.com