

军人人体尺寸数据库抽测样本量的确定

夏鹏泽 曾长松 姜志华 王永成

(总后军需装备研究所,北京,100010)

摘要:计算建立中国军人服装用人体尺寸数据库所需抽测的样本量。并根据实际建库男、女军人样本量,计算了百分位数的估计量所能达到的绝对精度。

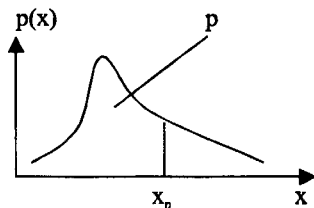
关键词:人体测量 数据库 样本量 相对精度 绝对精度 服装

中图分类号:TS 932

使用BM-1人体尺寸非接触测量仪对全军男、女军人人体进行抽样测量,建立准确反映我国军人人体特征的中国军人服装用人体尺寸数据库,是对我国军人人体特征进行深入研究、确立完善的军服号型体系,改进军服设计、方便军服生产供应的数据基础。而正确确定抽样测量样本量,对既能避免因抽样过多造成人力物力的浪费,又能满足实际工作的需要是极其重要的。

1 人体测量需估计的目标量

在人体尺寸抽样调查中,我们所关心的重点并不是样本某项尺寸的均值、方差等通常所遇到的统计量。因为样本均值对总体均值的估计即使达到了相当精确的程度,与正确分析军人人体特征、研制标准人体模型、制订服装号型标准,使它能符合大多数人的情况和满足大多数人的需要并无直接的联系。而我们更感兴趣的是人体各项尺寸的分布情况。因此,我们所估计的目标量主要是这些尺寸的百分位数。



p分位数图

记总体某个尺寸变量 X 的 p 分位数为 x_p ($0 < p < 1$), 则下式成立:

$$P(X \leq x_p) = p \quad (1)$$

若记变量 X 的密度函数为 $p(x)$, 则分位数 x_p 表示 x_p 左右的概率密度曲线下的面积为 p (如图所示)。

举例说,如果抽样结果显示,有75%的人身高不超过1730 mm,则可记为 $p = 0.75$, 变量身高 X 的75%分位数 $x_{0.75} = 1730$ mm。应用中,较重要的分

位数有 $x_{0.025}, x_{0.05}, x_{0.10}, x_{0.20}, x_{0.50}, x_{0.80}, x_{0.90}, x_{0.95}, x_{0.975}$ 等。其中, $x_{0.50}$ 是中位数。

2 百分位数估计量 \hat{x}_p 的相对精度

根据所测样本,可按一定方法对 x_p 进行估计,记估计量为 \hat{x}_p 。对一般的估计量 θ ,精度的提法有绝对精度和相对精度两种。这两种精度都是在一定的概率意义以下,例如对于给定的置信水平 $1 - \alpha = 95\%$,绝对精度 Δ 即是满足下式的量,

$$P\{|\theta - \theta| \leq \Delta\} = 0.95 \quad (2)$$

而通常的相对精度,即是指满足下式的 r ,

$$P\{|\theta - \theta| / \theta \leq r\} = 0.95 \quad (3)$$

上两式中的 θ 都是被估计参数的真值。按这两种定义不易确定群体人体尺寸精度与样本量的关系,而且也不一定能够满足我们的需要。

根据冯士雍等人的观点,对于一个很小的数 d (例如1%),取置信水平 $1 - \alpha = 95\%$,使满足

$$P(x_{p-d} < \hat{x}_p < x_{p+d}) = 0.95 \quad (4)$$

我们称 d 为估计量 \hat{x}_p 的相对精度。例如,若调查结果显示,1700 mm 是样本身高的50%分位数,如果要求身高的估计量的精度 $d = 1\%$,则在所估计的总体中,应有不少于49%、不多于51%的人体的身高低于1700 mm。可见,相对精度 d 对于正确确定人体尺寸的分布规律,是重要且实用的。用它作为人体测量需估计的目标量的相对精度是科学的。

3 样本量的确定

样本量的确定取决于对估计量精度的要求,对精度要求愈高,所需样本量愈大。在精度一定时,在不同的分位点上,样本量的估计结果也不相同。但根据冯士雍等人的研究得知,以估计50%分位数(即平均值)所要求的样本量最大。

设要求 x_p 的相对精度为 d ,对于 $p = 0.50$,用 $\hat{x}_{0.50} = \bar{x}$ 估计 $x_{0.50}$,由于人体尺寸可近似为服从正态分布 $N(\mu, \sigma^2)$, $x_{0.50} = \mu$, $\bar{x} \sim N(\mu, \sigma^2/n)$,取1

$-\alpha = 0.95$, 则

$$P(x_{0.50} - \mu_{1-\sigma/2} \sqrt{n} < x_{0.50} < x_{0.50} + \mu_{1-\sigma/2} \sqrt{n}) = 0.95$$

即

$$P(x_{0.50} - 1.96\sigma/\sqrt{n} < x_{0.50} < x_{0.50} + 1.96\sigma/\sqrt{n}) = 0.95 \quad (5)$$

根据相对精度 d 的定义, 样本量 n 应满足

$$x_{0.50} - 1.96\sigma/\sqrt{n} = x_{0.50-d}$$

或

$$x_{0.50} + 1.96\sigma/\sqrt{n} = x_{0.50+d}$$

由于正态分布的对称性, 对于确定的 d , 由以上两式确定的 n 值相等。即

$$n = \{1.96/[(x_{0.50+d} - x_{0.50})/\sigma]\}^2 = (1.96/\mu_{0.50+d})^2 \quad (6)$$

例如给定 $d = 1\%$ 时, $n = (1.96/0.02507) \approx 6113$ 。

对于不同的 d 值, 所需的 n 值如表 1 所示。

表 1 用 \bar{x} 估计 $x_{0.50}$ 时, 不同的相对精度所需的简单随机抽样的样本量

$d(\%)$	0.5	1	1.1	1.2	1.3	1.4	1.5	1.6
n	24457	6113	5052	4245	3617	3118	2716	1886

由于我们在对军人人体型规律进行研究确定军服号型时, 考虑到供应上的方便等原因, 往往确定男军服与女军服两个系列, 而在男、女军人中不再对干部和战士进行区分。因此, 此次人体尺寸测量工作, 我们将全体军人分为男、女两个总体。由于男军人数量较多, 是军队的主体。因此取男军人人体尺寸分位数估计量的相对精度 $d = 1\%$, 女军人人体尺寸分位数估计量的相对精度 $d = 1.5\%$ 。由表 1 可知, 男、女军人简单随机抽样样本量分别为 6113 人和 2716 人。

但是在全军范围内, 对男、女军人进行简单随机抽样的方法进行人体测量是不可能的。我们采用的是整群抽样的方法。在总样本量相等的情况下, 整群抽样比简单随机抽样精度差。对于整群抽样, 实际样本量

$$n_0 = n[1 + (M - 1)\rho] \quad (7)$$

式中, M 为平均群体大小, 即一个测量单位的人数; ρ 为群内相关系数。根据部队特点, 我们取男军人 $M = 80$ 、女军人 $M = 50$ 。又根据全国成年人(工业人口)四川省人体试测数据分析结果(中国标准化与信息分类编码研究所), $\rho = 0.0075$ 。考虑到在实际测量中可能出现的问题, 再增加 5~10% 的余量, 因此共需实测男军人。

$$n_0 = 6113 \times [1 + (80 - 1) \times 0.0075] \times 1.10 \approx 11000(\text{人})$$

需实测女军人

$$n_0 = 2716 \times [1 + (50 - 1) \times 0.0075] \times 1.05 \approx 3900(\text{人})$$

科学的抽样方法, 与正确确定样本量同样重要, 本文不作具体论述。

4 数据库实际达到的绝对精度

所谓绝对精度, 是指估计量 \hat{x}_p 与真值 x_p 之间的最大绝对差 Δ 。取 $1 - \alpha = 95\%$, 绝对精度 Δ 满足

$$P(|\hat{x}_p - x_p| \leq \Delta) = 0.95 \quad (8)$$

Δ 值不仅与简单随机抽样样本量有关, 还与变量 X 的标准差 σ 有关, 当然也与分位数 p 和 \hat{x}_p 的估计方法有关。

4.1 简单随机抽样样本量的计算与男、女军人人体主要部位尺寸统计量

此次抽样调查, 实际建库样本量男军人 12058 人、女军人 3964 人。前面讲过, 实际建库军人样本量是以整群抽样方法取得的, 整群抽样样本精度较差, 不能直接用于绝对精度的计算, 而应将其转换成简单随机抽样样本量 n 。根据式(7)

$$n = n_0/[1 + (M - 1)\rho] \quad (9)$$

由此得男、女军人简单随机样本量分别为 7572 人、2899 人。

从“中国军人服装用人体尺寸数据库”求得, 男、女军人人体主要部位尺寸统计量见表 2。

4.2 绝对精度的计算

下面, 我们以 50%、5% 分位数为例, 来计算数据库对男、女军人主要部位尺寸百分位数的估计所能达到的绝对精度。

4.2.1 50% 分位数(即平均值)绝对精度的计算

根据式(5), 取 $1 - \alpha = 0.95$

$$\Delta = |\hat{x}_{0.50} - x_{0.50}| = 1.96\sigma/\sqrt{n} \quad (10)$$

我们用样本标准差估计总体标准差 σ , 得男、女军人主要部位尺寸估计值 $\hat{x}_{0.50}$ 的绝对精度见表 1。

4.2.2 p 分位数($0 < p < 1, p \neq 0.5$)绝对精度的计算

当用简单随机样本的分位数 \hat{x}_p 估计 x_p 时, \hat{x}_p 的渐近分布为

$$N(x_p, [\sigma^2 p(1 - p)]/[n\varphi^2(u_p)]) \quad (11)$$

则

$$\Delta = |\hat{x}_p - x_p| = 1.96 \sqrt{p(1 - p)}/\varphi(u_p) \sqrt{n} \quad (12)$$

我们用样本标准差估计总体标准差 σ , 令 $p = 0.05$, 得男、女军人主要部位尺寸估计值 $\hat{x}_{0.05}$ 的绝对精度见表 2。

表 2 男、女军人人体主要部位尺寸统计量及绝对精度

测量项目	均值	标准差	$\hat{x}_{0.50}$ 的 Δ	$\hat{x}_{0.05}$ 的 Δ
身高(男)	1707	53	1	3
胸围(男)	932	57	1	3
腰围(男)	757	74	1	4
臀围(男)	931	49	1	2
身高(女)	1618	52	2	4
胸围(女)	881	66	2	5
腰围(女)	705	74	3	6
臀围(女)	917	56	2	4

注:表中单位为 mm

5 结束语

经研究得知,在人体各部位尺寸中,以腰围标准差最大。同时,由于女军人抽测样本量较少,因此,在本数据库中,以女军人腰围的绝对误差最大。但其 50%、5%分位数的绝对误差仍不大于 6 mm,该数据库完全能够满足有关服装方面的使用要求。

参 考 文 献

- [1] 冯士雍、孙山泽、毕健:人体测量抽样方案目标量的估计及样本量的确定,《应用概率统计》,1989,5(4),11