Measurement Agreement of Manufacturing and Customer Testing

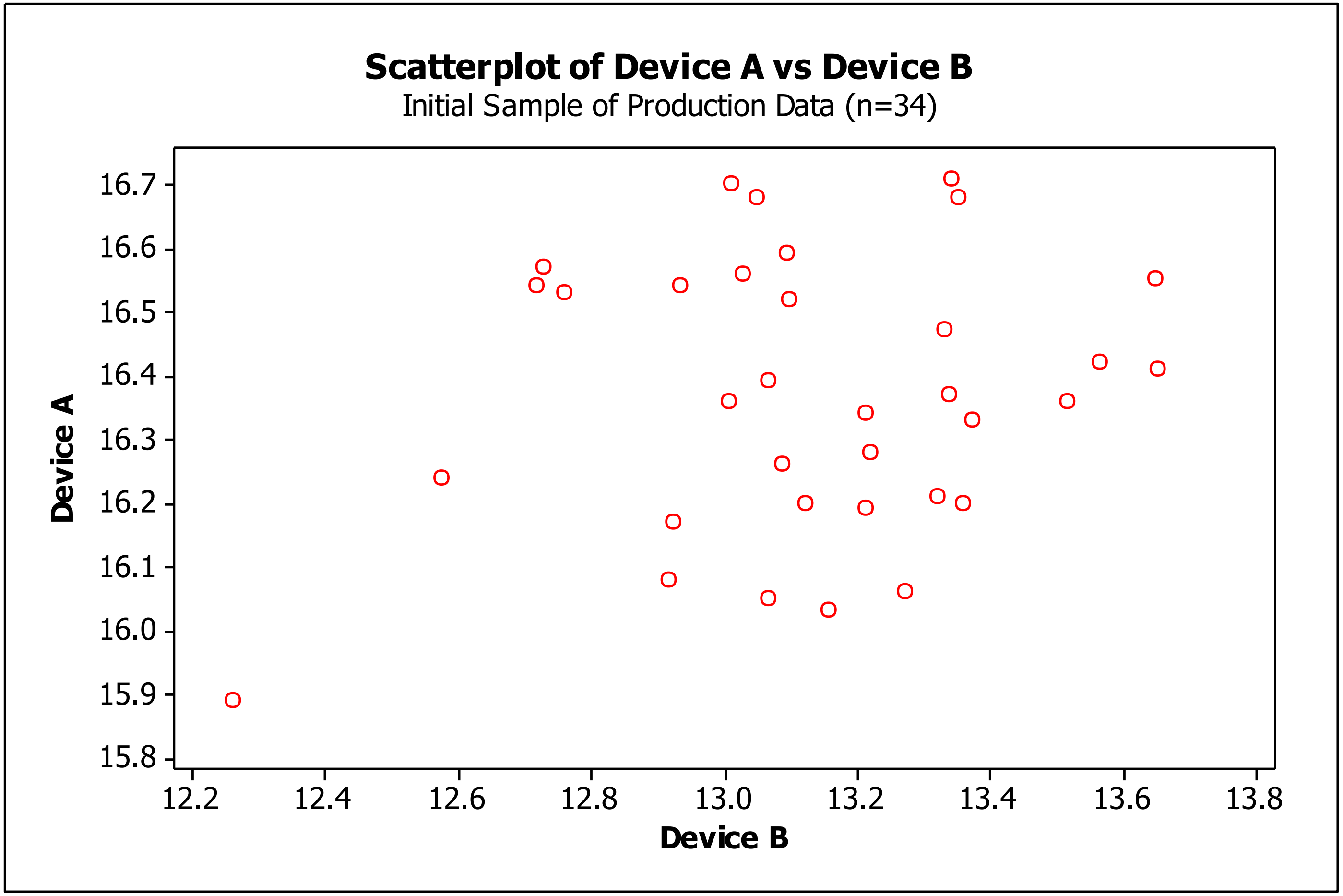
Jon Lindenauer
Weyerhaeuser Company

Problem Statement

- A manufacturing team would like to measure a key quality characteristic with device A. The customer measures the same quality characteristic with device B.
- Device A takes 10 to 15 minutes to complete the measurement cycle, while device B takes 45 to 55.
- The manufacturer would like to replace device B with A.

Initial Product Testing

- The product testing involved taking samples of production material. These samples are put in a large plastic bag and sealed.
- In the first stage of both tests, sub-sample material is altered by the prepping procedure.
- The sub-sample material is then destroyed by the second stage of each test.



Analysis of initial data (n = 34) from new product trials showed negligible correlation r = .20 (p-value = 0.27) between device A and B.

Why No Correlation?

- This destructive test does <u>not</u> allow measuring the same experimental unit twice.
- The samples were collected from an in control process.
  - The sample values were only a small part of the working range of the measurement devices.
- The approximate working range of device A is 15 to 25 and device B is 12 to 222.
  - The existing data range for device A was 20.9 to 21.7 and 17.3 to 18.7 for device B.

Solution

- An experiment to gather and test the mill data in a more controlled manner.
  - The material requirements for testing make it prohibitive to create lab sample.
  - We could only use material from prime production, representing only a small working range of material.
- We may need to use a method other than correlation to analyze how the two devices match each others measurements.

Measurement Agreement

- Altman and Bland (1983) and Bland and Altman (1986, 1999) discuss an alternative method to correlation analysis of two measurement devices. They prefer assessing the agreement between two measurement systems.
  - Assess if measurement systems agree on measuring the same or similar experimental units.
  - If they do agree, we can use one device to replace the other.

Bland and Altman Critique of Correlation Analysis

- *r* measures the strength of association, not the agreement between 2 testing devices.
- *r* depends on the range of the quantity measured.
  - In general, wider range means greater correlation.
- Some test devices may produce high correlation but not be in good agreement.

Customer Interaction

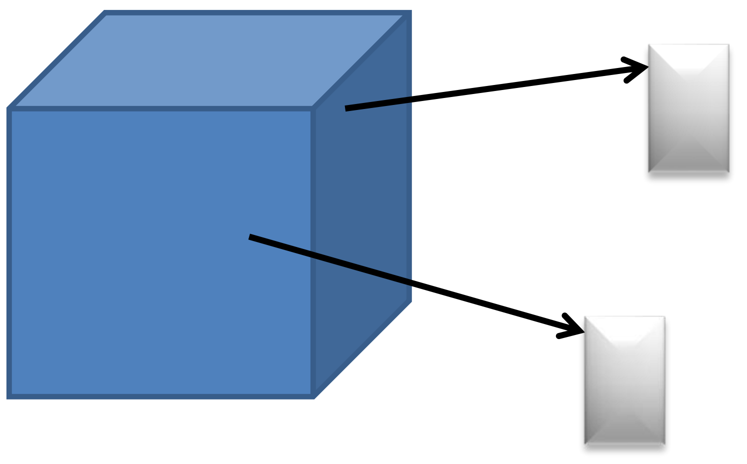
- Communication with the customer was key to the acceptance of the Bland and Altman measurement agreement method.
- The main concern of the customer was would they know if the material was out of the specification.

Customer Interaction

- To allay their concern, it was agreed that the lower tolerance limit around device A measurements conform to the product specification lower limit.
  - *Target minus 1.4*
  - The lower tolerance limit will be discussed later in the presentation.

Experiment Sampling

- Randomly select 10 units from prime production runs.
- 2 sample bags of material collected from each unit.



Experiment Sampling

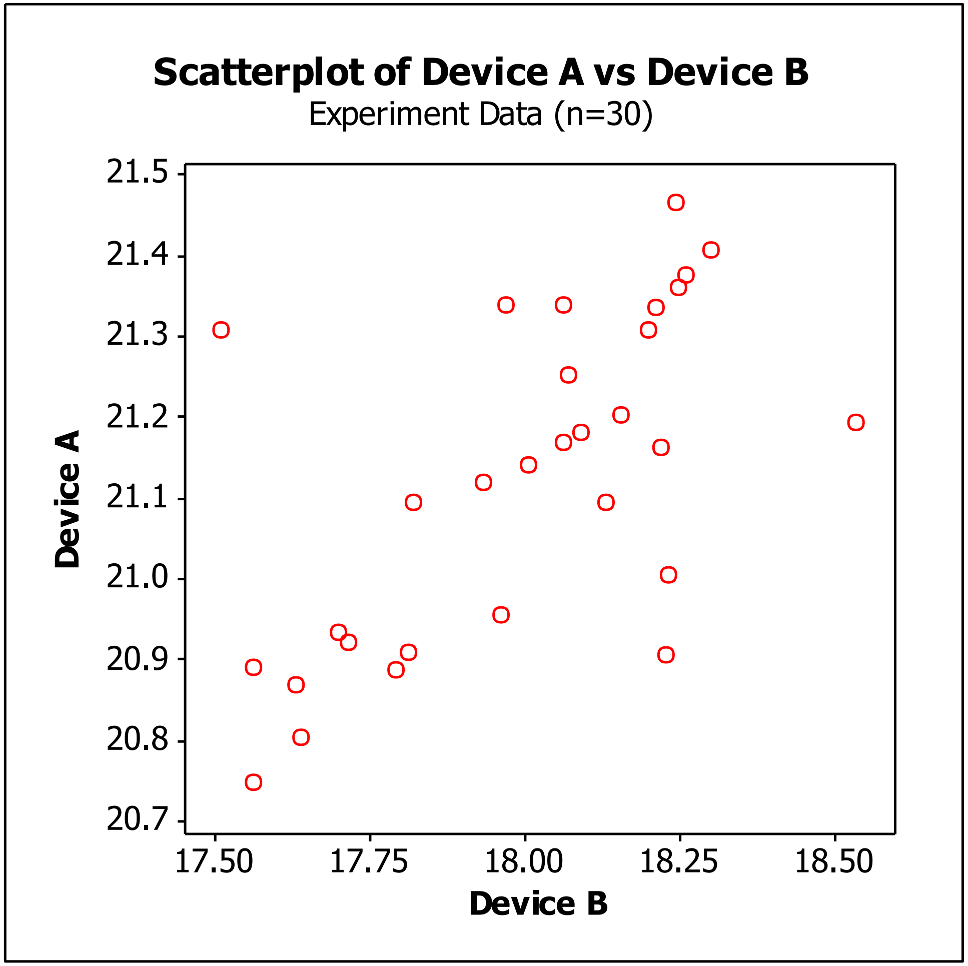
- The 2 sample bags are the sampling units. The bags from the same unit were each randomly tested at different times to represent independent replications.
- The first set of tests for device A and device B are matched from the same sample bag. The second set of tests for device A and device B are matched from the same sample bag.
- 3 skilled operators performed all the testing.
- Total of n = 30 for each device or 30 matched pairs.

Measurement Agreement Procedure

- Altman and Bland (1983) and Bland and Altman (1986 and 1999) discuss the procedure for measurement agreement analysis.
  - Plot A vs. B
  - Create the Agreement (Bland and Altman) plot.
  - Check for stable variability.
  - Assess the precision of the estimates.
  - Check the acceptance criteria.
  - Time constraints caused the stability and precision to be skipped. The paper contains this analysis.

Plot the Two Devices Measures

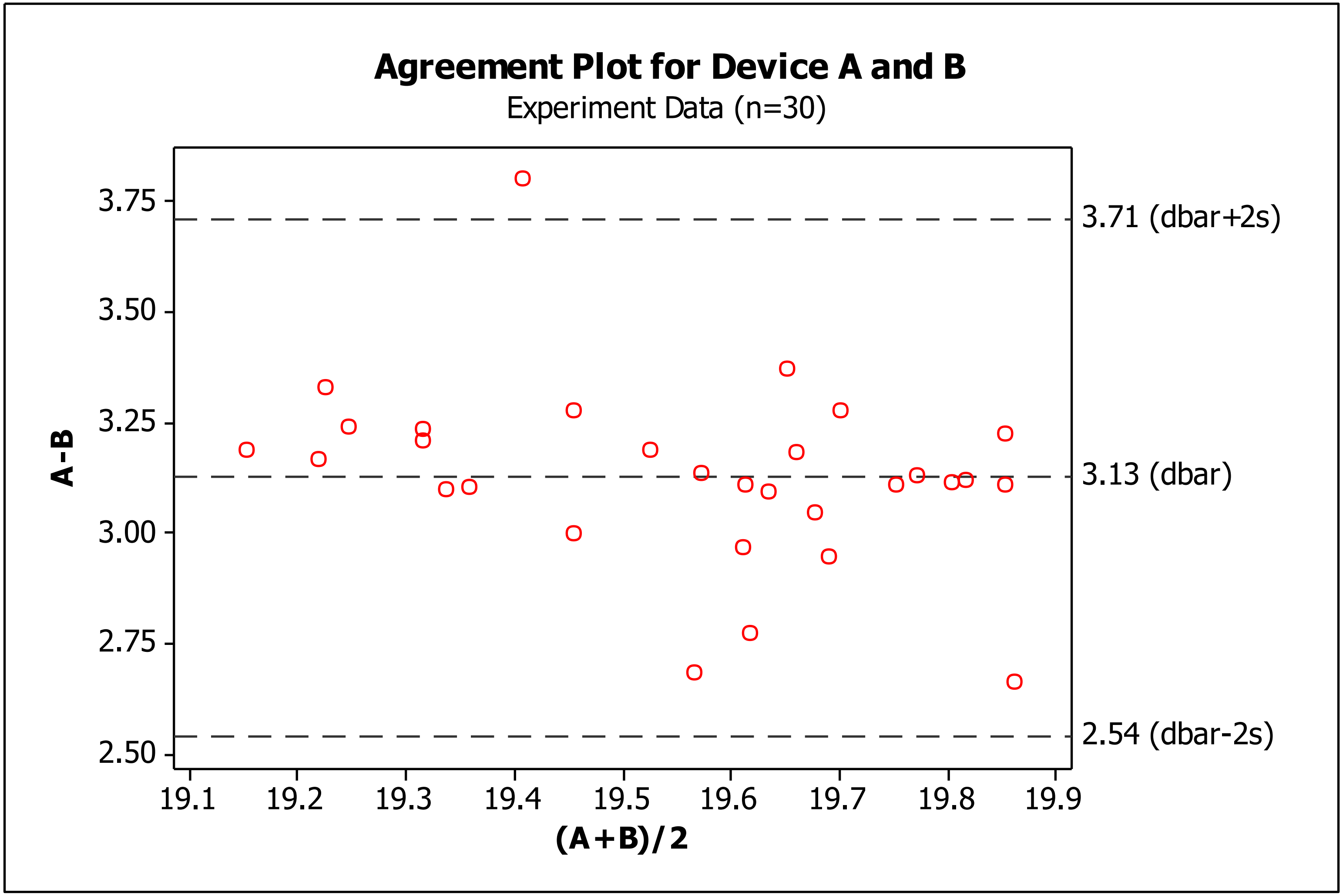
- The usual scatter plot of device A (y-axis) against device B (x-axis).
- Note that device A always measures about 3 units higher than device B.



The plot shows some relationship between device A and B, *r*= 0.628 (p-value = 0.000), but not strong.

Agreement (Bland and Altman) Plot

- The difference between matched pairs of test results for on the y-axis $(A - B)$
- The mean of the devices for each matched pair on the x-axis $(A_1 + B_1)/2$
- A horizontal line representing the mean of the differences $\bar{d}$
- Upper and lower 95% limits of agreement around the mean difference $(\bar{d} \pm 1.96s)$



The correlation of A-B and (A+B)/2 is r = -0.332 (p-value = 0.073) and we can assume there is no systematic bias.

Acceptable Difference Criteria

- What constitutes an acceptable difference in agreement between the measurement systems?
  - This depends on the application.
- Bland and Altman (1986) imply that good agreement should be decided up front and based on the following ratio:

$$\left\{\left((\bar{d}+2s)-(\bar{d}-2s)\right)\Big/\text{Mean of }\left[(A_1+B_1)/2\right]_i\right\}$$

Acceptable Difference Criteria

- It was agreed that the acceptable difference criteria < 10% defined good agreement

- Our difference is <u>acceptable</u>:

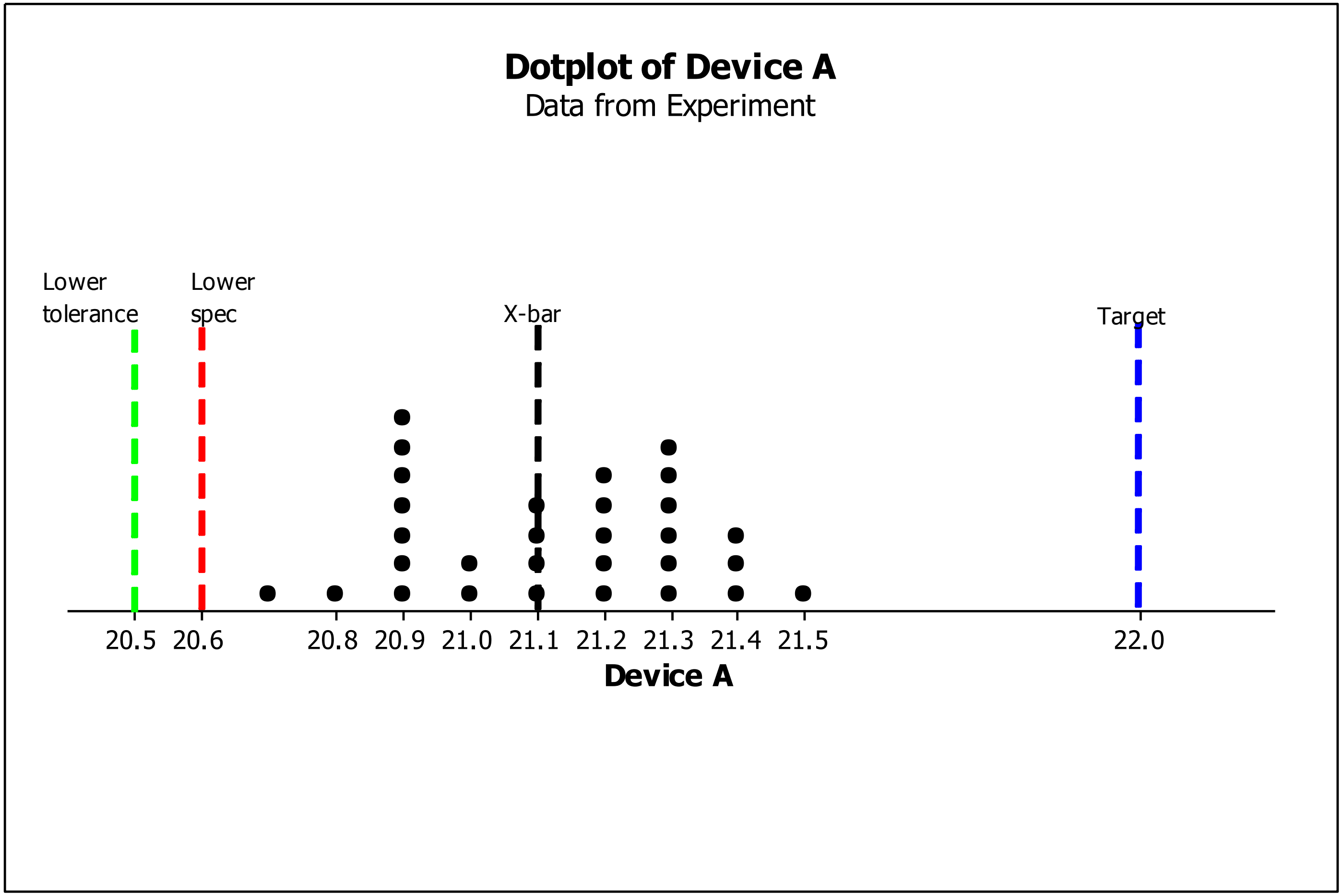
$$(3.71 - 2.54)/19.56 = .059 \Rightarrow 5.9\%$$

Tolerance Intervals

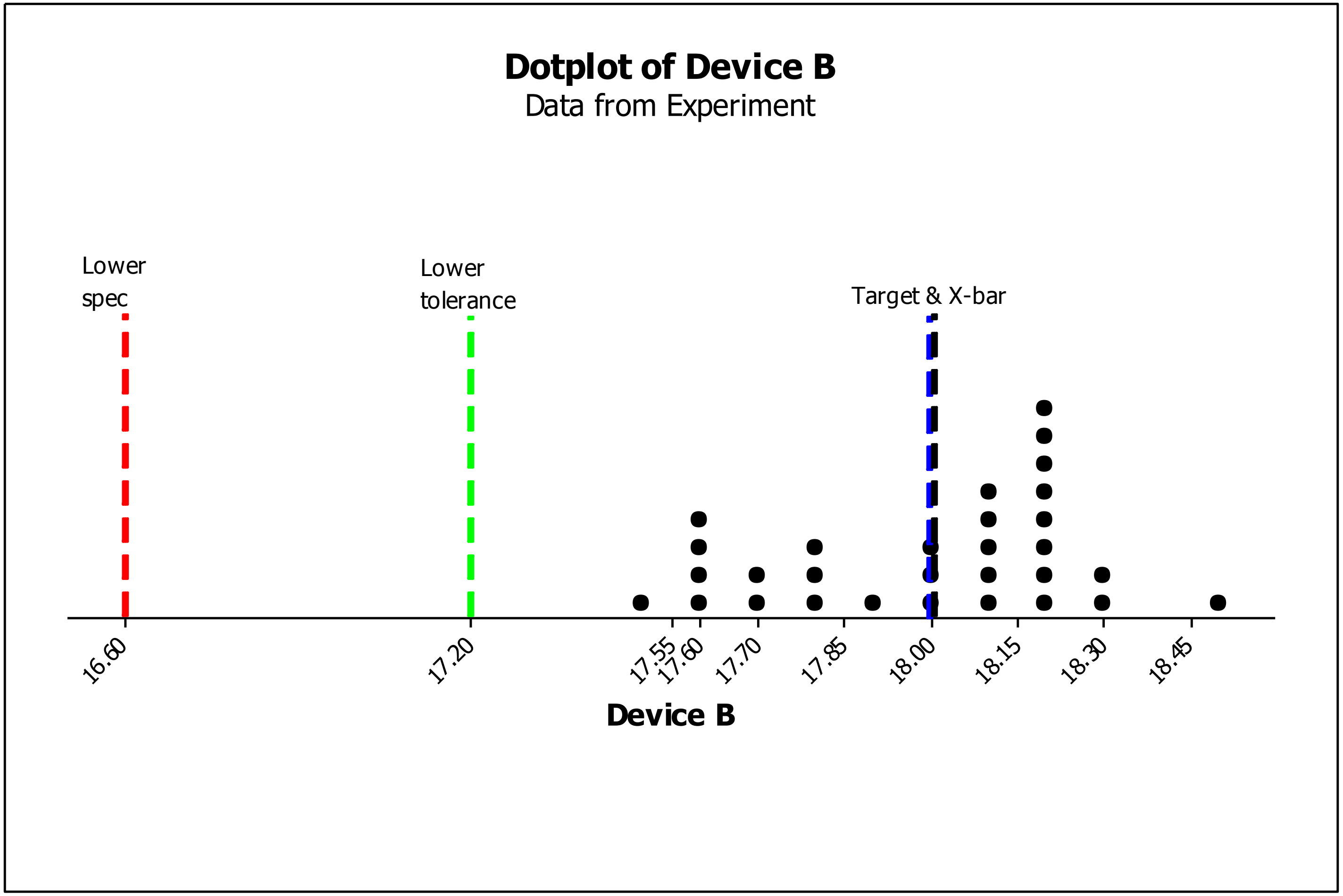
- Tolerance intervals of a process are limits that cover a percentage of the measured product quality characteristic distribution (95%, 99%, etc.) with a given confidence level (90%, 95%,etc.).
- Specification limits smaller than the tolerance limits will result in an unnecessarily high amount of rejects.
- One-sided tolerance limits can be calculated. (Montgomery, 2005)

Tolerance Intervals

- The customer agreed on a one-sided lower limit with 95% confidence that it will be exceeded by at least 99% of the incoming product.
- The one-sided lower tolerance limit was calculated with 95% confidence that the interval, $\bar{x} - Ks$, covers at least 99% of the distribution. The constant K , $\bar{x}$ and s were calculated by Minitab® statistical software.



The dot plot shows that for the 30 sample pairs tested, the target is 22, x-bar=21.1 and the lower tolerance limit of 20.5 is less than the lower specification limit of 20.6 .



The dot plot shows that for the 30 sample pairs tested, the target is 18, x-bar = 18 and the lower tolerance limit of 20.5 is less than the lower specification limit of 20.6 .

Comments

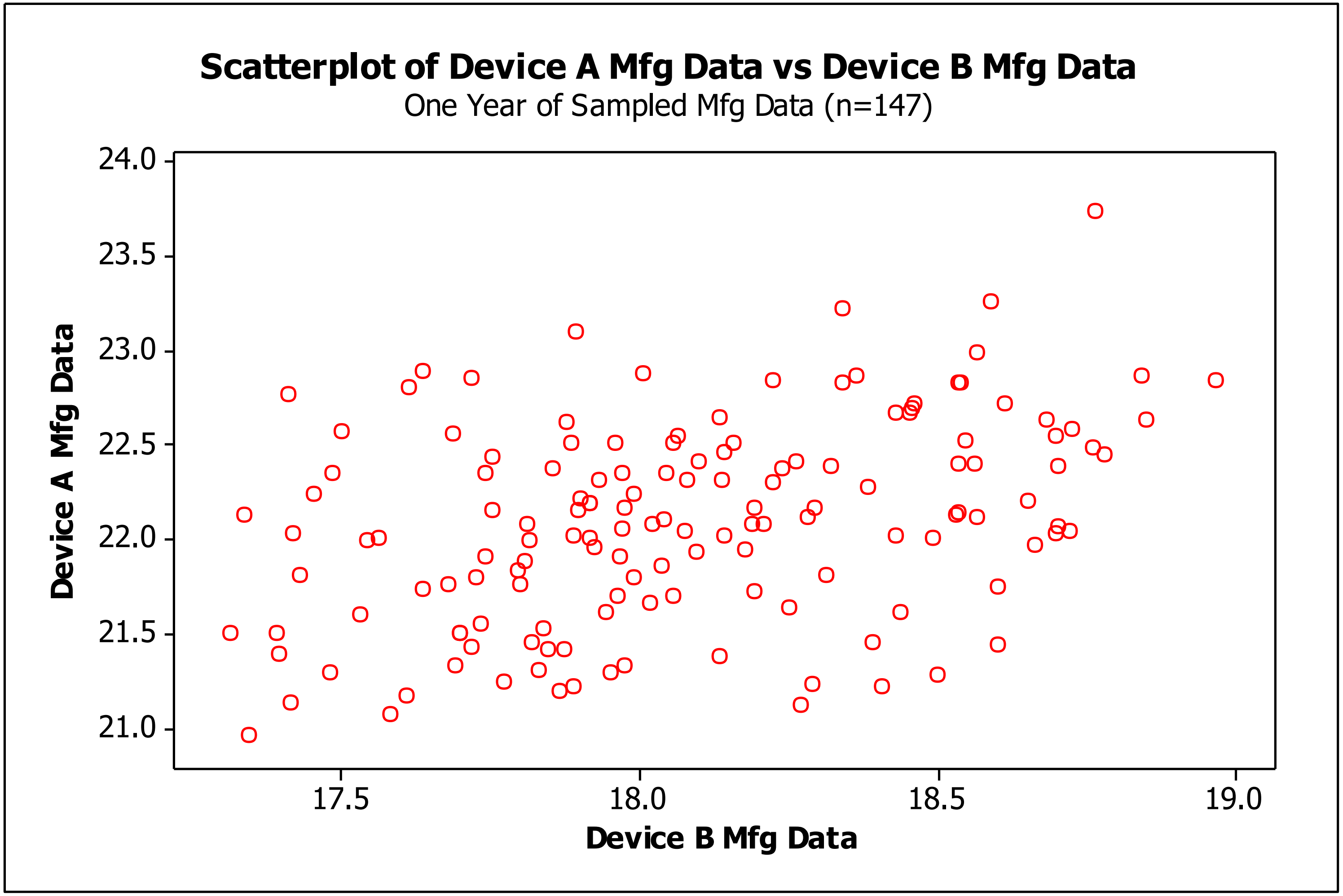
- The lower tolerance limit was lower than the lower specification limit for device A, which could cause excessive rejections.
  - Device B was the opposite.
- The customer wanted the spec limits for device A set at Target minus 1.4. They choose a target 4 units higher (22) than the device B target (18).
  - Note that this was “insurance” as device A is normally 3 units higher than device B.

Comments

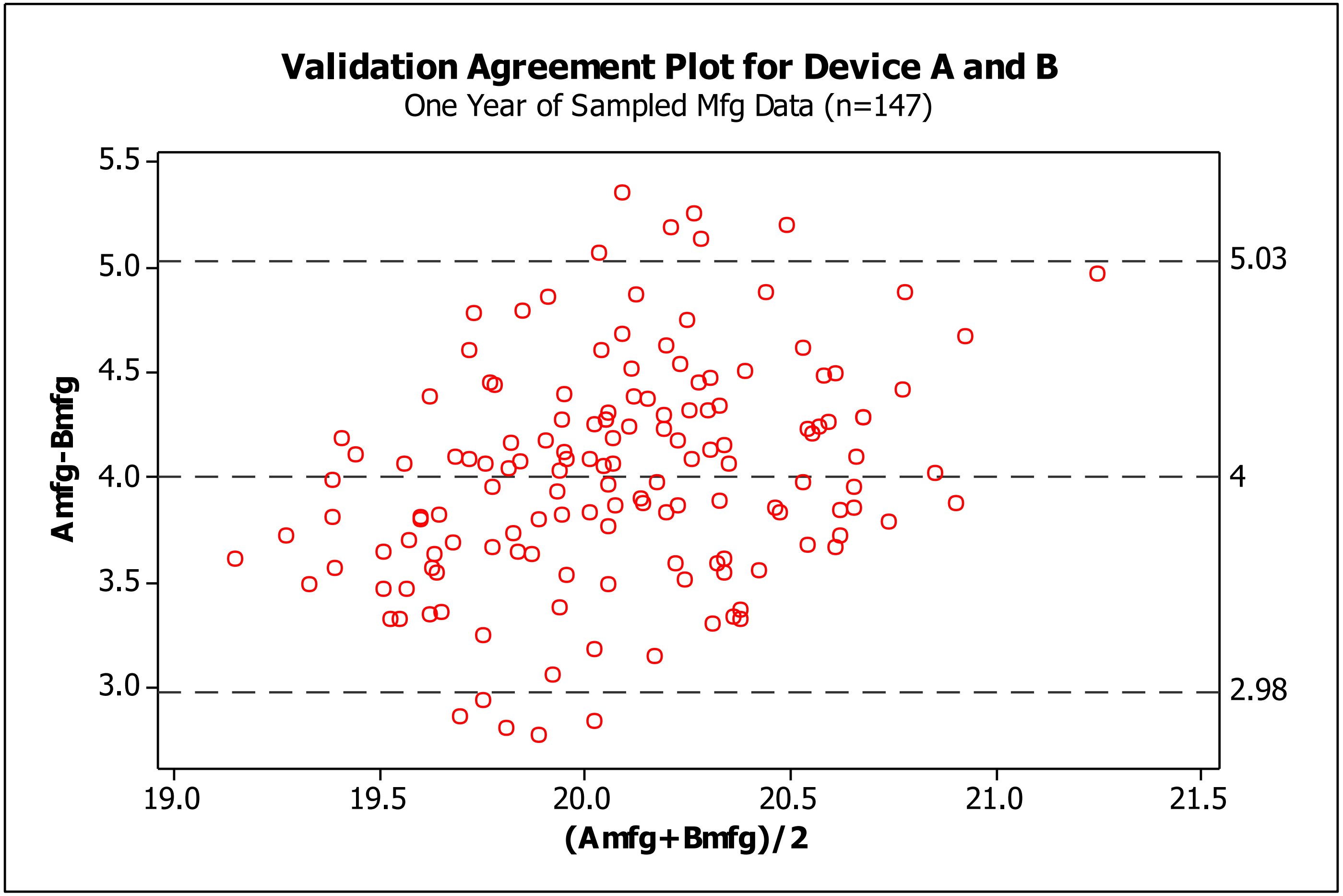
- The results of the Agreement plot and the dot plots of each device, with the specs and tolerance limits was acceptable.
- It was decided to use device A as a surrogate for device B.
- The agreement between the two devices was periodically checked throughout the following year.

Method Validation

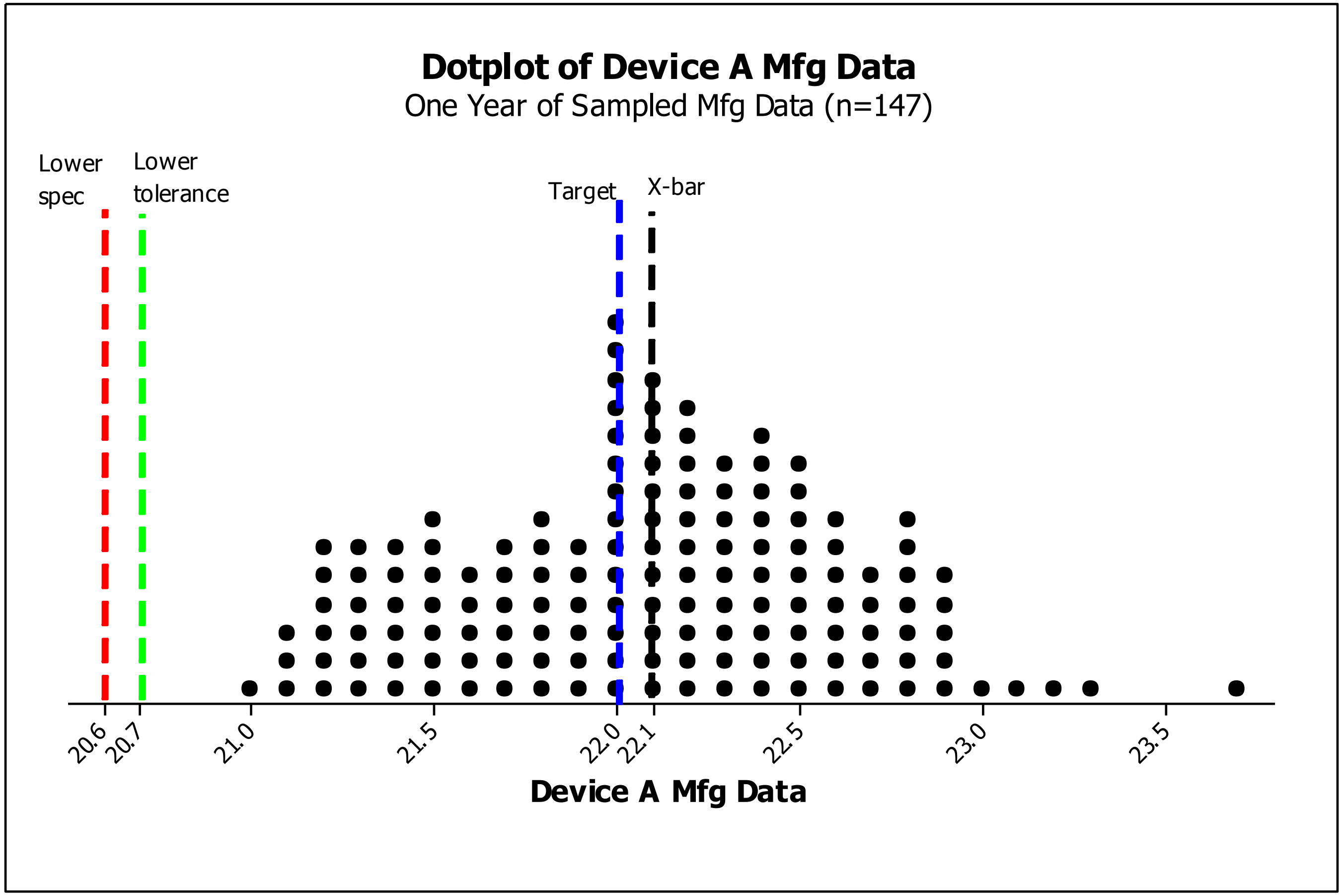
- Over the course of the next year, 147 bales were sampled at the mill for testing on both device A and device B for validation of the measurement agreement.



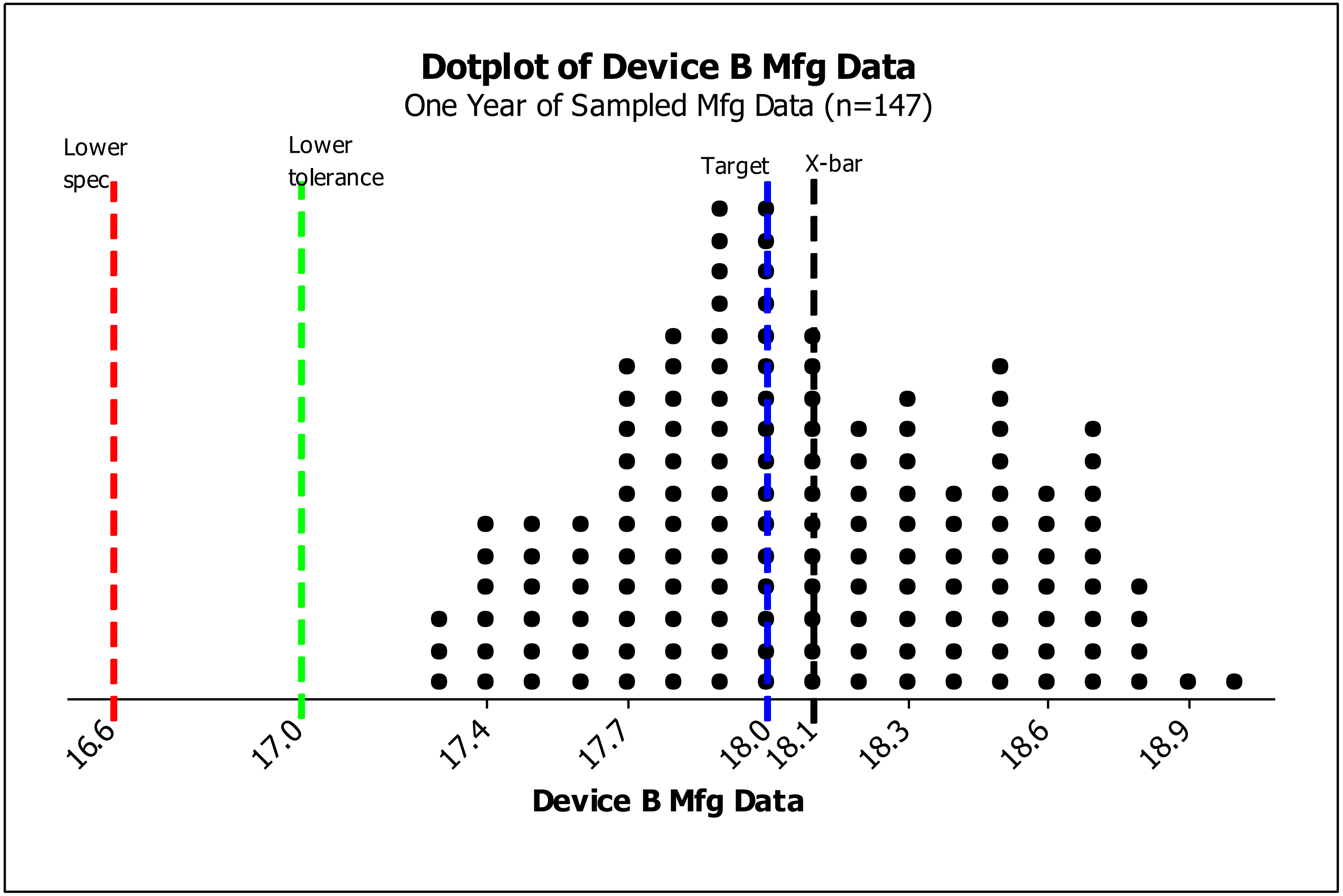
Correlation of Device A Mfg Data and Device B Mfg Data is *r* = 0.40 (p-value = 0.000)



The mill data shows more variability in the differences and a higher overall mean difference between device A and device B. The correlation of Amfg-Bmfg and (Amfg+Bmfg)/2 is r = 0.316 (p-value = 0.000). This was not an issue.



Dot plot of mill data for device A .The data conforms to the lower specification. The lower tolerance limit is greater than the lower spec limit, which is desired.



Dot plot of device B for the mill data. The data conforms to the lower specification. The lower tolerance limit is greater than the lower spec limit, which is desired.

Conclusions

- The acceptance criteria for good agreement was 10.2% for the manufacturing data.
  - slightly higher than 10%, but acceptable.
- The mill data (n=147) centers very near the target and conforms to the lower tolerance limit and the lower specification limit for both device A and device B.
- This validates the analysis and results of the Bland and Altman measurement agreement study: device A may be used as a surrogate for device B.

Questions?

Contact info: Jon Lindenauer, Weyerhaeuser

Jon Lindenauer, Weyerhaeuser
Statistician and ASQ Certified Quality Engineer
jon.lindenauer@weyerhaeuser.com