Part 1

Lecture 1b GAMES AND CONCEPTS

1

YATES' CORRECTION FACTOR FOR CONTINUITY

The curve superimposed over a Binomial histogram is the Gaussian probability distribution and can be used to approximate Binomial probabilities.

$P(B_4 \geq 3) = 0.2500 + 0.0625 = 0.3125$

$P(\text{Normal} \geq 3) = 0.1587$ Poor approximation

$P(\text{Normal} \geq 2.5) = 0.3085$ Good approximation

Jacob Bernoulli 1654 – 1705

A flip of a coin is called
a Bernoulli trial

Frank Yates 1902 – 1994

Introduced the Continuity
Correction of 0.5

GAUSSIAN PROBABILITY DENSITY FUNCTION

$$f(p) = \frac{e^{-\frac{(p-P)^2}{\left(\frac{2\times P\times Q}{n}\right)}}}{\sqrt{\frac{2\pi \times P \times Q}{n}}}$$

$$\pi = 3.142 \quad e = 2.716 \quad Q = 1 - P$$

p is the proportion of sick subjects in a random sample
P is proportion of sick subjects in a population sample

In our example the possible proportions were
0.000, 0.0625, 0.25, 0.375, 0.25, 0.0625, 1.000

Single cohort DESIGN

A researcher tests the curative ability of a new drug.

She follows 10 patients and records their **sex, age** and **weight** (outcome predictors).

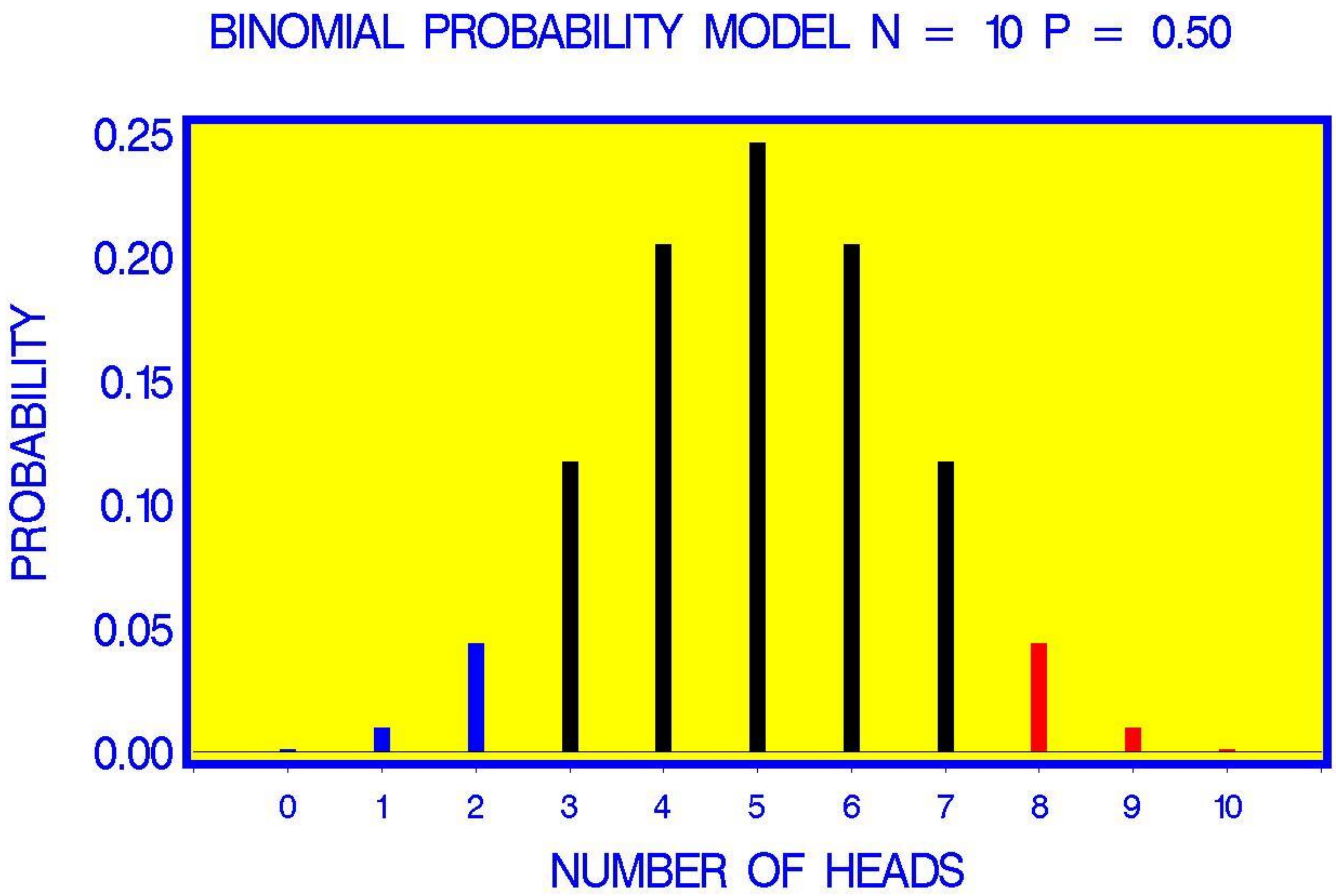
In 8 of the 10 patients the new drug is successful. In 2 patients the drug has no effect.

We begin by assuming that the probability of a cure is the same for each patient and is 0.5. This assumption is called the **NULL HYPOTHESIS** or the **CHANCE HYPOTHESIS**.

Therefore the probability of getting an 8/2 split can be obtained using the Binomial probability distribution with P = 0.5

Binomial Distribution n = 10 P = 0.5

h	$P(B_{10} = h)$	$P(B_{10} \le h)$		$P(B_{10} \ge h)$
0	0.00098	0.00098		1.00000
1	0.00977	0.01074		0.99902
2	0.04395	0.05469	<<	0.98926
3	0.11719	0.17188		0.94531
4	0.20508	0.37695		0.82813
5	0.24609	0.62305		0.62305
6	0.20508	0.82813		0.37695
7	0.11719	0.94531		0.17188
8	0.04395	0.98926	>>	0.05469
9	0.00977	0.99902		0.01074
10	0.00098	1.00000		0.00098


BINOMIAL PROBABILITY MODEL N = 10 P = 0.50
PROBABILITY
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10
NUMBER OF HEADS

```
TITLE1  "  FREQ PROCEDURE IS FIRST PROGRAM RUN  " ;

DATA  FINAL ; INPUT CURE $  N  @@ ;
DATALINES ;
NO   2  YES  8
RUN ;

PROC  FREQ  DATA = FINAL ; WEIGHT  N ;
TABLES  CURE  /  BINOMIAL  ;  EXACT  BINOMIAL  ;  RUN ;
```

CURE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	2	20.00	2	20.00
YES	8	80.00	10	100.00

Fundamental characteristics of a SAS program

(1) Each SAS statement ends in a semi-colon (;)

(2) The DATA statement gives the name “FINAL” to the dataset

(3) The INPUT statement defines the variables

- The CURE variable has character values – NO and YES, and must be followed by a $ sign
- The N variable is a Numeric Variable
- The data paragraph ends with a “RUN ;” statement

(5) The procedure paragraph starts with the word PROC . In this example the FREQ procedure is called to create a frequency table.

(6) The WEIGHT statement tells the program that 2 subjects were not cured and 8 patients were cured. If we did not use the WEIGHT statement then we would type in 10 lines of data, two NOs and 8 YESs.

(7) The TABLE statement defines a table of counts, in this case, a one dimensional table of 2 not cured and 8 cured patients.

(8) We will use the BINOMIAL probability distribution to calculate an exact p value. We are testing the hypothesis that the unknown probability P = 0.5, that is, the two drugs are of equal efficacy.