140.644: Statistical Machine Learning **Name:** _____

Instructor: Yiqun Chen TA: Junwei (Ivy) Sun

Quiz 2 — Practice — Regression and Classification

Not graded. Bring questions to class.

This is a shortened sample so you know what to expect. The real quiz has **4 multiple choice** and **2 questions in Part B** (the second has subparts), and is worth 100 points; this one has 2 and 2. The style, the difficulty, and the point values per question are the same. No calculator is needed. You may leave answers in terms of $\exp(\cdot)$ or $\log(\cdot)$.

Part A: Multiple Choice [20 points]

Circle the single best answer. **(10 points)** each.

1. A binary outcome records whether a patient visits the emergency department within 7 days. To fit a logistic regression in which $P(Y = 1 \mid X)$ is the probability of an emergency-department visit, the outcome should be coded as:

   (A) $Y = 0$ for no visit and $Y = 1$ for a visit

   (B) $Y = 1$ for no visit and $Y = 0$ for a visit

   (C) $Y = 7$ for a visit and $Y = 0$ for no visit

   (D) $Y = 1$ for either outcome

2. Which of the following is *not* a linear model in its parameters?

   (A) $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$

   (B) $Y = \beta_0 + \beta_1 \log(X_1) + \beta_2 X_1 X_2 + \epsilon$

   (C) $Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \epsilon$, where D_1, D_2 are indicators

   (D) $Y = \beta_0 + \beta_1 \exp(\beta_2 X) + \epsilon$

Part B: Reading an Analysis [40 points]

Let $Y = 1$ mean a patient misses a scheduled follow-up visit and $Y = 0$ mean the patient attends. Referral source has three categories: self-referred, primary-care referral, and emergency- department referral. Self-referral is the reference category, so the fitted logistic regression uses D_{primary} and D_{ED}, which equal 1 only for the named category:

$$\log\left(\frac{p}{1-p}\right) \;=\; \beta_0 + \beta_1(\text{age}) + \beta_2 D_{\text{primary}} + \beta_3 D_{\text{ED}}.$$

1. **(20 points)** Compare two 50-year-old patients: one is self-referred and the other is referred from the emergency department. State the two indicator values for each patient. Then write how much larger the emergency-department patient's predicted **log-odds** of missing the visit are, in terms of the β's.

2. **(20 points)** The clinic predicts a missed visit when $\hat{p} \geq t$. For the patients below, what is the misclassification rate at $t = 0.60$?

Patient	Fitted probability $\hat{p}$	Observed outcome
A	0.12	Attended
B	0.38	Missed
C	0.63	Missed
D	0.71	Attended
E	0.82	Missed