统计与应用研究
2020 年 6 月第 2 卷第 2 期

The Application of Innovative Teaching Thought in Mathematical Statistics Teaching

Huang Weidong

Guangxi Normal University, Guilin

Abstract: Based on the interval estimation and single normal population, the selection of statistics quantile in mean interval estimation was discussed. The problem of variance interval estimation was also studied. The ideas and necessity of innovative education in the teaching of statistics were put forward.

Key words: Mathematical statistical; Interval estimation; Innovative education

Received: 2020-04-22; Accepted: 2020-05-07; Published: 2020-05-09

文章引用：黄卫东．创新性教学思想在数理统计教学中的运用［J］．统计与应用研究，2020，2（2）：59–65.
https://doi.org/10.35534/rsa.0202009

创新性教学思想在数理统计教学中的运用

黄卫东

广西师范大学，桂林

邮箱：wjd.2007@hotmail.com

摘　要：以区间估计的教学为例，讨论了单个正态总体均值区间估计统计量的选择问题和单个正态总体方差区间估计分位点的选择问题，提出了在数理统计课程教学中进行创新性教学的思想和必要性。

关键词：数理统计；区间估计；创新性教育

收稿日期：2020-04-22；录用日期：2020-05-07；发表日期：2020-05-09

1　引言

数理统计是统计学专业的一门主干课程，它充分利用概率论的思想和方法，

根据所获得的数据资料，对所关心的问题作出估计与检验。数理统计的应用相当广泛，已成为诸多学科从事科学研究的必不可少的数学工具。因此提高数理统计课程的教学质量，具有十分重要的实际意义。作者从事数理统计的教学多年，并在教学中不断地进行探索与实践，积累了一定的教学经验。以下作者将以区间估计的教学为例，介绍创新性教育的一些体会。

2　单个正态总体均值区间估计问题的教学探讨

在实际教学中，发现不少学生对总体方差已知且样本方差可求时，均值区间估计问题统计量的选择有一定的困惑，一部分学生认为选择统计量$U=\dfrac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$和$T=\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n-1}}}$都可以，其中$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$，$S^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$，以下同。对于这个问题，作者在教学中进行了较为深入的讨论，先看一道例题。

例 1　假设物体的重量服从正态分布 $N(\mu, 11^2)$，现用某仪器测量该物体的重量，重复测量 5 次，数据如下（单位：克）：1250，1265，1245，1260，1275，给定置信概率 1−α 为 0.95，试求物体重量的区间估计。

解法 1　选取$U=\dfrac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}\sim N(0,1)$为统计量，由$P(-u_{\frac{\alpha}{2}}<U<u_{\frac{\alpha}{2}})=1-\alpha$，得 μ 的 1−α 置信区间为

$$\left(\overline{X}-u_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}},\ \overline{X}+u_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}\right) \tag{1}$$

由 $\overline{X}$=1259，σ_0=11，α=0.05，$u_{0.025}$=1.96，n=5，代入（1）式得物体重量 μ 的 0.95 区间估计为（1249.4，1268.6）。

解法 2　选取$T=\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n-1}}}\sim t(n-1)$为统计量，由$P(-t_{\frac{\alpha}{2}}^{n-1}<T<t_{\frac{\alpha}{2}}^{n-1})=1-\alpha$，得 μ 的 1−α 置信区间为

$$\left(\overline{X}-t_{\frac{\alpha}{2}}^{n-1}\frac{s}{\sqrt{n-1}},\ \overline{X}+t_{\frac{\alpha}{2}}^{n-1}\frac{s}{\sqrt{n-1}}\right) \tag{2}$$

由 $\overline{X}$=1259，s^2=114，α=0.05，$t_{0.025}^{(4)}$=2.78，n=5，代入（2）式得物体重量 μ 的 0.95 区间估计为（1244.2，1273.8）。

https://doi.org/10.35534/rsa.0202009

分析对于解法 1 和解法 2，置信度都为 0.95，下面分别计算其精度（即区间的长度），

因此在本题中，选择统计量$U=\dfrac{\bar{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$比选择统计量$T=\dfrac{\bar{X}-\mu}{\frac{s}{\sqrt{n-1}}}$要好，因为在同样的置信概率下，前者的精度更高。

那么，这是否具有普遍性呢？作者对这个问题进行了一定的研究性教学，力求使学生在掌握基本知识的同时，也能拓宽其思维，逐步树立其创新思想，为日后的科学研究工作奠定基础．为此，作者在 Excel 里对不同的样本容量 n 进行了量的计算。

设总体$X\sim N(\mu,\sigma_0^2)$，σ_0^2已知，$\alpha=0.05$，X_1，X_2，…，X_N 为其样本，若选择统计量$U=\dfrac{\bar{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$时得到的区间长度记为 L_1，选择统计量$T=\dfrac{\bar{X}-\mu}{\frac{s}{\sqrt{n-1}}}$时得到的区间长度记为 L_2，以下分别计算 EL_1 和 EL_2。

易知$EL_1=2u_{\frac{\alpha}{2}}\dfrac{\sigma_0}{\sqrt{n}}$，又

$$\frac{ns^2}{\sigma_0^2}\sim\chi^2(n-1),E\left(\sqrt{\frac{ns^2}{\sigma_0^2}}\right)=\int_0^\infty\sqrt{x}\frac{1}{\left(2^{\frac{n-1}{2}}\right)r\left(\frac{n-1}{2}\right)}x^{\frac{n-2}{2}}\mathrm{e}^{-\frac{x}{2}}\mathrm{d}x=$$

$$\frac{1}{\left(2^{\frac{n-1}{2}}\right)r\left(\frac{n-1}{2}\right)}\int_0^\infty 2^{\frac{n}{2}}y^{\frac{n}{2}-1}\mathrm{e}^{-y}\mathrm{d}y=\sqrt{2}\frac{1}{r\left(\frac{n-1}{2}\right)}\int_0^\infty 2^{\frac{n}{2}}y^{\frac{n}{2}-1}\mathrm{e}^{-y}\mathrm{d}y=\sqrt{2}\frac{r\left(\frac{n}{2}\right)}{r\left(\frac{n-1}{2}\right)},$$

从而$ES=\sqrt{\dfrac{2}{n}}\dfrac{r\left(\frac{n}{2}\right)}{r\left(\frac{n-1}{2}\right)}\sigma_0$，所以$EL_2=2t_{\frac{\alpha}{2}}^{(n-1)}\dfrac{ES}{\sqrt{n-1}}=2t_{\frac{\alpha}{2}}^{(n-1)}\sqrt{\dfrac{2}{n(n-1)}}\dfrac{r\left(\frac{n}{2}\right)}{r\left(\frac{n-1}{2}\right)}\sigma_0$。

表 1 是经 Excel 计算得出的对应于不同样本容量 n 的平均区间长度对比，取 σ 0=1，α =0.05。

表 1　平均区间长度对比表

Table 1　Average interval length table

n 取值	EL_1	EL_2	n 取值	EL_1	EL_2	n 取值	EL_1	EL_2
1	3.92	–	11	1.18	1.31	30	0.357	0.370
2	2.77	14.34	12	1.13	1.24	40	0.309	0.317
3	2.26	4.38	13	1.08	1.17	50	0.277	0.282
4	1.96	2.93	14	1.04	1.13	60	0.253	0.257

续表

n 取值	EL_1	EL_2	n 取值	EL_1	EL_2	n 取值	EL_1	EL_2
5	1.75	2.36	15	1.01	1.13	70	0.234	0.237
6	1.60	1.96	16	0.98	1.04	80	0.219	0.221
7	1.48	1.69	17	0.95	1.01	90	0.206	0.208
8	1.38	1.61	18	0.92	0.98	100	0.196	0.197
9	1.30	1.57	19	0.89	0.94	110	0.186	0.188
10	1.23	1.39	20	0.87	0.92	120	0.178	0.180

根据表 1 中数据可以看出，当样本容量 n 较小时，从区间估计的精度来考虑，选择统计量$U=\dfrac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$比选择统计量$T=\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n-1}}}$要好，但随着 n 的增大，其精度会越来越接近，当 $n \geqslant 50$ 时，这种差异已经微乎其微了，因此当样本容量 n 较大时（如 $n \geqslant 50$），选择统计量$U=\dfrac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$和$T=\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n-1}}}$都是可行的，特别是当我们对总体 $\sigma=\sigma_0$ 的假定把握不大（即 $P(\sigma=\sigma_0)<1$）且样本容量 n 较大时，选择统计量$T=\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n-1}}}$将会更好些。

3　单个正态总体方差区间估计问题的教学探讨

对总体均值 $\mu=\mu_0$ 已知，样本均值可求时统计量的选择问题，采取与上述讨论类似的方法可以得到如下的结果：对于同样的置信度，当样本容量 n 较小时，如果从精度方面来考虑，选择统计量$\chi_1^2=\dfrac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2}$比选择统计量$\chi_2^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{\sigma^2}$要好，当样本容量 n 较大时（如 $n \geqslant 50$），选择统计量$\chi_1^2=\dfrac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2}$和$\chi_2^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{\sigma^2}$都是可行的，特别是当我们对总体 $\mu=\mu_0$ 的假定把握不大（即 $P(\mu=\mu_0)<1$）且样本容量 n 较大时，选择统计量$\chi_2^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{\sigma^2}$将会更好些。下面，作者将从另一个角度来探讨这个问题。

设总体 $X \sim N(\mu, \sigma^2)$，总体均值 $\mu=\mu_0$ 已知，X_1，X_2，…，X_n 为其样本，样本容量 n 较小，置信概率为 $1-\alpha$，由前面的讨论，求 σ^2 的区间估计时，

https://doi.org/10.35534/rsa.0202009

我们通常选择 $\chi_1^2=\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2}\sim\chi^2(n)$ 为统计量。

由 $P(\chi^2_{1-\frac{\alpha}{2}}(n)<\chi_1^2<\chi^2_{\frac{\alpha}{2}}(n))=1-\alpha$，得到 σ^2 的 1–α 置信区间为 $\left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{\frac{\alpha}{2}}(n)},\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right)$，其区间长度 $L_1=\sum_{i=1}^{n}(X_i-\mu_0)^2\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n)}-\frac{1}{\chi^2_{\frac{\alpha}{2}}(n)}\right)$，进一步计算得 $EL_1=n\sigma^2\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n)}-\frac{1}{\chi^2_{\frac{\alpha}{2}}(n)}\right)$。

众所周知，分位点 $\chi^2_{1-\frac{\alpha}{2}}(n)$ 和 $\chi^2_{\frac{\alpha}{2}}(n)$ 的选择，一般不会使得求出来的 σ^2 的区间长度最短，因为它依赖于 n 与 α 的选取。那

怎么来解决这个问题呢？在创新性教育实践中，作者采用了以下的方法，对这类区间估计问题进行了一些修正。

由上述讨论知，σ^2 的 1–α 置信区间为（T_1，T_2），其中 $T_1=\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{\frac{\alpha}{2}(n)}}$，$T_2=\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{1-\frac{\alpha}{2}(n)}}$。另外，选择 $U=\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$

为统计量，由 $P\left(\left|\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right|<u_{\frac{\alpha}{2}}\right)=1-\alpha, P\left(\sigma>\frac{\sqrt{n}\,|\bar{X}-\mu_0|}{u_{\frac{\alpha}{2}}}\right)=1-\alpha$，即 $P\left(\sigma^2>\frac{n(\bar{X}-\mu_0)^2}{(u_{\frac{\alpha}{2}})^2}\right)=1-\alpha$，因此 σ^2 的 1–σ 单边置信区间为（T_3，$+\infty$），其中 $T_3=\frac{n(\bar{X}-\mu_0)^2}{(u_{\frac{\alpha}{2}})^2}$。对 $0<\alpha<1$，$n\geqslant 2$，容易验证 $ET_3\geqslant ET_1$。

故 σ^2 的 $1-\alpha^*$ 置信区间为（T_3，T_2），其中 $\alpha^*=1-\frac{3}{2}\alpha$。

实践表明，采用这种方法求 σ^2 的区间估计，虽然置信度有所降低，但其精度却有较大的提高，在具体问题中有一定的应用价值。

4 结论

数理统计课程的教学，没有固定的教学模式，是一个逐步探索的过程 . 对于数理统计课程中某些内容的教学，可以根据实际情况超越教学大纲的要求，延伸教学内容，适当进行一些研究性的教学活动，这对于学生知识的掌握以及后

续专业课程的学习都有非常重要的意义。

参考文献

[1] 陈希孺. 数理统计引论 [M]. 北京：科学出版社，1999：17–35.

[2] 茆诗松，程依明，濮晓龙. 概率论与数理统计教程 [M]. 北京：高等教育出版社，2004：153–165.

[3] 刘韶跃，彭向阳，熊雄，等. 数理统计 [M]. 湘潭：湘潭大学出版社，2009：135–142.

https://doi.org/10.35534/rsa.0202009