

IDS 702 Unit 3 Quiz - KEY

Fall 2025

First name:	_____	Last name:	_____
Net ID:	_____		

You are not permitted to use any materials (class notes, internet, etc.) on this assessment. Use of unauthorized materials or conferring with classmates will result in a 0 on the quiz and a report to the Office of Student Conduct.

I hereby state that I have not communicated with or gained information in any way from my classmates or unauthorized materials during this assessment, and that all work is my own.

Signature: _____

Multiple Choice

Circle the letter to mark your answer choice. Each question has one correct answer.

1. What is a link function in the context of generalized linear models?
 - a. A function that links the predicted outcome Y to the generalized outcome Y
 - b. A function that links the outcome Y to the expected value of the outcome Y
 - c. **A function that links the linear predictor to the expected value of the outcome Y**
 - d. A function that links the linear predictor directly to the outcome Y

2. Which assumption(s) of linear regression are also assumed for generalized linear models?
 - i. Linearity
 - ii. Independence
 - iii. Normality
 - iv. Equal variance
 - a. ii only
 - b. **i and ii**
 - c. i, ii, and iv
 - d. i, ii, iii, and iv

3. Which of the following models does **not** use the logit link?
 - a. Logistic regression
 - b. Multinomial regression
 - c. Ordinal regression
 - d. **Poisson regression**

4. Odds are defined as
 - a. The probability of the outcome
 - b. **The probability of the outcome divided by one minus the probability of the outcome**
 - c. The probability of the predictors divided by the probability of the outcome
 - d. The probability of the outcome divided by the probability of the predictors

5. Which of the following is a correct statement about the value of generalized linear models?
 - a. GLMs nearly always generate the most accurate predictions compared to other types of models
 - b. **GLMs allow us to address inferential research questions about the relationship between predictors and an outcome**
 - c. GLMs allow us to directly model outcome values without distributional assumptions
 - d. GLMs are easier to assess than other types of models because we can use predicted values instead of residuals

6. How does the method of estimation for GLMs relate to diagnostic tools that we use to assess and compare models?
 - a. Coefficient estimates are obtained via ordinary least squared estimation, so we can use metrics like R^2 for assessment and comparison.
 - b. **Coefficient estimates are obtained via maximum likelihood estimation, so we can use metrics based on the fitted likelihood for assessment and comparison.**
 - c. Coefficient estimates are obtained via hypothesis testing, so we can use p-values for assessment and comparison.
 - d. Coefficient estimates are obtained via minimizing residuals, so we can use metrics based on residuals for assessment and comparison.

7. What is an offset term?
 - a. The coefficient estimates on the exponential scale in a logistic regression model
 - b. An interaction term on the multiplicative scale in any generalized linear model
 - c. A variable whose effect is fixed to satisfy the proportional odds assumption in an ordinal regression model
 - d. **An additional term in a Poisson regression model that allows us to model rates of occurrence rather than only counts**

8. You fit a logistic regression model and obtain a coefficient estimate for a binary predictor of 0.4. Noting that $\exp(0.4) = 1.5$, which of the following is **not** a correct interpretation of this estimate?
 - a. The log odds of the outcome increase by 0.4 for $x_1 = 1$ compared to the reference group
 - b. The group for which $x_1 = 1$ experiences 1.5 times higher odds of the outcome compared to the reference group
 - c. **The group for which $x_1 = 1$ is 1.5 times more likely to experience the outcome than the reference group**
 - d. The odds of the outcome are 50% higher in the group for which $x_1 = 1$ than the reference group.
 - e. All interpretations are correct

9. Briefly define the equidispersion assumption, including which generalized linear model relies on this assumption.

Poisson regression: mean=variance (single parameter λ)

10. An airline distributes a survey to its customers at two locations and asks you to analyze the results. The outcome variable is satisfaction, defined as *very satisfied*, *satisfied*, *neutral*, and *dissatisfied*. This outcome is regressed on 3 predictor variables: *flight delay time* (numeric), *ticket type* (economy, comfort, or first), and *location* (RDU or CLT). How many coefficients, including intercept(s), would be estimated with each of the following models? The models are provided below for your reference.

Ordinal regression model:

$$\log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) = \beta_{0j} - \beta_1 x_{i1} - \dots - \beta_p x_{ip}$$

where $j = 1, \dots, J - 1$ and $i = 1, \dots, n$

Multinomial regression model:

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \beta_{0j} + \beta_{1j} x_{i1} + \dots + \beta_{pj} x_{ip}$$

where $j = 1, \dots, J - 1$ and $i = 1, \dots, n$

The ordinal model would have ____7____ coefficient estimates

This model has a unique intercept for each level of the outcome minus one, but same slope coefficients for each predictor, so we have $J - 1 = 3$ intercepts + $p = 4$ predictor terms (1 for flight delay time, 2 for ticket type, and 1 for location)=7 total coefficient estimates

The multinomial model would have ____15____ coefficient estimates

Here we have unique intercepts and slope coefficient estimates for each level of the outcome minus one, so $(J - 1) \times (p + 1) = 3 \times 5 = 15$ total coefficient estimates

Bonus

In what year did each of the following occur? (Partial credit for correct decade)

The Titanic sank: 1912

The Titanic wreckage was discovered: 1985

The Titanic film (directed by James Cameron) was released: 1997