6 章 標本分布 (Sampling Distributions)

東京大学

2014

Sampling

- In this chapter, we study how we can evaluate the **sampling uncertainty**. (Remember that we assume that we can *only* observe a sample not population, which is the core assumption of the **statistical inference**. See Chapter 1.)
- Before pursuing this, we must repeat an important warning: How we *collect* data is at least as important as how we analyze it. In particular, a sampling should be *representative* of the population, and **random sampling** is often the best way to achieve this.
- If a sample is not random, it may be so **biased** that it is worse than useless!

Example

: Consider a telephone survey of consumer attitudes toward fast foods.

- What if a survey of residences is conducted from 9am to 5pm?
  Answer: It will miss [   ], the very people most likely to use and appreciate fast foods.
- Thus, even if the phone list were randomly selected, we see the responses would still not be random.
- A better approach would be to select a smaller random sample of residences, and then phone back as often as necessary to get a complete, or at least a very high, response rate.
- A truly random sample of 20 replies like this might be much better than a biased sample ten times as large. A large but biased sample may look good because of its size, but in fact, it just consists of the same bias being repeated over and over!

Random Sampling

Definition

The **population** *is the total collection of objects or people to be studied, from which a sample is to be drawn.*

a. If you wish to predict an election, the population of interest might be all the American voters.

b. If you wish to estimate the average height of students in University of Tokyo, the population is all students in this university.

Remark: Note that (we assume) the population is **not** observable, but we can draw a sample from the population!

Remark: If the population is observable, we can just calculate the exact mean from the population, of course. There is no need to sample in this case. (Ex.) Census

The Random Sample

Definition

A sample is called a **random sample** *if each individual in the population is equally likely to be chosen every time we draw an observation.*

For example, we could take a random sample of 5 students in a class of 100 men in the following way:

- **Draw chips from a bowl**: Record each student's height on a chip, mix all these 100 chips in a large bowl, and then draw the sample of $n = 5$ chips.

Sampling with or without replacement

- There are two ways of sampling depending on whether or not we replace each chip before drawing the next, **sampling with replacement** and **sampling without replacement**.
- In small populations such as 100 people, if we draw a sample with replacement, later chips are completely independent of each chip drawn earlier. On the other hand, if each chip is **not** replaced, the probabilities involved in the draw of later chips will change, that is, later chips are dependent on each chip drawn earlier.

However, . . .

In **large populations**, even if we sample without replacement, it is practically the same as with replacement, so that we still essentially have **independence**.

We will consider the cases where the observations are independent below unless otherwise noted.

Sampling Distribution

Let's consider the case where we try to estimate the population mean μ by the sample mean $\bar{X}$.
We hope the sample mean $\bar{X}$ is a close estimate of the population mean μ.
An important question is how $\bar{X}$ varies from sample to sample.

Definition

The distribution of $\bar{X}$ is called the **sampling distribution**.

The sampling distribution tells us how $\bar{X}$ varies from sample to sample, from which we can get important information on how close $\bar{X}$ comes to μ.

There are two ways we can study the sampling distribution.

1. We could repeat the process of drawing sample and estimating the sample mean $\bar{X}$ over and over using a computer, and build up the sampling distribution. It is called **Monte Carlo** sampling.

2. A more precise and useful (but often more difficult) alternative is to derive **mathematical formulas** for the sampling distribution of $\bar{X}$. Once we have derived such formulas (as we do in the next section) they can be applied broadly to a whole multitude of sampling problems.

Moments of the Sample Mean

If we take a random sample of observations from this population and calculate the sample mean $\bar{X}$, how good will $\bar{X}$ be as an estimator of its target μ?
To answer this, we can calculate the mean and the variance of the $\bar{X}$.

$$E[\bar{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i]$$
$$= [\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad]$$

It tells us that **on average, the sample mean $\bar{X}$ will be “on target”, that is, equal to μ.**

Next, we calculate the variance of $\bar{X}$.

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}[Var(X_1) + Var(X_2) + \cdots + Var(X_n)] \\
&= \frac{1}{n^2}[n\sigma^2] = \frac{\sigma^2}{n}
\end{aligned}
$$

and

$$\text{Standard deviation of } \bar{X} = [\qquad\qquad\qquad]$$

Remark: Note that we used the fact that the n observations $X_1, \ldots, X_n$ are **independent**.

The standard deviation of $\bar{X}$ is commonly called the **standard error**, or **SE**:

$$\text{Standard error of } \bar{X}, \; SE = [\qquad\qquad\qquad]$$

This is sometimes called "**Square Root Rule**". This formula shows explicitly that the larger the value of n, the [ ] SE becomes. It then adds precision to the simple idea that the larger the sample, the more accuracy $\bar{X}$ estimates the population mean μ.

The Shape of the Sampling Distribution

In the previous section, we found the expected value and standard error of $\bar{X}$. The remaining issue is the shape of the sampling distribution.
There is a fundamentally important theorem called **Central Limit Theorem** on the distribution of $\bar{X}$.

> If the population follows the normal distribution, or the sample size is large (often $n = 10$ or 20 will be large enough), then **in either case** the sampling distribution has an approximately normal shape.

Our conclusion so far on random sampling may be summarized into one statement:

> **The Normal Approximation Rule**: In random samples of size n, the sample mean $\bar{X}$ fluctuates around the population mean μ with a standard error of $\sigma/\sqrt{n}$ (where σ is the population standard deviation). Therefore, as n increases, the sampling distribution of $\bar{X}$ concentrates more and more around the target μ. It also gets closer and closer to the normal distribution.

Example 6-2: A population of men on a large midwestern campus has a mean height $\mu = 69$ inches, and a standard deviation $\sigma = 3.22$ inches. If a random sample of $n = 10$ men is drawn, what is the chance the sample mean $\bar{X}$ will be within 2 inches of the population mean μ?

Example 6-3:
a. Suppose a large class in statistics has marks normally distributed around a mean of 72 with a standard deviation of 9. Find the probability that an individual student drawn at random will have a mark over 80.
b. Find the probability that a random sample of 10 students will have an average mark over 80.
c. If the population were not normal, what would be your answer to part **b**?

Example 6-4: A ski lift is designed with a total load limit of 10,000 pounds. It claims a capacity of 50 persons. Suppose the weights of all the people using the lift have a mean of 190 pounds and a standard deviation of 25 pounds. What is the probability that a random group of 50 persons will total more than the load limit of 10,000 pounds?

Proportions (Percentages)

We can apply the **Normal Approximation Rule** to the **binomial data** (e.g. the political poll data where there are only two parties, See Chapter 1).
Suppose that we are interested in the **proportion** of the population who will vote for Republican, π.

Question:
(i) What is $\bar{X}$ in this case?
(ii) What are the μ and σ^2 in this case?

Answer:
(i) We can treat each data as

$$X_i = \begin{cases} 1 \text{ if the individual votes Republican} \\ 0 \text{ otherwise} \end{cases}$$

Therefore, the sample proportion, P, can be written as

$$P = [\qquad\qquad\qquad].$$

Remark: Such random variables are called **dummy variables**, an indispensable concept in applied statistics.
(ii)

$$\begin{aligned} \mu &= E(X_i) = [\qquad\qquad\qquad\qquad\qquad] \\ \sigma^2 &= [\qquad\qquad\qquad\qquad\qquad] \end{aligned}$$

The Normal Approximation Rule for Proportions: In random samples of size n, the sample proportion P fluctuates around the population proportion π with a standard error of $\sqrt{\pi(1-\pi)/n}$.
Therefore, as n increases, the sampling distribution of P concentrates more and more around the target π. It also gets closer and closer to the normal distribution.

Example 6-6: Of your first 15 grandchildren, what is the chance there will be more than 10 boys? (We assume that the probability of a boy is 50 % here.)

6-16: In a large production run of millions of electronic chips, only 2% are defective. What is the chance that of 1,000 chips pulled off the assembly line, 40 or more would be defective?

6-19: In the 1988 U.S. Presidential election, 53.9% of the voters were for Bush. If a Gallup poll of 1000 voters have randomly sampled from the population, what is the chance it would have erroneously predicted Bush to have a minority?

7 章 母数の推定 (Parameter Estimation)

東京大学

2014

Point Estimation

- It is essential to remember that the population mean μ and variance σ^2 are constants (though generally unknown).

These are called **population parameters**.

- By contrast, the sample mean $\bar{X}$ and sample variance s^2 are random variables. Each varies sample to sample, according to its sampling distribution.

A random variable such as $\bar{X}$ and s^2, which is calculated from the observations in a sample, is called **sample statistic**.

There can be several estimators to estimate the population parameter.
For example...
How good is the sample mean $\bar{X}$ as an estimator of μ? Would the sample median be better?
To answer such questions, we now develop **criteria** for judging a "good" estimator.

Efficiency of Unbiased Estimators

We already have noted that the sample mean $\bar{X}$ is, on average, exactly on its target μ. We therefore call $\bar{X}$ an **Unbiased Estimator** of μ.
To generalize, we consider any population parameter θ and denote its estimator by U. If, on average, U is exactly on target, it is called an unbiased estimator.
More formally, we define:

Definition

U is an unbiased estimator of θ if

$$E(U) = \theta.$$

Of course, an estimator V is called biased if $E(V)$ is different from θ. In fact, bias is defined as this difference:

Definition

$$Bias \equiv E(V) - \theta.$$

Example 7-1: Suppose each of the 200,000 adults in a city under study has eaten a number of X of fast-food meals in the past week. However, a residential phone survey on a week-day afternoon misses those who are working -the very people most likely to eat fast foods. As shown the table, this leaves a small subpopulation who would respond, especially small for higher value of X.

	Whole Target Population		Subpopulation Responding	
X = Number of meals	Freq. f	Rel. Freq. f/N	Freq. f	Rel. Freq. f/N
0	100,000	0.5	38,000	0.76
1	40,000	0.2	6,000	0.12
2	40,000	0.2	4,000	0.08
3	20,000	0.1	2,000	0.04
	200,000	1	50,000	1

a. What is the mean μ of the whole target population, and the mean μ_R of the subpopulation who would respond?

b. A random sample of 200 phone calls will bring a response of about 50, whose average $\bar{R}$ will be used to estimate μ. What is its bias?

Efficient Estimators (Minimum Variance)

As well as being on target on the average, we also would like the distribution of an estimator to be highly concentrated – that is, to have a small variance.

This is the notion of **efficiency.**

We define the relative efficiency of two unbiased estimators:

Efficiency of V compared to W $\equiv \frac{Var(W)}{Var(V)}$

Example 1: Let's compare the **sample mean** $\bar{X}$ and the **sample median** $Med(X)$ to estimate the center of a **normal** population.
In sampling from a normal population, it can be shown that for large samples,

$$Var(Med(X)) \approx 1.57\sigma^2/n.$$

Hence,

$$\text{Efficiency of } \bar{X} \text{ relative to } Med(X) \approx \frac{1.57\sigma^2/n}{\sigma^2/n} = 1.57.$$

We conclude that, in estimating the center of a normal population, the sample mean $\bar{X}$ is about 57% **more efficient** than the sample median $Med(X)$.

Example 2: Let's compare the **sample mean** $\bar{X}$ and the **sample median** $Med(X)$ to estimate the center of a population with thicker tails than the normal, which is called the **Laplace** distribution.
In sampling from a Laplace population, it can be shown that for large samples,

$$Var(Med(X)) \approx 0.5\sigma^2/n.$$

Hence,

$$\text{Efficiency of } \bar{X} \text{ relative to } Med(X) \approx \frac{0.5\sigma^2/n}{\sigma^2/n} = 0.5.$$

We conclude that, in estimating the center of a Laplace population, the sample mean $\bar{X}$ is about 50% **less efficient** than the sample median $Med(X)$.
This indicates that, when a population has thicker tails (i.e., outlying observation are likely to occur), the sample mean has larger variance while the variance of the sample median does not increase much because it ignores the distant outliers.

Efficiency of Biased and Unbiased Estimators

Now suppose we are comparing both biased and unbiased estimators. How can we make precise the notion of being "closest to the target overall?"

The important criterion of for judging a estimator V is

Definition

Mean squared error (MSE) $\equiv E[(V-\theta)^2]$.

In fact, it can be shown that MSE is a combination of variance and bias:

$$MSE = (\text{Variance of estimator}) + (\text{its bias})^2. \tag{1}$$

We choose the estimator that [ ] this MSE.
We give a general definition of the relative efficiency of two estimators:

For any two estimators–whether biased or unbiased–

$$\text{Efficiency of } V \text{ compared to } W \equiv \frac{MSE(W)}{MSE(V)}$$

Proof of (1):

$$
\begin{aligned}
MSE &= E[(V-\theta)^2] \\
&= E[(V-E[V]+E[V]-\theta)^2] \\
&= E[(V-E[V])^2]+(E[V]-\theta)^2+2E[V-E[V]](E[V]-\theta) \\
&= (\text{Variance of estimator}) + (\text{its bias})^2
\end{aligned}
$$

Example 7-3: In Example 7-1, recall the phone survey of 50 response from 200 calls, that had a serious nonresponse bias. In addition, the average response $\bar{R}$ has variability too.

a. To measure how much $\bar{R}$ fluctuates around its target μ overall, calculate its MSE.

b. If the sample size was increased fivefold, how much would the MSE be reduced?

c. A second statistician takes a sample survey of only $n = 20$ phone calls, with persistent follow-up until he gets a response. Let this small but unbiased sample have a sample mean denoted by $\bar{X}$. What is its MSE?

d. In trying to publish his results, the second statistician was criticized for using a sample only $1/10$ as large as the first. In fact, his sample size $n = 20$ was labeled "ridiculous". What defense might he offer?

Consistency: Eventually on Target

Like efficiency, **consistency** is one of the desirable properties of estimators. A **consistent** estimator is one that concentrates in a narrower and narrower band around the target as sample size n increases indefinitely. One of the conditions that makes an estimator consistent is if its MSE approaches zero in the limit:

One of the conditions that makes an estimator consistent is:
if its bias and variance **both** approaches zero as the sample size n increases.

Example 7-4

a. Is $\bar{X}$ a consistent estimator of μ?

b. Is P a consistent estimator of π?

c. Is the average response $\bar{R}$ in Example 7-1 (based on a 25% response rate) a consistent estimator of μ?

Confidence Intervals

- Point estimates do not give us any information about the reliability of it.
- It is important to know how reliable the estimates are.
- The **standard error (SE)** is one of the important measure of it.
- However, the sampling distributions of statistics give more information about it. We can evaluate the "probability" that the true parameters are included in some interval.

A Single Mean: Theory (We assume σ^2 is KNOWN here)

How can we construct the **confidence interval** so that it includes the true mean with a particular "probability"? (It is common to choose 95% confidence.)

- Note that $Pr(|Z| < [\qquad]) = 0.95$ from Table IV.
- From the CLT or Normal approximation rule, we have $\bar{X} \dot{\sim} N(\mu, \frac{\sigma^2}{n})$.

These facts imply (as we discussed in Chapter 6)

$$Pr(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 1.96) \approx 0.95.$$

We can rewrite

$$
\begin{aligned}
&0.95 \approx Pr\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 1.96\right) \\
&= Pr\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) \\
&= Pr\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right). \qquad (2)
\end{aligned}
$$

Hence we have

$$Pr(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) \approx 0.95. \tag{3}$$

Remark: We note that μ is population constant. It is a probability statement about the random interval $\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}$ to $\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$. The implication of the equation (3) is as follows:

Suppose the statistician draws samples and calculates confidence intervals again and again, each time from a different random sample. The statement (3) indicates that, **in the long run,** 95% **of the intervals constructed this way will bracket the true mean** μ.

The 95% confidence interval is written as

$$\mu = \bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}. \tag{4}$$

If we wish to be more confident, for example, 99% confident, then the interval must be large enough to encompass 99% of the probability. Since this leaves .005 in each tail, the 99% confidence interval would become

$$\mu = \bar{X} \pm [\qquad] \frac{\sigma}{\sqrt{n}}.$$

Thus the confidence interval becomes [wider] than the 95% one.
In general,...
Let $z_{\alpha/2}$ be the value of $\alpha/2$ percentile of the standard normal distribution.
The $1-\alpha$ confidence interval is written as

$$\mu = \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \tag{5}$$

8-1 Make the correct choice in each square bracket.

a. The sample mean [$\bar{X}$, μ] is an unbiased estimate of the population mean [$\bar{X}$, μ] – assuming the sample is [random, very large].

b. $\bar{X}$ fluctuates from sample to sample with a standard deviation equal to [σ/n, $\sigma/\sqrt{n}$], which is also called the [standard error SE, population standard deviation].

c. If we make an allowance of about [$\sqrt{n}$, 2] standard errors on either side of $\bar{X}$, we obtain an interval wide enough that it has a 95% chance of covering the target μ. This is called the 95% confidence interval for [$\bar{X}$, μ].

d. A statistician who constructed a thousand of these 95% confidence intervals over his lifetime would miss the target [practically never, about 50 times, about 950 times]. Of course, he [would, would not] know just which times these were.

e. For greater confidence such as 99%, the confidence interval must be made [narrower, wider].

Small sample t

- In the previous section it was assumed, quite unrealistically, that a statistician knows the true population standard deviation σ.
- In this section, we consider the practical case where σ **is unknown**.
- With σ unknown, the statistician wishing to evaluate the confidence interval (5) must use some estimate of σ – with the most obvious candidate being the sample standard deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.
- But the use of s introduces **an additional source of unreliability, especially if the sample is small.**
- To retain 95% confidence, we must therefore widen the interval.
- We do so by replacing the $z_{.025}$ value taken from the standard normal distribution with a larger $t_{.025}$ value taken from a similar distribution called **Student's t distribution**.

When we substitute s and the compensating $t_{.025}$ into (4), we obtain:

95% confidence interval for the population mean

$$\mu = \bar{X} \pm t_{.025} \frac{s}{\sqrt{n}}. \tag{6}$$

The value $t_{.025}$ is listed in the shaded column of Table V, and is tabulated according to the degree of freedom (d.f.):

$$\begin{aligned} \text{d.f.} &\equiv \text{amount of information used in calculation } s^2 \\ &\equiv \text{divisor in } s^2 \end{aligned}$$

In calculating s^2, d.f. $= n - 1$.
For example, if the sample size is $n = 4$, we read down Table V to d.f. $=$ [ ], which gives $t_{.025} =$ [   ] to use in (6).

In practice, when do we use the normal z table, and when the t table?

- In the rare case when σ is known, the normal z value in (4) is appropriate.
- In the usual case when σ is estimated with s, the t value in (6) is appropriate –regardless of sample size.
- However, **as the sample size grows larger than 100, say, the normal z becomes a good approximation to the t.** (For example, as we read down the t-table , for d.f. $= 120$ we should use $t_{.025} = 1.98$; but using $z_{.025} = 1.96$ is an excellent approximation.)
- So in practice when σ is unknown, t needs to be used only for small samples ($n < 100$).

Example 8-2: From a large class, a random sample of 4 grades were drawn: 64, 66, 89, and 77. Calculate a 95% confidence interval for the whole class mean μ.

Difference in Two Means, Independent Samples: If Population Variances are Known (In Theory)

Two population means are commonly compared by forming their difference:

$$(\mu_1 - \mu_2).$$

This difference is the population target to be estimated. A reasonable estimate of this is the corresponding difference in sample means:

$$(\bar{X}_1 - \bar{X}_2).$$

Using a familiar argument, we can construct the 95% confidence interval around the estimate:

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm z_{.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

because $Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. (Note that two samples are assumed to be **independent** here.)

If Population Variance is Unknown (In Practice)

- In practice, the population σ is not known, and has to be replaced with an estimate, customarily denoted by s_p.
- We assume here that σ_1 and σ_2 are known to have a common value, say σ. (This is a rather strong assumption, but we assume it at this point.)

Then $z_{.025}$ has to be replaced with the broader value $t_{.025}$, and so:

95% confidence interval, using independent samples, when both populations have the same underlying variance:

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{.025} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

How do we derive the estimate s_p?

- Since both populations have the same variance σ^2, it is appropriate to pool the information from both samples to estimate it.
- So our estimate is called the **pooled variance** s_p^2.
- we add up all the squared deviation form both samples, and then divide by the total d.f. in both samples, $(n_1 - 1) + (n_2 - 1)$.

$$s_p^2 = \frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}$$

where X_1 (or X_2) represents the typical observations in the first (or second) sample. In this case, the d.f. for t is

$$\text{d.f.} = (n_1 - 1) + (n_2 - 1).$$

Example 8-3: From a large class, a random sample of 4 grades were drawn: 64, 66, 89, and 77. From a second large class, an independent sample of 3 grades were drawn: 56, 71, and 53. Calculate a 95% confidence interval for the difference between the two class means, $\mu_1 - \mu_2$.

Difference in Two Means, Matched Samples

- In the previous section we were using **independent** samples, for instance, a sample of students' grades in the fall was compared to a fresh sample of students' grades in the spring.
- In this section we will consider **dependent** samples, called **matched** or **paired** samples.

Analysis based on Individual Difference

Suppose a comparison of fall and spring grades is done using the **same students** both times. Then the paired grades (spring X_1 and fall X_2) for each of the students can be set out as follows:

	Observed Grades	
	X_1	X_2
Student	(Spring)	(Fall)
Trimble	64	57
Wilde	66	57
Giannos	89	73
Ames	77	65

- The natural first step is to see how each student changed; that is calculating the difference $D = X_1 - X_2$, for each students.
- Once these differences are calculated, we can proceed to treat the four differences D now as a single sample, and analyze them just as we analyze any other single sample.

95% confidence interval, using matched samples

$$\Delta = \bar{D} \pm t_{.025} \frac{s_D}{\sqrt{n}}.$$

For our sample, as we calculate $\bar{D} = [\quad\quad]$, $s_D = [\quad\quad\quad]$, and d.f.$= n - 1 = [\quad\quad]$, we obtain

$$\Delta = [\quad\quad] \pm [\quad\quad\quad\quad\quad\quad]$$

Remark: Comparing to Example 8-3, we find that the matched-pair approach gives **a much more precise interval estimate**. So, pairing is obviously a desirable feature to design into any experiment, where feasible.

Proportion: Large Sample Formula

Confidence intervals for proportions (percentages) are very similar to means. We simply use the appropriate form of the normal approximation rule, and so obtain the 95% confidence interval for π:

95% confidence interval for the population, for large n

$$\pi = P \pm 1.96\sqrt{\frac{P(1-P)}{n}}$$

For this to be a good approximation, the sample size n ought to be large enough so that at least 5 successes and 5 failures turn up.

Difference in Two Proportions: Large Samples

Just as we derived the confidence interval to compare two means, we could similarly derive the confidence interval to compare two population proportions:

95% confidence interval for the difference in proportions, for large n_1 and n_2, and independent samples

$$(\pi_1 - \pi_2) = (P_1 - P_2) \pm 1.96\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$