Question

Which of the following is ***not*** true?

- *a.* Logistic regression, Poisson regression, and gamma regression are all examples of generalized linear models.
- *b.* In generalized linear models, the mean and variance are unrelated.
- *c.* Canonical link functions are commonly used in link functions for many exponential family distributions.
- *d.* Although the canonical link function for the gamma distribution is the inverse, modelers often use the logarithm link function.

Correct.
In generalized linear models, the mean and variance are related. In other words, the variance of the response can be expressed as a function of its mean.

Question

Why can you not use Ordinary Least Squares regression (OLS) for a count of rare events with a skewed distribution?

- *a.* OLS regression assumes normal distribution of errors
- *b.* OLS regression assumes constant variance
- *c.* OLS regression can produce both positive and negative predicted values
- *d.* all of the above

Correct.
Ordinary Least Squares regression (OLS) assumes normal distribution of errors and constant variance. OLS regression can produce both positive and negative predicted values.

Question

Is overdispersion a problem in ordinary least squares regression?

- *a.* yes
- *b.* no

Correct.
Ordinary least squares regression assumes normal distribution. The normal distribution has a separate parameter, the variance (σ^2), to describe variability.

Question

Failing to correct for overdispersion results in which of the following? Select all that apply.

- *a.* underestimated parameter estimates
- *b.* overestimated parameter estimates
- *c.* underestimated standard errors for parameter estimates
- *d.* overestimated test statistics, and therefore, a too small *p*-value

Correct.
Overdispersion leads to underestimates of the standard errors of the parameter estimates, overestimates of the test statistics, which increase the Type I error rate, and liberal *p*-values.

Question

You want to model the rate of car insurance claims by geographic zone. The offset variable is which of the following?

- *a.* the number of claims
- *b.* the area of the geographic zone
- *c.* the number of insured in each geographic zone
- *d.* the population in the geographic zone

Correct.
The log of the measure of exposure is the offset variable. In this example, the rate of car insurance claims would be the count of car insurance claims divided by the number of insured in each geographic zone.

Question

A gamma regression model can be used for positive values with large means that are skewed to the right.

- *a.* yes
- *b.* no

Correct.
You use a gamma regression model when the response variable has continuous positive values, is highly skewed to the right, and has variances that are proportional to the squared mean.

Question

If you use PROC GLIMMIX with a log-link function, the predicted value is on the log scale. Do you need to request the predicted values with the ILINK option in order to obtain unbiased estimates of the means on the original scale?

- *a.* yes
- *b.* no

Correct.
The keyword ILINK requests that SAS provides the predicted values on the original scale of the data.