

Доверительный интервал: общие принципы и значение. Методика расчета

Директор Сотрудничающего Центра ВОЗ по
статистике и анализу здоровья населения,
Д.э.н., профессор А.Е.Иванова
Ivanova-home@yandex.ru

Общие принципы

- Идея статистического оценивания параметров генеральной совокупности по выборочным данным сводится к тому, что выборочная характеристика какого-либо параметра (например, показателя смертности) является не точным, а приближенным значением – *оценкой* – этого же параметра в генеральной совокупности (как правило, значение в генеральной совокупности неизвестно).
- Возникает вопрос: как сильно отклоняется эта оценка от истинного значения? В частности, нельзя ли указать такую величину *ошибки*, которая "практически достоверно" (т.е. с вероятностью, близкой к единице) гарантировала бы, что выборочная оценка не отличается от неизвестного значения более чем на величину этой ошибки?
- То есть, нельзя ли указать вокруг выборочного значения такой интервал, который бы с заданной (достаточно высокой) вероятностью – *доверительной вероятностью* – "накрывал" бы истинное значение этого параметра?
- Этот интервал в математической статистике называется *доверительным интервалом*; его величина зависит как от доверительной вероятности (т.е. надежности оценивания), так и от объема выборки.
- *Доверительная вероятность P* – это степень уверенности в том, что доверительный интервал действительно будет содержать истинное (неизвестное) значение параметра в генеральной совокупности.
- Величина доверительного интервала $X \pm t \times se$ пропорциональна ошибке выборки se
Кроме того, величина доверительного интервала зависит от параметра t , который выбирается, исходя из требуемого уровня надежности.
- Значение параметра $t = 1$ соответствует уверенности 67%, $t = 2$ – уверенности 95%, $t = 3$ – уверенности 99,7%.
- На практике для построения доверительного интервала берется, как минимум, удвоенная ошибка выборки, чтобы обеспечить уверенность не менее 95%. Таким образом, 95%-доверительный интервал: диапазон значений, внутри которого (с вероятностью 95%) лежат истинные значения (изучаемого показателя) всей популяции.

Стандартные ошибки и доверительные интервалы для возрастных и общих коэффициентов смертности

$$CI = \frac{100000}{P} \cdot (D \pm 1,96 \cdot \sqrt{D})$$

Величину \sqrt{D} в формуле расчета доверительных интервалов необходимо заменить на

$$\sqrt{D \cdot \left(1 - \frac{D}{P}\right)}$$

- Расчет доверительных интервалов для повозрастных и общих коэффициентов смертности (рассчитанных на 100000 населения)
- Если коэффициенты смертности превышают величину 10000 на 100000 (что часто наблюдается для возрастов старше 65 лет),

Пример расчета

	Популяция А		
	Численность	Число умерших	Коэффициент смертности на 1000
Дети	10000	80	8,0
Взрослые	15000	165	11,0
Пожилые	25000	375	15,0
Всего	50000	620	12,4

$$CI = \frac{100000}{P} \cdot (D \pm 1,96 \cdot \sqrt{D})$$

$$CI (M_{ch}) = (1000/10000) * (80 \pm 1.96 * \sqrt{80}) = 0.1 * (80 \pm 17.53) = 8.0 \pm 1.75$$

$$\sqrt{D \cdot \left(1 - \frac{D}{P}\right)}$$

$$CI (M_{old}) = (1000/25000) * (375 \pm 1.96 * \sqrt{375 * (1 - 375/25000)}) = 0.04 * (375 \pm 37.67) = 15 \pm 1.51$$

Стандартные ошибки и доверительные интервалы для стандартизованных коэффициентов

$$SE(ADR) = \sqrt{\sum_i \left(\frac{P_{is}}{P_s}\right)^2 \times \left(\frac{D_i}{P_i}\right) \times \left(\frac{1 - D_i/P_i}{P_i}\right)}$$

- Стандартная ошибка для коэффициента, стандартизованного прямым способом

$$se_{SMR} = \frac{\sqrt{D}}{\frac{1}{1000} \cdot \sum M_{si} \cdot P_{ui}}$$

- Стандартная ошибка для стандартизованного отношения смертности (косвенный и обратный способ)

Пример расчета

	Популяция А		Популяция "А"+"Б"	Произведе- ние возрастной смертности стандарта на численность изучаемого населения
	Числен- ность	Фактичес- кое число умерших		
Дети	10000	80	9,4	94000
Взрослые	15000	165	11,5	172500
Пожилые	25000	375	15,3	382500
Всего	50000	620	12,1	649000

$$se_{SMR} = \frac{\sqrt{D}}{\frac{1}{1000} \cdot \sum M_{si} \cdot P_{ui}}$$

$$SE_{SMR} = \sqrt{620 / (649000 / 1000)} = 0.04$$

$$CI_{SMR} = 0.96 \pm 1.96 * 0.04 = 0.96 \pm 0.08$$

Индекс стандартизации (Стандартизованное отношение смертности) – 0,96

Пример расчета

	Популяция А			Популяция Б			Стандартная популяция
	Численность	Число умерших	Коэффициент смертности на 1000	Численность	Число умерших	Коэффициент смертности на 1000	
Дети	10000	80	8	25000	250	10	35000
Взрослые	15000	165	11	15000	180	12	30000
Пожилые	25000	375	15	10000	160	16	35000
Всего	50000	620	12,4	50000	590	11,8	100000

Стандартизованный коэффициент смертности: популяция А = 11,35

$$SE(ADR) = \sqrt{\sum_i \left(\frac{P_{is}}{P_s}\right)^2 \times \left(\frac{D_i}{P_i}\right) \times \left(\frac{1 - D_i/P_i}{P_i}\right)}$$

$$SE_{ADR} = \sqrt{[(35000/100000)^2 * (80/10000) * ((1-80/10000)/10000) + (30000/100000)^2 * (165/15000) * ((1-165/15000)/15000) + (35000/100000)^2 * (375/25000) * ((1-375/25000)/25000)] = 0.48}$$

$$CI_{ADR} = 11.35 \pm 1.96 * 0.48 = 11.35 \pm 0.95$$