# Math 140
# Introductory Statistics

Next test on Oct 19th

# Health care in America

About 30% of young American adults ages 19 to 29 don’t have health insurance.

Suppose you take a random sample of ten American adults in this age group.

What is the probability that at least one of them doesn’t have health insurance?

Take ten people, probability at least ONE does not have h.I?

# Let's think

P(at least one DOES NOT have h.i) + P(all have it) = 1

This means, by moving over
P(all have it) to the other side

P(at least one DOES NOT have health insurance) =

1 - P(all have it)

# Let's think

P(at least one DOES NOT have health insurance) =

1 - P(all have it) =

1 - P(1st has it AND 2nd has it AND .. 10th has it)

# Let's think

P(at least one DOES NOT have health insurance) =

1 - P(all have it) =

1 - P(1st has it AND 2nd has it AND .. 10th has it) =

1 - P(1st has it) * P(2nd has it)… *P(10th has it)

Since they are independent

# Let's think

P(at least one DOES NOT have health insurance) =

1 - P(all have it) =

1 - P(1st has it AND 2nd has it AND .. 10th has it) =

1 - P(1st has it) * P(2nd has it)… *P(10th has it) =

1 - 0.7 * 0.7 * 0.7 … *0.7

ten times =

$1 - (0.7)^{10}$

= 0.972

# A sad story - Sally Clark

2 of her kids died of sudden infant death syndrome
Assume these are independent events and calculate

P(baby 1 died AND baby 2 dies)

Assuming P(baby dies) = 1/8500

# A sad story - Sally Clark

If the events were independent

P(baby 1 died AND baby 2 dies) =
P(baby 1 died) * P(baby2 died | baby 1 died)
= P(baby 1 died) * P(baby 2 died) =
1/8500 * 1/8500

1 in 70 million

In the UK there are only about
200,000 second births per year

She was sentenced to life in prison

# A sad story - Sally Clark

The Royal Statistical Society of the UK
argued that two babies dying in the same family
ARE NOT independent

and concluded that the previous analysis does not apply.

P(baby 1 died and baby 2 dies) =
P(baby 1 died) * P(baby2 died | baby 1 died)
= 1/8500 * 1/100

This translates to one or two per year for the UK data

# A sad story - Sally Clark

Sally Clark was released from prison

She died after 4 years.

Her family says she never recovered from the miscarriage of justice.

# 6.1 Probability distributions

Probability distribution =
Possible outcomes of a chance process

The probability distribution allows us to find probabilities for any outcome

We have three ways of specifying a population:

1. List of all (individual) units
2. Frequency Table
3. Relative Frequency or Proportion Table

Mean? SD?

# List of units

| Number | Type | Value $x$ | $x - \mu$ |
|---|---|---|---|
| 1 | Penny | 1 ¢ | -3 |
| 2 | Penny | 1 ¢ | -3 |
| 3 | Penny | 1 ¢ | -3 |
| 4 | Penny | 1 ¢ | -3 |
| 5 | Penny | 1 ¢ | -3 |
| 6 | Nickel | 5 ¢ | 1 |
| 7 | Nickel | 5 ¢ | 1 |
| 8 | Nickel | 5 ¢ | 1 |
| 9 | Dime | 10 ¢ | 6 |
| 10 | Dime | 10 ¢ | 6 |
| | Total = 10 coins | Sum = 40 cents | |

$$\mu = \text{population mean} = \frac{\sum x}{n}$$

$$\mu = \frac{1+1+1+1+1+5+5+5+10+10}{10} = 4$$

# List of units

| Number | Type | Value $x$ | $x - \mu$ |
|---|---|---|---|
| 1 | Penny | 1 ¢ | -3 |
| 2 | Penny | 1 ¢ | -3 |
| 3 | Penny | 1 ¢ | -3 |
| 4 | Penny | 1 ¢ | -3 |
| 5 | Penny | 1 ¢ | -3 |
| 6 | Nickel | 5 ¢ | 1 |
| 7 | Nickel | 5 ¢ | 1 |
| 8 | Nickel | 5 ¢ | 1 |
| 9 | Dime | 10 ¢ | 6 |
| 10 | Dime | 10 ¢ | 6 |
| | Total = 10 coins | Sum = 40 cents | |

$$\mu = \text{population mean} = \frac{\sum x}{n}$$

$$\mu = \frac{1+1+1+1+1+5+5+5+10+10}{10} = 4$$

$$\sigma_n = \text{SD} = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

$$\sigma_n = \sqrt{\frac{9+9+9+9+9+1+1+1+36+36}{10}} =$$

$$\sigma_n = \sqrt{\frac{120}{10}} = \sqrt{12} \approx 3.4641$$

# Make list from data

| | | Second Die | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| First Die | 1 | 1, 1 | 1, 2 | 1, 3 | 1, 4 | 1, 5 | 1, 6 |
| | 2 | 2, 1 | 2, 2 | 2, 3 | 2, 4 | 2, 5 | 2, 6 |
| | 3 | 3, 1 | 3, 2 | 3, 3 | 3, 4 | 3, 5 | 3, 6 |
| | 4 | 4, 1 | 4, 2 | 4, 3 | 4, 4 | 4, 5 | 4, 6 |
| | 5 | 5, 1 | 5, 2 | 5, 3 | 5, 4 | 5, 5 | 5, 6 |
| | 6 | 6, 1 | 6, 2 | 6, 3 | 6, 4 | 6, 5 | 6, 6 |

Construct the probability distribution for

1) The sum of the two dice
2) The larger number on the two dice

# List for the Sum of the data

| | | Second Die | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| First Die | 1 | 1, 1 | 1, 2 | 1, 3 | 1, 4 | 1, 5 | 1, 6 |
| | 2 | 2, 1 | 2, 2 | 2, 3 | 2, 4 | 2, 5 | 2, 6 |
| | 3 | 3, 1 | 3, 2 | 3, 3 | 3, 4 | 3, 5 | 3, 6 |
| | 4 | 4, 1 | 4, 2 | 4, 3 | 4, 4 | 4, 5 | 4, 6 |
| | 5 | 5, 1 | 5, 2 | 5, 3 | 5, 4 | 5, 5 | 5, 6 |
| | 6 | 6, 1 | 6, 2 | 6, 3 | 6, 4 | 6, 5 | 6, 6 |

| Possibilities | Probability | |
|---|---|---|
| Sum = 2 | (1,1) | 1/36 |
| Sum = 3 | (1,2) or (2,1) | 2/36 |

You do the rest

# List for the Sum of the data

If we add them we should always get 1, since this represents all possibilities

| Sum of Two Dice, $x$ | Probability, $P$ |
|---|---|
| 2 | 1/36 |
| 3 | 2/36 |
| 4 | 3/36 |
| 5 | 4/36 |
| 6 | 5/36 |
| 7 | 6/36 |
| 8 | 5/36 |
| 9 | 4/36 |
| 10 | 3/36 |
| 11 | 2/36 |
| 12 | 1/36 |
| Total | 1 |

# Do also for larger number

| Larger number | | Probability |
|---|---|---|
| Is 1 | (1,1) | 1/36 |
| Is 2 | (1,2),(2,1),(2,2) | 3/36 |

You do the rest

# Do also for larger number

| Larger Number, $x$ | Probability, $p$ |
|---|---|
| 1 | 1/36 |
| 2 | 3/36 |
| 3 | 5/36 |
| 4 | 7/36 |
| 5 | 9/36 |
| 6 | 11/36 |
| Total | 1 |

# We can calculate

Probability that the sum of number is 3 = 2/36
Probability that the larger number is 3 = 5/36
Etc etc

What we get after tossing the dice is a
random variable
depends on chance - may change from trial to trial

We call it X.
For example, if we care for the SUM of numbers

$P(X=3) = 2/36 = 1/18$
$P(X=7) = 6/36 = 1/6$

# Smoking and Lung cancer

| Lung Cancer Cases | Proportion |
|---|---|
| Smoking responsible | 0.87 |
| Smoking not responsible | 0.13 |

Suppose two lung cancer patients are randomly selected

What is the probability distribution of

X- the number of patients with lung cancer caused by smoking

# Smoking and Lung cancer

For 2 sick people, either smoking was cause of disease or not

4 possibilities

| | |
|---|---|
| Not caused by smoking | Not caused by smoking |
| Not caused by smoking | Caused by smoking |
| Caused by smoking | Not caused by smoking |
| Caused by smoking | Caused by smoking |

# Smoking and Lung cancer

Not caused by smoking

P=0.13

Not caused by smoking

P= 0.13

Independent events

P(both patients had cancer not caused by smoking) =

0.13*0.13 = 0.0169 ~ less than 2%

# Smoking and Lung cancer

| Number Caused by Smoking, $x$ | Probability, $p$ |
|---|---|
| 0 | |
| 1 | |
| 2 | |

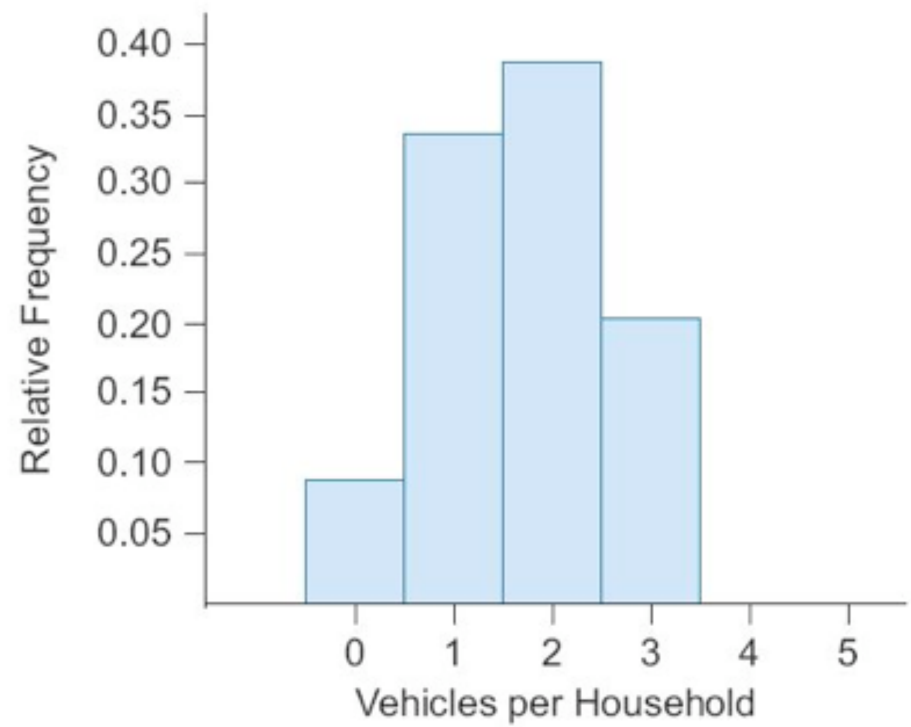
You fill it out

# Smoking and Lung cancer

| Number Caused by Smoking, $x$ | Probability, $p$ |
|---|---|
| 0 | 0.0169 |
| 1 | $0.1131 + 0.1131 = 0.2262$ |
| 2 | 0.7569 |

# Building a parking lot

| Vehicles per Household, *x* | Proportion of Households, *p* |
|---|---|
| 0 | 0.087 |
| 1 | 0.331 |
| 2 | 0.381 |
| 3 | 0.201 |




What is the probability that a home will have two or more cars?
( Assume no one has 4 )

# Building a parking lot

| Vehicles per Household, *x* | Proportion of Households, *p* |
|---|---|
| 0 | 0.087 |
| 1 | 0.331 |
| 2 | 0.381 |
| 3 | 0.201 |

What is the probability that a home will have two or more cars?

$P(X=2) = 0.381+0.201 = 0.582$

How about calculating the probability
That two randomly selected homes have NO cars?

# Building a parking lot

P(two randomly selected homes have NO cars) =

P(1st 0 cars) * P(2nd 0 cars | 1st 0 cars)=

Independent events =

P(1st 0 cars) * P(2nd 0 cars)
= 0.087 * 0.087 = 0.008

Less than 1%

| Vehicles per Household, $x$ | Proportion of Households, $p$ |
|---|---|
| 0 | 0.087 |
| 1 | 0.331 |
| 2 | 0.381 |
| 3 | 0.201 |

# Building a parking lot

P(exactly one car in a duplex) =
Take two homes, one has a car, the other has zero cars

= P(1 car in 1st house AND 0 cars in 2nd house OR
0 cars in first house AND 1 car in 2nd house)

A = 1 car in 1st house AND 0 cars in 2nd house
B = 0 cars in 1st house AND 1 car in 2nd house

# Building a parking lot

A = 1 car in 1st house AND 0 cars in 2nd house
B = 0 cars in 1st house AND 1 car in 2nd house

These are disjoined!

Recall P(A or B) = P(A) +P(B) - P(A and B)
Here P(A and B) = 0
They are disjoined

# Building a parking lot

P(1 car in 1st house AND 0 cars in 2nd house OR
0 cars in first house AND 1 car in 2nd house)

=

= P (1 car in 1st house AND 0 cars in 2nd house ) +
P (0 cars in first house AND 1 car in 2nd house)

P(1 car in 1st house AND 0 cars in 2nd house OR
0 cars in first house AND 1 car in 2nd house)

=

= P (1 car in 1st house AND 0 cars in 2nd house ) +
P (0 cars in first house AND 1 car in 2nd house)
(disjoined)

=

P(1 car) * P(0 cars) + P(0 cars) * P(1 car)
(independent)

=

0.331 * 0.087 + 0.087 * 0.331 = 0.058

# Make the full chart for duplexes

| Total Number of Vehicles, *x* | Probability, *p* |
|---|---|
| 0 | |
| 1 | 0.058 |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |

# Make the full chart for duplexes

| Total Number of Vehicles, *x* | Probability, *p* |
|---|---|
| 0 | 0.008 |
| 1 | 0.058 |
| 2 | 0.176 |
| 3 | 0.287 |
| 4 | 0.278 |
| 5 | 0.153 |
| 6 | 0.040 |

# Practice and hk

Page 284

P1, P2, P3, E1, E2, E3, E5, E4, E6, E7

Try E2 first

6 computers, 3 are broken, you get to sample only 2

Find $P(X=0)$, $P(X=1)$, $P(X=2)$

$X$= number of sampled computers that are broken