From Berger, R.L., 2015. Estimation: Point and Interval. In: James D. Wright (editor-in-chief), International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Vol 8. Oxford: Elsevier. pp. 16–20.
ISBN: 9780080970868

Estimation: Point and Interval

Roger L Berger[†], Arizona State University, Phoenix, AZ, USA

This article is a revision of the previous edition article by G. Casella[†], R.L. Berger, volume 10, pp. 4744–4749, © 2001, Elsevier Ltd.

Abstract

In parametric statistical inference, knowledge about a population parameter yields knowledge about the entire population. Thus, methods of estimating population parameters are cornerstones to statistical analysis. Point estimators provide a single value as an estimate of a parameter. Set estimators provide a set of possible values. Set estimators quantify uncertainty about the parameter through the size of the set and the probability of the set covering the parameter. Various methods for deriving point and set estimators and various methods for evaluating estimators are discussed.

Introduction

When sampling from a population described by a density or mass function $f(x|\theta)$, knowledge of θ yields knowledge of the entire population. Hence, it is natural to seek a method of finding a good estimator of the point θ, that is, a good *point estimator*. However, a point estimator alone is not enough for a complete inference, as a measure of uncertainty is also needed. For that, we use a *set estimator* in which the inference is the statement that $\theta \in C$ where $C \subset \Theta$, Θ is the parameter space (set of all possible values of θ), and $C = C(\boldsymbol{x})$ is a set determined by the value of the data, $\boldsymbol{X} = \boldsymbol{x}$, observed. If θ is real-valued, then we usually prefer the set estimate C to be an interval. Our uncertainty is quantified by the size of the interval and its probability of covering the parameter.

Point Estimators

In many cases there will be an obvious or natural candidate for a point estimator of a particular parameter. For example, the sample mean is a natural candidate for a point estimator of the population mean. However, when we leave a simple case like this, intuition may desert us, so it is useful to have some techniques that will give reasonable candidates for consideration. Those that have stood the test of time include these.

Method of Moments

The method of moments (MOM) is, perhaps, the oldest method of finding point estimators, dating back at least to Karl Pearson in the late 1800s. One of the strengths of MOM estimators is that they are usually simple to use and almost always yield some sort of estimate. In many cases, unfortunately, this method yields estimators that may be improved upon.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample from a population with density or mass function $f(\boldsymbol{x}|\theta_1, \ldots, \theta_k)$. MOM estimators are found by equating the first k sample moments to the corresponding k population moments. That is, we define the sample moments by $m_j = \sum_{i=1}^{n} X_i^j / n$ and the population moments by $\mu_j(\theta_1, \ldots, \theta_k) = E_{\boldsymbol{\theta}} X^j$ for $j = 1, \ldots, k$. Then we set $m_j = \mu_j(\theta_1, \ldots, \theta_k)$ and solve for $\theta_1, \ldots, \theta_k$. This solution is the MOM estimator of $\theta_1, \ldots, \theta_k$.

Maximum Likelihood

For a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ from $f(\boldsymbol{x}|\theta_1, \ldots, \theta_k)$, the likelihood function is defined by

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k)$$

The values of $\theta_1, \ldots, \theta_k$ that maximize this function are those parameter values for which the observed sample $\boldsymbol{x}$ is most likely and are called the maximum likelihood estimators (MLEs). If the likelihood function is differentiable (in θ_i), the MLEs can often be found by solving

$$\frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}|\boldsymbol{x}) = 0, \quad i = 1, \ldots, k$$

where the vector with coordinates $\frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}|\boldsymbol{x})$ is called the *score function* (see Schervish, 1995: Section 2.3).

Example: If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a sample from a Bernoulli (p) population, the likelihood function is

$$L(p|\boldsymbol{x}) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

and differentiating $\log L(p|\boldsymbol{x})$ and setting the result equal to zero gives the MLE $\widehat{p} = \sum_{i=1}^{n} x_i / n = \bar{x}$, the sample mean. This is also the MOM estimator.

If instead we have a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ from a binomial (k, p) population, where p is known and k is unknown, the likelihood function is

$$L(k|\boldsymbol{x}, p) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}$$

and the MLE must be found by the numerical maximization. The MOM gives the closed form solution

$$\widehat{k} = \frac{\bar{x}^2}{\bar{x} - \sum_i (x_i - \bar{x})^2 / n},$$

which can take on negative values. This illustrates a shortcoming of the MOM, one not shared by the MLE. Another, perhaps more serious shortcoming of the MOM estimator is that it may not be based on a sufficient statistic (*see* Statistical Sufficiency), which means it could be inefficient in not using all of the available information in a sample. In contrast, both MLEs and Bayes estimators (*see* Bayesian Statistics) are based on sufficient statistics.

[†] Deceased.

International Encyclopedia of the Social & Behavioral Sciences, Second Edition, 2015, 16–20

Bayes

In the Bayesian paradigm a random sample $X_1, \ldots, X_n$ is drawn from a population indexed by θ, where uncertainty about θ can be described by a probability distribution (called the *prior distribution*). After the sample is taken, the prior distribution is updated with this sample information. The updated prior is called the *posterior distribution*.

If we denote the prior distribution by $\pi(\theta)$ and the sampling distribution by $f(\boldsymbol{x}|\theta)$, then the posterior distribution, the conditional distribution of θ given the sample, $\boldsymbol{x}$, is

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)},$$

where $m(x) = \int f(x|\theta)\pi(\theta)\mathrm{d}\theta$ is the marginal distribution of $\boldsymbol{x}$.

Example: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample from a Bernoulli (p) population. Then $Y = \sum_i X_i$ is binomial (n, p). If p has a beta (α, β) prior distribution, that is

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1},$$

the posterior distribution of p given y is

$$\begin{aligned}\pi(p|y) &= \frac{f(y|p)\pi(p)}{m(y)} \\ &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}p^{y+\alpha-1}(1-p)^{n-y+\beta-1},\end{aligned}$$

which is a beta distribution with parameters $y+\alpha$ and $n-y+\beta$. The posterior mean, a common Bayes estimator of p, is

$$\widehat{p}_B = \frac{y+\alpha}{\alpha+\beta+n}.$$

Evaluating Point Estimators

There are many methods of deriving point estimators (robust methods, least squares, estimating equations, invariance), but the three in Section Point Estimators are among the most popular. No matter what method is used to derive a point estimator, it is important to evaluate the estimator using some performance criterion.

One way of evaluating the performance of a point estimator W of a real-valued parameter θ is through its *mean squared error* (MSE), defined by

$$\mathrm{MSE}_\theta W = E_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (E_\theta W - \theta)^2.$$

Defining the bias of a point estimator by $\mathrm{Bias}_\theta W = \mathrm{E}_\theta W - \theta$ yields $\mathrm{MSE}_\theta W = \mathrm{Var}_\theta W + (\mathrm{Bias}_\theta W)^2$. An estimator whose bias is identically (in θ) equal to zero is called *unbiased*.

For an unbiased estimator we have $\mathrm{MSE}_\theta W = \mathrm{Var}_\theta W$; if an estimator is unbiased, its MSE is equal to its variance. If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a sample from a population with mean θ and variance σ^2, the sample mean is an unbiased estimator of θ $(E_\theta \overline{X} = \theta)$, and

$$\mathrm{MSE}_\theta \overline{X} = \mathrm{Var}_\theta \overline{X} = \frac{\sigma^2}{n}.$$

Controlling bias does not guarantee that MSE is minimized. In particular, it is sometimes the case that a trade-off occurs between variance and bias. For example, in sampling from a normal population with variance σ^2, the usual unbiased estimator of σ^2, $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$, has $\mathrm{MSE}_\sigma S^2 = 2\sigma^4/(n-1)$. An alternative estimator of σ^2 is the MLE $\widehat{\sigma}^2 = \sum_{i=1}^n (X_i - \overline{X})^2/n = (n-1)S^2/n$. This is a biased estimator of σ^2 with

$$\mathrm{MSE}_\sigma \widehat{\sigma}^2 = \frac{2n-1}{n^2}\sigma^4 < \frac{2}{n-1}\sigma^4 = \mathrm{MSE}_\sigma S^2,$$

showing that $\widehat{\sigma}^2$ has smaller MSE than S^2. Thus, by trading off variance for bias, the MSE is improved.

Measuring performance by the squared difference between the estimator and a parameter is a special case of a function called a *loss function*. The study of the performance and optimality of estimators evaluated through loss functions is a branch of decision theory. In addition to MSE, based on squared error loss, another popular loss function is absolute error loss, $L(\theta, W) = |W-\theta|$. Both of these loss functions increase as the distance between θ and W increases, with minimum value $L(\theta, \theta) = 0$. That is, the loss is minimum if the estimator is correct.

In a decision theoretic analysis an estimator is evaluated through its *risk function*. For an estimator W of θ, the risk function is $R(\theta, W) = \mathrm{E}_\theta L(\theta, W)$; the risk function is the average loss. If the loss is squared error, the risk function is the MSE.

Using squared error loss, the risk function (MSE) of the binomial Bayes estimator of p is

$$\begin{aligned}\mathrm{MSE}_p \widehat{p}_B &= \mathrm{Var}_p \widehat{p}_B + \left(\mathrm{Bias}_p \widehat{p}_B\right)^2 \\ &= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2.\end{aligned}$$

In the absence of good prior information about p we might choose α and β to make the risk function of $\widehat{p}_B$ constant (called an *equalizer rule*). The solution is $\alpha = \beta = \sqrt{n/4}$, yielding

$$\mathrm{MSE}_p \widehat{p}_B = \frac{n}{4(n+\sqrt{n})^2}.$$

We can also use a Bayesian approach to the problem of loss function optimality, where we would use the prior distribution to compute an average risk, $\int_\Theta R(\theta, W)\pi(\theta)\mathrm{d}\theta$, known as the *Bayes risk*. We then find the estimator that yields the smallest value of the Bayes risk. Such an estimator is called the Bayes rule with respect to a prior π.

To find the Bayes decision rule for a given prior π we write the Bayes risk as

$$\int_\Theta R(\theta, W)\pi(\theta)\mathrm{d}\theta = \int_x \left[\int_\Theta L(\theta, W(x))\pi(\theta|x)\mathrm{d}\theta\right] m(x)\mathrm{d}x,$$

where the quantity in square brackets is the expected value of the loss function with respect to the posterior distribution, called the *posterior expected loss*. It is a function only of $\boldsymbol{x}$ and not a function of θ. Thus, for each $\boldsymbol{x}$, if we choose the estimate $W(\boldsymbol{x})$ to minimize the posterior expected loss, we will minimize the Bayes risk, and $W(\boldsymbol{x})$ is the Bayes rule.

For squared error loss, the posterior expected loss is minimized by the mean of the posterior distribution. For absolute error loss, the posterior expected loss is minimized by the median of the posterior distribution. If $\boldsymbol{X} = (X_1, \ldots, X_n)$

is a sample from a normal population, with mean θ and variance σ^2 ($n(\theta, \sigma^2)$), and the prior on θ is $n(\mu, \tau^2)$, the posterior mean is

$$E(\theta|x) = \frac{\tau^2}{\tau^2 + (\sigma^2/n)}\bar{x} + \frac{\sigma^2/n}{\tau^2 + (\sigma^2/n)}\mu.$$

Because the posterior distribution is normal, it is symmetric and the posterior mean is the Bayes estimator for both squared error and absolute error loss. The posterior mean in our binomial/beta example, $\hat{p}_B = (y + \alpha)/(\alpha + \beta + n)$, is the Bayes estimator for squared error loss.

We can loosely group evaluation criteria into large sample (asymptotic) methods and small sample methods. Our calculations using MSE and risk functions illustrate small sample methods. In large samples, MLEs typically perform very well, being asymptotically normal and efficient, that is, attaining the smallest possible variance. Other types of estimators that are derived in a similar manner (for example, M-estimators – *see* Robustness in Statistics) also share good asymptotic properties. For a detailed discussion see Casella and Berger (2001), Lehmann (1999), Stuart et al. (1999), or Lehmann and Casella (1998: Chapter 6).

Interval Estimation

Reporting a point estimator of a parameter θ only provides part of the story. The story becomes more complete if an assessment of the error of estimation is also reported. Informally, this can be accomplished by giving an estimated standard error of the estimator, and, more formally, this becomes the reporting of an interval estimate. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, an *interval estimate* of a parameter θ is a pair of functions, $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$, for which the inference $\theta \in (L(x), U(x))$ is made. The *coverage probability* of the random interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ is the probability that $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ covers the true parameter, θ, and is denoted by $P_\theta(L(\boldsymbol{X}) < \theta < U(\boldsymbol{X}))$.

By definition the coverage probability depends on the unknown θ, so it cannot be reported. What is typically reported is the *confidence coefficient*, the infimum of the coverage probabilities, $\inf_\theta P_\theta(L(X) < \theta < U(X))$.

If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a sample from a population with mean θ and variance σ^2, a common interval estimator for θ is

$$\bar{x} - 2\frac{s}{\sqrt{n}} < \theta < \bar{x} + 2\frac{s}{\sqrt{n}},$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation. The validity of this interval can be justified from the Central Limit Theorem, because

$$\frac{\overline{X} - \theta}{S/\sqrt{n}} \to n(0, 1),$$

the standard normal distribution. We then see that the coverage probability (and confidence coefficient) of this interval is approximately 95%.

The above interval is a large sample interval, because its justification is based on an asymptotic property. There are many methods for constructing interval estimators that are valid in small samples, including these.

Inverting a Test Statistic

There is a correspondence between acceptance regions of hypothesis tests (*see* Hypothesis Testing in Statistics) and confidence sets, summarized in the following theorem.

Theorem 1: For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of H_0: $\theta = \theta_0$. For each sample point $\boldsymbol{x}$, define a set $C(\boldsymbol{x})$ in the parameter space by

$$C(x) = \{\theta_0 : x \in A(\theta_0)\}.$$

Then the random set $C(\boldsymbol{X})$ is a confidence set with confidence coefficient $1 - \alpha$. Conversely, let $C(\boldsymbol{X})$ be a $1 - \alpha$ confidence set. For each $\theta_0 \in \Theta$, define

$$A(\theta_0) = \{x : \theta_0 \in C(x)\}.$$

Then $A(\theta_0)$ is the acceptance region of a level α test of H_0: $\theta = \theta_0$.

Example: If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a sample from an $n(\theta, \sigma^2)$ population, with σ^2 known, the level α test of H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$ will accept the null hypothesis, if $|\bar{x} - \theta_0|/(\sigma/\sqrt{n}) < z_{\alpha/2}$. This inequality can be equivalently written as

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \theta_0 < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

This interval of θ values, for which the null hypothesis is accepted at level α, is a $1 - \alpha$ confidence interval for θ.

Pivotal Intervals

An elegant method of constructing set estimators is the use of pivotal quantities (Barnard, 1949). A random variable, $Q(\boldsymbol{X}, \theta) = Q(X_1, \ldots, X_n, \theta)$, is a pivotal quantity (or pivot), if the distribution of $Q(\boldsymbol{X}, \theta)$ is independent of all parameters. If we find a set C such that $P(Q(\boldsymbol{X}, \theta) \in C) = 1 - \alpha$, then the set $C(\boldsymbol{X}) = \{\theta: Q(\boldsymbol{X}, \theta) \in C\}$ has coverage probability $1 - \alpha$.

In *location* and *scale* cases, once we calculate the sample mean $\overline{X}$ and the sample standard deviation S, examples of pivots are:

Form of pdf	*Type of pdf*	*Pivot*
$f(x - \theta)$	Location	$\overline{X} - \theta$
$f(x/\sigma)/\sigma$	Scale	S/σ
$\frac{f((x-\theta)/\sigma)}{\sigma}$	Location–scale	$\frac{\overline{X} - \theta}{S}$

In general, differences are pivotal for location problems, while ratios (or products) are pivotal for scale problems. *see* Fiducial and Structural Statistical Inference.

Example: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample from an exponential (λ) population, and let $T = \sum_{i=1}^n X_i$. T has a gamma (n, λ) distribution. This gamma pdf is

$$f(t|\lambda) = \Gamma(n)\left(\frac{t}{\lambda}\right)^{n-1} e^{-t/\lambda}\frac{1}{\lambda}.$$

In this gamma pdf t and λ appear together as t/λ, and, in fact, the gamma (n, λ) pdf is a scale family in λ. If $Q(T, \lambda) = 2T/\lambda$, then $Q(T, \lambda)$ has a gamma $(n, 2)$ (also called χ^2_{2n}) distribution,

which does not depend on λ. Hence, the quantity $Q(T, \lambda)$ is a pivot, and a $1-\alpha$ pivotal interval is

$$\frac{2T}{\chi^2_{2n,\alpha/2}} < \lambda < \frac{2T}{\chi^2_{2n,1-\alpha/2}},$$

where $P(\chi^2_{2n} > \chi^2_{2n,\alpha}) = \alpha$.

Bayesian Intervals

If $\pi(\theta|\boldsymbol{x})$ is the posterior distribution of θ given $\boldsymbol{X} = \boldsymbol{x}$, then for any set $A \subset \Theta$ the posterior probability of A is

$$P(\theta \in A|x) = \int_A \pi(\theta|x)\mathrm{d}\theta.$$

If $A = A(\boldsymbol{x})$ is chosen so that this posterior probability is $1-\alpha$, then A is called a $1-\alpha$ *credible set* for θ.

The interpretation of the Bayes interval estimator is different from the classical intervals. In the classical approach to assert 95% coverage is to assert that in 95% of repeated experiments the realized intervals will cover the true parameter. In the Bayesian approach a 95% coverage means that the probability is 95% that the parameter is in the realized interval. In the classical approach the randomness comes from the repetition of experiments, while in the Bayesian approach the randomness comes from uncertainty about the value of the parameter (summarized in the prior and posterior distributions).

Example: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample from an Poisson (λ) population, and assume that λ has a gamma (a, b) prior, where a is an integer. The posterior distribution of λ is

$$\pi(\lambda|x) \text{ is gamma}\left(a + \sum_{i=1}^{n} x_i, \frac{b}{nb+1}\right).$$

Thus the posterior distribution of $[2(nb+1)/b]\lambda$ is $\chi^2_{2(a+\sum_i x_i)}$, and a $1-\alpha$ Bayes credible interval for λ is

$$\frac{\chi^2_{2(a+\sum_i x_i),1-\alpha/2}}{2(nb+1)/b} < \lambda < \frac{\chi^2_{2(a+\sum_i x_i),\alpha/2}}{2(nb+1)/b}.$$

We can form a Bayes credible set by using the *highest posterior density* (HPD) region of the parameter space, also. That is, choose c so that the set $C(x) = \{\lambda: \pi(\lambda|x) > c\}$ satisfies

$$1-\alpha = P(\lambda \in C(x)|x).$$

Such a construction is optimal in the sense of giving the shortest interval for a given $1-\alpha$, although if the posterior is multimodal the set may not be an interval.

Other Intervals

We have presented two-sided parametric confidence intervals that are constructed to cover a parameter. Other types of intervals include one-sided intervals, distribution-free intervals, prediction intervals, and tolerance intervals.

One-sided intervals are those in which only one endpoint is estimated, such as $\theta \in (L(\boldsymbol{X}), \infty)$. Distribution-free intervals are intervals whose probability guarantee holds with few (or no) assumptions on the underlying population. The other two interval definitions, together with the usual confidence interval, provide us with a hierarchy of inferences, each more stringent than the previous.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample from a population with cdf $F(\boldsymbol{x}|\theta)$. If $C(\boldsymbol{x}) = (L(\boldsymbol{x}), U(\boldsymbol{x}))$ is an interval, for a specified value $1-\alpha$ it is a:

1. *confidence* interval if, for all θ, $P_\theta(L(X) < \theta < U(X)) \geq 1-\alpha$;
2. *prediction* interval if, for all θ, $P_\theta(L(X) < X_{n+1} < U(X)) \geq 1-\alpha$, where $\boldsymbol{X}_{n+1}$ is an independent, new observation;
3. *tolerance* interval if, for all θ and for a specified value p, $P_\theta(F(U(X)|\theta) - F(L(X)|\theta) \geq p) \geq 1-\alpha$.

So a confidence interval covers a parameter, a prediction interval covers a new observation, and a tolerance interval covers a proportion of the population. Each gives a different inference with the appropriate one being dictated by the problem at hand. Vardeman (1992) discusses the importance of each of these intervals.

Conclusions

Point estimation is one of the cornerstones of statistical analysis and the basic element on which many inferences are based. Inferences using point estimators gain statistical validity when they are accompanied by an interval estimate, providing an assessment of the uncertainty. We have mainly discussed parametric point and interval estimation, where we assume that the underlying model is correct. Such an assumption can be questioned, and considerations of nonparametric or robust alternatives can address this (*see* Robustness). For more on these subjects see, for example, Boos and Stefanski (2013), Hettmansperger and McKean (1998), or Staudte and Sheather (1990). Full treatments of parametric point and interval estimation can be found in Casella and Berger (2001), Stuart et al. (1999), or Schervish (1995).

Acknowledgment

This article is a revised version of an article on the same topic in the first edition of the *International Encyclopedia of the Social and Behavioral Sciences*, which was coauthored with the late George Casella, University of Florida.

See also: Bayesian Statistics; Likelihood in Statistics; Quantile Regression; Statistical Identification and Estimability; Statistical Sufficiency.

Bibliography

Barnard, G.A., 1949. Statistical inference (with discussion). Journal of the Royal Statistical Society, Series B 11, 115–139.

Boos, D.D., Stefanski, L.A., 2013. Essential Statistical Inference, Theory and Methods. Springer, New York.

Casella, G., Berger, R.L., 2001. Statistical Inference, second ed. Wordsworth/Brooks Cole, Pacific Grove, CA.

Hettmansperger, T.P., McKean, J.W., 1998. Robust Nonparametric Statistical Methods. Wiley, New York.

Lehmann, E.L., 1999. Elements of Large-Sample Theory. Springer-Verlag, New York.
Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, second ed. Springer-Verlag, New York.
Schervish, M.J., 1995. Theory of Statistics. Springer-Verlag, New York.
Staudte, R.G., Sheather, S.J., 1990. Robust Estimation and Testing. John Wiley, New York.
Stuart, A., Ord, J.K., Arnold, S., 1999. Kendall's Advanced Theory of Statistics, Vol. 2A, Classical Inference and the Linear Model, sixth ed. Wiley, London.
Vardeman, S.B., 1992. What about the other intervals? The American Statistician 46, 193–197.