Lesson 13

Tuesday 3/12/24

Inference

Conclusions = Data + Assumptions (Manski, 2011)

Applications

Extrapolate

Estimate

Express uncertainty

Probability

Relative frequency version

$$p(\text{event}) = \frac{\text{\# of times an event occurs}}{\text{\# of times the event could have occurred}}$$

the proportion or fraction of times we would expect an event to occur (expressed on a [0,1] scale.

Example: probability that someone drawn at random from the population gets arrested at least once by age 25.

Probability

Degree of belief version

p(event) = the likelihood or chance an event occurs expressed on the [0,1] scale.

Example: a jury decides that someone is liable in a civil lawsuit by a preponderance of the evidence (meaning it is more probable than not that the defendant is liable).

Bounding Rule

Probabilities must be in the [0,1] interval; a probability of zero means the event is impossible; a probability of one means the event is certain.

<u>Example</u>: p(age at prison release = 8) is zero; means there is no chance that someone age 8 could be released from prison.

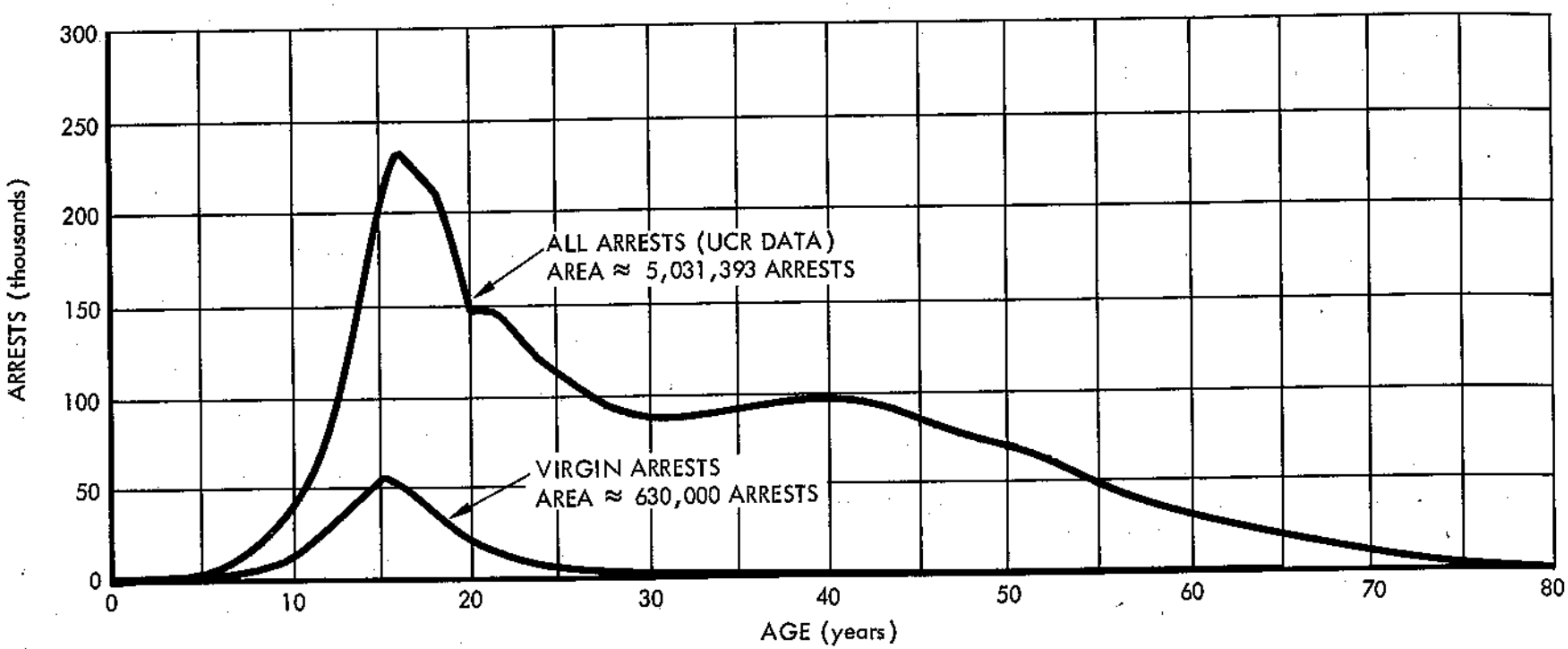
Complements

If p(a) is the probability that event *a* occurs, then p(not a) is 1 - p(a). We say that p(not a) is the complement of p(a).

Example: A sample of people were all arrested at age 16. For each of these people, the arrest can be classified as either a "first-time" arrest or a "recidivist" arrest (but not both). We can say that p(recidivist) is the complement of p(first time).

Age and Arrest (Christensen's Chart, 1967)

FIGURE J-2. 1965 ARRESTS BY AGE FOR ALL NONTRAFFIC OFFENSES



Restricted Addition Rule

If *a* and *b* are mutually exclusive events then $p(\text{a}) + p(\text{b}) = p(\text{a+b})$. In words, $p(\text{a+b})$ is the probability that *a* or *b* occurs → p(a or b)

Example: A sample of people were all arrested at age 16. For each of these people, the arrest can be classified as a "first-time" arrest or a "recidivist" arrest. These two types of arrests are mutually exclusive and exhaustive. So, *p*(first time) + *p*(recidivist) = 1. In words, we can say that the probability that an arrest is either a first-time or a recidivist arrest is 1.

General Addition Rule

If *a* and *b* are <u>not</u> mutually exclusive events then:

$$p(\text{a+b}) = p(\text{a}) + p(\text{b}) - p(\text{a and b})$$

<u>Example</u>: A sample of people are released from prison. Each of these people had been sentenced to prison for one or more of the following 4 offense types: (1) violent; (2) property; (3) drug; or (4) other offense type. What is the probability that someone drawn at random from this sample had been serving time for <u>either</u> a violent <u>or</u> property offense?

$$p(\text{v+pr}) = p(\text{v}) + p(\text{pr}) - p(\text{v and pr})$$

Numerical Example of General Addition Rule

Example: consider a sample of people who have been convicted of domestic violence. We follow each of these people for 3 years and document arrests for new crimes against the same victim. Here are our data:

	Violent=No	Violent=Yes	Total
Property=No	51	35	86
Property=Yes	36	18	54
Total	87	53	140

What is the probability that someone drawn at random was arrested for either a violent or a property offense?

Numerical Example of General Addition Rule (Continued)

	Violent= No	Violent= Yes	Total
Property=No	51	35	86
Property=Yes	36	18	54
Total	87	53	140

p(violent) = 53/140 = 0.379

p(property) = 54/140 = 0.386

p(v & pr) = 18/140 = 0.129

p(v+pr) = 53/140 + 54/140 - 18/140 =

= 0.379 + 0.386 - 0.129 = 0.636

Union and Intersection Notation

$$p(a+b) = p(a) + p(b) - p(a\ \underline{\text{and}}\ b)$$

$$=$$

$$p(a \cup b) = p(a) + p(b) - p(a \cap b)$$

union
or

intersection
and

Restricted Multiplication Rule

If events *a* and *b* are independent, then $p(a \text{ and } b) = p(a) \times p(b)$

a = fair coin flip → heads
b = fair coin flip → tails

$p(a \text{ and } b) = p(HT) = 0.5 \times 0.5 = 0.25$

Simulation Demonstrating Validity of Restricted Multiplication Rule

Advanced – not on exam

```
> a <- vector()
> b <- vector()
>
> for(i in 1:10000000){
+   a[i] <- ifelse(runif(n=1)>0.5,"H","T")
+   b[i] <- ifelse(runif(n=1)>0.5,"H","T")
+   }
>
> table(a,b)
   b
a         H       T
  H 2501200 2500166
  T 2498130 2500504
> mean(a=="H")*mean(b=="T")
[1] 0.2501018
>
```

Textbook, pp. 161-162

Experiment: flip a fair coin two times and count the number of heads arising from the 2 flips. The sample space for this experiment is: 0 heads, 1 head, 2 heads.

Repeat this experiment k times.

	k = 10	k = 100	k = 1000	k = 10,000	k = 100,000
# heads = 0	3	20	262	2472	24919
# heads = 1	5	59	504	5036	50074
# heads = 2	2	21	234	2492	25007

Joint Probability Assuming Independence

	Weapon = No	Weapon = Yes	Total
Crime = No	335	11	346
Crime = Yes	70	16	86
Total	405	27	432

p(carry a weapon <u>and</u> no criminal involvement)

Restricted Multiplication Rule:

$p(\mathrm{w}\ \&\ \mathrm{c}) = p(\mathrm{w}) \times p(\mathrm{nc}) = 27/432 \times 346/432$

$= 0.063 \times 0.801 = 0.050$

General Multiplication Rule

	Weapon = No	Weapon = Yes	Total
Crime = No	335	11	346
Crime = Yes	70	16	86
Total	405	27	432

p(carry a weapon and no criminal involvement)

$$p(a \ \& \ b) = p(a) \times p(b|a)$$

$$p(w \ \& \ nc) = p(w) \times p(nc|w) = 27/432 \times 11/27$$

$$= 0.063 \times 0.407 = 0.026$$

Binomial Probability Distribution

$$p(y = r) = \binom{N}{r} p^r q^{(N-r)}, \text{for } r = 0, 1, \ldots \text{ N}$$

which can also be written as:

$$p(y = r) = \frac{N!}{r!(N-r)!} p^r q^{(N-r)}, \text{for } r = 0, 1, \ldots \text{ N}$$

where *p* is the probability of the event of interest occurring on any given experiment, *q* = 1 - *p*, *y* is the observed number of events, *r* represents the outcomes in the sample space, and *N* is the number of experiments.

Coin Flipping Experiment Revisited

Experiment: flip a fair coin $N = 2$ times and count the number of heads arising from the 2 flips. The sample space for this experiment is: r=0 heads, r=1 head, and r=2 heads.

$$p(y=0) = \frac{2!}{0!(2-0)!} 0.5^r 1 - 0.5^{(2-0)} = 0.5^2 = 0.25$$

$$p(y=1) = \frac{2!}{1!(2-1)!} 0.5^1 1 - 0.5^{(2-1)} = 2 \times 0.5 \times 0.5 = 0.5$$

$$p(y=2) = \frac{2!}{2!(2-2)!} 0.5^2 1 - 0.5^{(2-2)} = 0.5 \times 0.5 = 0.25$$

Compare to what we got before...