Sections 3.6, 4.1, and 4.2

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

3.6 Binomial random variable

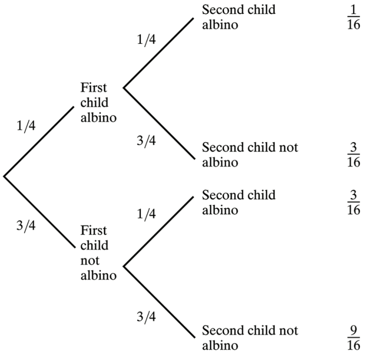
- $\boxed{\text{Independent-trials model}}$ A series of n independent trials is conducted. Each trial results in success or failure. The probability of success is equal to p for each trial, regardless of the outcomes of the other trials.
- The **binomial distribution** defines a discrete random variable Y that counts the number, out of the n trials, exhibiting a certain trait with probability p in the "independent trials model."

Example 3.6.1 Albinism

- If both parents carry the gene for being albino, each kid they have has a $p = 0.25$ chance of being albino. Each child has the same chance of being albino independent of whether the other children are albino.
- Let Y count the number of kids out of two that are albino. Y can be 0, 1, or 2.

Probability tree for albinism

Probability tree for albinism among two children of carriers of the gene for albinism.



Albino example, cont'd

- Let the four possible experimental outcomes for the first/second child be albino/albino, albino/not, not/albino, not/not.
- $Y = 0$ corresponds to not/not, $Y = 1$ corresponds to either albino/not or not/albino, and $Y = 2$ corresponds to albino/albino.
- $\Pr\{Y = 0\} = \Pr\{\text{not/not}\} = \frac{9}{16}$.
- $\Pr\{Y = 1\} = \Pr\{\text{albino/not}\} + \Pr\{\text{not/albino}\} = \frac{3}{16} + \frac{3}{16} = \frac{6}{16}$.
- $\Pr\{Y = 2\} = \Pr\{\text{albino/albino}\} = \frac{3}{16} + \frac{3}{16} = \frac{1}{16}$.

Probability distribution in tabular form

Table 3.6.1 Probability distribution for number of albino children

Number of		
Albino	*Nonalbino*	Probability
0	2	$\frac{9}{16}$
1	1	$\frac{6}{16}$
2	0	$\frac{1}{16}$
		1

Y is binomial with $p = 0.25$ and $n = 2$.

Binomial distribution formula

def'n A binomial random variable Y with probability p and number of trials n has the probability of j successes (and $n-j$ failures) given by

$$\Pr\{j \text{ successes}\} = \Pr\{Y = j\} = {}_nC_j \; p^j \; (1-p)^{n-j}.$$

The **binomial coefficient** ${}_nC_j$ counts the number of ways to order j "successes" and $n-j$ failures. For example, if $n = 4$ and $j = 2$ then ${}_4C_2 = 6$ because there's 6 orderings

$$SSFF \;\; SFSF \;\; SFFS \;\; FSSF \;\; FSFS \;\; FFSS$$

Binomial coefficient, formal definition

The **binomial coefficient is**

$$ {}_nC_j = \frac{n!}{j!(n-j)!} $$

where $x!$ is read "x factorial" given by

$$ x! = x(x-1)(x-2)\cdots(3)(2)(1). $$

The first few are

$$
\begin{aligned}
0! &= 1 \\
1! &= 1 \\
2! &= (2)(1) = 2 \\
3! &= (3)(2)(1) = 6 \\
4! &= (4)(3)(2)(1) = 24 \\
5! &= 120
\end{aligned}
$$

Binomial probabilities

- There are a lot of formulas on the previous slide.
- It's possible to compute probabilities like $\Pr\{Y = 2\}$ by hand using the formulas and Table 2 on p. 615.
- For Y binomial with n trials and probability p, R computes $\Pr\{Y = j\}$ easily using `dbinom(j,n,p)`
- Use R for your homework!

Example 3.6.4 Mutant cats!

- Study in Omaha, Nebraska found $p = 0.37$ have a mutant trait.
- Randomly draw $n = 5$ cats and count Y, the number of mutants.
- Y is binomial with $p = 0.37$ and $n = 5$. Let's have R find the probability of $Y = 0$, $Y = 1$, $Y = 2$, $Y = 3$, $Y = 4$, $Y = 5$:

```
> dbinom(0,5,0.37)
[1] 0.09924365
> dbinom(1,5,0.37)
[1] 0.2914298
> dbinom(2,5,0.37)
[1] 0.3423143
> dbinom(3,5,0.37)
[1] 0.2010418
> dbinom(4,5,0.37)
[1] 0.05903607
> dbinom(5,5,0.37)
[1] 0.006934396
```

Probability distribution for $n = 5$ and $p = 0.37$

Table 3.6.3 Binomial distribution with $n = 5$ and $p = 0.37$

Number of		
Mutants	Nonmutants	Probability
0	5	0.10
1	4	0.29
2	3	0.34
3	2	0.20
4	1	0.06
5	0	0.01
		1.00

$\boxed{\text{Questions}}$ What is $\Pr\{Y \leq 2\}$? $\Pr\{Y > 2\}$? $\Pr\{2 \leq Y \leq 4\}$?

Mean and standard deviation of binomial random variable

Let Y be binomial with n trials and probability p.

$$\mu_Y = n\ p$$
$$\sigma_Y = \sqrt{n\ p\ (1-p)}$$

Example: for mutant cats, $\mu_Y = 5(0.37) = 1.85$ cats and $\sigma_Y = \sqrt{5(0.37)(0.63)} = 1.08$ cats.

Coming up: normal distribution

- The binomial disribution is **discrete**. Since it is discrete, a binomial distribution is described with a simple table of probabilities.
- There are other widely used discrete distributions, including the Poisson and geometric random variables.
- The next random variable we will talk about is the most widely used of all random variables: the normal distribution.
- Unlike the binomial, the normal distribution is **continuous**, and therefore has a density.

Section 4.1 Normal curves

- "Bell-shaped curve"
- The normal density curve defines a continuous random variable Y.
- Normal curves approximate lots of real data densities (examples coming up).
- A normal curve is defined by the mean μ and standard deviation σ.
- We will also find that sample means $\bar{Y}$ are approximately normal in Chapter 5. So are sample proportions $\hat{p}$ (more later).
- Let's look at some real data examples...

Serum cholesterol in $n = 727$ 12–14 year-old children

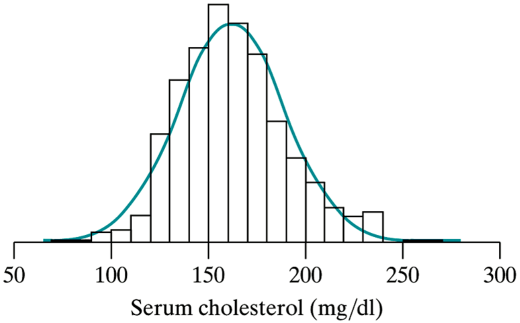


Figure 4.1.1 Distribution of serum cholesterol in 727 12- to 14-year-old children

Normal fit to cholesterol data with $\mu = 162$ mg/dl and $\sigma = 28$ mg/dl.

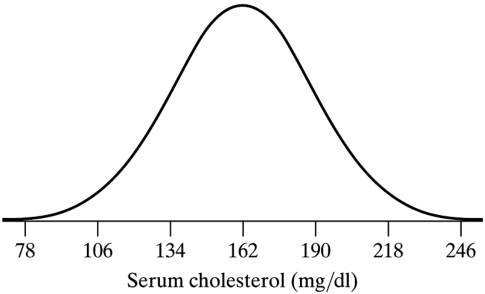
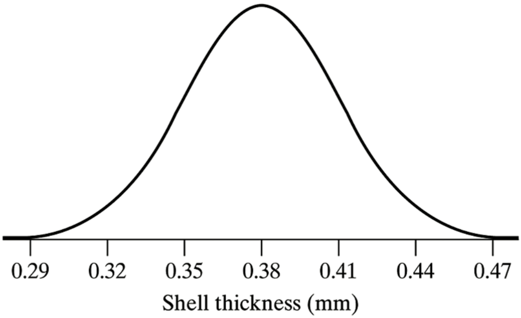


Figure 4.1.2 Normal distribution of serum cholesterol, with $\mu = 162$ mg/dl and $\sigma = 28$ mg/dl

Normal distribution of eggshell thickness

Shell thicknesses of White Leghorn hens. $\mu = 0.38$ mm & $\sigma = 0.03$ mm



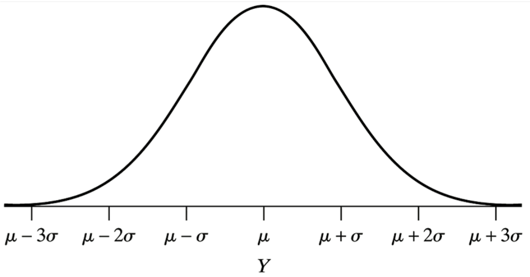
4.2 Normal density functions

- The density function is given by

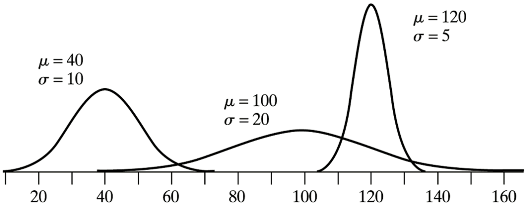
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- All normal curves have the same shape. They have a mode at μ and are more spread out – flatter – the larger σ is.
- Almost all of the probability is contained between $\mu - 3\sigma$ and $\mu + 3\sigma$.
- The area under every normal density is one.
- If Y has a normal density with mean μ and standard deviation σ, we can write $Y \sim N(\mu, \sigma)$.

Normal curve with mean μ and standard deviation σ



Three normal curves with different means and standard deviations



Discussion

- Introduced two random variables, binomial and normal. binomial is discrete, normal continuous.
- Binomial has a probability *table* with $\Pr\{Y = j\}$ for $j = 0, 1, \ldots, n$, normal has density function $f(x)$.
- Binomial sometimes written $Y \sim \text{bin}(n, p)$
- Normal sometimes written $Y \sim N(\mu, \sigma)$.
- R computes probabilities for both.
- Next lecture we'll discuss how to get probabilities for normal random variables.