Stevens Institute of Technology

SYS-601 Practice Exam #1 Solutions

1.1 Data Types and Terminology

(a) Consider the `mtcars` R dataset which gives information about automobiles listed in the 1974 issue of *Motor Trend* magazine. It includes 32 observations of 10 variables including the four below. *Select* the best corresponding data type for each variable.

(i) `mpg`: Miles per US gallon (MPG)

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

Zero miles per gallon is meaningful and dividing makes sense, we can say one automobile has twice as good of fuel efficiency as another.

(ii) `cyl`: Number of cylinders

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

Zero cylinders is meaningful and dividing makes sense, we can say one automobile has twice as as many cylinders as another.

(iii) `wt`: Weight (1000 pounds)

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

Zero weight is meaningful and dividing makes sense, we can say one automobile has twice the weight of another.

(iv) `am`: Transmission (0: automatic, 1: manual)

(A) X Nominal (B) Ordinal (C) Interval (D) Ratio

It is not possible to order automatic versus manual, even though a numerical coding scheme is used.

(b) Consider the following observations from a simple experiment of rolling a six-sided die. There are six elementary events per roll: ⚀, ⚁, ⚂, ⚃, ⚄, or ⚅.
Select the best relationship between derived events X and Y defined below if:

(i) X: roll is ⚀ or ⚁ Y: roll is ⚁ or ⚂

(A) Collectively Exhaustive
(B) X Equally Likely
$P(X) = P(Y) = 1/3$
(C) Complementary
(D) Mutually Exclusive

(ii) X: roll is ⚀ or ⚁ or ⚂ or ⚃ Y: roll is ⚃ or ⚄ or ⚅

(A) X Collectively Exhaustive
$P(X \cup Y) = 1$
(B) Equally Likely
(C) Complementary
(D) Mutually Exclusive

(iii) X: roll is ⚀ or ⚁ or ⚂ Y: roll is ⚀ or ⚁ or ⚂

(A) Collectively Exhaustive
(B) X Equally Likely
$P(X) = P(Y) = 1/2$
(C) Complementary
(D) Mutually Exclusive

(iv) X: roll is ⚀ or ⚁ or ⚂ Y: roll is ⚄ or ⚅

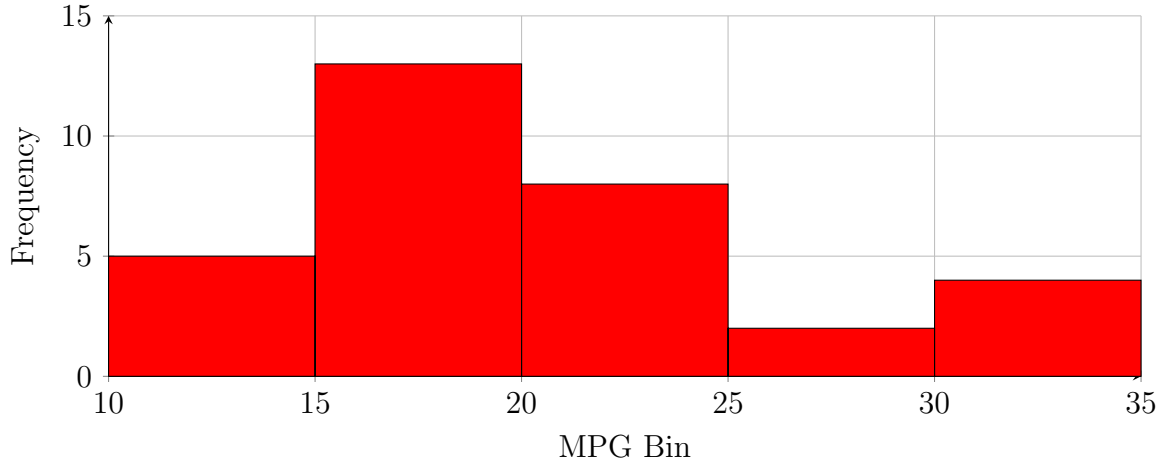
(A) Collectively Exhaustive
(B) Equally Likely
(C) Complementary
(D) X Mutually Exclusive
$P(X \cap Y) = 0$

1.2 Descriptive Statistics

Consider the 32 *sorted* `mpg` variable (miles per US gallon) of the `mtcars` data set reproduced below (note: this dataset is also available via the Canvas midterm module as a .csv file).

10.4	14.7	15.5	17.8	19.2	21.4	22.8	30.4
10.4	15.0	15.8	18.1	19.7	21.4	24.4	30.4
13.3	15.2	16.4	18.7	21.0	21.5	26.0	32.4
14.3	15.2	17.3	19.2	21.0	22.8	27.3	33.9

(a) Using the space below, sketch a histogram of MPG for automobiles in this dataset. Choose an appropriate bin size and label the axes.



(b) What is the mean MPG of automobiles in this dataset?

The mean is $\bar{x} = (1/N)\sum_{i=1}^{N} x_i = 20.1$ miles per gallon.

(c) What is the median MPG of automobiles in this dataset?

The median is 19.2 miles per gallon.

(d) What is the sample standard deviation of MPG for automobiles in this dataset?

The sample std. deviation is $s_x = \sqrt{(1/(N+1))\sum_{i=1}^{N}(x_i - \bar{x})^2} = 6.0$ miles per gallon.

(e) What is the inter-quartile range of MPG for automobiles in this dataset?

The IQR is $Q_3 - Q_1 = P_{75} - P_{25} = 7.5$ miles per gallon.

1.3 Probability Theory

Consider the 32 *sorted* `am` (transmission, 0: automatic, 1: manual) and `cyl` (number of cylinders) variables from the `mtcars` data set reproduced below (note: this dataset is also available via the Canvas midterm module as a .csv file).

0, 4	0, 6	0, 8	0, 8	0, 8	1, 4	1, 4	1, 6
0, 4	0, 6	0, 8	0, 8	0, 8	1, 4	1, 4	1, 6
0, 4	0, 6	0, 8	0, 8	0, 8	1, 4	1, 4	1, 8
0, 6	0, 8	0, 8	0, 8	1, 4	1, 4	1, 6	1, 8

(a) Complete the following probability matrix for transmission (automatic, manual) and number of cylinders (4, 6, 8).

Transmission (`am`)	Number Cylinders (`cyl`) 4	6	8	$\sum$
Automatic (0)	$3/32 = 0.09$	$4/32 = 0.13$	$12/32 = 0.38$	$19/32 = 0.59$
Manual (1)	$8/32 = 0.25$	$3/32 = 0.09$	$2/32 = 0.06$	$13/32 = 0.41$
$\sum$	$11/32 = 0.34$	$7/32 = 0.22$	$14/32 = 0.44$	

(b) What is the marginal probability of an automobile having automatic transmission?

$$P\{\texttt{am} = 0\} = 19/32 = 0.59$$

(c) What is the joint probability of an automobile having 6 cylinders and automatic transmission?

$$P\{\texttt{cyl} = 6 \cap \texttt{am} = 0\} = 4/32 = 0.13$$

(d) What is the union probability of an automobile having 4 or 6 cylinders?

$$P\{\texttt{cyl} = 4 \cup \texttt{cyl} = 6\} = 11/32 + 7/32 = 0.56$$

(e) What is the conditional probability of an automobile having automatic transmission if it has 6 cylinders?

$$P\{\texttt{am} = 0 | \texttt{cyl} = 6\} = 4/7 = 0.57$$

1.4 Discrete Random Variables

Consider the following probability mass function (PMF) for the random variable X, the number of children in families purchasing a new automobile.

x:	0	1	2	3	4
$P(x)$:	0.20	0.30	0.25	0.15	0.10

(a) What is the mean of X?

$$\mu = \sum_{i=1}^{N} x \cdot P(x) = 1.65 \text{ children}$$

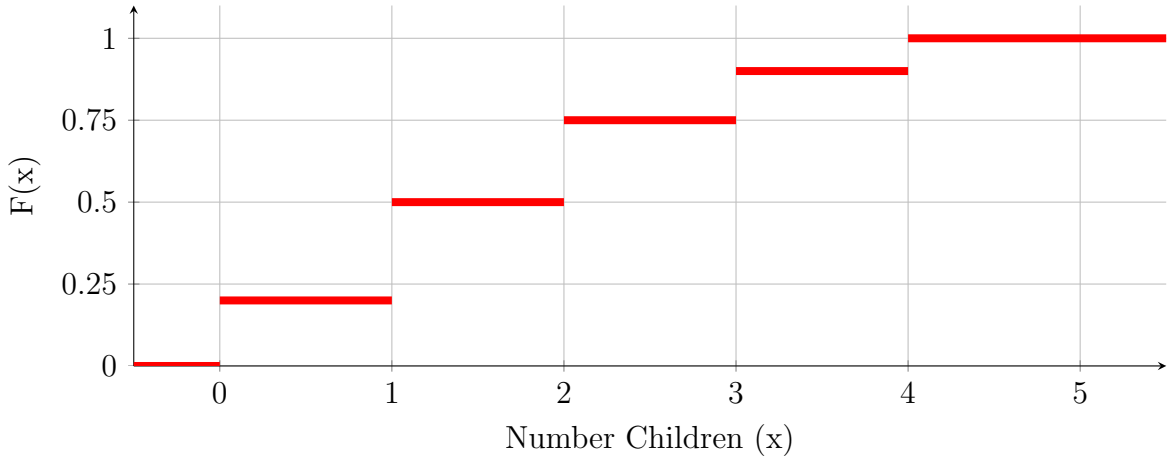
(b) What is the variance of X?

$$\sigma^2 = \sum_{i=1}^{N} (x - \mu)^2 \cdot P(x) = 1.53 \text{ children}^2$$

(c) Complete the cumulative distribution function (CDF) values $F(x)$ for the table below.

x:	0	1	2	3	4
$F(x)$:	0.2	0.5	0.75	0.9	1.0

(d) Using the space below, sketch a CDF plot for the random variable X. Label the axes.



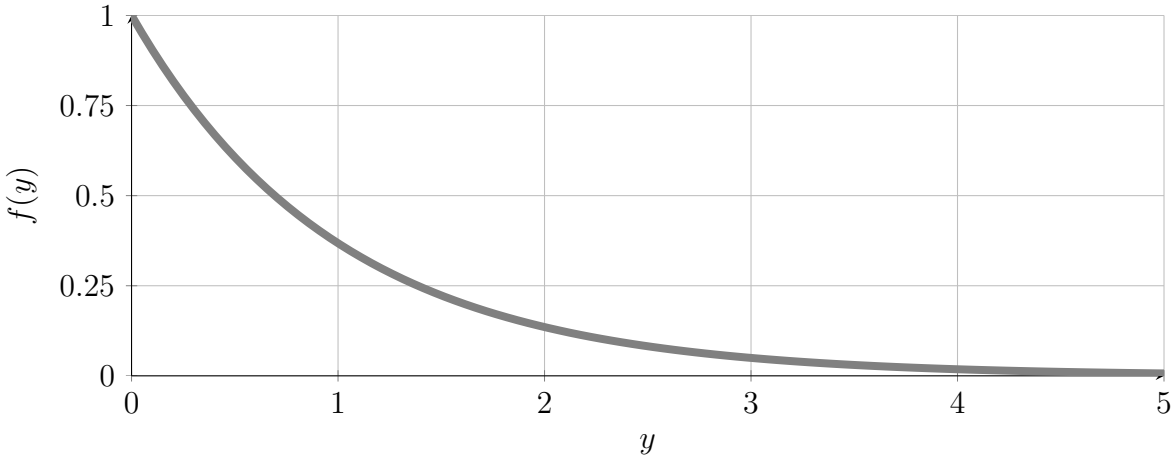
(e) What is the probability a family purchasing an automobile has more than two children?

$$P(x > 2) = 1 - F(2) = 1 - 0.75 = 0.25$$

1.5 Continuous Random Variables

The random variable Y measures the amount of time (in years) between corrective maintenance actions for an automobile. Consider the following probability density function (PDF):

$$f(y) = e^{-y}$$



(a) What is the mean of Y?

$$\mu = \int_0^\infty y f(y) dy = 1 \text{ year (Note: this is an exponential distribution with } \lambda\text{=1)}$$

(b) What is the variance of Y?

$$\sigma^2 = \int_0^\infty (y-\mu)^2 f(y)dy = 1 \text{ year}^2 \text{ (Note: this is an exponential distribution with } \lambda=1)$$

(c) Using a trapezoidal area approximation (**not** the CDF equation), estimate $F(1)$.

$$F(1) \approx \frac{1}{2}(b_1 + b_2) \cdot h = \frac{1}{2}(1 + 0.375) \cdot 1 = 0.6875$$

(d) What is the CDF equation $F(y)$?

$$F(y) = \int_0^y f(i)di = 1 - e^{-y} \text{ (Note: this is an exponential distribution with } \lambda=1)$$

(e) What is the 95th percentile Y value, y_{95}? (*Hint:* $F(y_{95}) = 0.95$)

$$F(y_{95}) = 1 - e^{-y_{95}} = 0.95 \implies y_{95} = -\ln 0.05 \approx 3.00 \text{ years}$$