

eAppendix

Selected methods for percentile estimation and their use in popular software

Method A: $p(n+1)$

Method A, cited for example by Altman¹ and CLSI², calculates an index $p(n+1)$. This method is attributed³ to Gumbel⁴ and Weibull⁵ and is widely available in software packages such as JMP (SAS Institute Inc, Cary, NC) (only method),⁶ NCSS (NCSS, Kaysville, UT) (recommended method),⁷ Minitab (Minitab Inc., State College PA),⁸ SAS (SAS Institute Inc, Cary, NC) (optional method),⁹ Statistica (StatSoft, Inc., Tulsa, OK) (optional method)¹⁰ and SPSS (IBM Corp., Somers, NY) (default method)¹¹.

Method B: $0.5+pn$

Method B, cited for example by Armitage et al.¹² and Lentner,¹³ calculates an index $0.5+pn$, and is attributed³ to Hazen.¹⁴ Method B is used in the software programs Mathematica (Wolfram Research Inc., Champaign, IL) (optional method)¹⁵ and MedCalc (MedCalc Software bvba, Mariakerke, Belgium) (only method, except for reference interval estimation).¹⁶

Method C: $p(n-1)+1$

Method C uses $p(n-1)+1$ and is attributed^{3,17} to Gumbel.⁴ This method is implemented in both Microsoft Excel 2003¹⁸ and 2007 (Microsoft Corp., Redmond, WA)¹⁹ as well as in the Calc program, part of the OpenOffice suite by Sun Microsystems (Oracle Corp., Redwood Shores, CA), although this does not seem to be documented. Method C is the default value in R²⁰ and is also among the methods available in Statistica (StatSoft, Inc., Tulsa, OK).¹⁰

Method D: $p(n+1/3)+1/3$

Method D uses $p(n+1/3)+1/3$ and is proposed by Hyndman and Fan³ as the best choice among many other methods including the above. We are not aware of any available software implementations of this method.

References

1. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991
2. CLSI. *Defining, establishing, and verifying reference intervals in the clinical laboratory: approved guideline - third edition*. CLSI Document C28-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
3. Hyndman RJ, Fan Y. Sample quantiles in statistical packages. *Am Stat*. 1996;50:361-365.
4. Gumbel EJ. La Probabilité des Hypothèses. *Comptes Rendus de l'Académie des Sciences* (Paris). 1939; 209; 645-647.
5. Weibull W. The Phenomenon of Rupture in Solids. *Ingeniörs Vetenskaps Akademien Handlingar*. 1939;153:17.
6. SAS Institute Inc. JMP® 8 Statistics and Graphics Guide. Cary, NC: SAS Institute Inc.; 2008.
7. NCSS. User's Guide I, Kaysville, Utah: NCSS, 2007.
8. Minitab Inc. Minitab 15 Methods and Formulas help file. 2007.
9. SAS Institute Inc. Base SAS(R) 9.2 Procedures Guide: Statistical Procedures, Second Edition. Available at: http://support.sas.com/documentation/cdl/en/procstat/63032/HTML/default/procstat_univariate_sect028.htm. Accessed March 15, 2011.
10. StatSoft Inc. STATISTICA 10 Electronic manual: Percentile Calculation Options in STATISTICA. 2011

11. SPSS Inc. SPSS 17 Algorithms help file. 2008.
12. Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research*. 4th ed. Oxford: Blackwell Science; 2002.
13. Lentner C (ed). *Geigy Scientific Tables*, 8th edition, Volume 2. Basel: Ciba-Geigy Limited; 1982.
14. Hazen A. Storage to be Provided in Impounding Reservoirs for Municipal Water Supply (with discussion), *Transactions of the American Society of Civil Engineers*. 1914;77:1539-1669.
15. Weisstein EW. Quantile. From MathWorld - A Wolfram Web Resource. Available at: <http://mathworld.wolfram.com/Quantile.html>. Accessed March 15, 2011.
16. MedCalc Software bvba: MedCalc for Windows. Software manual. Mariakerke: MedCalc Software bvba, 2010. Available at: www.medcalc.org/download/medcalcmanual.pdf. Accessed March 15, 2011.
17. Langford E. Quartiles in elementary statistics. *Journal of Statistics Education*. 2006; 14. Available at: <http://www.amstat.org/publications/jse/v14n3/langford.html>. Accessed March 15, 2011.
18. Microsoft Corp. Office 2003 Percentile function documentation. Available at: <http://office.microsoft.com/en-us/excel/HP052092111033.aspx>. Accessed March 15, 2011.
19. Microsoft Corp. Office 2007 Percentile function documentation. Available at: <http://office.microsoft.com/en-us/excel/HP100625291033.aspx>. Accessed March 15, 2011.
20. Frohne I, Hyndman RJ. R version 2.8.1 documentation: Sample Quantiles. The R Foundation for Statistical Computing, 2008.

Table 1: Accuracy of percentiles calculated in samples with various sizes from normally distributed population data. Results are presented as average percentile, the average of the relative differences (%) and their standard deviation (SD).

Method	Sample size	75 th percentile		95 th percentile	
		Average	% Difference (SD)	Average	% Difference (SD)
A: p(n+1)	Population (n=10 ⁶)	0.67		1.64	
	20	0.71	4.95 (43.99)	1.85	12.20 (31.06)
	60	0.68	1.64 (25.87)	1.71	3.81 (17.07)
	120	0.68	0.74 (18.35)	1.68	1.85 (11.88)
	250	0.68	0.44 (12.70)	1.66	0.92 (8.07)
	500	0.68	0.21 (9.03)	1.65	0.47 (5.78)
	1000	0.67	0.11 (6.39)	1.65	0.24 (4.07)
B: pn+0.5	Population (n=10 ⁶)	0.67		1.64	
	20	0.67	-1.03 (43.43)	1.64	-0.56 (25.48)
	60	0.67	-0.29 (25.72)	1.64	-0.52 (15.81)
	120	0.67	-0.13 (18.31)	1.64	-0.33 (11.37)
	250	0.67	-0.11 (12.70)	1.64	-0.19 (8.06)
	500	0.67	-0.02 (9.05)	1.64	-0.07 (5.72)
	1000	0.67	-0.04 (6.38)	1.64	-0.02 (4.05)
C: p(n-1)+1	Population (n=10 ⁶)	0.67		1.64	
	20	0.63	-6.84 (43.29)	1.43	-12.97 (24.17)
	60	0.66	-2.20 (25.64)	1.57	-4.83 (15.45)
	120	0.67	-1.12 (18.32)	1.60	-2.50 (11.39)
	250	0.67	-0.56 (12.72)	1.63	-1.21 (7.95)
	500	0.67	-0.28 (9.04)	1.64	-0.59 (5.71)
	1000	0.67	-0.10 (6.41)	1.64	-0.28 (4.06)
D: p(n+1/3)+1/3	Population (n=10 ⁶)	0.67		1.64	
	20	0.68	0.84 (43.47)	1.71	3.80 (27.05)
	60	0.68	0.43 (25.70)	1.66	0.92 (16.10)
	120	0.68	0.21 (18.21)	1.65	0.44 (11.52)
	250	0.67	0.07 (12.77)	1.65	0.21 (8.08)
	500	0.67	0.05 (9.01)	1.65	0.12 (5.69)
	1000	0.67	-0.01 (6.40)	1.65	0.07 (4.07)

Table 2. Accuracy of percentiles calculated in samples with various sizes from log-normally distributed population data. Results are presented as average percentile, the average of the relative differences (%) and their standard deviation (SD).

Method	Sample size	5 th percentile		25 th percentile		75 th percentile		95 th percentile	
		Average	% Difference (SD)	Average	% Difference (SD)	Average	% Difference (SD)	Average	% Difference (SD)
A: p(n+1)	Population (n=10 ⁶)	0.02		0.21		4.72		44.16	
	20	0.03	18.92 (141.23)	0.25	19.54 (88.17)	6.60	39.88 (112.00)	180.43	308.61 (1766.28)
	60	0.02	6.04 (69.28)	0.22	6.27 (43.72)	5.26	11.51 (47.29)	64.24	45.49 (122.56)
	120	0.02	2.97 (46.85)	0.22	3.06 (29.85)	4.99	5.66 (30.93)	52.76	19.49 (59.86)
	250	0.02	1.60 (31.58)	0.21	1.50 (20.31)	4.84	2.64 (20.65)	48.09	8.92 (35.00)
	500	0.02	0.82 (22.28)	0.21	0.83 (14.24)	4.77	1.15 (14.25)	46.01	4.20 (23.12)
	1000	0.02	0.42 (15.69)	0.21	0.42 (10.03)	4.75	0.64 (10.04)	45.13	2.21 (16.00)
B: pn+0.5	Population (n=10 ⁶)	0.02		0.21		4.72		44.16	
	20	0.04	83.31 (179.54)	0.28	30.36 (95.54)	6.05	28.25 (101.85)	114.69	159.73 (973.07)
	60	0.03	24.10 (74.87)	0.23	9.24 (44.82)	5.10	8.18 (45.46)	54.44	23.28 (92.47)
	120	0.03	11.74 (48.61)	0.22	4.64 (30.18)	4.91	4.00 (30.34)	48.43	9.69 (52.26)
	250	0.02	5.36 (32.68)	0.22	2.18 (20.44)	4.80	1.79 (20.42)	45.98	4.14 (33.40)
	500	0.02	2.77 (22.48)	0.21	1.12 (14.34)	4.76	0.95 (14.17)	45.16	2.27 (22.52)
	1000	0.02	1.44 (15.76)	0.21	0.58 (10.02)	4.74	0.43 (9.97)	44.64	1.09 (15.70)
C: p(n-1)+1	Population (n=10 ⁶)	0.02		0.21		4.72		44.16	
	20	0.06	147.81 (244.30)	0.30	41.31 (103.48)	5.46	15.75 (90.13)	48.07	8.87 (174.14)
	60	0.03	42.47 (85.06)	0.24	12.60 (46.32)	4.95	4.84 (44.02)	44.30	0.33 (71.00)
	120	0.03	20.77 (52.38)	0.22	6.25 (30.74)	4.83	2.45 (29.93)	44.18	0.05 (47.04)
	250	0.02	9.79 (33.15)	0.22	2.95 (20.57)	4.77	1.04 (20.20)	44.19	0.08 (31.47)
	500	0.02	4.82 (22.90)	0.21	1.46 (14.31)	4.74	0.53 (14.23)	44.24	0.19 (22.16)
	1000	0.02	2.46 (15.92)	0.21	0.78 (10.08)	4.73	0.30 (10.02)	44.21	0.12 (15.60)
D: p(n+1/3)+1/3	Population (n=10 ⁶)	0.02		0.21		4.72		44.16	
	20	0.04	61.98 (163.85)	0.27	26.87 (92.65)	6.26	32.62 (107.23)	138.42	213.47 (1367.82)
	60	0.03	18.11 (72.10)	0.23	8.28 (44.54)	5.16	9.27 (46.15)	57.48	30.18 (100.50)
	120	0.02	8.55 (47.85)	0.22	3.94 (29.90)	4.93	4.40 (30.42)	49.96	13.15 (54.94)
	250	0.02	4.32 (32.35)	0.22	2.08 (20.46)	4.82	2.09 (20.52)	46.69	5.74 (33.82)
	500	0.02	2.15 (22.37)	0.21	1.07 (14.22)	4.77	1.11 (14.25)	45.44	2.91 (22.72)
	1000	0.02	1.07 (15.72)	0.21	0.54 (10.01)	4.74	0.52 (10.01)	44.79	1.43 (15.79)

Figures

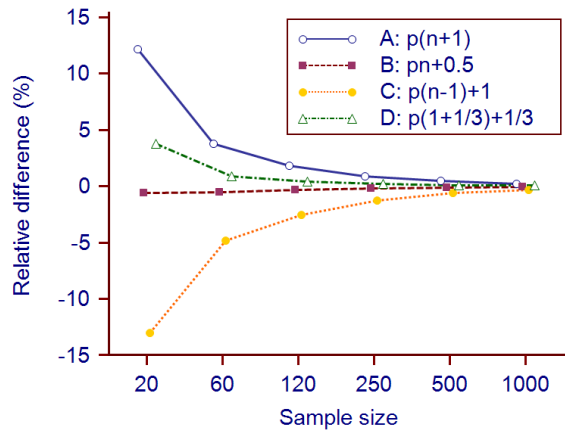


Figure 1: Average of relative differences of estimated 95th percentile with true population percentile in samples from normally distributed population data.

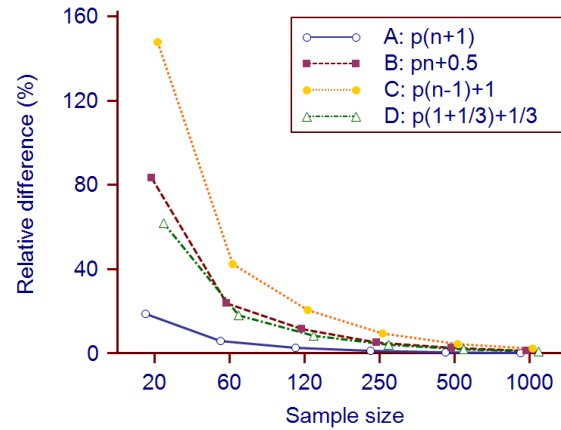


Figure 2: Average of relative differences of estimated 5th percentile with true population percentile in samples from log-normally distributed population data.

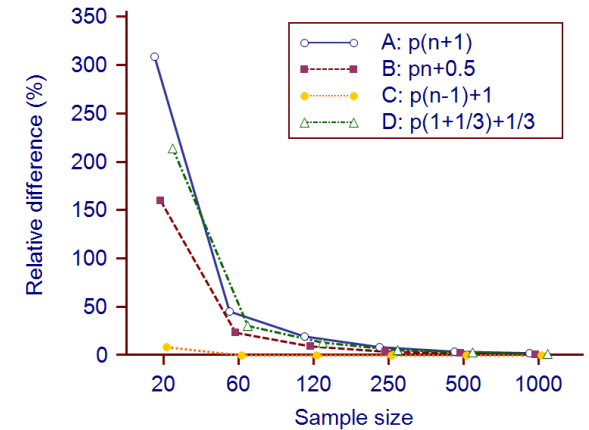


Figure 3: Average of relative differences of estimated 95th percentile with true population percentile in samples from log-normally distributed population data.