Math 3339
Section 21155 - SR 117 - MW 1 - 2:30pm

Bekki George: bekki@math.uh.edu

University of Houston

Sections 8.1-8.2, 4.1-4.2
Office Hours: Mondays 11am - 12:30pm, Tuesdays 3-4pm
(also available by appointment)
Office: 206 PGH
Course webpage: www.casa.uh.edu
Poppers
1. A company estimates that 60% of the adults in the U.S. have seen its TV commercial and that if an adult sees the commercial, there is a 15% chance that the adult will buy its product. What is the probability that an adult chosen at random in the U.S. will have seen the company’s commercial and will have bought its product?
Televisions and Life Expectancy

<table>
<thead>
<tr>
<th>Country</th>
<th>Life Exp.</th>
<th>Per TV</th>
<th>Country</th>
<th>Life Exp.</th>
<th>Per TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>44</td>
<td>200</td>
<td>Mexico</td>
<td>72</td>
<td>6.6</td>
</tr>
<tr>
<td>Australia</td>
<td>76.5</td>
<td>2</td>
<td>Morocco</td>
<td>64.5</td>
<td>21</td>
</tr>
<tr>
<td>Cambodia</td>
<td>49.5</td>
<td>177</td>
<td>Pakistan</td>
<td>56.5</td>
<td>73</td>
</tr>
<tr>
<td>Canada</td>
<td>76.5</td>
<td>1.7</td>
<td>Russia</td>
<td>69</td>
<td>3.2</td>
</tr>
<tr>
<td>China</td>
<td>70</td>
<td>8</td>
<td>S. Africa</td>
<td>64</td>
<td>11</td>
</tr>
<tr>
<td>Egypt</td>
<td>60.5</td>
<td>15</td>
<td>Sri Lanka</td>
<td>71.5</td>
<td>28</td>
</tr>
<tr>
<td>France</td>
<td>78</td>
<td>2.6</td>
<td>Uganda</td>
<td>51</td>
<td>191</td>
</tr>
<tr>
<td>Haiti</td>
<td>53.5</td>
<td>234</td>
<td>U.K.</td>
<td>76</td>
<td>3</td>
</tr>
<tr>
<td>Iraq</td>
<td>67</td>
<td>18</td>
<td>U.S.</td>
<td>75.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Japan</td>
<td>79</td>
<td>1.8</td>
<td>Vietnam</td>
<td>65</td>
<td>29</td>
</tr>
<tr>
<td>Madagascar</td>
<td>52.5</td>
<td>92</td>
<td>Yemen</td>
<td>50</td>
<td>38</td>
</tr>
</tbody>
</table>

a) Which of the countries listed has the fewest people per television set? Which has the most? What are those numbers?

b) Create a scatter plot. Does there appear to be an association?
Example - continued

c) Compute the value of the correlation coefficient between life expectancy and people per television.

d) Since the association is so strongly negative, one might conclude that simply sending television sets to the countries with lower life expectancies would cause their inhabitants to live longer. Comment on that argument.

e) If two variables have a correlation close to +1 or −1, indicating a strong linear relationship, does it follow that there must be a cause-and-effect relationship between them?
Example - continued

This example illustrates a very important distinction between association and causation. Two variables may be strongly associated without a cause-and-effect relationship existing between them. Often the explanation is that both variables are related to a third variable not being measured; this variable is often called a lurking or confounding variable.

f) In this case, suggest a confounding variable that is associated with both a country’s life expectancy and the prevalence of televisions in the country.
R Commands

Enter your lists:
> xList=c(..)
> yList=c(..)

To do a scatterplot, the command is plot(x, y) or plot(y~x)
> plot(xList, yList)

The correlation is found by using cor(x,y)
> cor(xList, yList)
[1] 0.8779736
To find the LSRL, we will use the `lm()` function. The parameter for this function is called the model formula. The model formula is \( y \sim x \) (read as \( y \) is modeled by \( x \)).

```r
> yp=lm(yList~xList)
> yp
```

Call:
```
lm(formula = yList ~ xList)
```

Coefficients:
```
(Intercept)    xList
  67.283     6.784
```

So, the LSRL is
The function `abline()` after the `plot()` command will make the LSRL appear on the graph:

```r
> plot(xList, yList)
> abline(yp)
```

`residuals()` will find and list all residual values

```r
> res=residuals(yp)
> res
```

```r
  1   2   3   4   5
```

```r
> plot(xList, res)
```

You can see many plots by using `plot` with your model.

```r
> plot(yp)
```
Example

Transforming Data - see data set link below notes on casa
Random Variables

Suppose an experiment is conducted. A **random variable** is a function that assigns values to the outcomes of the experiment.

Example: A coin is flipped resulting in either heads or tails and the random variable $X$ is defined by

$$X(\text{heads}) = 1 \quad X(\text{tails}) = 0$$

We say that the random variable $X$ indicates heads.

Example: A fair six sided die is tossed, let the random variable $X$ indicate the event that an even number is rolled. What is $X(2)$? $X(3)$?
Random Variables

Example: Recall the gas station problem, with 2 gas stations having six pumps each. Define a random variable \( X = \) total number of pumps in use at the two stations What are the possible values of \( X \)?

Define a random variable \( Y = \) total number of pumps not currently being used at the two stations What are the possible values of \( Y \)? How are \( X \) and \( Y \) related?
Example: Suppose pump 1 at gas station number one is observed as the next customer begins to use it. Let $T =$ the length of time the customer is at the pump. What are the possible values of $T$?
Random Variables

There are two types of random variables:
1. Discrete Random Variables: A random variable whose possible values are either finite or may be listed.
2. Continuous Random Variables: A random variable whose possible values consists of an interval of values or several intervals of values and for which the probability of any one value is zero.

For continuous random variables $X$, they take on all values in an interval of numbers. In fact, the probability of $X$ equaling an individual number is 0! The probability distribution of $X$ is described by a density curve. The probability of any event is the area under the curve and above the values of $X$ that make up the event.

Ex: Which of the above variables were discrete? Which were continuous?
Discrete Random Variables

Def: The **probability mass function** (pmf) of a discrete rv is defined for every number \( x_1 \) by \( f(x_i) = P(X = x_i) \).

Example: Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

<table>
<thead>
<tr>
<th>Lot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Defective</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

One of these lots is to be randomly selected for shipment to a particular customer. Let \( X = \) number of defectives in the selected lot. What are the possible values of \( X \)? Determine the values of the pmf.
The probabilities that a customer selects 1, 2, 3, 4, or 5 items at a convenience store are 0.32, 0.12, 0.23, 0.18, and 0.15, respectively.

Construct a probability distribution for the data:

<table>
<thead>
<tr>
<th>X</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Find $P(1.0 < X < 3.0)$
Discrete Random Variables

Properties of $f$:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\sum_{x} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$, where $A \subset \mathbb{R}$ is a discrete set.

Suppose we toss a fair coin 10 times. Let $X = \text{number of heads in the 10 tosses}$. What are the possible values of $X$?

How many heads do we expect to get in the 10 tosses?

What is $f(0)$? $f(1)$?
Example: Suppose you are given the following distribution table:

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.15</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Find the following:

1. $P(X = 4)$

2. $P(X < 2)$

3. $P(2 < X \leq 5)$

4. $P(X > 3)$
Given the following sampling distribution:

<table>
<thead>
<tr>
<th>X</th>
<th>−16</th>
<th>−13</th>
<th>−7</th>
<th>11</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{1}{25}$</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{7}{100}$</td>
<td></td>
</tr>
</tbody>
</table>

3. $P(X = 14) =$

4. $P(-10 < X < 11) =$
Discrete Random Variables

The Cumulative Distribution Function:
Def: The cumulative distribution function (cdf) $F(x)$ of a discrete rv $X$ with pmf $f(x)$ is defined for every number $x$ by $F(x) = P(X \leq x)$.

For any number $x$, $F(x)$ is the probability that the observed value of $X$ will be at most $x$. 
Discrete Random Variables

Example: A store carries flash drives with either 1 GB, 2 GB, 4 GB, 8 GB, or 16 GB of memory. The table below gives the distribution of the rv $X$ = the amount of memory in a purchased drive.

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.05</td>
<td>0.10</td>
<td>0.35</td>
<td>0.40</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Determine the values of $F$ for the possible values of $X$.

What is the value of $F(2.5)$?

Sketch a graph of $F$. 
Discrete Random Variables

Example: Suppose the random variable $X$ takes on possible values $x = 0, 1, 2, 3$ and has pmf given by $f(x) = \frac{x + 1}{k}$, determine the value of $k$. 
5. Suppose the random variable $X$ takes on possible values $x = 0, 1, 2$ and has pmf given by $f(x) = \frac{2x + 3}{k}$, determine the value of $k$. 

(a) 15 
(b) 10 
(c) 15 
(d) 10 
(e) none of these