Making Probabilistic Statements

Using Univariate and Joint Distributions

UNIVERSITY OF DELAWARE

uide to E × | When We Say 70 Percent, It Reall × | +

://fivethirtyeight.com/features/when-we-say-70-percent-it-really-means-70-percent/

: The D... | bookdown: Authori... | The magick packag... | magick v1.0 | R packages | R/exams | Wiley GoToMeeting | Manufacturing

FiveThirtyEight

Politics Sports Science & Health Economics Culture

APR. 4, 2019, AT 5:16 PM

When We Say 70 Percent, It Really Means 70 Percent

By Nate Silver

Filed under Housekeeping

UNIVERSITY OF DELAWARE

1

A probabilistic statement

Probabilistic Statements – Bernoulli Distribution

Mathematically Precise Statements

$$X \sim Bernoulli(\theta = 90\%)$$

What values of X are possible?	**What is the probability $X = 0$?**	**What is the probability $X = 0.9$?**	**What is the probability $X \leq 1$?**

Probabilistic Statements – Bernoulli Distribution

Majority of Our Work Is Making Statements from Representative Samples

Estimates From **Representative Sample**

Trial #	1	2	3	4	5	6	7	8
Outcome	0	0	1	0	0	0	0	1

What is the probability $X = 0$?

What is the probability $X = 0.9$?

What is the probability $X \leq 0.9$?

UNIVERSITY OF DELAWARE

Making Probabilistic Statements About Distributions With Known CDF's

Cumulative Distribution Functions (CDF)

$CDF\ notation$:

$$F(x) = P(X \leq x)$$

UNIVERSITY OF DELAWARE

Assuming: $X \sim Bernoulli(\theta = 0.8)$

Bernoulli Distribution CDF

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ 1-\theta & if\ 0 \le x \le 1 \\ 1 & if\ x \ge 1 \end{cases}$$

$CDF\ notation$:
$F(x) = P(X \le x)$

What is the probability $X \le 0.9$?

Making Probabilistic Statements About Distributions With Known CDF's

$X \sim Bernoulli(\theta = 0.8)$

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ 1-\theta & if\ 0 \le x \le 1 \\ 1 & if\ x \ge 1 \end{cases}$$

$CDF\ notation$:
$F(x) = P(X \le x)$

What is the probability $X \le 0.9$?

$P(X \le 0.9) = F(0.9) = 1 - \theta = 0.1 = 10\%$

Making Probabilistic Statements About Distributions With Fancy Names

$X \sim Fleischhacker$ ← Zero-Parameter Distribution

$CDF\ notation$:
$F(x) = P(X \leq x)$

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ 3x^2 - 2x^3 & if\ 0 \leq x \leq 1 \\ 1 & if\ x \geq 1 \end{cases}$$

What is the probability $X \leq 7$?

What is the probability $X \leq \frac{1}{2}$?

UNIVERSITY OF DELAWARE

8

Making Probabilistic Statements About Distributions With Fancy Names

$X \sim Fleischhacker$

$CDF\ notation$:
$F(x) = P(X \leq x)$

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ 3x^2 - 2x^3 & if\ 0 \leq x \leq 1 \\ 1 & if\ x \geq 1 \end{cases}$$

What is the probability $X \leq \frac{1}{2}$?

$$\boxed{P\left(X \leq \frac{1}{2}\right) = F\left(\frac{1}{2}\right) = 3 \times \left(\frac{1}{2}\right)^2 - 2 \times \left(\frac{1}{2}\right)^3 = \frac{3}{4} - \frac{2}{8} = \frac{1}{2} = 50\%}$$

UNIVERSITY OF DELAWARE

9

Using R's CDF Functions

$F(x) =$ `pfoo`

`foo` is called a placeholder name in computer programming. The word `foo` itself is meaningless, but you will substitute more meaningful words in its place. In the examples here, `foo` will be replaced by an abbreviated probability distribution name like `binom` or `norm`.

$X \sim Normal(\mu = 30, \sigma = 10)$

What is the probability $X \leq 14$?

```
pnorm(q = 14, mean = 30, sd = 10)
```

UNIVERSITY OF DELAWARE

10

Using R's CDF Functions

$X \sim Normal(\mu = 30, \sigma = 10)$

What is the probability $X \leq 14$?

argument for x
(i.e. **q**uantile)

```
pnorm(q = 14, mean = 30, sd = 10)
```

`foo` replacement

parameters for the distribution
(note: mathematicians love Greek letters, but R does not always use the same notation.
μ=`mean` & σ=`sd`)

UNIVERSITY OF DELAWARE

11

Your turn to use R

pfoo

The number of months the average Amazon warehouse employee works until they quit or get fired follows a Weibull distribution with shape parameter 1 (i.e. $\alpha = 1$) and scale parameter equal to 10 (i.e. $\beta = 10$).

Question 1: What is the probability that a worker is employed with Amazon for less than one year?

Question 2: What is the probability that a worker is employed for more than 6 months?

Probabilistic Statements – Weibull Distribution

Estimates From Representative Sample

Obs #	1	2	3	4	5	6	7	8
Outcome	13.71	41.24	7.15	7.64	15.98	42.64	9.07	7.08

What is the probability $X \leq 12$?

What is the probability $12 \leq X \leq 24$?

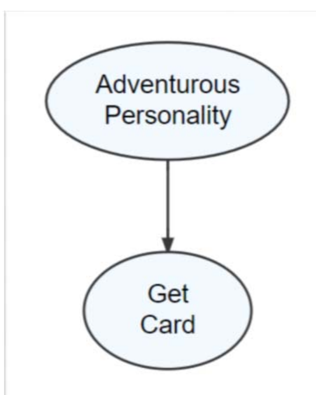
If Time and Energy Permit:

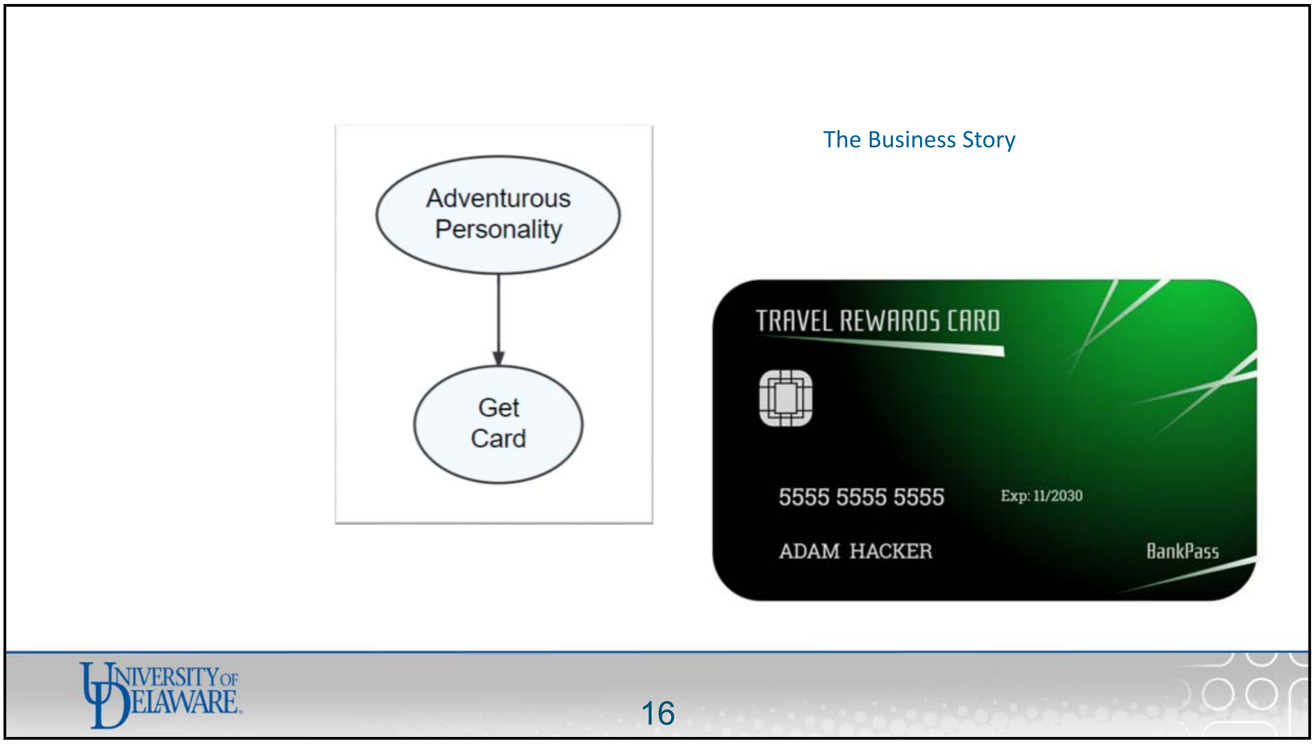
Making Probabilistic Statements

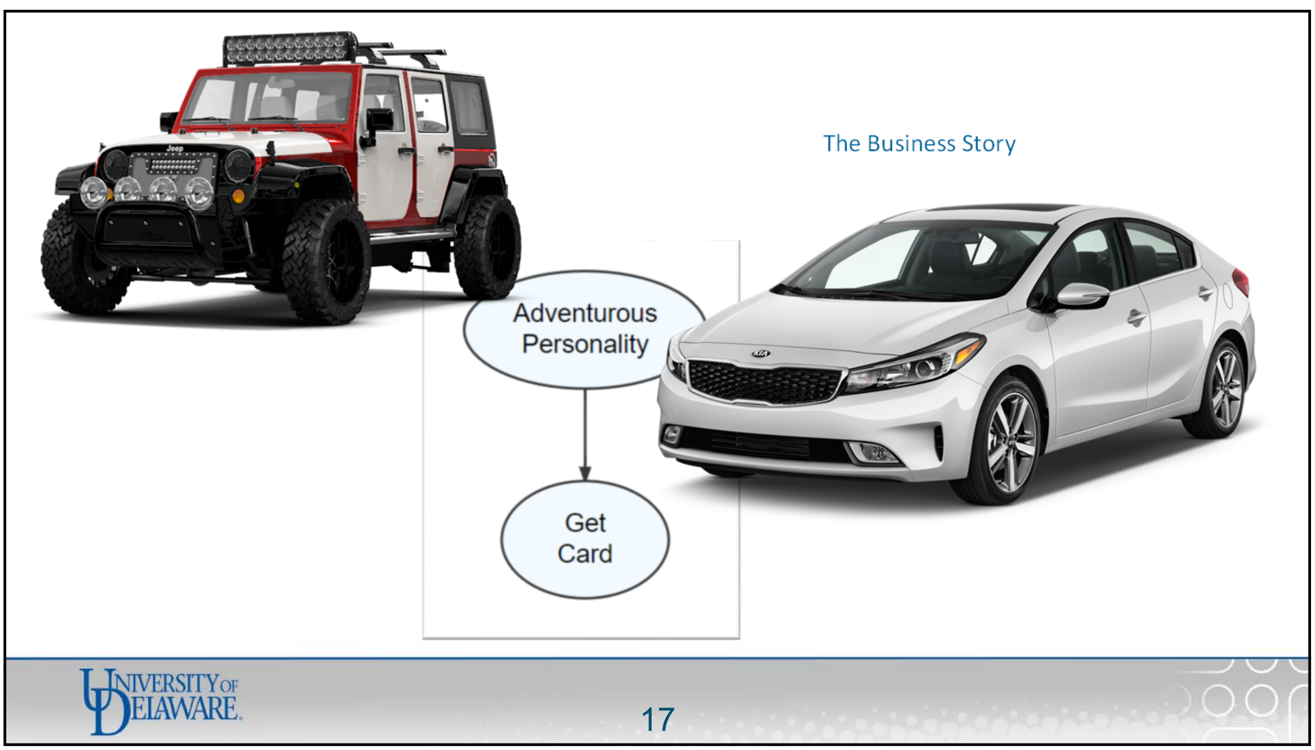
Using Joint Distributions

Joint Distributions

Making probabilistic statements over sets of random variables.

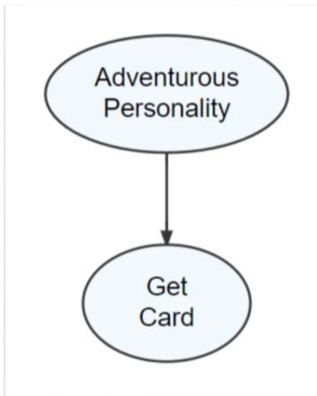



The Business Story
Adventurous Personality
Get Card
TRAVEL REWARDS CARD
5555 5555 5555
Exp: 11/2030
ADAM HACKER
BankPass
UNIVERSITY OF DELAWARE
16


The Business Story
Adventurous Personality
Get Card
UNIVERSITY OF DELAWARE
17

The DAG

Making The Math Story

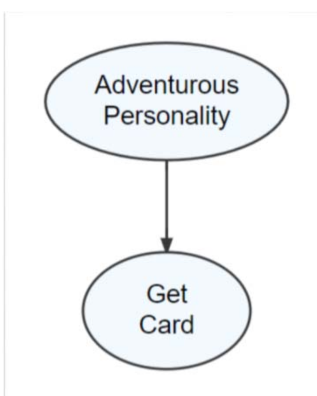


$$Y \equiv \begin{cases} 1, & Customer\ Owns\ Car\ Model\ \#1 \\ 2, & Customer\ Owns\ Car\ Model\ \#2 \\ \vdots & \vdots \\ K, & Customer\ Owns\ Car\ Model\ \#K \end{cases}$$

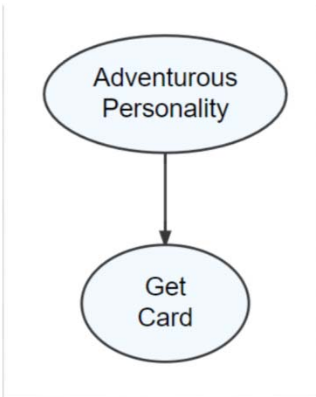
$$X \equiv \begin{cases} 0, & \text{customer does not get card} \\ 1, & \text{customer gets the card} \end{cases}$$

The DAG

Joint Distributions Are The Gold Standard of Data Analysis



X	Y	$P(X,Y)$
no	car model 1	??
no	car model 2	??
$\vdots$	$\vdots$	$\vdots$
no	car model K	??
yes	car model 1	??
yes	car model 2	??
$\vdots$	$\vdots$	$\vdots$
yes	car model K	??



Joint Distribution

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

UNIVERSITY OF DELAWARE

20

Joint Distribution

plug in values, get a probability

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

What is the probability that a potential customer drives a Jeep Wrangler and gets the card?

UNIVERSITY OF DELAWARE

21

Joint Distribution

marginal distribuiton

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

What is the probability that a potential customer drives a Jeep Wrangler?

Joint Distribution

marginal distribuiton

$$P(Y) = \sum_{x \in X} P(X = x, Y = y)$$

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

What is the probability that a potential customer drives a Jeep Wrangler?

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	1
1	Kia Forte	

Joint Distribution

marginal distribuiton

What is the probability that a potentiatl customer drives a Jeep Wrangler?

$$P(Y) = \sum_{x \in X} P(X = x, Y = y)$$

$$P(Y = Jeep) = \sum_{x \in X} P(X = x, Y = Jeep)$$

$$= P(X = 0, Y = Jeep) + P(X = 1, Y = Jeep)$$

$$= 5\% + 15\%$$

$$= 20\%$$

Joint Distribution

conditional distribuiton

x	y	$P(x, y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

What is the probability that a potential customer gets the card given that they drive a Jeep Wranlger?

Joint Distribution

conditional distribuiton

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

x	y	$P(x,y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	10%
1	Kia Forte	1%

What is the probability that a potential customer gets the card given that they drive a Jeep Wranlger?

UNIVERSITY OF DELAWARE

26

x	y	$P(x,y)$
0	Toyota Corolla	50%
0	Jeep Wrangler	5%
0	Subaru Outback	5%
0	Kia Forte	4%
1	Toyota Corolla	10%
1	Jeep Wrangler	15%
1	Subaru Outback	
1	Kia Forte	

What is the probability that a potential customer gets the card given that they drive a Jeep Wranlger?

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$$P(X=1|Y=Jeep) = \frac{P(X=1,Y=Jeep)}{P(Y=Jeep)}$$

$$= \frac{15\%}{P(X=0,Y=Jeep)+P(X=1,Y=Jeep)}$$

$$= \frac{15\%}{5\%+15\%} = \frac{15\%}{20\%} = 75\%$$

UNIVERSITY OF DELAWARE

x	y	*Count of Instances From Representative Sample*
0	Toyota Corolla	312
0	Jeep Wrangler	5
0	Subaru Outback	10
0	Kia Forte	20
1	Toyota Corolla	167
1	Jeep Wrangler	45
1	Subaru Outback	5
1	Kia Forte	5

Representative Sample Joint Distribution

What is the probability that a potential customer gets the card given that they drive a Jeep Wranlger?

Class Exercise