# Week 1 - Introduction to GLMs

## Introduction to GLMs

1. During the Severe Acute Respiratory Syndrome (SARS) outbreak of 2003, some researchers believed that the treatment Ribavirin may be helpful in preventing death due to SARS. Consider a statistical model with the dosage level of Ribavirin as a continuous predictor and fatality ("Death" or "No Death") as a response. This model violates the standard linear regression assumptions.
   1. True
2. Generalised linear models (GLMs) extend the linear regression framework to allow for non-normal responses, such as counts.
   1. True
3. Standard linear regression is a type of generalized linear model (GLM).
   1. True
4. The response of a generalized linear model (GLM) is the random component.
   1. True
5. The link function in a generalized linear model (GLM) connects the random response to a linear combination of predictor variables.
   1. True
6. A generalized linear model (GLM) includes:
   1. A random response component.
   2. A link function.
   3. A systematic component consisting of a linear combination of discrete or continuous predictors, and fixed parameters.
7. Generalized linear models require that the response comes from the exponential distribution.
   1. False

## Binomial Regression

1. The range (output) of the binomial regression link function is the interval .
   1. False
2. The domain (input) of the binomial regression link function is the interval .
   1. True
3. The "probit" link function is , where is the inverse of the standard normal cdf, is the probability of success from the binomial response, and is the linear predictor.
   1. True
4. The likelihood function for binomial regression is the joint pmf of the response, but interpreted as a function of the parameters of the model (with the response data fixed).
   1. True
5. The likelihood function and the log-likelihood function:
   1. are both maximized at the same input/parameter value.
6. Let event have probability of occurrence. Then the odds in favour of is defined as: .
   1. True
7. Suppose that the probability of contracting a virus is . What are the odds of contracting ?
8. Consider data on the survival of patients who had undergone surgery for breast cancer. The data consists of a response (survival status after five years) and two predictors (the age of the patient at the time of the operation, and the number of cancerous auxiliary nodes detected):
   1. : Age of patient in years at time of operation (**predictor**)
   2. : Number of cancerous axillary nodes detected (**predictor**)
   3. : Survival status (**response**): 0 = the patient survived 5 years or longer; 1 = the patient died within 5 year

Suppose that a logistic regression model, with standardized predictors, correctly fits the data:

,

where is the probability of a patient surviving 5 years or longer, and

for .

Which of the following are correct?

1. For a fixed number of cancerous axillary nodes detected, a one standard deviation increase in age increases the odds of survival beyond 5 years by a multiplicative factor of , on average.
2. For a fixed number of cancerous axillary nodes detected, a one standard deviation increase in age increases the log-odds of survival beyond 5 years by , on average.
3. represents the mean log odds of surviving 5 years or longer for a person of (sample) mean age, and with the (sample) mean number of cancerous axillary nodes detected.
4. Consider a logistic regression model that uses data to estimate the probability that a client will default on a monthly credit card payment (defaulting on a payment means that the client fails to pay their bill by the deadline for the month in question.)
   1. : credit limit in dollars (**predictor**)
   2. : dollar amount of the bill statement one month prior (**predictor**)
   3. : dollar amount of the bill statement for two months prior (**predictor**)
   4. : dollar amount of the payment one month prior (**predictor**)
   5. : dollar amount of the payment two months prior (**predictor**)
   6. : default status (**response**): 0 = the client did not default on the payment for the month in question 5; 1 = the client did default on the payment for the month in question.

Suppose that a logistic regression model correctly fits the data:

,

where is the probability of default.

1. represents the mean log-odds of defaulting on a payment for a person with a $0 credit limit, a $0 bill statement for the last two months, and $0 in payments for the last two months.
2. is the average log-odds of default for a one-thousand dollar increase in credit limit, adjusting for the prior two months' bill statement and payment amounts.
3. represents the mean odds of defaulting on a payment for a person with a $0 credit limit, a $0 bill statement for the last two months, and $0 in payments for the last two months.

## Binomial Regression Inference

1. The maximum likelihood estimator is unbiased.
   1. False
2. As the sample size tends to infinity, the distribution of the maximum likelihood estimator becomes .
   1. True
3. Let be the parameter associated with predictor in a binomial regression model. For a reasonably large sample size , a standard normal "z-test" can be used to test whether should be in the model.
   1. True
4. Let be a random sample from a distribution with pdf , and let be the maximum likelihood estimator of . Then

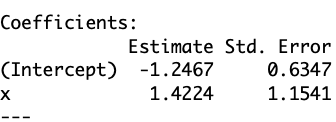
is an approximate 95% confidence interval for .

* 1. True

1. Goodness of fit metrics - such as the residual deviance - are only useful for the binomial regression with a relatively large number of trials (e.g., ).
   1. True
2. Consider a logistic regression fit an independent response and a single predictor variable . The linear predictor is:

.

Test vs  by computing the appropriate p-value, rounded to the hundredths place.



1. Consider a logistic regression fit an independent response and a single predictor variable . The linear predictor is:

.

Use maximum likelihood theory to construct an approximate 95% confidence interval for . Round all values to the hundredths place.

* 1. (-2.48, -0.02)