

Sampling Distributions

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison
Statistics 371

6th October 2005

Small Example

- ▶ Consider a cross of two heterozygotes, $Aa \times Aa$.
- ▶ The probability distribution of the genotypes of the offspring is as follows.

	Offspring Genotype		
cross	AA	Aa	aa
$Aa \times Aa$	0.25	0.50	0.25

- ▶ In a sample of size $n = 2$, let:

$$Y = \#\{\text{dominant offspring (AA or Aa)}\}$$

- ▶ $\hat{p} = Y/n$ is called the **sample proportion**.
- ▶ The sampling distribution is:

Y	\hat{p}	Prob.
0	0.0	0.0625
1	0.5	0.3750
2	1.0	0.5625

Navigation icons

Navigation icons

Sampling Distributions

- ▶ A **population** may be modeled as a box with numbered or colored balls.
- ▶ We think of a **sample** of data as having been selected at random from this population.
- ▶ From each sample, we can calculate a **sample statistic** such as a sample mean.
- ▶ The **sampling distribution** of the sample mean is the collection of all possible sample means that could occur by random sampling (of a given sample size n).
- ▶ The textbook refers to the thought exercise of considering all the ways a sample could have turned out as a **meta-experiment**.

Navigation icons

Larger Example

1. For the previous cross, what is the probability that exactly 15 of 20 offspring are dominant?
 - ▶ The number of dominant offspring Y will have a binomial distribution with $n = 20$ and $p = 0.75$.

$$\mathbb{P}\{Y = 15\} = \frac{20!}{15!5!} (0.75)^{15} (0.25)^5 \doteq 0.2023$$

- ▶ In R,


```
> dbinom(15, 20, 0.75)
[1] 0.2023312
```

Navigation icons

1. What is the probability that \hat{p} is within 0.05 of p ?

- ▶ Translate the probability to a binomial question.

$$\begin{aligned}
 \mathbb{P}\{0.70 \leq \hat{p} \leq 0.80\} &= \mathbb{P}\{0.70 \leq Y/20 \leq 0.80\} \\
 &= \mathbb{P}\{14 \leq Y \leq 16\} \\
 &= \mathbb{P}\{14\} + \mathbb{P}\{15\} + \mathbb{P}\{16\} \\
 &\doteq 0.5606
 \end{aligned}$$

▶ In R,

```

> sum(dbinom(14:16, 20, 0.75))
[1] 0.5606259

```

- ▶ Now consider a population where each individual is associated with a quantitative variable.
- ▶ We can compute the sample mean from each sample.
- ▶ The **sampling distribution of \bar{Y}** is the collection of sample means from the meta-experiment of all possible samples of size n .

Increasing Sample Size

- ▶ How does $\mathbb{P}\{0.70 \leq \hat{p} \leq 0.80\}$ change as the sample size n increases?
- ▶ Here is a table.

n	$\mathbb{P}\{0.70 \leq \hat{p} \leq 0.80\}$
20	0.5606
40	0.6389
80	0.7554
160	0.8799
320	0.9671
640	0.997
1280	1
2560	1

Sampling Distribution of \bar{Y}

- ▶ The mean of the sampling distribution of \bar{Y} , $\mu_{\bar{Y}}$, is the same as the population mean. In symbols,

$$\mu_{\bar{Y}} = \mu .$$

- ▶ The standard deviation of the sampling distribution of \bar{Y} , $\sigma_{\bar{Y}}$, is smaller than the population standard deviation by a factor of \sqrt{n} . In symbols,

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} .$$

- ▶ If the sample size n is sufficiently large, the shape of the sampling distribution of \bar{Y} will be approximately normal. This is the **Central Limit Theorem**.
- ▶ If the population is normal, a sample size of 1 suffices.
- ▶ If the population is not normal, it depends on how the population differs from normality to determine if the normal approximation is reasonably accurate.

Example calculation

Suppose that the weights of seeds are approximately normal with a mean of 500 mg and a standard deviation of 150 mg. Find the probability that the sample mean is between 450 and 550 for a variety of sample sizes.

For $n = 4$, we have

$$\begin{aligned}\mathbb{P}\{450 \leq \bar{Y} \leq 550\} &= \mathbb{P}\left\{\frac{450 - 500}{150/\sqrt{4}} \leq \frac{\bar{Y} - 500}{150/\sqrt{4}} \leq \frac{550 - 500}{150/\sqrt{4}}\right\} \\ &= \mathbb{P}\{-0.67 \leq Z \leq 0.67\} \\ &= 0.5028\end{aligned}$$

from a normal table calculation.



Exercise 5.18

Assume that height of corn plants are normally distributed with a mean 145 cm and a standard deviation of 22 cm.

What proportion of plants are between 135 and 155 cm?

```
> pnorm(155, 145, 22) - pnorm(135, 145, 22)
```

```
[1] 0.3505637
```

Find $\mathbb{P}\{135 \leq \bar{Y} \leq 155\}$ when $n = 16$.

```
> pnorm(155, 145, 22/sqrt(16)) - pnorm(135, 145, 22/sqrt(16))
```

```
[1] 0.9309637
```

Find $\mathbb{P}\{135 \leq \bar{Y} \leq 155\}$ when $n = 36$.

```
> pnorm(155, 145, 22/sqrt(36)) - pnorm(135, 145, 22/sqrt(36))
```

```
[1] 0.993614
```

