
SPSS教程：计算率或率差的95%可信区间

作者：张耀文 来源：医咖会

本文原地址：<https://www.iikx.com/news/statistics/12380.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS教程：计算率或率差的95%可信区间

。临床研究中，经常要计算某个样本的率(proportion)，以及根据样本的率估算总体率的95%可信区间(Confidence Interval, CI)。样本率的计算很容易，那如何计算估算总体率的95%CI呢？

对于此类二项分布的数据，如果“数据量足够大”，一般可以使用正态近似法估算总体率的可信区间。

那如何才是“数据量足够大”呢？这个没有统一的标准。有学者推荐若以相对误差控制在10%左右为实际可接受标准，则使用正态近似法估计总体率95%可信区间的近似条件为：

- 1)当 $n \times p=5$ 时，样本率应在40%-60%之间;
- 2)当 $n \times p=10$ 时，样本率应在20%-80%之间;
- 3)当 $n \times p=15$ 时，样本率应在1%-99%之间。由于1%-99%几乎包括了二项分布率的实际应用范围，且在此条件下的可信度与1- 水平相当，故 $n \times p=15$ 为应用近似法的“理想条件” [1]。

1、单个率及其95%CI。

正态近似估计单个率的 95%CI 具体公式是 $p \pm Z_{\alpha/2} \sqrt{\frac{pq}{n}}$,

对应 95%CI 的公式则是 $p \pm 1.96 \sqrt{\frac{pq}{n}}$, 这里的 p 指关心的率,
 $q=1-p$ 。

假设有男性研究对象 200 人, 患糖尿病的人数为 73 人。

这个数据量是可以按照正态近似法计算的, 则

$$p=73/200=0.3650, \text{ 95\%CI 为 } 0.365 \pm 1.96 \sqrt{\frac{0.3650*0.6350}{200}}$$
$$= 0.3650 \pm 0.0667 = 0.2983 \text{ to } 0.4317。$$

也可以使用各种软件计算率及95%CI, 一个在线工具是VassarStats

(<http://vassarstats.net/index.html>)。

在VassarStats的主界面点击Proportions The Confidence Interval of a Proportion即可计算。不同的是该在线工具使用不是正态近似法, 而是Wilson法, 同时与手工计算取舍小数点不同, 所以计算结果略有差别。

The Confidence Interval of a Proportion

This unit will calculate the lower and upper limits of the 95% confidence interval for a proportion, according to two methods described by Robert Newcombe, both derived from a procedure outlined by E. B. Wilson in 1927 (references below). The first method uses the Wilson procedure without a correction for continuity; the second uses the Wilson procedure with a correction for continuity.

For the notation used here, n = the total number of observations and k = the number of those n observations that are of particular interest. Thus, if one observes 23 recoveries among 60 patients, $n = 60$, $k = 23$, and the proportion is $23/60 = 0.3833$.

To calculate the lower and upper limits of the confidence interval for a proportion of this sort, enter the values of k and n in the designated places, then click the «Calculate» button.

k =	<input type="text" value="73"/>	Proportion =	<input type="text" value="0.365"/>
n =	<input type="text" value="200"/>		
<input type="button" value="Reset"/>		<input type="button" value="Calculate"/>	

95% confidence interval: no continuity correction

Lower limit =

Upper limit =

95% confidence interval: including continuity correction

Lower limit =

Upper limit =

2、两个率的差及率差的95%CI。

具体公式是 $(p_1 - p_2) \pm Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ ，对应 95%CI

的公式则是 $(p_1 - p_2) \pm 1.96 \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ ，

这里的 p_1 、 p_2 指关心的两个率， $q_1=1-p_1$ 、 $q_2=1-p_2$ 。

假设有男性研究对象 200 人，患糖尿病的人数为 73 人；

女性研究对象 300 人，患糖尿病的人数为 85 人，

则 $p_1 - p_2 = 73/200 - 85/300 = 0.0817$ ，95%CI

为 $0.0817 \pm 1.96 \sqrt{\frac{0.3650 * 0.06350}{200} + \frac{0.2833 * 0.7167}{300}}$

$= 0.0817 \pm 0.0840 = -0.0023 \text{ to } 0.1657$ 。

使用 VassarStats 工具计算时，可在 VassarStats 的主界面点击 Proportions The Confidence Interval for the Difference Between Two Independent Proportions。

The Confidence Interval for the Difference Between Two Independent Proportions

This page will calculate the lower and upper limits of the 95% confidence interval for the difference between two independent proportions, according to two methods described by Robert Newcombe, both derived from a procedure outlined by E.B.Wilson in 1927 (references below). The first method uses the Wilson procedure without a correction for continuity; the second uses the Wilson procedure with a correction for continuity.

For the notation used here, n_a and n_b represent the total numbers of observations in two samples, A and B; k_a and k_b represent the numbers of observations within each sample that are of particular interest; and p_a and p_b represent the proportions k_a/n_a and k_b/n_b , respectively. Thus, if sample A shows 23 recoveries among 60 patients, $n_a=60$, $k_a=23$, and the proportion is $p_a=23/60=0.3833$. If sample B shows 18 recoveries among 72 patients, $n_b=72$, $k_b=18$, and the proportion is $p_b=18/72=0.2500$. The difference between the two proportions is $\text{diff}=p_a-p_b=0.3833-0.2500=0.1333$.

Note that this method is recommended only for the case where p_a-p_b is equal to or greater than zero. Thus, the sample with the larger proportion should be designated as Sample A and the one with the smaller proportion should be designated as Sample B. To calculate the lower and upper limits of the confidence interval for a difference of this sort, enter the values of k and n for samples A and B in the designated places, then click the «Calculate» button.

Larger Proportion	Smaller Proportion
Sample A	Sample B
$k_a =$ <input type="text" value="73"/>	$k_b =$ <input type="text" value="85"/>
$n_a =$ <input type="text" value="200"/>	$n_b =$ <input type="text" value="300"/>
$p_a =$ <input type="text" value="0.365"/>	$p_b =$ <input type="text" value="0.2833"/>
$p_a - p_b =$ <input type="text" value="0.0817"/>	

95% confidence interval: no continuity correction

Lower limit =

Upper limit =

95% confidence interval: including continuity correction

Lower limit =

Upper limit =

3、SPSS计算。

SPSS没有现成的模块计算率或率差的95%CI，但可以通过编程界面实现。将数据录入SPSS并且指定95%CI对应的值 $z_{crit}=1.96$ (可取任意变量名)。

打开Syntax界面，输入以下程序即可计算完成。

最终得到的结果如下。

实际上，SPSS的Syntax中编写程序，还是使用正态近似法的计算公式，还不如手工(或使用EXCEL计算)方便。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发