Statistical Aspects of a Reasonable Test Plan

Testing and Certification Workshop
December 10, 2009

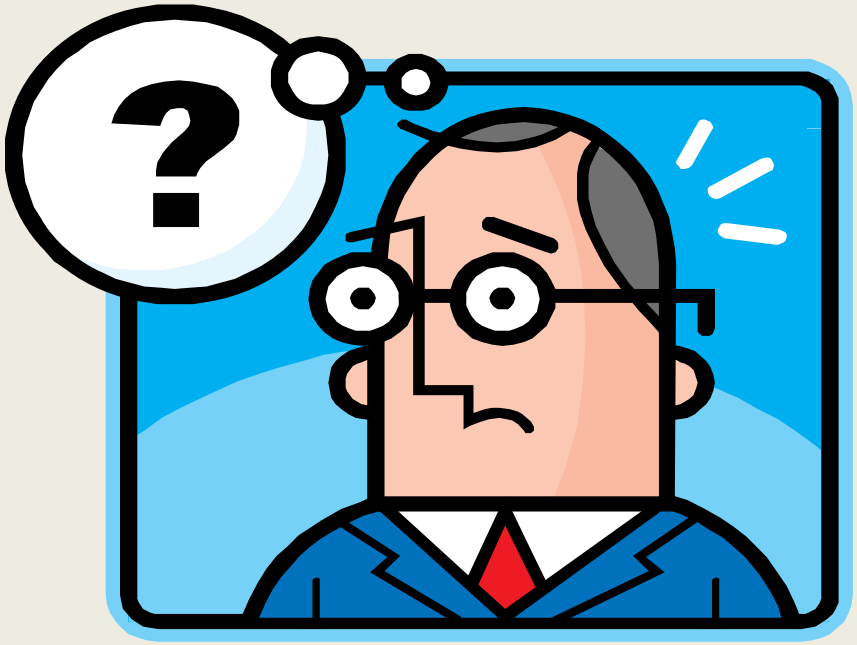
Michael A. Greene, Ph. D.
Mathematical Statistician
Division of Hazard Analysis, U. S. CPSC

This presentation was prepared by the CPSC staff, has not been reviewed or approved by, and may not necessarily reflect the views of the Commission. Because this work was prepared in the authors' official capacity, it is in the public domain and may be freely copied.

Outline of Talk

1. Overview
2. General Conformity Certification, Assurance
3. Prediction Interval/Bound Defined
4. Computing a Prediction Interval/Bound and Sample Sizes.
    1. Quantitative (Normal Distribution Theory)
    2. Qualitative (Pass/Fail)

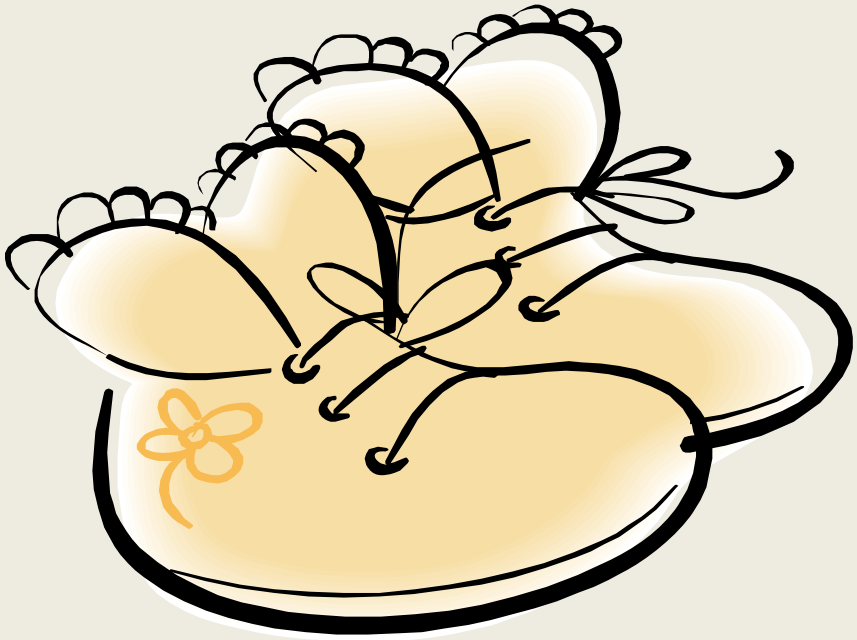
1. Overview:
The Most Often Asked Question



How Many Items
Should I Test?

3?
5?
12?
18?
20?
30?
50?

One Size Doesn’t Fit Anyone



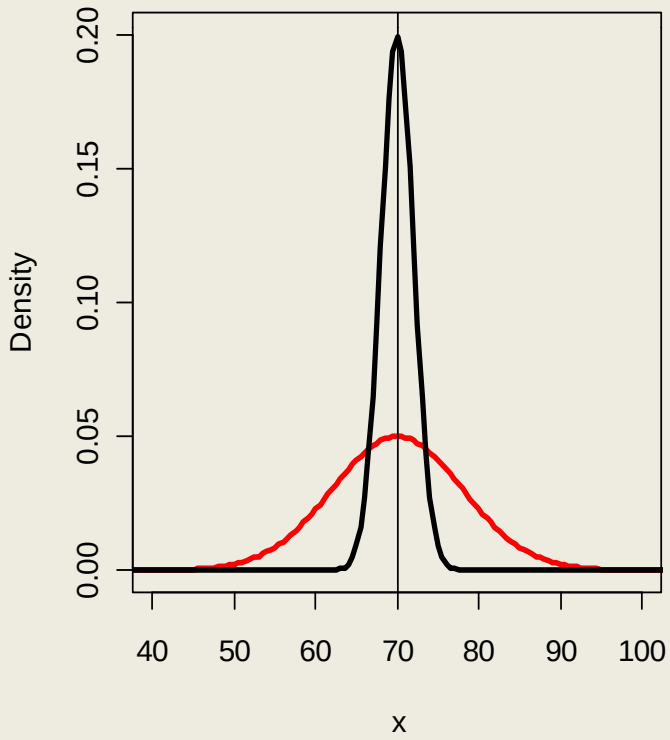
- There is no fixed number that is appropriate for all products.
- It depends.

A Principle

For Choosing Sample Sizes

- *The larger the sample, the lower the variability of an estimate.*

Comparing
Small and Large Samples



How Much Precision?

Estimating Something?

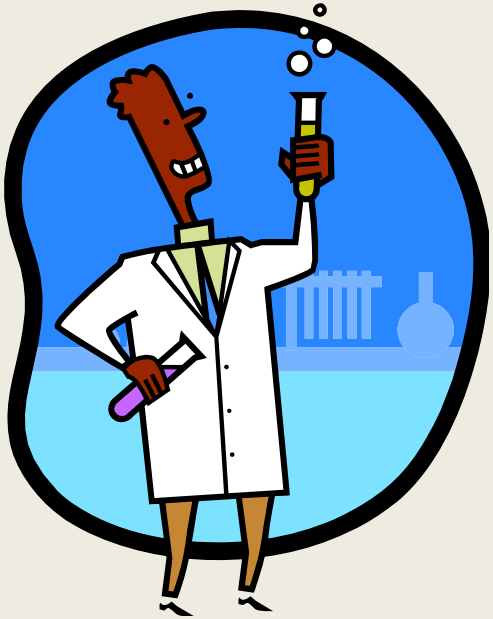
- We aren't estimating a mean, or something like that, are we?
- No. Just that the production run complies with "… all rules, bans, standards or regulations applicable to the product…" *with some probability, not 100% certainty*

Larger Samples, Lower Variability

- The same principle works here.
- Everything else held constant, the more items tested in the sample, the higher the probability of complying…

Approach

- Explore the effect of different sample sizes to determine what works when.



2. General Conformity Certification

- Sec. 14 General Conformity Certification
- Special requirements for children's products and testing of *random samples* for continued compliance.

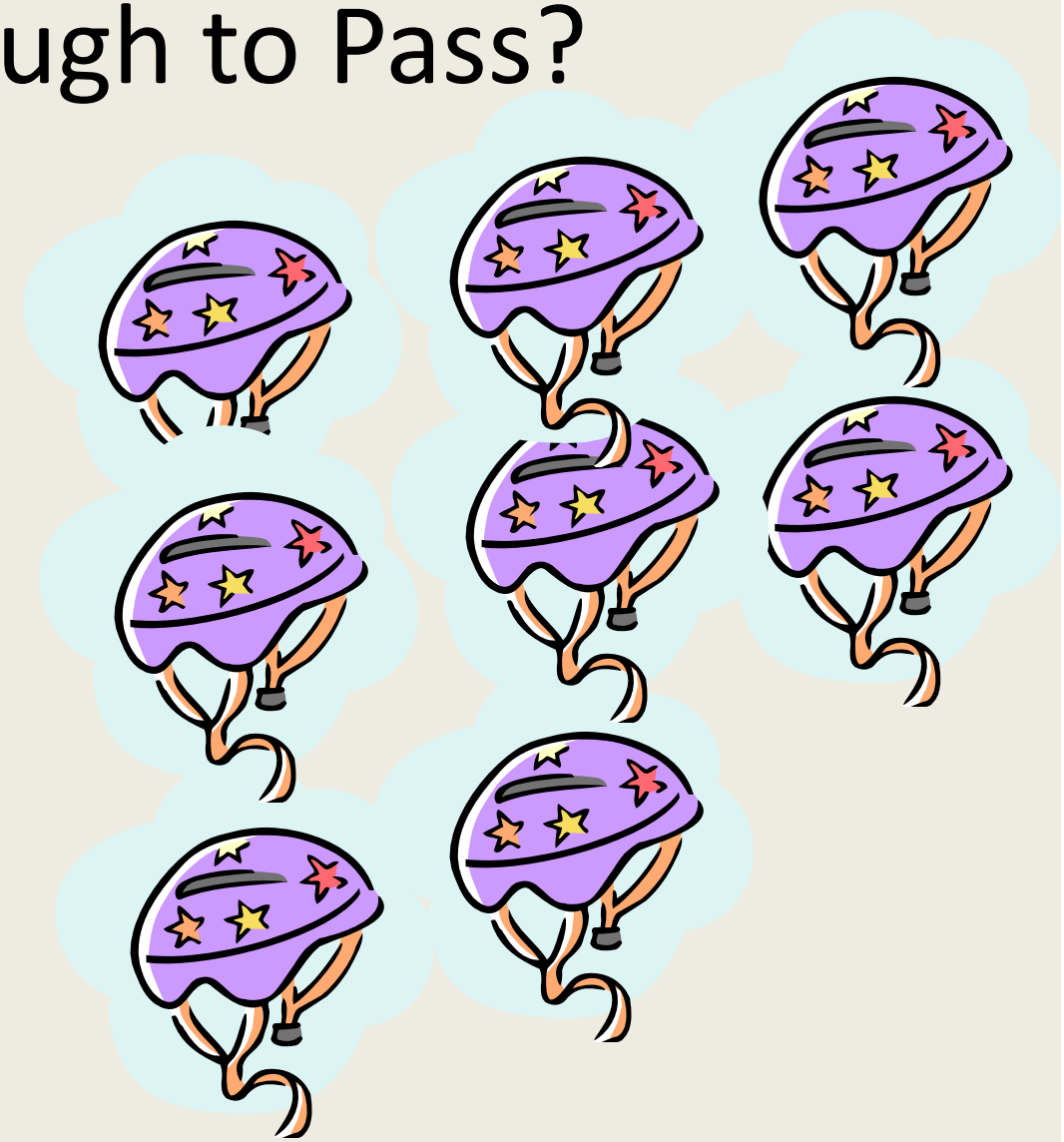
Is It Enough to Pass?

--I sent 8 helmets off for testing. They all passed.

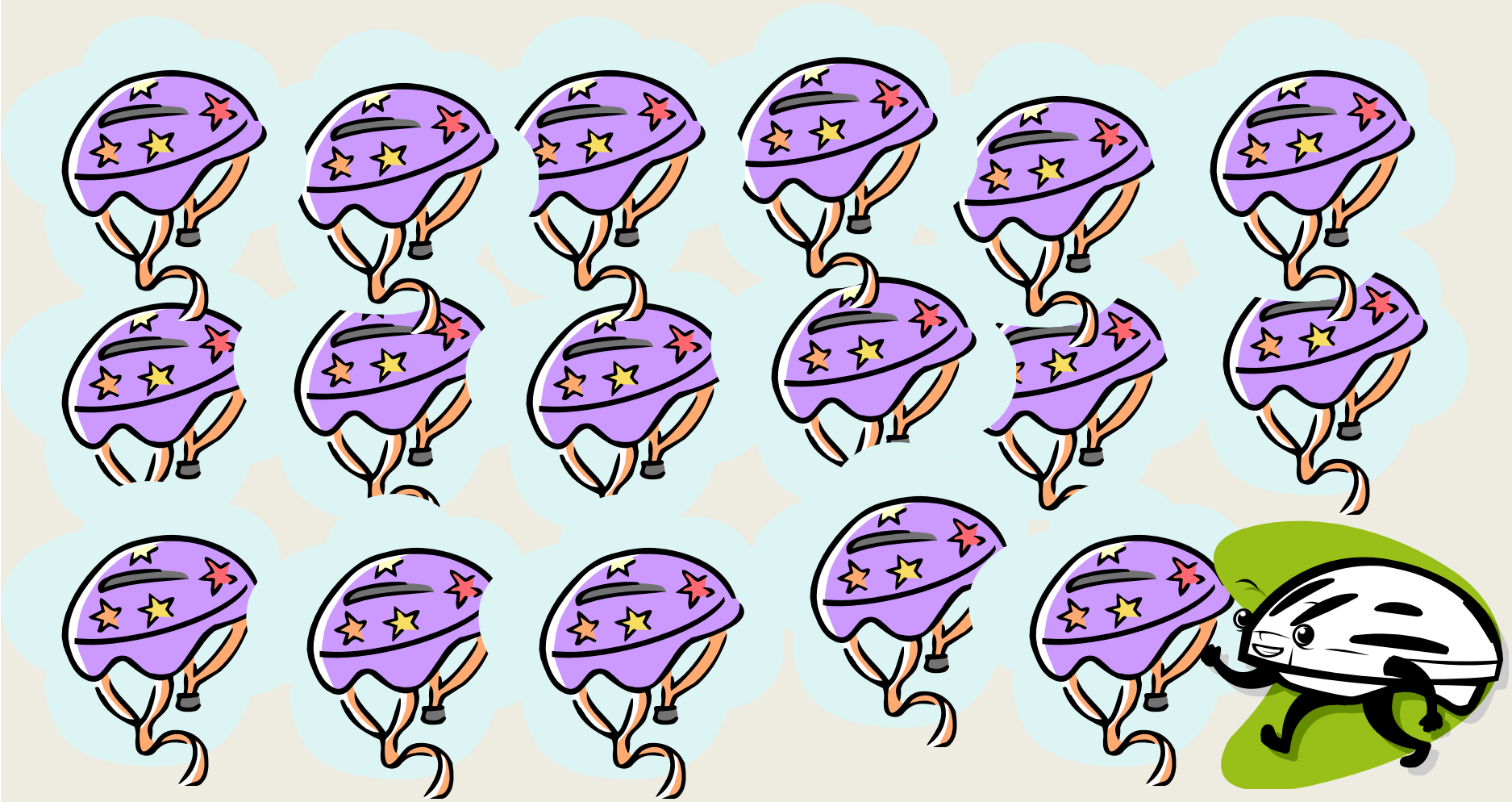
That's all I have to do. Right?

--Maybe not.

Unless you only made 8 helmets!



Production



Notes

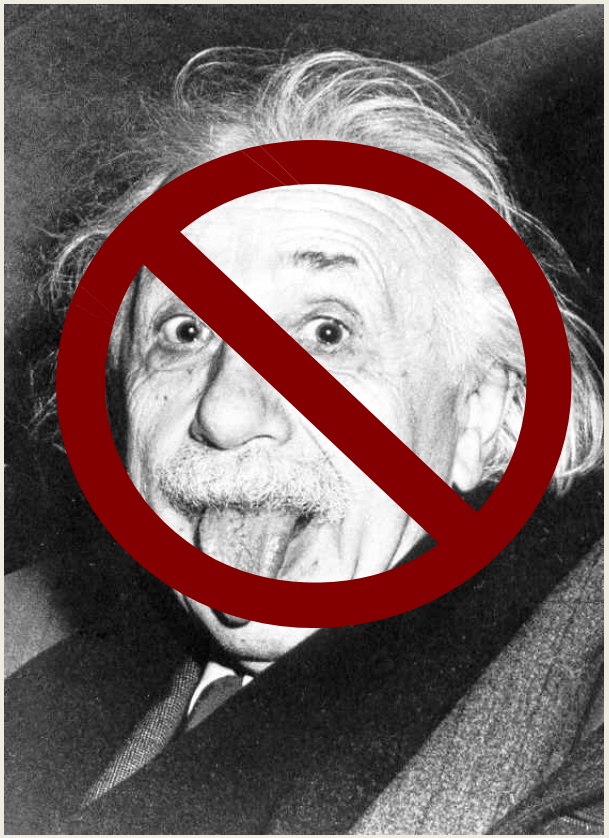
Assuming that a reasonable testing program is in place, in order to show that a product complies with all rules…

If any of the items in the test sample fails, it has not been shown that the product (production lot) complies. Corrective action may be needed.

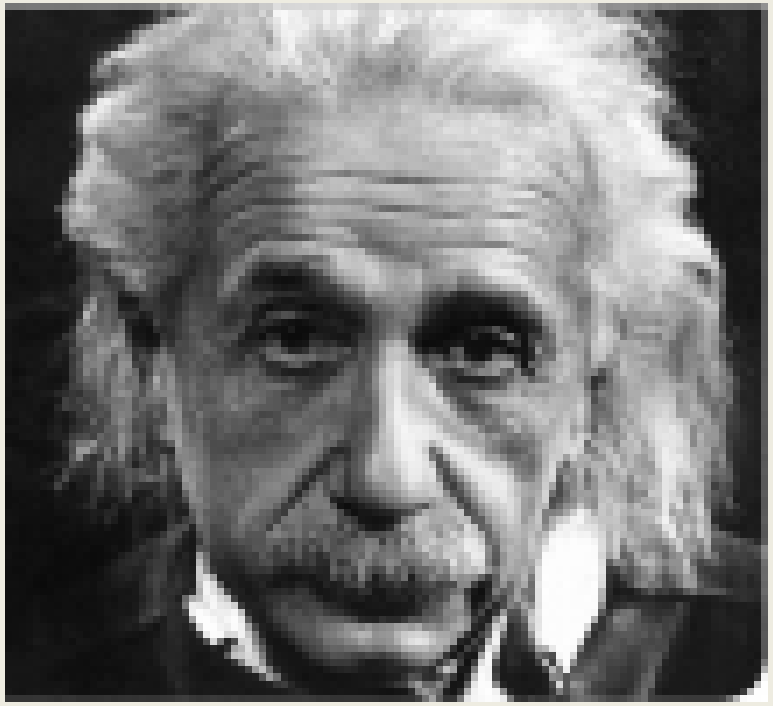
If none of the items in the test sample fail, the *sample* complies, but more may be required to demonstrate that the *production run* complies.

3. Ways to Provide Assurance

Quote Authority



Do Statistics



Assurance: A Prediction Bound

- The setup
  - Items are produced in lots of known size.
  - Some items (a sample), are *randomly selected* from the lot for testing.
  - Test results have been received for each of the items.
  - From the test results we can make a statement about the probability that the entire production lot meets a certain requirement (e.g., 300 ppm). This is the prediction bound approach.

Assurance: A Prediction Bound Cont'd

- A prediction bound/interval is some *number* where we can say with a certain *probability* that *all* the items in that lot have measurements that are at or below that *number*.

- We usually know the *number*. The number is in the regulations. We are interested in the *probability*.

- Example : There is a 99.9% chance that all items in a production lot of 1,000 items have less than 300 ppm of lead.

Assurance: Interpretation

- The items in the production lot all have lead content below 300 ppm or there is at least one item with more than 300 ppm. The values do not matter. The lot is either good or bad!

- In the long run, using this method, 999 lots in 1,000 should be good. One lot will be bad. There is no way to know which one is bad.

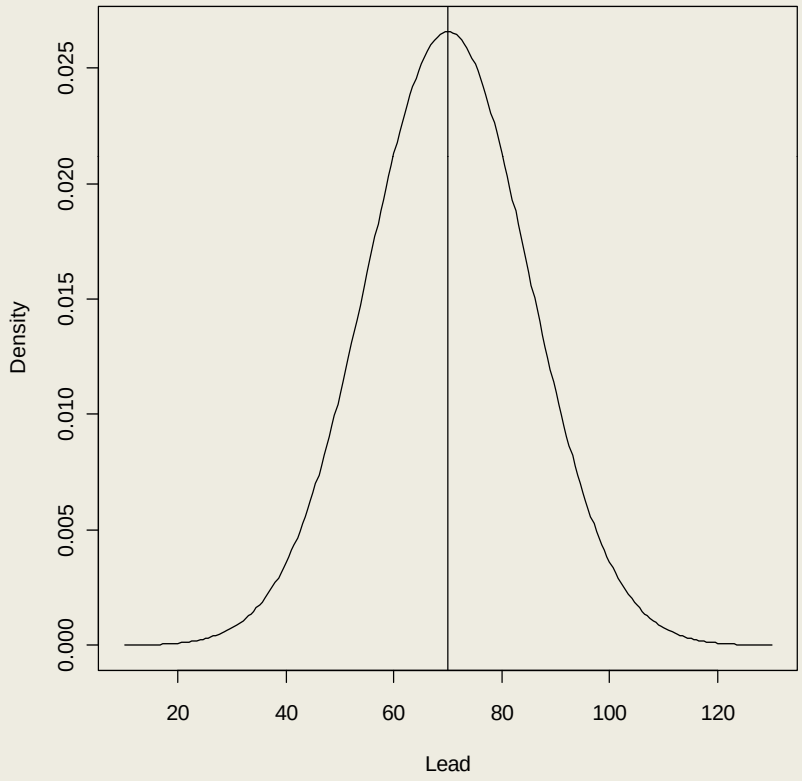
4. Computing a Prediction Bound

- A random sample of 10 toy items from a lot of 1,000 toys was tested for lead. The average lead content in the sample was measured at 70 ppm. The standard deviation was 15 ppm.
- *Assumption*: The lead content follows a *normal distribution*.
- *Question*: What is the probability that the lead content for every toy in the production run of 1,000 toys is below 300 ppm, i.e., the production lot is good?

4.1 Answer

- Inputs: sample size=10, mean is 70, standard deviation is 15, lot size is 1,000. Bound is 300 ppm.

- Results: *99.993% or on average 1 bad lot in 14,200 runs.*



How Did He Figure That Out?

- There must be a trick

The Formula. Normal Theory Case

- Formula to compute the production bound *y*

$$y = \bar{x} + s * t(1 - \alpha/m, n - 1) * \sqrt{1 + 1/n}$$

We set y = 300 ppm and solve for 1-α, the probability. Sample size (*n)=10,* standard deviation *(s) =15,* mean*(* $\bar{x}$ *)= 70,* and lot size *(m)=1000. t* is the *t distribution.*

Solution: 1-alpha = 99.993% or 1 bad lot in 14,200 runs.

- This formula is known to be conservative (understates the probability by a little).

- Source: Hahn and Meeker (1991), p. 63. See References.

What Happens With Different Values?

- Original Case. Sample size=10, mean is 70 ppm, s is 15 ppm, lot size is 1,000. Bound is 300 ppm. Results: 99.993% or 1 bad lot in 14,200 lots on average.

- Suppose the mean is 170 ppm. 99.15%, 1 in 118 lots.
- Suppose the mean is 220 ppm. 67.1%, 1 in 3 lots.

- Mean is 70 ppm, lot size is 10,000. 99.93%, 1 in 1,420 lots

Same Case, Smaller Sample Sizes

- Original Case. Sample size=10, mean is 70 ppm, *s* is 15 ppm, lot size is 1,000. Bound is 300 ppm. Results: *99.993% or 1 bad lot in 14,200 lots*

- Sample size is 5. Probability is 92.5%, 1 in 13.
- Sample size is 7. Probability is 99.6%, 1 in 278.

Factors in Sample Size Planning Using the Formula

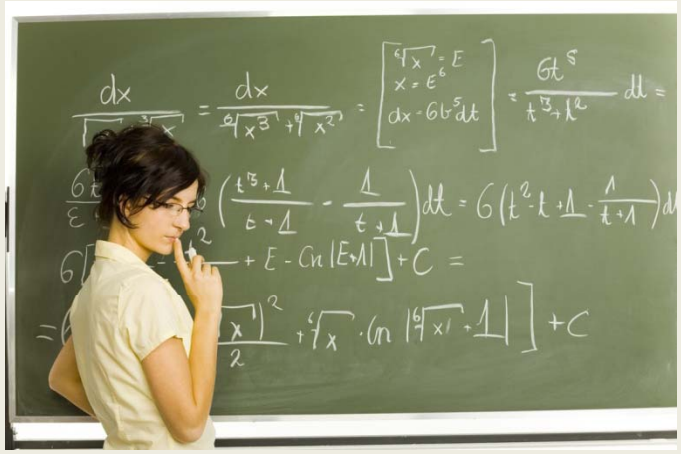
Larger Sample Size

- High risk of injury/death
- *Large production Run (m)*
- *High variability (s)*
- *High mean,* $\bar{x}$
- Non-destructive tests
- No similar products
- Non-compliance is not obvious
- Relevant consumer complaints

Smaller Sample Size

- Low risk of injury/death
- *Small production Run*
- *Low variability*
- *Low mean*
- Destructive tests
- Similar products
- Non-compliance is obvious
- No relevant complaints

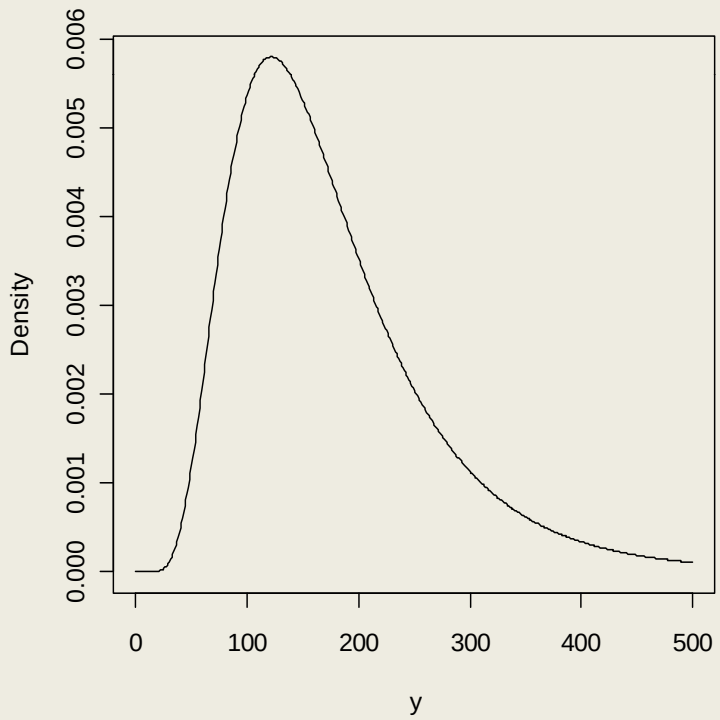
Review



- Probability (assurance about the lot) goes down with
  - Mean closer to 300 ppm
  - Larger Standard Deviation
  - Larger production lot
  - Smaller sample size

Normal Distribution Doesn't Hold?

Lognormal Distribution



Remedy

- Transform the data and
- Test the transformation to see if the normal distribution works.
- If so, use the formula
- If not, (a) try some other transformation or (b) use the Pass Fail (Qualitative) Approach.

Products with Quantitative Data Applicable to the Prediction Bound Approach

- ATV pitch stability coefficient/ATV stopping requirements.
- Baby bath seat tilt angle
- Lead content
- Lead in paint
- Phthalate content
- Bicycle helmet impact attenuation test
- Bicycle stopping distance
- Noise from cap guns
- Flame spread time for wearing apparel
- Dive stick angles
- Mattress combustion temperature
- Small part circumference

What Is the Appropriate Prediction Probability?

Higher Probability

- High risk of injury/death
- *Large production Run*
- *High variability*
- *High mean*
- Non-destructive tests
- No similar products
- Non-compliance is not obvious
- Relevant consumer complaints

Lower Probability

- Low risk
- *Small production Run*
- *Low variability*
- *Low mean*
- Destructive
- Similar products
- Non-compliance is obvious
- No relevant complaints

Example on next page

Production Lot Size Example

- A manufacturer's test results indicate that his production lot has a 90% chance that all items are below 300 ppm. His lot is 500 items.
- A second manufacturer has the same results but for a lot of 50,000 items.
- 90% for 50,000 items does not provide as much assurance as 90% for 500 items.

4.2 Qualitative: Pass/Fail Data

- Test results only indicate if each item in the sample has passed or failed.

- General principle. Since we do not know if the pass was close to the threshold or far from the threshold, there is additional uncertainty.
- That uncertainty translates into a larger sample size, lower probability ….. *and no statement about a totally compliant production lot.*

Setup

Two Cases

- Case 1: Large Production Lot
- Case 2: Small Production Lot

The Formula Box

Procedure

1. Define p = as the proportion noncompliant in the lot.
2. We want an upper bound for p depending on the desired probability 1-α. (95%, 99%).
3. Randomly sample k items.
4. Find the largest value of p such that $(1-p)^k \geq \alpha$.
5. Solution: $p \approx 3/k$ for 95%, 4.6/k for 99%.

- Source: Jovanovic and Levy (1997).

Example

- Lot size = 10,000. k= sample size = 50, $1-\alpha=0.95$ or 0.99.
- Solution $p = 3/50 =$ 0.06 for 95% and 0.092 for 99%.
- Interpretation: With 95% probability the lot contains no more than 6% (600) non-compliant items. For 99%, 920 non-compliant items or fewer.

More Examples

- Sample 100 items, all compliant. *$p = 3/100 = 0.03$ or no more than 3% non-compliant items* with 95% probability.

- Sample 25 items, all compliant. *$p = 3/25 = 0.12$ or no more than 12% non-compliant items* with 95% probability.

Small Production Lot Example

- In the last set of examples, we assumed that the probability of selecting a non-compliant item didn't change as additional items were selected. OK for large production lots but may not be good for small lots…

The Final Formula

- With α as 0.05 or 0.01, solve the hypergeometric distribution for *m, the number of non-compliant items in the lot of m+n items.* Sample is *k* as before.

$$\frac{\binom{m}{0}\binom{n}{k}}{\binom{m+n}{k}} \geq \alpha$$

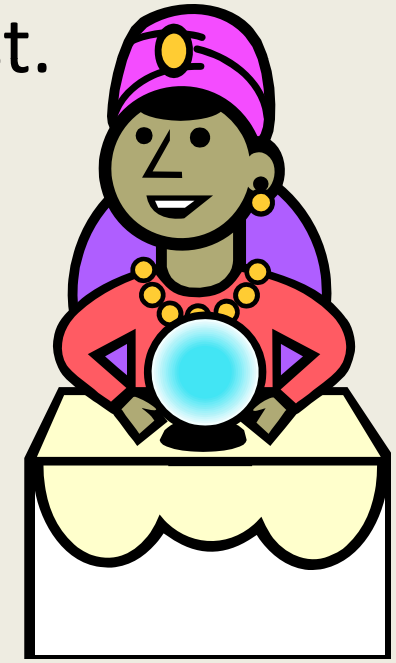
Examples

- Lot size is 200 products.
  - Sample 50 items, all compliant. *m = 10 or p= (10/200=) 5% with 95% probability.*
  - Sample 25 items, all compliant. *m =21 or p=10.5% with 95% probability.*

Note: the 3/*k* relationship is approximately correct in this situation.

5. Sample Size Planning

The statistician's nightmare because you have to know something about the test results in order to know how many items to test.



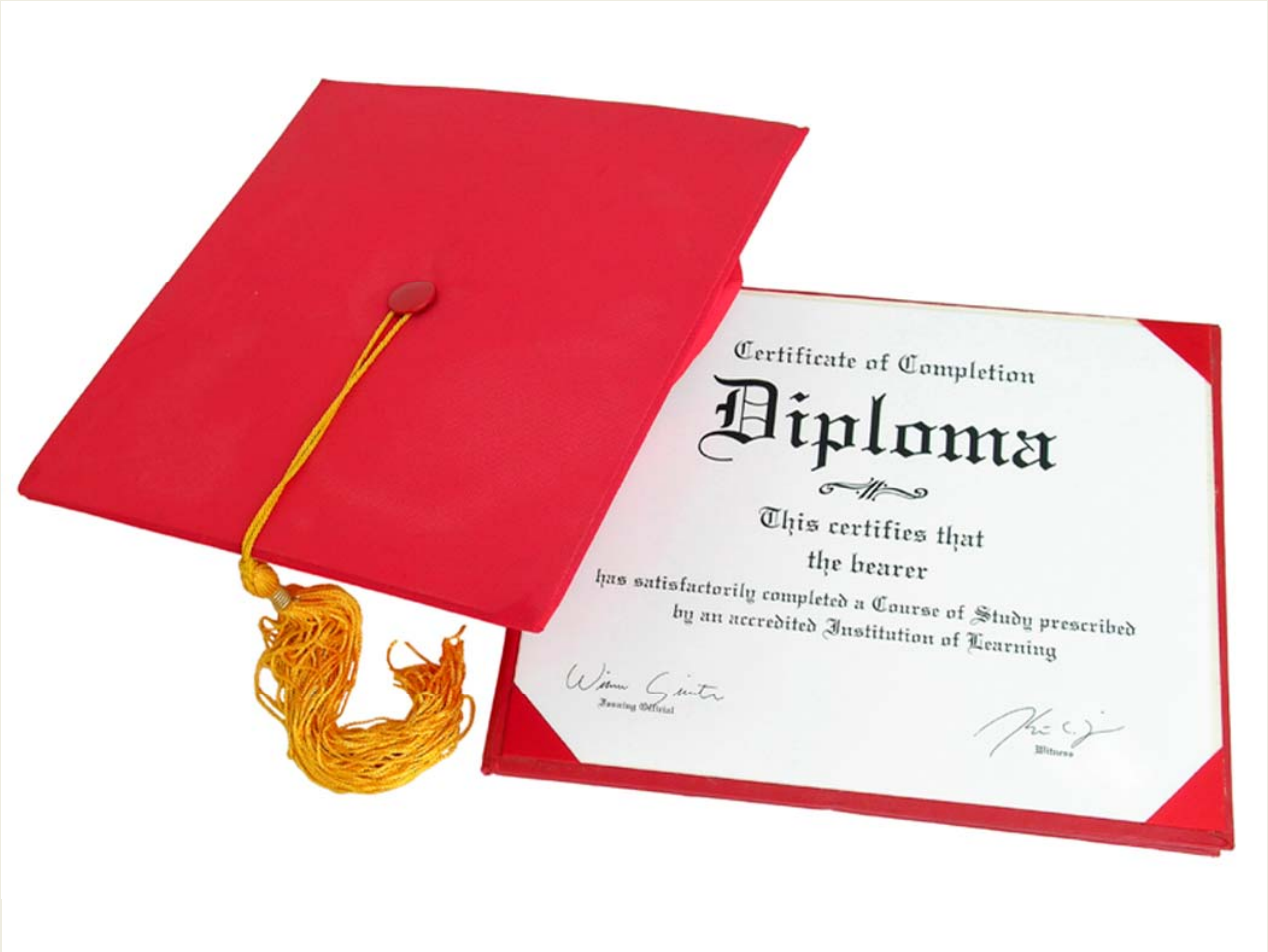
Suggested Procedure

- Pick the right model (normal, qualitative…) and the appropriate formula from the formula box.
- Use prior knowledge or small experiment to fill in values for the product (mean, standard deviation, proportion of non-compliant items, etc.).
- Take a guess at the sample size.
- Make a computation using that sample size. Find the probability the lot complies, or the upper bound on the proportion not complying.
- Increase or decrease the sample size as required.
- Then test the sample and analyze the data.

Our Trip

We've Come a Long Way

Review

- Factors to consider in analyzing test data.
- Quantitative and Pass/Fail
- Prediction bound based on normal theory
- Upper limit on proportion or number of non-compliant products in a lot.
- Sample size principles.

THANKS. Hope This Was Helpful