Thoughts on Intervals Case

Gary W. Oehlert, lightly rev. S. Weisberg

School of Statistics
University of Minnesota

February 22, 2008

Introduction

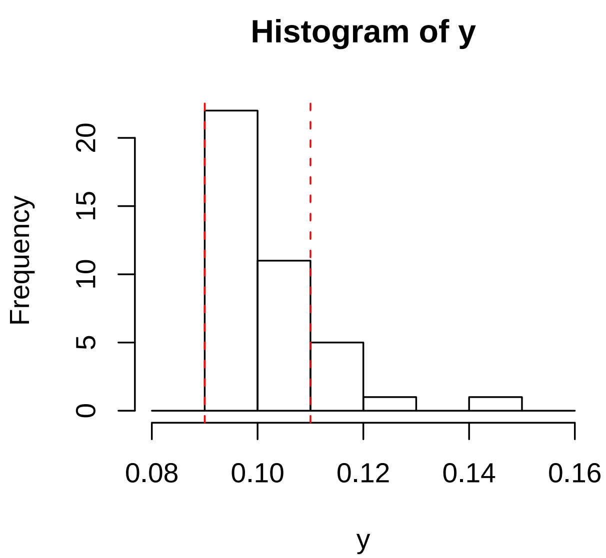
Sample of 40 fibers has mean of .102 mm and SD of .01 mm.

Fibers should have diameter between .09 mm and .11 mm.

Employee does irrelevant 99% CI for the mean.

What can we say about fibers and tolerances?

The data



Good Enough?

Version 1 Estimate fraction of acceptable fibers.

p is fraction of fibers in tolerance.

Take a sample of size N, let k be in tolerance.

$$\hat{p} = \frac{n}{N}$$

Use binomial inference (or asymptotic normal).

Good Enough

Version 2 Acceptance Sampling.

Several versions. Simplest says:

1. Take random sample of size N.
2. Find k, the number of defectives.
3. Reject entire lot if $k > C$.

C depends on N and the fraction of defectives p we're willing to tolerate.

Operating Characteristic

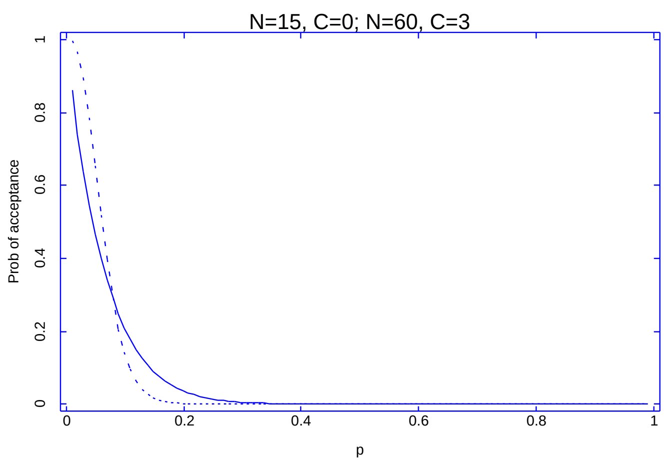
Assume a fraction p of defectives and a sample of size N.

Then k is Bin(N, p).

Look at $P(k \leq C)$ as a function of p.

For example, consider $N = 15, C = 0$ and $N = 60, C = 3$. N controls steepness, C controls location.

Operating Characteristic, $N = 15, C = 0, N = 60, C = 3$



Choice of N and C

Let α be a desired confidence, and let β be a maximum allowable defective rate.

For every (α, β) we can find (N, C) so that if $k \leq C$ we can say:

> *We are at least $(1 - \alpha)$ confident that the acceptable fraction is at least $(1 - \beta)$.*

Table

N	$100(1-\beta)$	Confidence $(1-\alpha)$ .50	.60	.70	.80	.90	.95	.99
$C=0$	75	3	4	5	6	9	11	17
	80	4	5	6	8	11	14	21
	85	5	6	8	10	15	19	29
	90	7	9	12	16	22	29	44
	95	14	18	24	32	45	59	90
	99	69	92	120	161	230	299	459
$C=1$	75	7	8	10	11	15	18	24
	80	9	10	12	14	18	22	31
	85	11	13	16	19	25	30	42
	90	17	20	24	29	38	46	64
	95	34	40	49	59	77	93	130

Intervals

$X_1, X_2, \ldots, X_n$ are *iid* from F (assume it has a density for simplicity). Let $\mathbf{X}$ denote the vector of all data.
Confidence interval

$$P[L(\mathbf{X}) < \mu < U(\mathbf{X})] = 1-\alpha$$

Prediction interval

$$P[L(\mathbf{X}) < X_{n+1} < U(\mathbf{X})] = 1-\alpha$$

Tolerance interval

$$P[F(U(\mathbf{X})) - F(L(\mathbf{X})) > 1-\beta] \geq 1-\alpha$$

Tolerance Interval

A tolerance interval captures at least $1-\beta$ of the population with probability at least $1-\alpha$.

For comparable coverage, CI generally shorter than PI, which is generally shorter than TI.

Two types: parametric (usually based on normal) and nonparametric (based on order statistics).

NIST web page has a nice introduction.

Normal tolerance interval

Sort of looks like CI or PI

$$\bar{x} \pm K_{\alpha,\beta,n} s$$

Howe (1969, JASA) gives the approximation

$$K_{\alpha,\beta,n} = \sqrt{\frac{n-1}{\chi^2_{\alpha,n-1}}}\sqrt{(1+\frac{1}{n})}\,\Phi^{-1}(1-\beta/2)$$

where Φ^{-1} is the inverse standard normal cumulative and $\chi^2_{\alpha,n-1}$ is the lower α percent point of chisquare with $n-1$ df.

(The rightmost two factors give a prediction interval.)

Distribution Free Tolerance Intervals

Normal-based intervals will be **very poor** when data do not follow normal distribution, and more non-normal data doesn't help. The CLT is not working for you.

We can compute tolerance intervals based on order statistics that only assume that the distribution is continuous.

Distribution-free interval is longer than parametric normal interval when data are normal.

Order Statistics

Sort data into order statistics:

$$x_{(1)} < x_{(2)} < \cdots < x_{(n)}$$

Our interval will have the form

$$(x_{(r)}, x_{(n+1-r)})$$

E.g., $r = 1$ gives us the range of the data.

Trick is to find n, r, α, β combinations that work.

Will need large n to get small α or β.

Betas to the rescue

David (1981 *Order Statistics*) explains all.
If $U(\mathbf{X}) = x_{(n+1-r)}$ and $L(\mathbf{X}) = x_{(r)}$, the

$$\begin{aligned} P[F(U(\mathbf{X})) - F(L(\mathbf{X})) > 1-\beta] &= P[F(x_{(n+1-r)}) - F(x_{(r)}) > 1-\beta] \\ &= P[u_{(n+1-r)} - u_{(r)} > 1-\beta] \\ &= P[u_{(n+1-2r)} > 1-\beta] \end{aligned}$$

This is just the probability that a uniform order statistic is in a certain range, which can be computed as an incomplete beta integral.

Table of r for Tolerance Intervals

	$\alpha = .25$				$\alpha = .1$				$\alpha = .05$			
	β				β				β			
m	.2	.1	.05	.025	.2	.1	.05	.025	.2	.1	.05	.025
15	1											
20	1				1							
25	2				1				1			
30	2	1			1				1			
35	2	1			2				1			
40	3	1			2	1			2			
45	3	1			3	1			2			
50	4	1			3	1			3	1		
55	4	2	1		3	1			3	1		
60	5	2	1		4	1			3	1		
65	5	2	1		4	2			4	1		
70	6	2	1		5	2			4	1		
75	6	3	1		5	2			4	1		
80	7	3	1		5	2	1		5	2		
85	7	3	1		6	2	1		5	2		
90	7	3	1		6	2	1		6	2		
95	8	3	1		7	3	1		6	2	1	
100	8	4	1		7	3	1		7	2	1	
110	9	4	2	1	8	3	1		7	3	1	
120	10	5	2	1	9	4	1		8	3	1	