

Lecture 8.1, MATH-57091 Probability and Statistics for High-School Teachers.

Artem Zvavitch

Department of Mathematical Sciences, Kent State University

October 12 - October 15, 2014.

Sampling Proportions from a Finite Population

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q).

Sampling Proportions from a Finite Population

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q). Let p denote the proportion of the population having property Q .

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q). Let p denote the proportion of the population having property Q . So $N \cdot p$ members of the population have property Q and $N \cdot (1 - p)$ do not.

Suppose that 50 out of 800 students in Aurora High School are left-handed.

Sampling Proportions from a Finite Population

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q). Let p denote the proportion of the population having property Q . So $N \cdot p$ members of the population have property Q and $N \cdot (1 - p)$ do not.

Suppose that 50 out of 800 students in Aurora High School are left-handed. If left-handedness is the characteristic of interest (i.e. "property Q ") then $N = 800$ and $p = 50/800 = 1/16$.

Sampling Proportions from a Finite Population

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q). Let p denote the proportion of the population having property Q . So $N \cdot p$ members of the population have property Q and $N \cdot (1 - p)$ do not.

Suppose that 50 out of 800 students in Aurora High School are left-handed. If left-handedness is the characteristic of interest (i.e. "property Q ") then $N = 800$ and $p = 50/800 = 1/16$.

A sample of size n is said to be a **random sample** if it is chosen in a manner so that each of the possible subsets of size n of our population is equally likely to be selected.

Sampling Proportions from a Finite Population

Consider a population of size N in which certain members have a particular characteristic of interest (satisfy a certain property Q). Let p denote the proportion of the population having property Q . So $N \cdot p$ members of the population have property Q and $N \cdot (1 - p)$ do not.

Suppose that 50 out of 800 students in Aurora High School are left-handed. If left-handedness is the characteristic of interest (i.e. "property Q ") then $N = 800$ and $p = 50/800 = 1/16$.

A sample of size n is said to be a **random sample** if it is chosen in a manner so that each of the possible subsets of size n of our population is equally likely to be selected.

One of the ways to select a random sample is first to select (equally likely) one member out of N after pick (again equally likely) one member out of $N - 1$ left and so on...

Sampling Proportions from a Finite Population

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise.

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

We also note that

$$P(X_i = 1) = \frac{Np}{N} = p \text{ and } P(X_i = 0) = 1 - p.$$

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

We also note that

$$P(X_i = 1) = \frac{Np}{N} = p \text{ and } P(X_i = 0) = 1 - p.$$

NOTE that the random variables $X_1, X_2, X_3, \dots, X_n$ are not independent, indeed if we know that $X_1 = 1$, i.e. that the first selected member in the sample satisfy the property Q then when we select the second one we select it out of $N - 1$ members and the number of members satisfying property Q is now $pN - 1$

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

We also note that

$$P(X_i = 1) = \frac{Np}{N} = p \text{ and } P(X_i = 0) = 1 - p.$$

NOTE that the random variables $X_1, X_2, X_3, \dots, X_n$ are not independent, indeed if we know that $X_1 = 1$, i.e. that the first selected member in the sample satisfy the property Q then when we select the second one we select it out of $N - 1$ members and the number of members satisfying property Q is now $pN - 1$

$$P(X_2 = 1 | X_1 = 1) = \frac{Np - 1}{N - 1} \neq p$$

i.e.

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

We also note that

$$P(X_i = 1) = \frac{Np}{N} = p \text{ and } P(X_i = 0) = 1 - p.$$

NOTE that the random variables $X_1, X_2, X_3, \dots, X_n$ are not independent, indeed if we know that $X_1 = 1$, i.e. that the first selected member in the sample satisfy the property Q then when we select the second one we select it out of $N - 1$ members and the number of members satisfying property Q is now $pN - 1$

$$P(X_2 = 1 | X_1 = 1) = \frac{Np - 1}{N - 1} \neq p$$

i.e.

$$P(X_2 = 1 | X_1 = 1) \neq P(X_2 = 1).$$

As we done a number of times before let X_i , $i = 1, \dots, n$ be a random variable such that it is equal to 1 if i 's member of our sample satisfy property Q and zero otherwise. Thus the sample means is

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

We also note that

$$P(X_i = 1) = \frac{Np}{N} = p \text{ and } P(X_i = 0) = 1 - p.$$

NOTE that the random variables $X_1, X_2, X_3, \dots, X_n$ are not independent, indeed if we know that $X_1 = 1$, i.e. that the first selected member in the sample satisfy the property Q then when we select the second one we select it out of $N - 1$ members and the number of members satisfying property Q is now $pN - 1$

$$P(X_2 = 1 | X_1 = 1) = \frac{Np - 1}{N - 1} \neq p$$

i.e.

$$P(X_2 = 1 | X_1 = 1) \neq P(X_2 = 1).$$

Which is not so good for our applications.

It turns out that we can still assume that X_i are independent when we sample from the "huge" population.

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

Thus if consider population of size N then the error we done assuming that first two samples are in depended is about $1/N$. If we repeat this assumption for all n elements in our sample than we will accumulate an error but it will still be very small if N is huge comparing to n .

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

Thus if consider population of size N then the error we done assuming that first two samples are in depended is about $1/N$. If we repeat this assumption for all n elements in out sample than we will accumulate an error but it will still be very small if N is huge comparing to n .

We will assume that population we study is huge compering to the size of the sample.

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

Thus if consider population of size N then the error we done assuming that first two samples are in depended is about $1/N$. If we repeat this assumption for all n elements in out sample than we will accumulate an error but it will still be very small if N is huge comparing to n .

We will assume that population we study is huge compering to the size of the sample. Then

$$X = X_1 + X_2 + \dots + X_n,$$

becomes approximately binomial with parameters n and p .

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

Thus if consider population of size N then the error we done assuming that first two samples are in depended is about $1/N$. If we repeat this assumption for all n elements in our sample than we will accumulate an error but it will still be very small if N is huge comparing to n .

We will assume that population we study is huge compering to the size of the sample. Then

$$X = X_1 + X_2 + \dots + X_n,$$

becomes approximately binomial with parameters n and p . Thus

$$\mathbb{E}X = np \text{ and } SD(X) = \sqrt{np(1-p)}.$$

It turns out that we can still assume that X_i are independent when we sample from the "huge" population. Indeed if we would "assume" independence than the error we make is

$$p - \frac{Np-1}{N-1} = \frac{Np-p-Np+1}{N-1} = \frac{1-p}{N-1}.$$

Thus if consider population of size N then the error we done assuming that first two samples are in depended is about $1/N$. If we repeat this assumption for all n elements in our sample than we will accumulate an error but it will still be very small if N is huge comparing to n .

We will assume that population we study is huge compering to the size of the sample. Then

$$X = X_1 + X_2 + \dots + X_n,$$

becomes approximately binomial with parameters n and p . Thus

$$\mathbb{E}X = np \text{ and } SD(X) = \sqrt{np(1-p)}.$$

and from $\bar{X} = \frac{X}{n}$ we get

$$\mathbb{E}\bar{X} = p \text{ and } SD(\bar{X}) = \sqrt{\frac{p(1-p)}{n}}.$$

A reminder: If X is a Binomial random variable with parameters n and p , then

A reminder: If X is a Binomial random variable with parameters n and p , then

$$\frac{X - np}{\sqrt{np(1-p)}}.$$

approaches the standard normal distribution as n approaches ∞ .

A reminder: If X is a Binomial random variable with parameters n and p , then

$$\frac{X - np}{\sqrt{np(1-p)}}.$$

approaches the standard normal distribution as n approaches ∞ . From here we get that

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}.$$

approaches the standard normal distribution as n approaches ∞ .

Suppose that exactly 46 percent of the population favors a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favor this candidate?

Suppose that exactly 46 percent of the population favors a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favor this candidate?

Solution: If X is the number who favor the candidate in our random sample, then X is binomial with parameters $n = 200$ and $p = .46$ and our goal is to estimate $P(X \geq 100)$. As we done before when we jump from discrete to continuous it is better to consider to estimate with $-.5$:

Suppose that exactly 46 percent of the population favors a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favor this candidate?

Solution: If X is the number who favor the candidate in our random sample, then X is binomial with parameters $n = 200$ and $p = .46$ and our goal is to estimate $P(X \geq 100)$. As we done before when we jump from discrete to continuous it is better to consider to estimate with $-.5$:

$$P(X \geq 100) = P(X \geq 99.5)$$

Suppose that exactly 46 percent of the population favors a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favor this candidate?

Solution: If X is the number who favor the candidate in our random sample, then X is binomial with parameters $n = 200$ and $p = .46$ and our goal is to estimate $P(X \geq 100)$. As we done before when we jump from discrete to continuous it is better to consider to estimate with $- .5$:

$$\begin{aligned} P(X \geq 100) &= P(X \geq 99.5) \\ &= P\left(\frac{X - 200 * .46}{\sqrt{200 * .46 * .54}} \geq \frac{99.5 - 200 * .46}{\sqrt{200 * .46 * .54}}\right) = P\left(\frac{X - 92}{7.0484} \geq \frac{99.5 - 92}{7.0484}\right) \end{aligned}$$

Suppose that exactly 46 percent of the population favors a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favor this candidate?

Solution: If X is the number who favor the candidate in our random sample, then X is binomial with parameters $n = 200$ and $p = .46$ and our goal is to estimate $P(X \geq 100)$. As we done before when we jump from discrete to continuous it is better to consider to estimate with $- .5$:

$$\begin{aligned} P(X \geq 100) &= P(X \geq 99.5) \\ &= P\left(\frac{X - 200 * .46}{\sqrt{200 * .46 * .54}} \geq \frac{99.5 - 200 * .46}{\sqrt{200 * .46 * .54}}\right) = P\left(\frac{X - 92}{7.0484} \geq \frac{99.5 - 92}{7.0484}\right) \\ &\approx P(N(0,1) > 1.0641) \approx 0.144. \end{aligned}$$