

- Identify the most plausible distribution that would be used to model the following random variables. Make sure to specify any parameters.
  - The number of free throws a player makes given two attempts. The players free throw percentage is .72.
  - The number of customers that enter a bank between 1 and 2 pm. The average number of customers in this time period is 15.
  - A person is asked to guess what the next card picked from a deck of 52 cards is. Let  $X$  be the event that they guess correctly and be 0 if they guess wrong and 1 if they guess correctly.
  - The number of shots it takes an archer to hit the bulls-eye on a given target. On average they hit the bulls-eye every 7 out of 10 shots.
- For the given random variables calculate the quantities of interest
  - Let  $X \sim \text{Binomial}(15, .2)$ . Find  $E[X]$ ,  $SD[X]$ ,  $P(X \leq 2)$ .
  - Let  $X \sim \text{Pois}(3)$ . Find  $E[X]$ ,  $SD[X]$ ,  $P(X \geq 2)$ .
  - Let  $X \sim \text{Geometric}(.2)$ . Find  $E[X]$ ,  $SD[X]$ ,  $P(X = 2 \text{ or } X \geq 5)$ .
- Use the following data to answer the given questions:  
 1 1 1 1 2 2 2 2 2 3 5 7 8 10 11
  - Make a dotplot of this data
  - Find the mean and standard deviation of the data
  - Find the median and the IQR
  - What is the mode of this data?
  - Find  $Q(.3)$
  - Make a boxplot of this data
  - Would you prefer to use the mean and standard deviation or would you prefer to use the median and IQR to summarize the center and spread of this data?
- Let  $X$  be a discrete random variable with the following CDF
 
$$F(x) = \begin{cases} 0 & : x < 0 \\ .2 & : 0 \leq x < 1 \\ .3 & : 1 \leq x < 2 \\ .7 & : 2 \leq x < 3 \\ .9 & : 3 \leq x < 4 \\ .95 & : 4 \leq x < 5 \\ 1 & : 5 \leq x \end{cases}$$
  - What is  $P(X \geq 3)$ ?
  - What is  $P(X < 2)$ ?
  - What is  $P(X = 4)$ ?

5. State whether the following are valid probability distributions. If they are not then state why they are not valid.

x	1	2	3	4	5
P(X=x)	.1	.2	.3	.4	.5

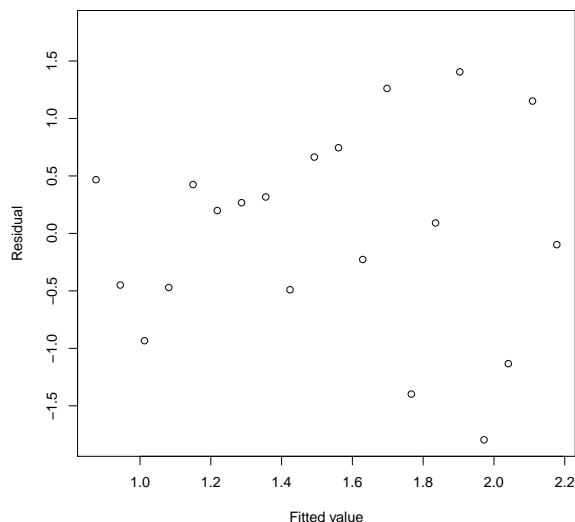
x	1	2	3	4	5
P(X=x)	.1	.2	.3	.4	0

x	1	2	3	4	5
P(X=x)	.1	-.2	.3	.7	.1

6. Say we have the following information about two variables  $X$  and  $Y$  and we want to predict  $Y$  using  $X$ .

Quantity	value
$n$	20
$\sum x$	10.00
$\sum y$	30.53
$\sum (x - \bar{x})^2$	1.84
$\sum (y - \bar{y})^2$	17.69
$\sum (x - \bar{x})(y - \bar{y})$	2.40

Statistic	N	Mean	St. Dev.	Min	Max
x	20	0.500	0.311	0.000	1.000



- What is the fitted least squares line for this data?
- Compute the sample correlation between  $x$  and  $y$  and interpret this value.
- What fraction of the raw variability in  $y$  is accounted for in the fitting of a line to the data?
- Make predictions for  $y$  at  $x = .2$  and for  $x = 1.3$ .

- (e) Which prediction do you trust more? Why?
  - (f) Are there any noticeable problems that are implied by the residual plot?
7. Suppose I have a random variable  $X$  with probability mass function  $f(x) = c/x$  for  $x = 2, 4, 6, 8, 10$  and  $f(x) = 0$  for all other values.
- (a) What value does  $c$  need to be to make this a valid probability mass function?
  - (b) Find the CDF corresponding to this probability mass function.
  - (c) What is  $P(X = 4)$ ?
  - (d) What is  $P(X \text{ is odd})$ ?
  - (e) What is  $E[X]$ ?
  - (f) What is  $E[3X + 2]$ ?
  - (g) What is  $SD[X]$ ?
8. Suppose I have a random variable  $X$  with probability mass function  $f(x) = c \sin(x)$  for  $0 \leq x \leq \pi$  and  $f(x) = 0$  for any other value of  $x$ .
- (a) What value does  $c$  need to be to make this a valid probability mass function?
  - (b) Find the CDF corresponding to this probability mass function.
  - (c) What is  $P(2 \leq X \leq 5)$ ?
  - (d) What is  $P(X = \pi/2)$ ?
  - (e) What is  $E[X]$ ?
  - (f) What is  $E[3X + 2]$ ?
  - (g) What is  $Q(.1)$ ?