Friday, Feb 3

Deriving a Sampling Distribution (the "Five-Step Method")

We can derive the probability distribution of a statistic (i.e., a *sampling distribution*) given the probability distribution of a single observation (i.e., a *population distribution*) by using the following steps.

1. Create the sample space which consists of all possible samples.
2. Compute the probability of each sample in the sample space.
3. Compute the value of the statistic for each sample in the sample space.
4. Create a table of the possible values of the statistic.
5. Compute the probability of each value of the statistic.

A **sample space** is the set of all possible samples of observations.

Sampling Distribution of a Mean

Example: Suppose we have a forest of trees of which 60% have a volume of 20 cubic feet, and 40% have a volume of 30 cubic feet. What would be the sampling distribution of the mean volume based on a random sample of the $n = 2$ trees? (Note: We are going to assume for now that we are *sampling with replacement*, meaning that we could select the same tree more than once.)

Table 1: Population Distribution

x	$P(x)$
20	0.6
30	0.4

Table 2: Sample Space

Sample	Probability	$\bar{x}$
20, 20	$0.6 \times 0.6 = 0.36$	20
20, 30	$0.6 \times 0.4 = 0.24$	25
30, 20	$0.4 \times 0.6 = 0.24$	25
30, 30	$0.4 \times 0.4 = 0.16$	30

Table 3: Sampling Distribution

$\bar{x}$	$P(x)$
20	0.36
25	$0.24 + 0.24 = 0.48$
30	0.16

Note: We are using two properties of probabilities from probability theory here.

1. The probability of two or more events happening together (e.g., A *and* B) equals the *product* of their probabilities *if* the events are *independent*, meaning that the probability of each event does not change depending on if the other events have or have not occurred.

2. The probability that at least one of two or more events happening (e.g., A *or* B) equals the *sum* of their probabilities *if* the events are *mutually exclusive*, meaning that the events cannot both occur at the same time.

Example: What is the sampling distribution of the mean of a sample of $n = 2$ observations of the throwing distance of the trebuchet?

Table 4: Population Distribution

x	$P(x)$
1	0.1
2	0.3
3	0.6

Table 5: Sample Space

Sample	Probability	$\bar{x}$
1, 1	0.01	1.0
1, 2	0.03	1.5
1, 3	0.06	2.0
2, 1	0.03	1.5
2, 2	0.09	2.0
2, 3	0.18	2.5
3, 1	0.06	2.0
3, 2	0.18	2.5
3, 3	0.36	3.0

Table 6: Sampling Distribution

$\bar{x}$	$P(\bar{x})$
1.0	0.01
1.5	0.06
2.0	0.21
2.5	0.36
3.0	0.36

Example: We can find the sampling distribution of any statistic in the same way. What is the sampling distribution of the sample variance (s^2) based on a sample of $n = 2$ observations?

Sampling Distribution of a Proportion

Example: What is the sampling distribution of the *proportion* of female platies preferring the yellow-tailed male from a sample of $n = 3$ observations?

Table 7: Population Distribution

x	$P(x)$
C	0.3
Y	0.7

Note: We will denote a proportion from a sample as $\hat{p}$.

The Binomial Distribution

Assume that each observation has a probability distribution with two possible values — a "success" (S) and a "failure" (F). Assume the probability of a success is p and the probability of a failure is thus $1 - p$. Finally assume that the observations are *independent* meaning that the probabilities of a success or failure of any one observation does not depend on the other observations.

x	$P(x)$
S	p
F	$1-p$

Example: Assume the following distribution of one observation for the platy preference.

x	$P(x)$
Y	0.7
C	0.3

Here we are defining Y as a success and C as a failure, so $p = 0.7$ and $1 - p = 0.3$.

The *sampling distribution* of the *number of successes* (s) in a sample will be a *binomial distribution.* The sampling distribution of s is given by the following equation.[1]

$$P(s) = \frac{n!}{s!(n-s)!}p^s(1-p)^{n-s}.$$

Two mathematical details to remember when using this formula:

1. The ! symbol is the *factorial operation.* For example,

$$\begin{aligned}5! &= 5 \times 4 \times 3 \times 2 \times 1 = 120,\\ 4! &= 4 \times 3 \times 2 \times 1 = 24,\\ 3! &= 3 \times 2 \times 1 = 6,\\ 2! &= 2 \times 1 = 2,\\ 1! &= 1,\\ 0! &= 1.\end{aligned}$$

   Note that $0! = 1$, which is perhaps not intuitive.

2. For powers remember that any number raised to the power of 1 is that number (i.e., $p^1 = p$ and $(1-p)^1 = 1-p$), and any number raised to the power of zero is one (i.e., $p^0 = 1$ and $(1-p)^0 = 1$).

[1] Sometimes we write this as

$$P(s) = \binom{n}{s}p^s(1-p)^{n-s},$$

because $\binom{n}{s} = \frac{n!}{s!(n-s)!}$. The $\binom{n}{s}$ is called the *binomial coefficient.* Also usually this formula is written with x in place of s, but I have used s to emphasize that the the formula computes the probability of the number of *successes.*

Sampling Distribution of a Proportion Revisited

Example: What is the sampling distribution of the *proportion* of female platies preferring the yellow-tailed male from a sample of $n = 3$ observations?

Table 8: Population Distribution

x	$P(x)$
C	0.3
Y	0.7

Table 9: Sampling Distribution

s	$\hat{p}$	$P(\hat{p})$
0	0	0.027
1	1/3	0.189
2	2/3	0.441
3	1	0.343

Note: We will denote a proportion from a sample as $\hat{p}$.

Here is how we can compute the probabilities in the sampling distribution of $\hat{p}$.

$$P(0) = \underbrace{\frac{3!}{0!(3-0)!}}_{1} \underbrace{0.7^0(1-0.7)^{3-0}}_{0.027} = 1 \times 0.027 = 0.027$$

$$P(1) = \underbrace{\frac{3!}{1!(3-1)!}}_{3} \underbrace{0.7^1(1-0.7)^{3-1}}_{0.063} = 3 \times 0.063 = 0.189$$

$$P(2) = \underbrace{\frac{3!}{2!(3-2)!}}_{3} \underbrace{0.7^2(1-0.7)^{3-2}}_{0.147} = 3 \times 0.147 = 0.441$$

$$P(3) = \underbrace{\frac{3!}{3!(3-3)!}}_{1} \underbrace{0.7^3(1-0.7)^{3-3}}_{0.343} = 1 \times 0.343 = 0.343$$

Note that the formula computes two parts — the number of samples that produce s successes out of n observations, and the probability of each sample. These can be seen when we look at the sample space.

Table 10: Sample Space

Sample	Probability	s	$\hat{p}$
Y, Y, Y	$0.7 \times 0.7 \times 0.7 = 0.343$	3	1
C, Y, Y	$0.3 \times 0.7 \times 0.7 = 0.147$	2	2/3
Y, C, Y	$0.7 \times 0.3 \times 0.7 = 0.147$	2	2/3
Y, Y, C	$0.7 \times 0.7 \times 0.3 = 0.147$	2	2/3
Y, C, C	$0.7 \times 0.3 \times 0.3 = 0.063$	1	1/3
C, Y, C	$0.3 \times 0.7 \times 0.3 = 0.063$	1	1/3
C, C, Y	$0.3 \times 0.3 \times 0.7 = 0.063$	1	1/3
C, C, C	$0.3 \times 0.3 \times 0.3 = 0.027$	0	0