1.9 The Kappa Technique

MBC 638

Data Analysis and Decision Making

2 of 72

"No two things are alike, but even if they were, we would still get different values when we measured them."

-D. Wheeler

3 of 72

Discrete Measurement Systems

4 of 72

Discrete Measurement Systems

Kappa Technique

$$K = \frac{P_{\text{observed}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$

5 of 72

Discrete Measurement Systems

Kappa Technique

$$K = \frac{P_{\text{observed}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$

Discrete Measurement Systems

Kappa Technique

$$K = \frac{P_{\text{observed}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$











Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B B G B B G B 5/12 7/12

You G G G G B B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

13 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 G
 B/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

14 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V 7 7/12 = 0.58

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

16 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 B
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 G
 8/12
 4/12

 V
 V
 V
 V
 V
 V
 7/12 = 0.58

17 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

18 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

19 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

21 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 G
 8/12
 4/12

 V
 V
 V
 V
 V
 V
 7/12 = 0.58

22 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

23 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

24 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G B B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

26 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 B
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 G
 8/12
 4/12

 V
 V
 V
 V
 V
 V
 V
 7/12 = 0.58

27 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V V 7 7/12 = 0.58

28 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 V
 V
 V
 V
 V
 V
 V
 7/12 = 0.58

29 of 72

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

• $P_{\text{observed}} = 0.58$

Between Operators: Reproducibility

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Me
 G
 G
 G
 B
 B
 G
 B
 B
 G
 B
 5/12
 7/12

 You
 G
 G
 G
 G
 B
 B
 B
 G
 B
 G
 8/12
 4/12

 Y
 Y
 Y
 Y
 Y
 Y
 Y
 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} =$

31 of 72

Between Operators: Reproducibility

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V 7 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67)$

32 of 72

Between Operators: Reproducibility

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33)$

Between Operators: Reproducibility

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$

34 of 72

Between Operators: Reproducibility

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47}$

35 of 72

Discrete Measurement Systems (cont.)

Kappa Technique

$$K = \frac{P_{\text{observed}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$

Discrete Measurement Systems (cont.)

Kappa Technique

$$K = \frac{P_{\text{observed}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$

• If K > 0.7, measurement system is good.

37 of 72

Between Operators: Reproducibility (cont.)

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47}$

38 of 72

Between Operators: Reproducibility (cont.)

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47} = 0.21$

Between Operators: Reproducibility (cont.)

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G B B B B G B G 7/12 4/12

V V V 7 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47} = 0.21$
- · Indicates that our measurement system is not good

40 of 72

Between Operators: Reproducibility (cont.)

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V 7 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47} = 0.21$
- Indicates that our measurement system is not good

Between Operators: Reproducibility (cont.)

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G B B B B G B G 7/12 4/12

V V V 7 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47} = 0.21$
- Indicates that our measurement system is not good

42 of 72

Between Operators: Reproducibility (cont.)

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Me G G G B B B G B B G B 5/12 7/12

You G G G G G B B B G B G G 8/12 4/12

V V V 7 7/12 = 0.58

- $P_{\text{observed}} = 0.58$
- $P_{\text{chance}} = (0.42)(0.67) + (0.58)(0.33) = 0.47$
- $K = \frac{0.58 0.47}{1 0.47} = 0.21$
- Indicates that our measurement system is not good

43 of 72

The disparity between quality ratings in the previous example is an illustration of a reproducibility problem.

44 of 72



Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree Day 1 G G G B B B B B B B G G 5/12 7/12

Day 2 G B G B B G B B B G B 4/12 8/12

47 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Day 1 G G G B B B B B B B G G 5/12 7/12

Day 2 G B G B B G B B B B G B 4/12 8/12 9/12 = 0.75

48 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Day 1 G G G B B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B B B G B 4/12 8/12

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad **Total Agree** Day 1 G G G B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B G B 4/12 8/12 9/12 = 0.75**/ / /** VVVV

50 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad **Total Agree** Day 1 G G G B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B G B 4/12 8/12 V V V V V 9/12 = 0.75

51 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad **Total Agree** Day 1 G G G B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B G B 4/12 8/12 9/12 = 0.75

52 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad **Total Agree** Day 1 G G G B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B G B 4/12 8/12 V V V V V 9/12 = 0.75

53 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad **Total Agree** Day 1 G G G B B B B B B G G 5/12 7/12 Day 2 G B G B B G B B B G B 4/12 8/12 9/12 = 0.75V V V V V

Within Operator: Repeatability

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Day 1
 G
 G
 G
 B
 B
 B
 B
 B
 G
 G
 5/12
 7/12

 Day 2
 G
 B
 G
 B
 B
 B
 B
 B
 G
 B
 4/12
 8/12

 V
 V
 V
 V
 V
 V
 V
 V
 9/12 = 0.75

55 of 72

Within Operator: Repeatability

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Day 1
 G
 G
 G
 B
 B
 B
 B
 B
 G
 5/12
 7/12

 Day 2
 G
 B
 G
 B
 B
 B
 B
 B
 G
 B
 4/12
 8/12

 V
 V
 V
 V
 V
 V
 V
 V
 9/12 = 0.75

56 of 72

Within Operator: Repeatability

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

 Day 1
 G
 G
 G
 B
 B
 B
 B
 B
 G
 G
 5/12
 7/12

 Day 2
 G
 B
 G
 B
 B
 B
 B
 B
 G
 B
 4/12
 8/12

 V
 V
 V
 V
 V
 V
 V
 V
 9/12 = 0.75

57 of 72

Within Operator: Repeatability

1 2 3 4 5 6 7 8 9 10 11 12 Good Bad Total Agree

Day 1 G G G B B B B B B B G G 5/12 7/12

Day 2 G B G B B G B B B B B G B 4/12 8/12

V V V V V V 9/12 = 0.75

• $P_{\text{observed}} = 0.75$

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$

59 of 72

Within Operator: Repeatability

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- K =

60 of 72

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$

 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 Good
 Bad
 Total Agree

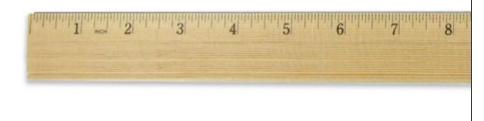
 Day 1
 G
 G
 G
 B
 B
 B
 B
 B
 G
 G
 5/12
 7/12

 Day 2
 G
 B
 G
 B
 B
 B
 B
 B
 G
 B
 4/12
 8/12

 V
 V
 V
 V
 V
 V
 V
 V
 9/12 = 0.75

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85

62 of 72



- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85

64 of 72

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85

66 of 72

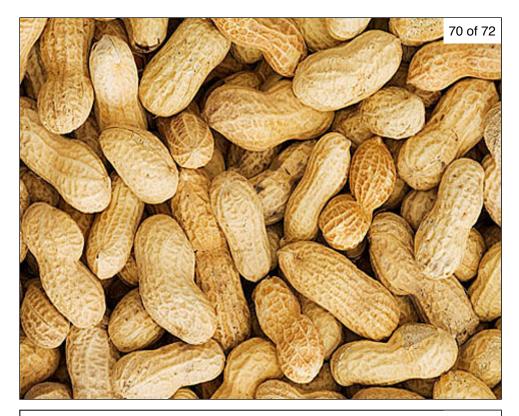
- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85 for a single reviewer
 - Higher than the standard for two person comparisons, > 0.7



Less variability is expected when a single person is executing a process.

69 of 72

- $P_{\text{observed}} = 0.75$
- $P_{\text{chance}} = (0.42)(0.33) + (0.58)(0.67)$
- $K = \frac{0.75 0.52}{1 0.52} = 0.48$
- Still too low: K should be ≥ 0.85
 - Higher than the standard for two person comparisons, > 0.7



Continuous Data Measurements

72 of 72

Continuous Data Measurements

The rule of thumb commonly used to determine if the measurement system is **capable** is to see if the measurement-to-total ratio is less than 10%:

$$\frac{\sigma_{\rm measurement}}{\sigma_{\rm total}} \le 0.10$$