

# Exam Briefing

- Date: 30 April 2025
- Time: 1PM – 3PM
- Venue: **TBA**
- **Study Week Consultation:**
  - Date: 15 April 2025 (8AM-2PM)
  - Venue: **TBA**
  - Not compulsory, **book 30 mins slots as a group with the TA if needed**

# Exam Briefing

- Format:
  - Closed Book with formula sheet provided
  - **Bring your Calculators!**
  - **Section A: 20 MCQs** (5 alternatives each) [40 marks]
    - Like Quiz
  - **Section B: 3 Structured Questions** [60 marks]
    - Short answers – like Assignment 2
    - Multiple parts
    - Both qualitative and quantitative questions

# Log base 2 Calculations

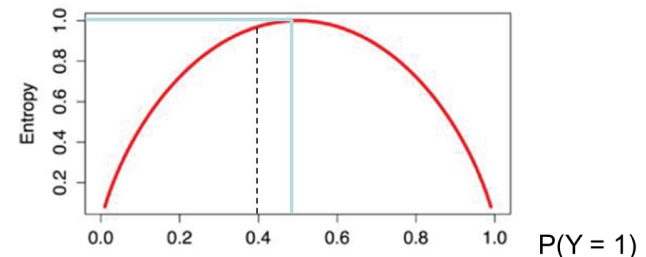
$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$$

$$\text{Entropy} = H(Y) = -\sum_{i=1}^n p_i \log_2(p_i)$$

$$= -[0.6 \times \log_2(0.6) + 0.4 \times \log_2(0.4)]$$

$$= -\left[0.6 \times \frac{\log_{10} 0.6}{\log_{10} 2} + 0.4 \times \frac{\log_{10} 0.4}{\log_{10} 2}\right]$$

$$= 0.9709506$$



# Coverage

- Questions will focus on testing understanding of concepts:
  - Introduction to Analytics
  - Exploratory Data Analytics
  - Healthcare Data Management
  - Risk Factor Analysis 1: Correlation Analysis, Linear Regression, etc
  - Risk Factor Analysis 2: Logistic Regression, Model Validation, etc
  - Causal Analysis (Propensity Score Matching, Bradford Hill Criteria, etc)
  - Unsupervised and Supervised Machine Learning
    - Donabedian and AHRQ Framework for Quality of Care
    - Clustering Analysis; CART Decision Trees; Advanced Decision Trees
    - Cross Validation; ML Workflow; Gini Impurity, Entropy, Etc
  - Healthcare Operations Management (Process Analysis, Little's Law)

# Coverage

- You will not be asked to write R codes on the spot but understanding of R outputs will be required.
  - For example: interpretation of R regression output and the feature importance chart from CART (Quality.Rmd)
  - Outputs that are covered in class
- Case Studies are excluded
- Assignment 2 (Part 2) Case Study is excluded

```
Call:
lm(formula = TenYearCHD ~ ., data = framinghamTrain)

Residuals:
    Min       1Q   Median       3Q      Max
-0.80807 -0.18579 -0.10575 -0.00967  1.04095

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.776e-01  9.446e-02  -7.174 9.54e-13 ***
male         5.230e-02  1.481e-02   3.532 0.000419 ***
age          7.455e-03  9.070e-04   8.220 3.21e-16 ***
education   -8.190e-03  6.770e-03  -1.210 0.226498
currentSmoker 1.002e-02  2.178e-02   0.460 0.645564
cigsPerDay   2.235e-03  9.329e-04   2.396 0.016665 *
BPMeds       4.503e-02  4.187e-02   1.075 0.282334
prevalentStroke 1.846e-01  9.160e-02   2.015 0.043987 *
prevalentHyp 1.141e-02  2.088e-02   0.547 0.584725
diabetes     -4.679e-02  5.102e-02  -0.917 0.359218
totChol      7.731e-05  1.624e-04   0.476 0.634109
sysBP        2.723e-03  5.932e-04   4.591 4.62e-06 ***
diaBP       -1.225e-03  9.690e-04  -1.264 0.206259
BMI          2.088e-03  1.853e-03   1.127 0.260004
heartRate    -6.279e-04  5.826e-04  -1.078 0.281172
glucose      1.689e-03  3.596e-04   4.698 2.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3399 on 2545 degrees of freedom
Multiple R-squared:  0.1105,    Adjusted R-squared:  0.1053
F-statistic: 21.08 on 15 and 2545 DF,  p-value: < 2.2e-16
```

