Proportions

Chapter 7

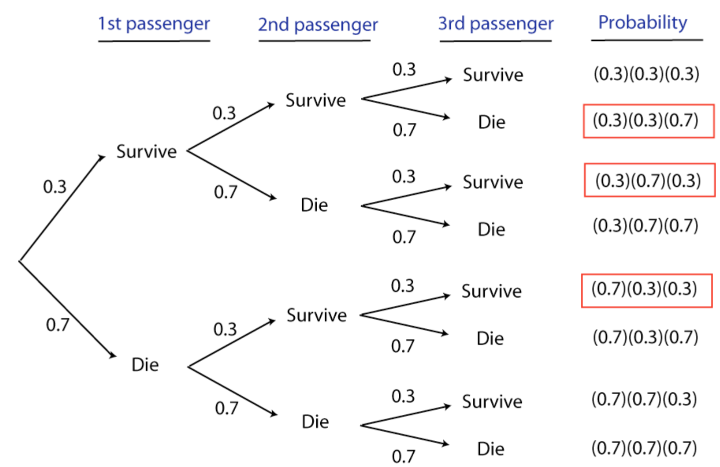
A *proportion* is the fraction of individuals having a particular attribute.

Example:
2092 adult passengers on the Titanic;

654 survived

Proportion of survivors = 654/2092
$\cong 0.3$

Probability that two out of three randomly chosen passengers survived the Titanic



Binomial distribution

n trials; p probability of success

The *binomial distribution* describes the probability of a given number of "successes" from a fixed number of independent trials.

$$\Pr[X] = \binom{n}{X} p^X (1-p)^{n-X}$$

$$\binom{n}{X} = \frac{n!}{X!(n-X)!}$$

n! = n × n–1 × n–2 ×... 3 × 2 × 1

6! = 6 × 5 × 4 × 3 × 2 × 1 = 720

0!=1
1!=1

Probability that two out of three randomly chosen passengers survived the Titanic

$$\Pr[2] = \binom{3}{2}(0.3)^2(1-0.3)^{3-2}$$

$$= \frac{3!}{2!\times 1!}(0.3)^2(0.7)^1$$

$$= 3(0.3)^2(0.7) = 0.189$$

Probability that two out of three randomly chosen passengers survived the Titanic

$$\Pr[2] = \binom{3}{2}(0.3)^2(1-0.3)^{3-2}$$

Number of ways to get 2 survivors out of 3 passengers

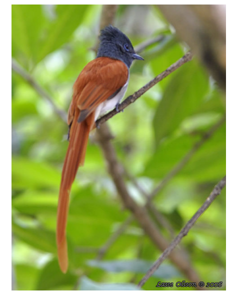
Probability of 2 survivors

Probability of 1 death

Example: Paradise flycatchers

A population of paradise flycatchers has 80% brown males and 20% white. Your field assistant captures 5 male flycatchers at random.

What is the chance that 3 of those are brown and 2 are white?



$p = 0.8$ $\quad n = 5$ $\quad X = 3$

$$\Pr[3] = \binom{5}{3} 0.8^3(1-0.8)^{5-3} = \frac{120}{6\times 2} 0.8^3(0.2)^2 = 0.205$$

In R:

```
> dbinom(x = 3, size = 5, prob = 0.8)
[1] 0.2048
```

Try at home:

What is the probability that 3 or more are brown?

Try at home:

What is the probability that 3 or more are brown?

$$\Pr[3\ \textit{or more are brown}] = \Pr[3] + \Pr[4] + \Pr[5]$$
$$= 0.205 + 0.410 + 0.328$$
$$= 0.943$$

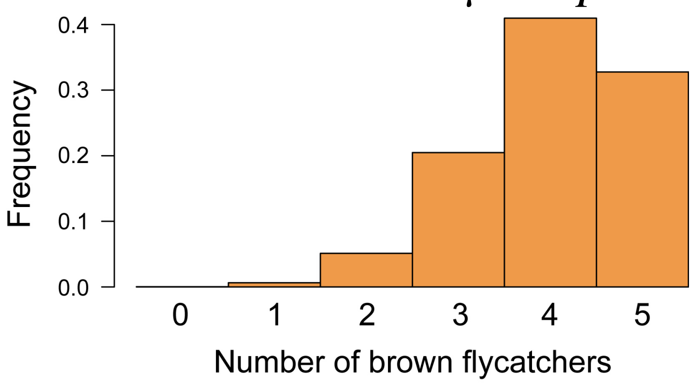
Properties of the binomial distribution: Mean and variance of number of successes

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

Binomial distribution for $p = 0.8$, $n = 5$

$$\mu = np$$



Proportion of successes in a sample

$$\hat{p} = \frac{X}{n}$$

The hat (^) shows that this is an estimate of p.

Properties of sample proportions

Mean: p

Variance: $\dfrac{p(1-p)}{n}$

Estimating a proportion

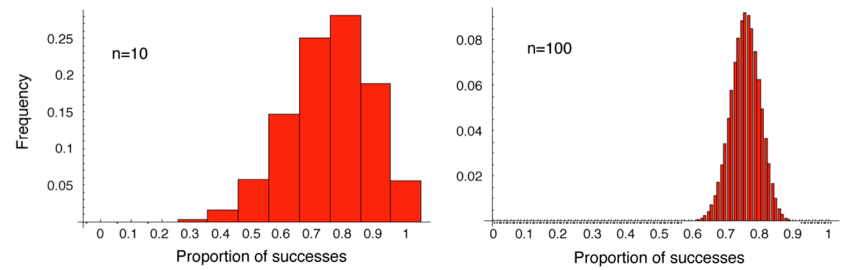
$$\hat{p} = \frac{X}{n}$$

Number of “successes” over total sample size

Standard error of the estimate of a proportion

$$\mathrm{SE}[\hat{p}] = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A larger sample has a lower standard error



95% confidence interval for a proportion

$$p' = \frac{X+2}{n+4}$$

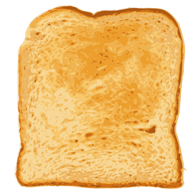
$$\left(p' - 1.96\sqrt{\frac{p'(1-p')}{n+4}}\right) \le p \le \left(p' + 1.96\sqrt{\frac{p'(1-p')}{n+4}}\right)$$

This is the Agresti-Coull confidence interval.

Example: Murphy's Law—Toast butter-side down

British students dropped 9821 pieces of toast. 6101 of these landed butter-side down; 3720 landed butter-side up.

How often does toast land butter-side down, as "predicted" by Murphy's Law?



https://www.goodfreephotos.com/

http://www.counton.org/thesum/issue-07/issue-07-page-05.htm

Example: Murphy's Law—Toast butter-side down

British students dropped 9821 pieces of toast. 6101 of these landed butter-side down; 3720 landed butter-side up.

How often does toast land butter-side down, as "predicted" by Murphy's Law?

$$\hat{p} = \frac{6101}{9821} = 0.621$$

Example: Murphy's Law—Toast butter-side down

What is the 95% confidence interval for this estimate?

$$p' = \frac{X+2}{n+4} = \frac{6101+2}{9821+4} = 0.621$$

Example:
Murphy's Law—Toast butter-side down

What is the 95% confidence interval for this estimate?

$$p' = \frac{X+2}{n+4} = \frac{6101+2}{9821+4} = 0.621$$

$$p' - 1.96\sqrt{\frac{p'(1-p')}{n+4}} < p < p' + 1.96\sqrt{\frac{p'(1-p')}{n+4}}$$

$$0.621 - 1.96\sqrt{\frac{0.621(1-0.621)}{9821+4}} < p < 0.621 + 1.96\sqrt{\frac{0.621(1-0.621)}{9821+4}}$$

$$0.611 < p < 0.631$$

Agresti-Coull confidence intervals in R

```
> library(binom)
> binom.confint(x = 6101, n = 9821, method = "ac")
```

```
         method    x    n      mean     lower     upper
1 agresti-coull 6101 9821 0.6212198 0.6115804 0.6307645
```

Hypothesis testing on proportions

The binomial test

Binomial test

The binomial test uses data to test whether a population proportion *p* matches a null expectation for the proportion.

H_0: The relative frequency of successes in the population is p_0 .

H_A: The relative frequency of successes in the population is not p_0 .

Example

Human volunteers were asked to try five types of meat products. On of these was dog food; the others were various types of paté. The volunteers were asked which of the five was their favorite.

Two out of 18 volunteers chose the dog food. Is there evidence that people prefer human food over dog food?

Bohannon, J., R. Goldstein, and A. Herschkowitsch. 2010. Can people distinguish pâté from dog food? *Chance* 23: 43–46.

Hypotheses

H_0: Dog food is chosen as best 20% of the time.

H_0: $p_0 = 0.2$.

H_A: Dog food is chosen as best different from 20% of the time.

H_A: $p_0 > 0.2$

$N = 18$, $p_0 = 0.2$,
$X = 2$ in the sample chose dog food

$$P = 2(\Pr[2] + \Pr[1] + \Pr[0])$$

$$= 2\left(\binom{18}{2}(0.2)^2(1-0.2)^{16} + \binom{18}{1}(0.2)^1(1-0.2)^{17} + \binom{18}{0}(0.2)^0(1-0.2)^{18}\right)$$

$$= 2(0.1722 + 0.0811 + 0.0180) = 0.543$$

Note: The capital *P* here is used for the *P*-value, in contrast to the population proportion with a small *p*.

$P = 0.54$

This is greater than the α value of 0.05, so we would *not* reject the null hypothesis.

It is plausible that people do not prefer paté over dog food.

Binomial test in R

```
binom.test(2, n = 18, p = 0.2)
```

```
     Exact binomial test

data:  2 and 18
number of successes = 2, number of trials = 18,
     p-value = 0.555
alternative hypothesis: true probability of
     success is not equal to 0.2
95 percent confidence interval:
 0.01375122 0.34712044
sample estimates:
probability of success
             0.1111111
```