TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Statistics Introduction
MATHematics MAGnification™

Dr. Ralph deLaubenfels

TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Teacher-Scholar Institute

Columbus, Ohio

2018

STATISTICS INTRODUCTION MAGNIFICATION

This is one of a series of very short books on math, statistics, and physics called "Math Magnifications." The "magnification" refers to focusing on a particular topic that is pivotal in or emblematic of mathematics.

OUTLINE

We will introduce the most fundamental ideas and terminology in statistics, those needed for understanding statistical constructions. We will illustrate them with fat content of burgers, lengths and moods of cats, gender of kittens, poisoning of cookies, marbles, human height and time spent sleeping, self esteem versus math, coin flipping, and various properties of wolverines.

The material presented here might seem long-winded, for what we are able to do so far. Our definitions anticipate depth of understanding for future delving into statistical activity.

The only prerequisites for this magnification are arithmetic, including fractions and percents, and an intuitive notion of probability, as measuring how likely something is to occur. More specific probability factoids will be introduced, when needed, in this magnification. See [5], [7], or [8] for more about probability; [8] uses calculus, but has a wide spectrum of interesting examples.

Statistics attempts to makes sense of data. The information age, burying us with a constant avalanche of data, has made this both more challenging and more necessary. See [5, Chapters I and II] for the organization and summary of data. We wish, in this introduction and future magnifications, to describe statistical *inference*, where we draw conclusions and make decisions from our data, always while buffeted by uncertainty.

The information age can easily degenerate into the *mis*information age. Statistics is subtle and slippery enough to contribute to this effectively. No longer must scientific fraud be done (for those so inclined) by such crudities as erasing numbers that don't fit your theory and scribbling in ones that do. Statistical abuse creates fraud that is much more difficult to unearth; see, for example, [10].

This potential evil use of statistics is sufficient reason to learn it well.

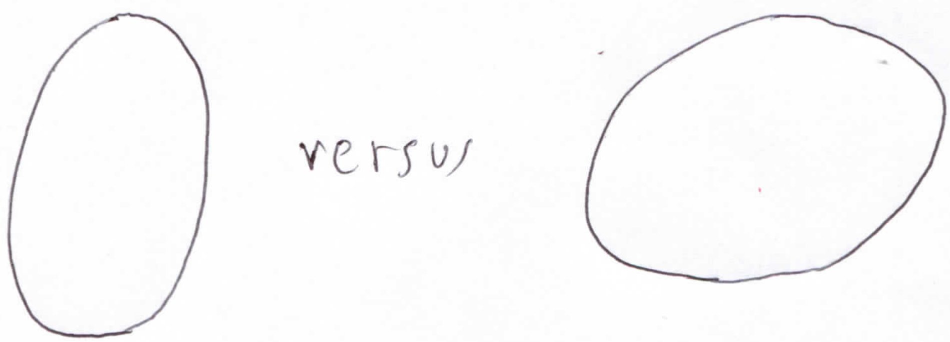
Statistics tends to be associated with social science and the fuzzier parts of medical science. This might explain a recurring loss of credibility. For example, one might take a survey to study the relationship between self esteem and math ability; the use of a survey already shows a confusion between reality and the perception of reality.

To defend the honor of statistics, we must begin with the scientifically and mathematically rigorous origins of statistical inference, used, in the late 18^{th} century, to study the following (see [2], [3], [9], or [11]).

1. the libration of the moon

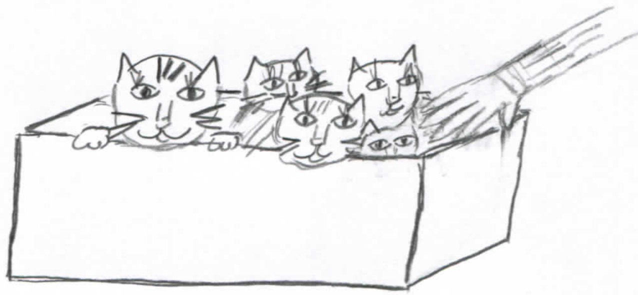
2. the effect of Jupiter on Saturn's orbit

3. the shape of the earth:



In this, and future magnifications on statistics, we hope to communicate that statistics is making explicit and quantified thought processes that we all perform instinctively, albeit often at a subconscious and inconsistent level. Terminology, sometimes extensive, is necessary to bring these processes out in the open, if only to expose their limitations.

Definitions 1. A **population** or **universe** is a set of things we care about. A **sample** is a subset of the population that we can work with and measure. There is also the verb **sampling**, meaning taking a sample from the population. The image here is repeatedly reaching into a box of kittens and pulling one out.



Example 2. Many of us care about fast-food burgers, if only as something sinful. With this mission in mind, our population could be

$$\{\text{fast-food burgers}\},$$

and our sample

$$\{\text{100 randomly chosen fast-food burgers}\}.$$

There are many things, as scientists, we could do with that sample: eat them ourselves, feed them to wolverines, measure the amount of fat in each burger, see how far they could be thrown.

Definition 3. **Statistical inference** uses a sample to make intelligent guesses about a population.

In our example, if each of 100 wolverines eats a member of our sample, and 83 of them have Wolverine Hyperactivity Disorder (WHD) within an hour of eating, we might conclude that fast-food burgers are likely to give wolverines WHD. Better yet, we might feel empowered to estimate *how* likely.

Populations are usually large, possibly infinite; in our example, our concern is not only with past and present burgers, but possible future burgers. Samples are relatively small. Thus statistical inference is arguably impossible. Any introduction to statistics should keep this in mind, and be prepared for ambiguities or uncertainties in conclusions drawn.

Definition 4. A **population parameter** is a population characteristic. Statistical inference often focuses on using a sample to estimate a population parameter.

Example 5. Let our population be the human population of the earth. One parameter for this population is the average length of time spent sleeping at night, in minutes.

Definitions 6. The **average** or **mean** of a set of numbers $x_1, x_2, \ldots x_n$ is $\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

For example, the mean of $1, 0, -3, 1, 9$ is

$$\frac{1}{5}(1 + 0 - 3 + 1 + 9) = 1.6.$$

If the set of numbers comes from a population, the average is denoted by the Greek letter μ (written "mu" and pronounced "mew"), and is called the **population mean**. If the set of numbers $x_1, x_2, \ldots$ comes from a sample, the average is denoted $\overline{x}$ and is called the **sample mean.**

See [5] or [8] for more about the mean.

Another popular parameter is *proportion.* Very generally, if Ω is a set, the **proportion** of things in Ω with a certain property is

$$\frac{\text{(the number of things in } \Omega \text{ with the property)}}{\text{(the number of things in } \Omega\text{)}}.$$

For example, if a sack contains 5 red marbles and 15 blue marbles, the proportion of red marbles in the sack is $\frac{5}{20} = \frac{1}{4}$ or 25%.

A proportion of things in a population is a **population proportion**, denoted p. A proportion of things in a sample is a **sample proportion**, denoted $\hat{p}$.

Example 7. Suppose we are concerned about the lengths of cats. Our population is then the set of all cats. Let's take a sample of ten cats and measure their lengths; let's say we get, in inches, the following lengths:

$$16, 9, 8, 10, 15, 17, 16, 9, 5, 5.$$

One population parameter of interest here is the average length, in inches, of all cats. We could estimate this with the sample mean

$$\overline{x} \equiv \frac{1}{10}(x_1 + x_2 + \cdots + x_{10}) = \frac{1}{10}(16 + 9 + 8 + 10 + 15 + 17 + 16 + 9 + 5 + 5) = 11.$$

Here's another possible perspective. Suppose I'm considering manufacturing cat crates 15 inches long. Of concern is then the (population) proportion of cats that are more than 15 inches long, since cats are not amenable to having their length constricted.

The population proportion p just mentioned may be estimated with the proportion of cats in the sample that are more than 15 inches long:

$$\hat{p} = \frac{3}{10} = 0.3 \text{ or } 30\%.$$

Probability Terminology 8. Most definitions and constructions in statistics are in terms of an intuitive idea, *probability.* Probability is then used to evaluate the accuracy of statistical inferences.

We mentioned at the beginning of this magnification the presence of uncertainty throughout statistical inference. Because probability measures uncertainty, its omnipresence is inevitable.

Convention insists that we denote $P(A)$ (reads "P of A") for the **probability** of an event A. See [5, Chapter III], [7, Definition 1.3 and Examples 1.4] or [8, Chapter 2].

Definition 9. (See [5, Chapter V], [7, especially Definition 1.11 and Examples 1.12], or [8, Chapter 3]). A **random variable** assigns a number to each outcome of an experiment.

Our interest is in the following experiment: randomly select a member of a specified population. Then each outcome corresponds to a member of the population and a random variable assigns a number to each member of the population.

All random variables in this magnification are defined on a population as we just described.

Random Variable Convention 10. Random variables are traditionally denoted by capital English letters X, Y, etc. The actual numerical values that a random variable can equal are traditionally denoted by lower case English letters x, y, etc.

For example, let our experiment be flipping a coin three times, and let X be the number of heads achieved. X is a random variable; prior to flipping, we don't know what X will equal. The possible values for X are $x = 0, 1, 2$, or 3.

In general, a random variable is a theoretical assertion of what could possibly happen. All we hope to know about a random variable is how likely its possible values are, but nothing has actually happened; we are viewing the population with an open mind, before specifying a particular member. For the random variable X of the previous paragraph, we are contemplating flipping a coin, and imagining all the possibilities. The lower-case letters x come up when we sample; for example, if we finally flex our fingers and flip the specified coin three times, getting a head, then another head, then a tail, we now have $x = 2$.

Definition 11. The **distribution** of a random variable X refers to all probabilities involving the possible values of X. This is determined by, for any real x, a, b, $P(X = x)$, meaning the probability that X equals x, and $P(a < X < b)$, meaning the probability that X is strictly between a and b.

Any probability distribution could arise in many different ways. For example, the number of *tails* when flipping a fair (heads and tails equally likely) coin has the same distribution as the number of heads.

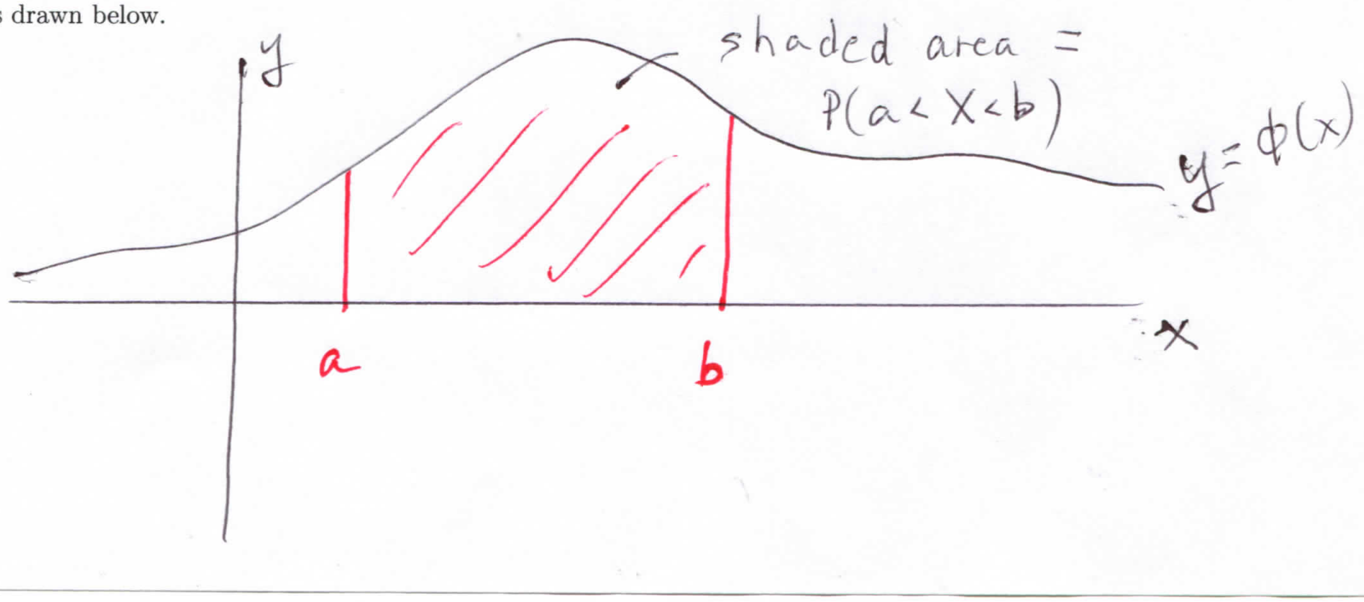
When dealing with random variables, we often choose to ignore its precise definition, and focus on the distribution of the random variable.

Definition 12. For our favorite types of random variables, we present their distributions in the following ways (see [7, Definitions 2.2 and APP.1] and [5, Chapter V]).

A **continuous** random variable X has a **probability density function (pdf)**. Denoting the pdf for X by ϕ, said pdf has the following property. For any real numbers a, b, with $a \leq b$,

$$P(a < X < b) = \text{ the area between } y = 0, y = \phi(x), x = a, \text{ and } x = b,$$

as drawn below.



A **discrete** random variable X has only a sequence $x_1, x_2, x_3, \dots$ of possible values. The **probability mass function (pmf)** lists the probabilities of X equaling each of those values:

$$p(x_k) \equiv P(X = x_k), \;\; k = 1, 2, 3, \dots.$$

Definition 13. We say that two random variables X and Y **have the same distribution** if

$$P(a < X \leq b) = P(a < Y \leq b)$$

for any real a and b.

If X and Y are discrete, having the same distribution is equivalent to having the same pmf; if X and Y are continuous, having the same pdf implies they have the same distribution. If one of X and Y is discrete and the other is continuous, then X and Y do not have the same distribution.

When two random variables have the same distribution, from a global probabilistic point of view they are the same. Yet they are *not* the same random variable. We have already mentioned that, when flipping a fair coin, counting the number of tails has the same distribution as counting the number of heads; however, for most flips of the coin, the number of heads is different than the number of tails. See also Examples 18 and 22.

Definition 14. Informally, a collection of random variables is **independent** if the value of one random variable has no effect on the distributions of the other random variables. See [8, page 201] for the exact definition.

A population parameter often appears explicitly in the pmf or pdf of a random variable.

Examples 15. (1) A random variable X has a **binomial** distribution if, for some $n = 1, 2, 3, \dots$, $0 \leq p \leq 1$,

$$X \equiv \text{ number of successes in } n \text{ independent trials,}$$

with $p \equiv P(\text{success})$ on each trial.

The prototype for binomial is flipping a coin n times, with $p = P(\text{heads})$ on each flip. But many definitions of "trial" and "success" are possible. The trial could be a cat having a kitten, with success defined to be whatever gender is desired (we avoid the trap of expressing a preference). The trial could be to take a cookie from a kindly cookie sales entity, with success defined to be "cookie is poisoned."

Here is the pmf for binomial X:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \;\; (k = 0, 1, 2, \dots, n).$$

The parameters needed for complete understanding of a binomial random variable are n and p.

(2) A **normal** random variable is continuous, with pdf

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \;\; (x \text{ real}),$$

where e is a fixed famous irrational number. Our favorite parameters here are μ (called the **mean** of the random variable) and σ (called the **standard deviation** of the random variable).

More generally, to estimate a population parameter θ, we choose a random variable whose values give *some* information about θ. Here are two examples, to match Example 2, Definitions 6, Example 7, and Example 22.

If θ is μ, the average volume, in decibels, of a wolverine on the attack, we would look at the random variable that calculates the number of decibels of a randomly chosen wolverine on the attack.

If θ is p, the proportion of wolverines that are rabid, we would look at the random variable that assigns, to any randomly chosen wolverine, a 1 if it is rabid and a 0 if it is not rabid. See Examples 7 and 22.

Discussion 16. In practice, we often know the general form of a random variable, so that complete omniscience would follow from knowing the values of a few parameters (in the prior Examples 15, see n and p in (1) and μ and σ in (2)). Statistical inference then uses the sample to estimate the population parameters of interest; the better the estimate, the closer we are to omniscience.

For example, we might believe that human height is normally distributed (see (2) in the previous Examples 15). If we knew μ and σ, we would completely understand human height. Our thirst for knowledge thus impels us to use statistical inference to estimate μ and σ.

Definition 17. It will be essential in the future, if only to get probability into the picture, that sampling be formulated at the level of random variables.

A **random sample** of size n from a random variable X is an independent sequence $X_1, X_2, \ldots, X_n$ of random variables such that, for $1 \leq k \leq n$, X_k has the same distribution as X (see Definition 11).

The corresponding random sample of numbers, denoted, as per Random Variable Convention 10,

$$x_1, x_2, \ldots x_n,$$

occurs when we sample in the physical sense of evaluating X on the sample points; x_1 is the measured value of X_1 on the first member of the sample, x_2 is the measured value of X_2 on the second member of the sample, etc.

Informally, X measures members of the population. The sequence of numbers $x_1, x_2, \ldots, x_n$ comes about from making said measurement on members of the sample.

The customary terminology presented here is not ideal. "Random sample" here might mean a sequence of random variables or a sequence of numbers, and it is different than the meaning of "sample" given in Definitions 1, since our Definition 17 involves applying a random variable X to a sample, to get a "random sample from X."

Example 18. Let's get back to the fast-food burgers of Example 2. If we wish to study the amount of fat in said burgers, we could define the random variable X to be the number of grams of fat in a randomly chosen fast-food burger.

For a random sample from X, we could plan to randomly choose 5 fast-food burgers, and define random variables

$$X_1 \equiv \text{number of grams of fat in the first burger chosen}$$

$$X_2 \equiv \text{number of grams of fat in the second burger chosen}$$

$$X_3 \equiv \text{number of grams of fat in the third burger chosen}$$

$$X_4 \equiv \text{number of grams of fat in the fourth burger chosen}$$

$$X_5 \equiv \text{number of grams of fat in the fifth burger chosen}$$

$X, X_1, X_2, \ldots, X_5$ are all random variables, because we don't know *which* burger or burgers will be chosen; different burgers will produce different values of $X, X_1, X_2, \ldots$. The random variables are a plan or blueprint of theoretical hamburgers to come; no physical hamburgers have materialized.

Note that each of $X_1, X_2, \ldots, X_5$ have the same distribution as X, since, in all cases, we are choosing from the universe (population) of all fast-food burgers.

For a random sample of numbers, we must measure the amount of fat in each of the 5 burgers. Let's say the first burger contains 10 grams of fat, the second 11.2, the third 7, the fourth 11.2, and the fifth 9. That is denoted

$$x_1 = 10, x_2 = 11.2, x_3 = 7, x_4 = 11.2, x_5 = 9.$$

For the numerical random sample we must get our feet wet, or, in this case, our fingers greasy.

Definitions 19. A **statistic** or **point estimator** for a population parameter θ is a function of a random sample from a random variable X, denoted $\hat{\theta}$; that is, $\hat{\theta}$ is calculated in a specified way from the random sample.

The corresponding function of the numerical random sample, with numbers $x_1, x_2, x_3, \ldots$ replacing random variables $X_1, X_2, X_3, \ldots$, is called a **point estimate** of θ. We are sorry to say that the estimate is also denoted $\hat{\theta}$.

These definitions are kept general (notice that neither X nor the calculation of $\hat{\theta}$ in terms of the random sample from X are specified) by not mentioning what we really want: our estimators or estimates of θ should be getting close to θ, as we sample more and more.

Example 20. For the fast-food burgers of Examples 2 and 18, we might focus, for the sake of headlines or other one-line bullet points, on the average or mean, denoted μ (see Definitions 6) number of grams of fat in a randomly chosen burger. For X as in Example 18, we could take the average of a random sample $X_1, X_2, \ldots, X_n$, denoted

$$\overline{X} \equiv \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

for our point estimator of μ; that is, $\hat{\mu} = \overline{X}$. For any random variable X, $\overline{X}$ is called the **sample mean**, and is the most popular point estimator of the population mean μ.

Unfortunately,

$$\overline{x} \equiv \frac{1}{n}(x_1 + x_2 + \cdots + x_n),$$

where $x_1, x_2, \ldots, x_n$ is the corresponding random sample of numbers as in Definition 17, is also called the sample mean; see Definitions 6.

The sample mean $\overline{X}$ is a random variable: different burgers will produce different values of $X_1, X_2, \ldots, X_n$, hence different averages.

For the numerical random sample of Example 18, we get

$$\overline{x} \equiv \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) = \frac{1}{5}(10 + 11.2 + 7 + 11.2 + 9) = 9.68$$

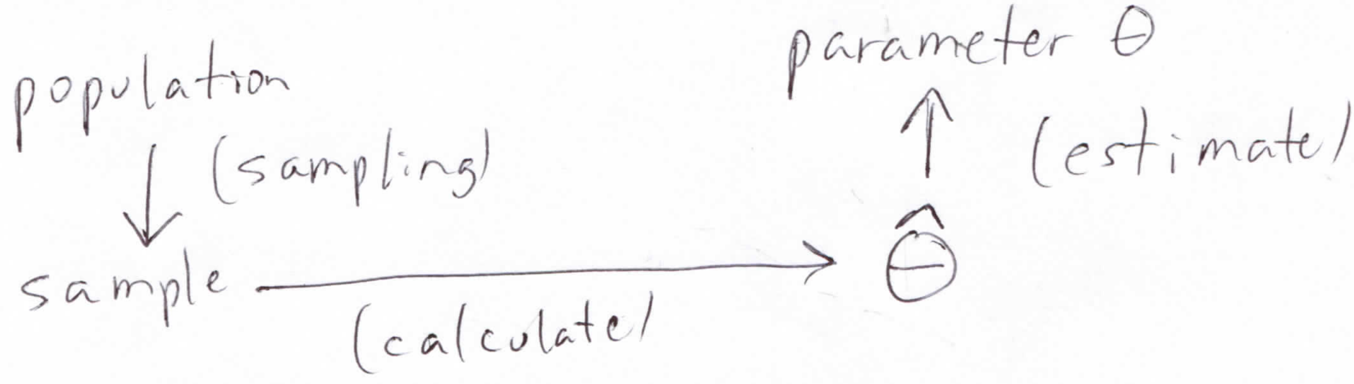
for our estimate of the average number of grams in a fast-food burger.

Statistical Inference on a Population Parameter 21. Here is our action plan, as the business world used to say (perhaps they still say it), for statistical inference on a population parameter θ.

1. get a random sample from a random variable X.
2. calculate a point estimator $\hat{\theta}$ of θ from the random sample.
3. use $\hat{\theta}$ to approximate θ.

Notice that there is no mention of how *good* the approximation is nor what point estimator should be used. We will address this in future magnifications.

Here is a sketch of this action plan, at the numerical level.



Example 22. Let's put Example 7 in the setting of random variables and Action Plan 21.

For estimating the average length of (all) cats, denoted μ, let X be the length, in inches, of a randomly chosen cat. As with Example 18, that dealt with fast-food burgers, we have the random sample

$$X_k \equiv \text{length, in inches, of } k^{th}\text{cat chosen},$$

for $k = 1, 2, 3, \ldots, 10$.

We use $\overline{X}$, the sample mean (see Definitions 6),

$$\frac{1}{10}\left(X_1 + X_2 + \cdots + X_{10}\right),$$

to approximate μ, the population mean. Since actual measurements have been made in Example 7, we may use the numerical random sample

$$x_1 = 16, x_2 = 9, x_3 = 8, x_4 = 10, x_5 = 15, x_6 = 17, x_7 = 16, x_8 = 9, x_9 = 5, x_{10} = 5$$

to get

$$\overline{x} \equiv \frac{1}{10}\left(x_1 + x_2 + \cdots + x_{10}\right) = \frac{1}{10}\left(16 + 9 + 8 + 10 + 15 + 17 + 16 + 9 + 5 + 5\right) = 11$$

as our approximation of μ.

For estimating the proportion of all cats that are more than 15 inches long (see the last paragraph of Examples 15), a relevant random variable is the discrete random variable Y with the following pmf.

y	0	1
$p(y)$	$(1-p)$	p

Here p is the desired population proportion of all cats that are more than 15 inches long. Note that Y is binomial (see Examples 15 (1)), with $n = 1$ and success defined to be "cat chosen is more than 15 inches long." Y is called a **Bernoulli** random variable, with the prototype of flipping a coin once and counting the number of heads; that is, $Y = 1$ if we get heads and $Y = 0$ if we get tails, while p is the probability of getting heads, on each flip.

Our random sample from Y is

$$Y_k = \begin{cases} 1 & \text{if the } k^{th} \text{ cat chosen is more than 15 inches long} \\ 0 & \text{if the } k^{th} \text{ cat chosen is less than or equal to 15 inches long} \end{cases}$$

for $k = 1, 2, 3, \ldots, 10$.

The most popular estimator of p, the population proportion, is the sample proportion (see Definitions 6)

$$\hat{p} = \text{proportion of the sampled cats that are more than 15 inches long;}$$

as a function of our random sample, as in Definition 17,

$$\hat{p} \equiv \frac{1}{10}\left(Y_1 + Y_2 + Y_3 + \cdots + Y_{10}\right).$$

Since

$$Y_k = \begin{cases} 1 & \text{if } X_k > 15 \\ 0 & \text{if } X_k \leq 15' \end{cases}$$

we may use the numerical random sample given in Example 7 (and listed above in this Example as a random sample from X), but change it to values of Y: since $x_1 = 16 > 15, y_1 = 1$, since $x_2 = 9 \leq 15, y_2 = 0$, etc.

$$y_1 = 1, y_2 = 0 = y_3 = y_4 = y_5, y_6 = 1 = y_7, y_8 = 0 = y_9 = y_{10},$$

so that the sample proportion $\hat{p}$, the proportion of "successes" (meaning being more than 15 inches long) in the sample of cats, equal to the proportion of ones in $\{y_1, y_2, \ldots, y_{10}\}$, is

$$\hat{p} = \frac{1}{10}(1 + 0 + 0 + 0 + 0 + 1 + 1 + 0 + 0 + 0) = \frac{3}{10}.$$

The random sample $\{Y_k\}_{k=1}^{10}$ represents a loss of information from the random sample $\{X_k\}_{k=1}^{10}$; all we notice about each of $x_1, x_2, \ldots$ is whether it is greater than 15.

Example 23. The choice of the sample mean $\overline{X}$ as an estimator of the population mean μ and the sample proportion $\hat{p}$ as an estimator of the population proportion p seems natural (see Definitions 6). But "good" (in senses that we will specify in a future magnification) choices of estimators are not always clear. Here is a popular example.

The **linear regression** model for a pair of random variables X and Y is the belief that, on average,

$$Y = a + bX,$$

for some numbers a and b.

Sampling is done from X and Y simultaneously; that is, we measure ordered pairs

$$(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n).$$

Here is what turns out to be the "best" (the moral superiority must be defined; we will address this in a later magnification) point estimators for the population parameters a and b:

$$\hat{b} \equiv \frac{[(x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y}) + \cdots + (x_n - \overline{x})(y_n - \overline{y})]}{[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2]}, \quad \hat{a} \equiv \left(\overline{y} - \overline{x}\hat{b}\right).$$

See [6, Chapter VI, Section E], for the techniques involved in getting $\hat{a}$ and $\hat{b}$.

Inverse relationship of statistics to probability 24.

Probability: population ⟶ sample

Statistics: sample ⟶ population

Example 25. If we know that 90% of (all) cats are aloof, and we take a random sample of 10 cats, probability tells us that the probability that exactly 9 of them are aloof is $10(0.9)^9(0.1) \sim 38.74\%$.

Conversely, if we don't know what proportion of all cats are aloof, and we have exactly 9 cats that are aloof, out of a random sample of 10 cats, statistics tells us to guess that 90% of all cats are aloof.

Laplace, one of the founders of statistical inference (see [3], [9], or [11]) called statistical inference *inverse probability*, by which he meant reasoning from effect to cause. Platonism (see, for example, [1, pages 67–68]), especially Plato's parable of the cave, where the physical world is the shadow on a cave wall, of the enduring patterns that he called *forms*, may be viewed as inverse probability; see [2, footnote 3 of Chapter 1] or [4, page 12].

HOMEWORK

1. Suppose the body temperature, in degrees Fahrenheit, of twelve randomly chosen wolverines, is

$$101.4, 99.0, 100.8, 99.3, 101.1, 105, 98.7, 100.2, 99.6, 102.3, 102, 102.$$

Use our favorite estimators (see Example 23) to approximate each of the following.

(a) The average body temperature of all wolverines; and

(b) the proportion of all wolverines that are hot, where "hot" is defined to be more than 102 degrees Fahrenheit.

2. See Homework Problem **1** and its answers below.

(a) Is the average body temperature of all wolverines 100.95 degrees Fahrenheit?

(b) If we take another random sample of wolverines, will their average body temperature be 100.95 degrees Fahrenheit?

(c) Are $\frac{1}{6}$ of all wolverines hot?

(d) If we take another random sample of wolverines, will $\frac{1}{6}$ of them be hot?

ANSWERS

1. (a) $\frac{1}{12}(101.4 + 99.0 + 100.8 + 99.3 + 101.1 + 105 + 98.7 + 100.2 + 99.6 + 102.3 + 102 + 102) = 100.95$.

(b) $\frac{2}{12} = \frac{1}{6}$.

2. Same answer for each of (a), (b), (c), and (d): Probably not; we know nothing about other wolverines besides the ones we sampled.

(d) is particularly problematic. If our sample contained seven wolverines, $\frac{1}{6}$ of them would consist of $\frac{7}{6}$, or one and one sixth, wolverines. We are mildly disturbed trying to visualize one sixth of a wolverine.

REFERENCES

1. W.S. Anglin and J. Lambek, "The Heritage of Thales," Springer, 1995.

2. A. I. Dale, "A History of Inverse Probability," Springer, second edition, 1999.

3. R. deLaubenfels, "The Victory of Least Squares and Orthogonality in Statistics," The Amer. Statistician 60 (2006), 315–321.

4. R. deLaubenfels, "Common Origins of Statistical Inference and Calculus," http://teacherscholarinstitute.com/Papers/OriginsStatisticalInferenceCalculus.pdf (2012).

5. R. deLaubenfels, "Fun with Introductory Probability," www.teacherscholarinstitute/Books/Probability.pdf (2014).

6. R. deLaubenfels, "Linear Algebra," or E Pluribus Unum, http://teacherscholarinstitute.com/FreeMathBooksHighschool.html (2017).

7. R. deLaubenfels, "Probability Introduction Magnification," http://teacherscholarinstitute.com/MathMagnificationsReadyToUse.html.

8. J. L. Devore, "Probability and Statistics for Engineering and the Sciences," Brooks/Cole, eighth edition, 2012.

9. A. Hald, "A History of Parametric Statistical Inference From Bernoulli to Fisher, 1713–1935," Springer, 2007.

10. D. Huff, "How to Lie with Statistics," W. W. Norton, 1954.

11. S. M. Stigler, "The History of Statistics: The Measurement of Uncertainty before 1900," the Belknap Press of Harvard University Press, Cambridge, MA, 1986.