Mix Design Test Variability and Round Robin

2006 North Central HMA Technical Workshop and Conference
10 - 12 January, 2006

Background

❖ Round robin study originally proposed in Omaha 2004 meeting to

➢ …address some of the HMA problems commonly occurring among NC states

➢ …determine which test procedures are causing differences in HMA and aggregate properties

➢ Move towards test standardization

Proposal

- Different states have different practices--Do these differences in lab practices cause significant differences in test results, mix design and performance???
  - Participating lab send aggregate and binder for one or two mixes to NCSC and other participating labs
  - All the participating labs run pre-determined set of aggregate and mix tests (one-point verification) on the same materials following the procedure routinely followed in their lab
  - Send data to NCSC for compilation and analysis

Intent

❖ Are all the labs producing similar results in spite of minor/major differences in followed practices?

❖ If not, which labs are similar/dissimilar, with respect to which property?

❖ What differences in practice may be contributing to this dissimilarity?

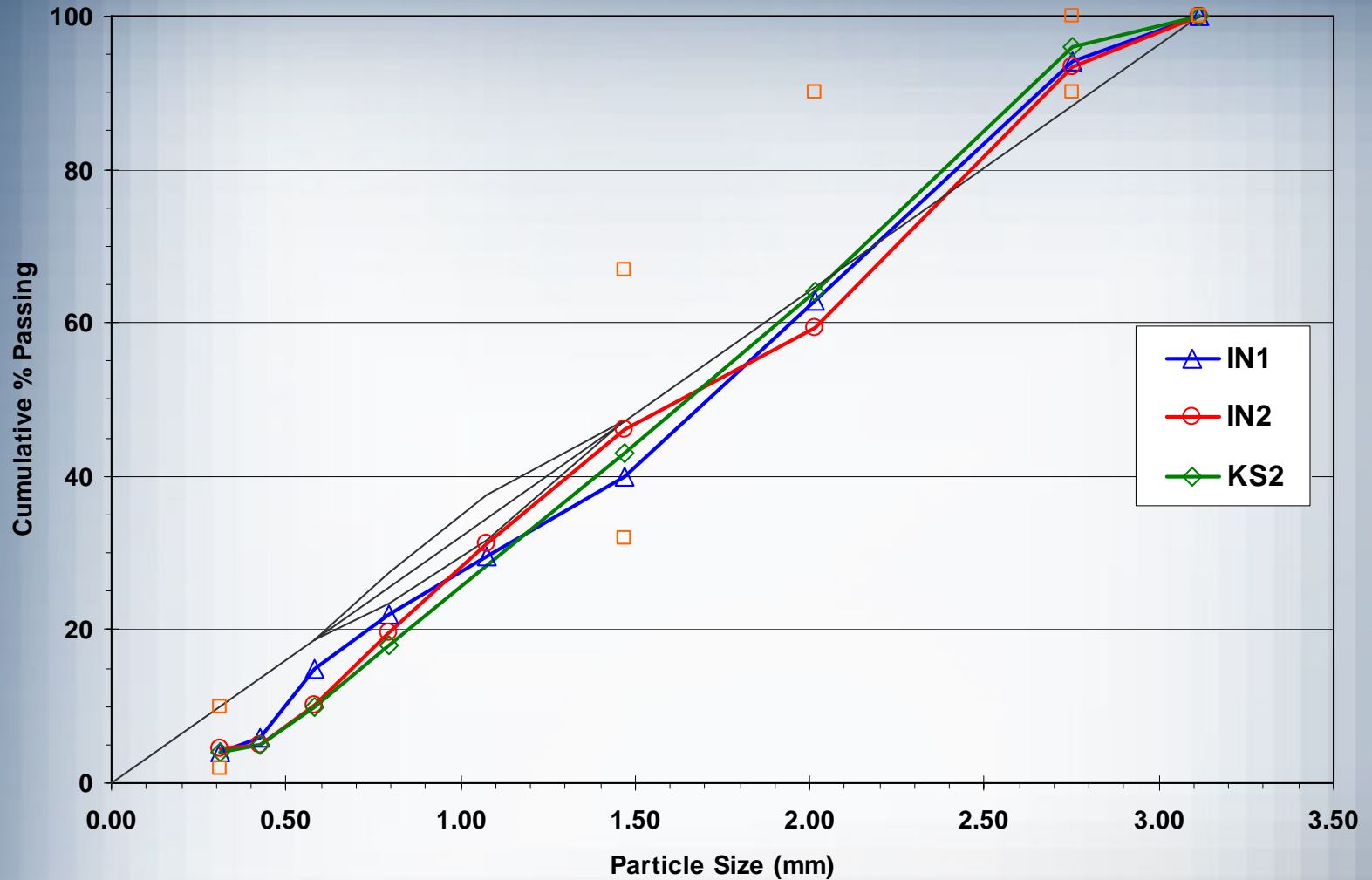❖ Were all the mixes tested statistically similar?

Labs and Mixes

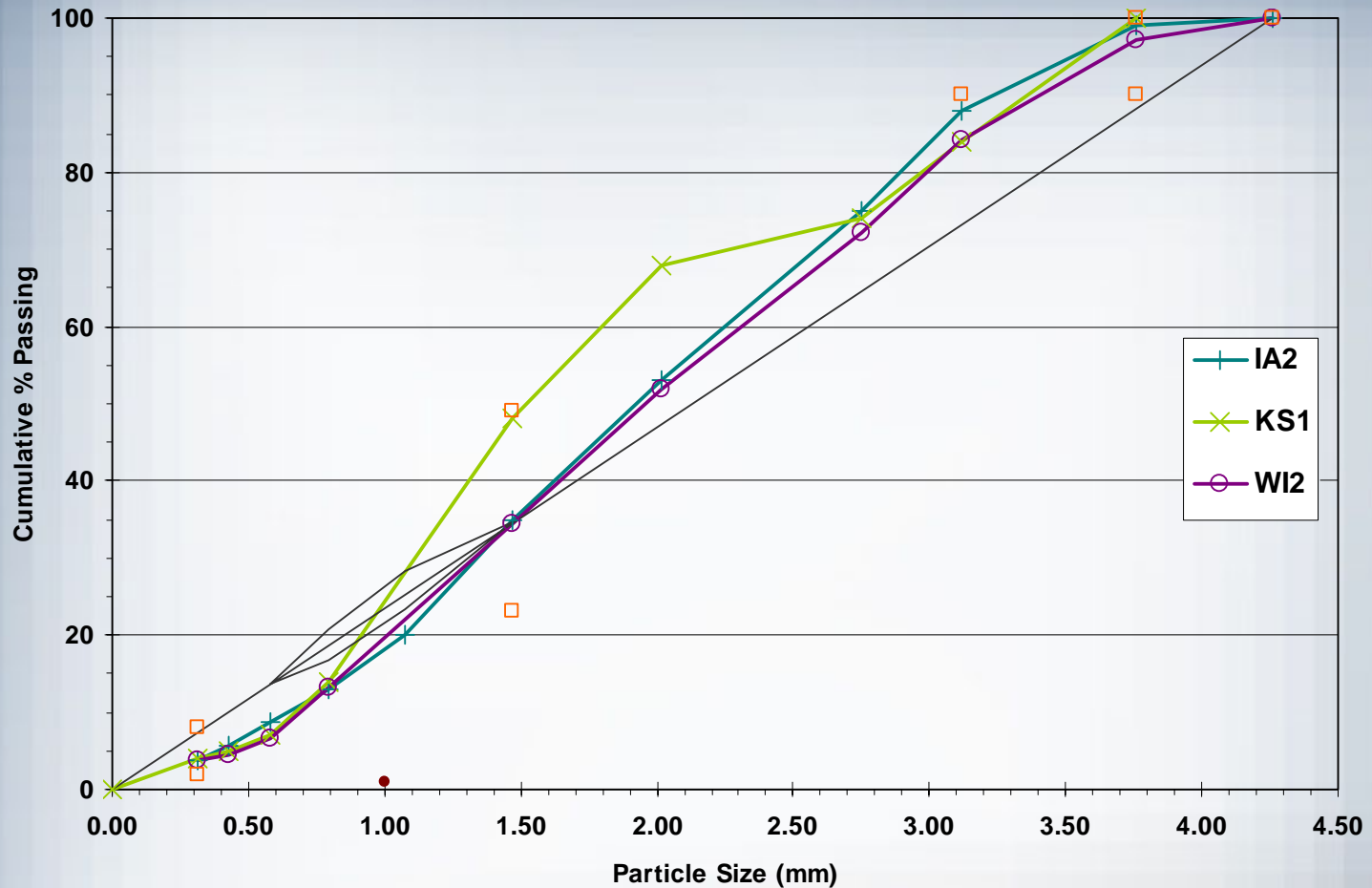| Participating labs	Mixes supplied
❖ Iowa DOT	IA1 and IA2
❖ Kansas DOT	KS1 and KS2
❖ Indiana DOT	IN1 and IN2
❖ Nebraska DOR	NE1
❖ Mathy Construction Co.	WI1 and WI2
❖ NC Superpave Center	

Aggregate Gradation -- 9.5 mm NMAS

Aggregate Gradation -- 12.5 mm NMAS
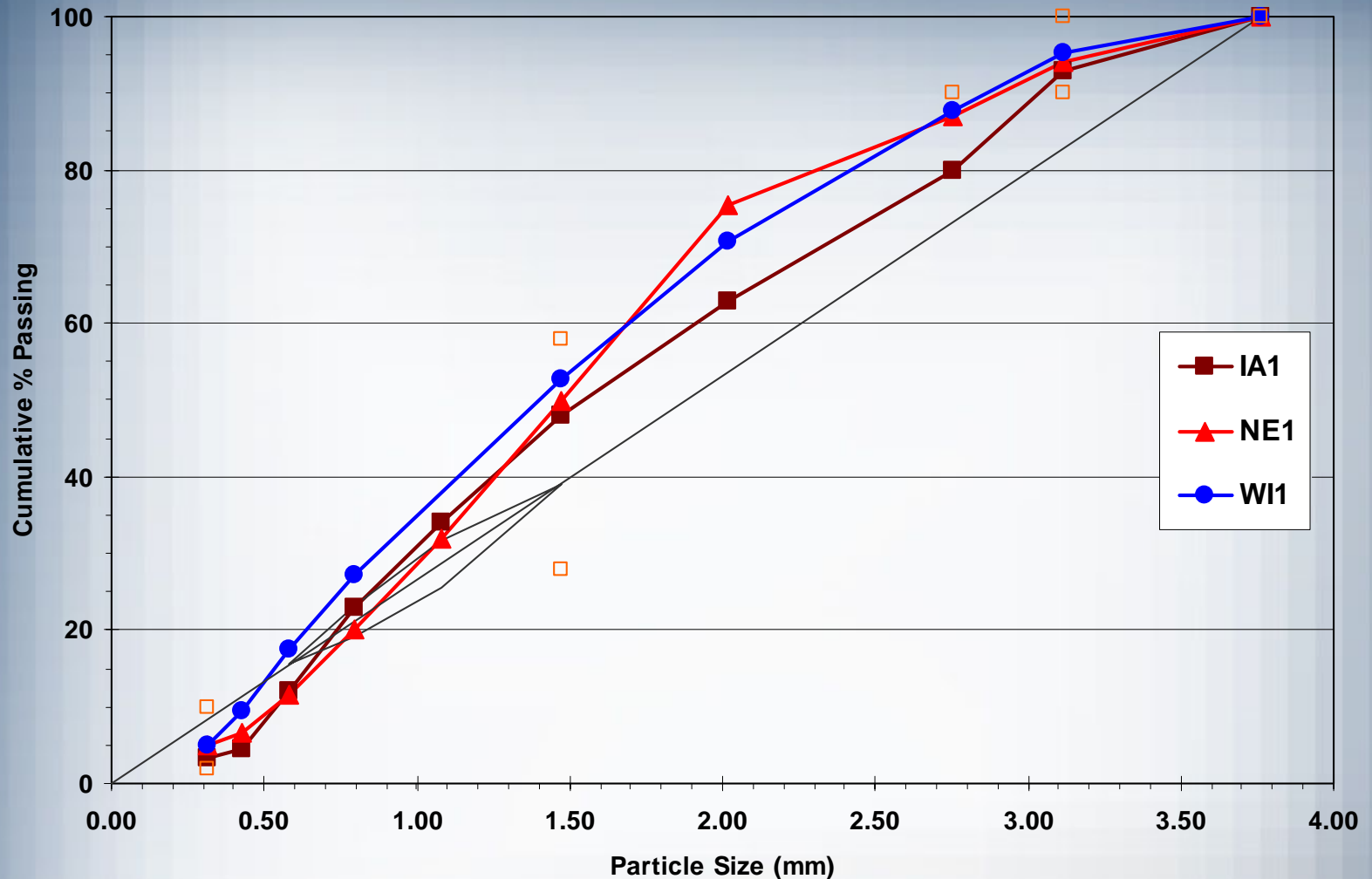
Aggregate Gradation -- 19 mm NMAS

Mix Details

| Mix ID	NMAS	Binder Grade	N_{des}	ESALs
IN1	9.5 mm	PG64-22	50	<0.3
IN2	9.5 mm	PG76-22	100	10 - 30
IA1	12.5 mm	PG64-28	96	10
IA2	19.0 mm	PG76-28	109	12.9
KS1	19.0 mm	PG76-28	100	21.3
KS2	9.5 mm	PG70-28	125	40
NE1	12.5 mm	PG70-28	96	10
WI1	12.5 mm	PG58-28	75	2
WI2	19.0 mm	PG58-28	100	3 -10

Properties Studied

Aggregate

- G_{sb}
- FAA

Mix

- G_{mm}
- G_{mb}
- V_a
- *VMA*
- *VFA*

Test Results and Analysis

Some Statistical Definitions…

Normal distribution = Gaussian distribution

Total probability = 1 (100%) =
Total area under the curve

area under the curve = probability

$\alpha/2$ $\alpha/2$
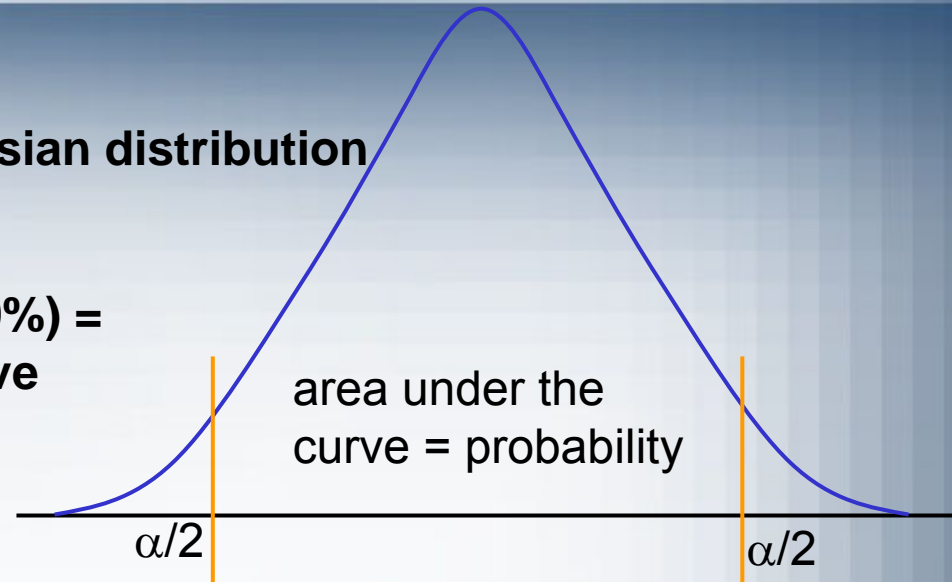
α value = significance level = criterion used to reject null hypothesis; typical values used in various research areas range from 2 to 10% (0.02 to 0.1)

p-value =probability value = probability of obtaining a statistic value (mean G_{mm}, here) that is significantly different from the value specified in the null hypothesis
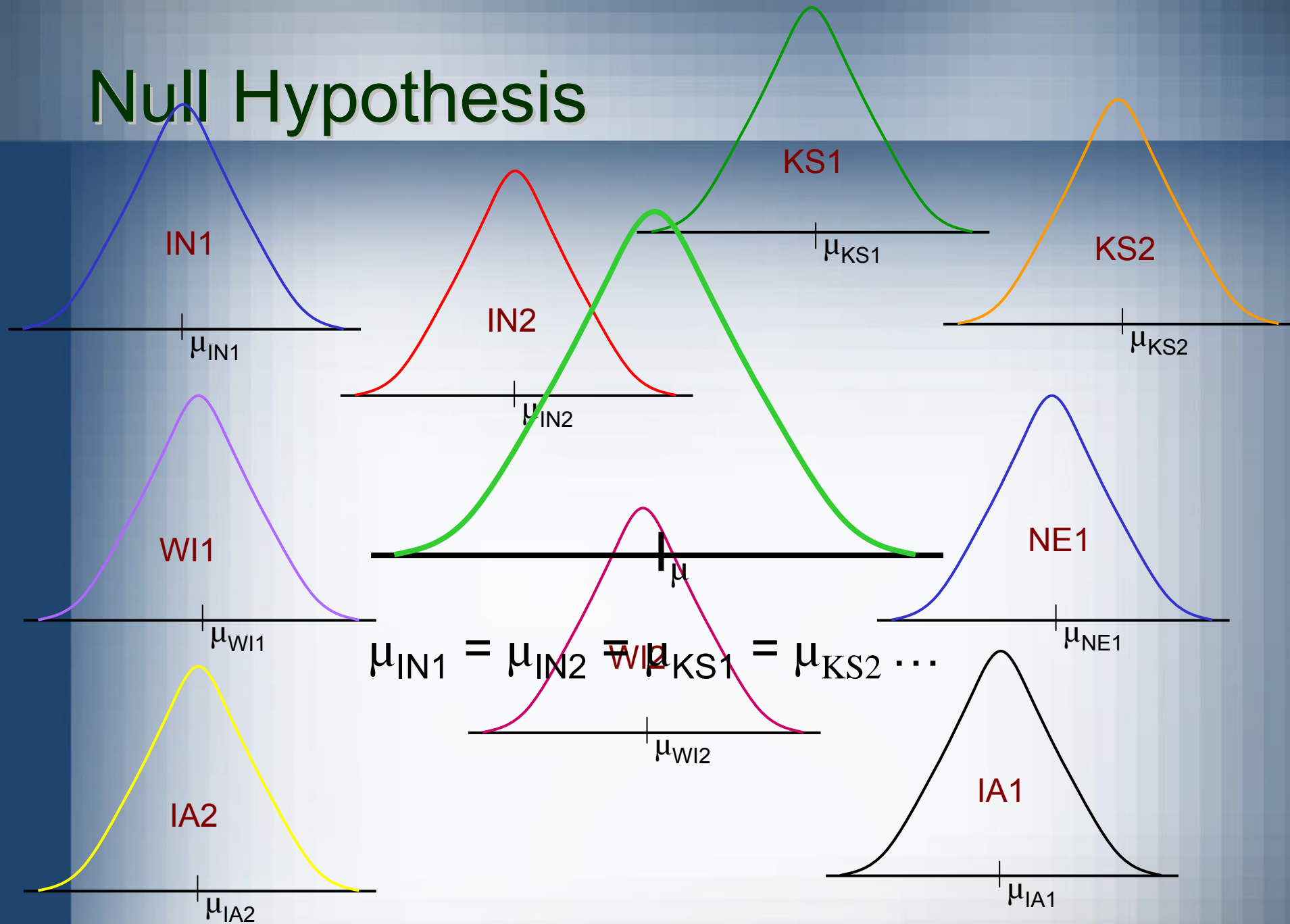
More Statistical Terms…

area under the curve = probability

$\alpha/2$ $\alpha/2$
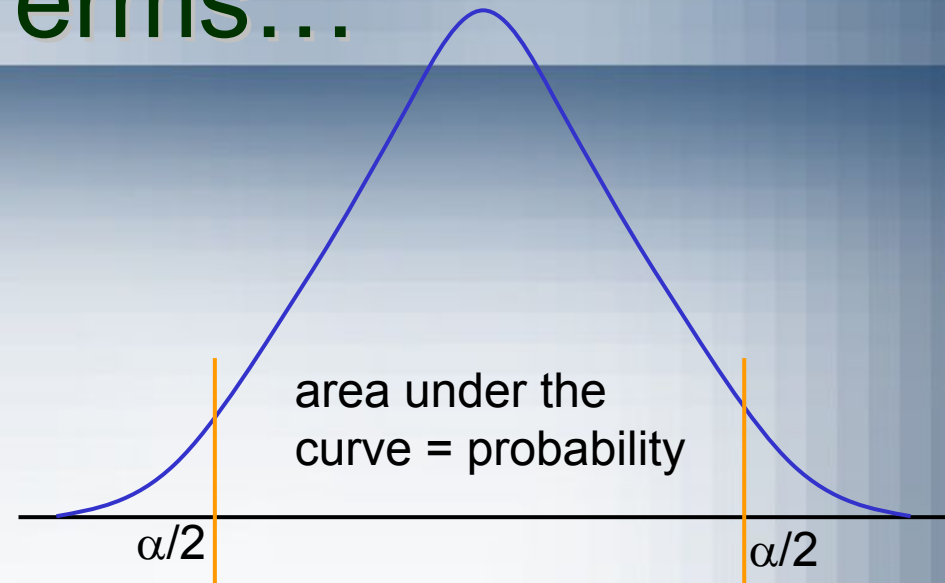
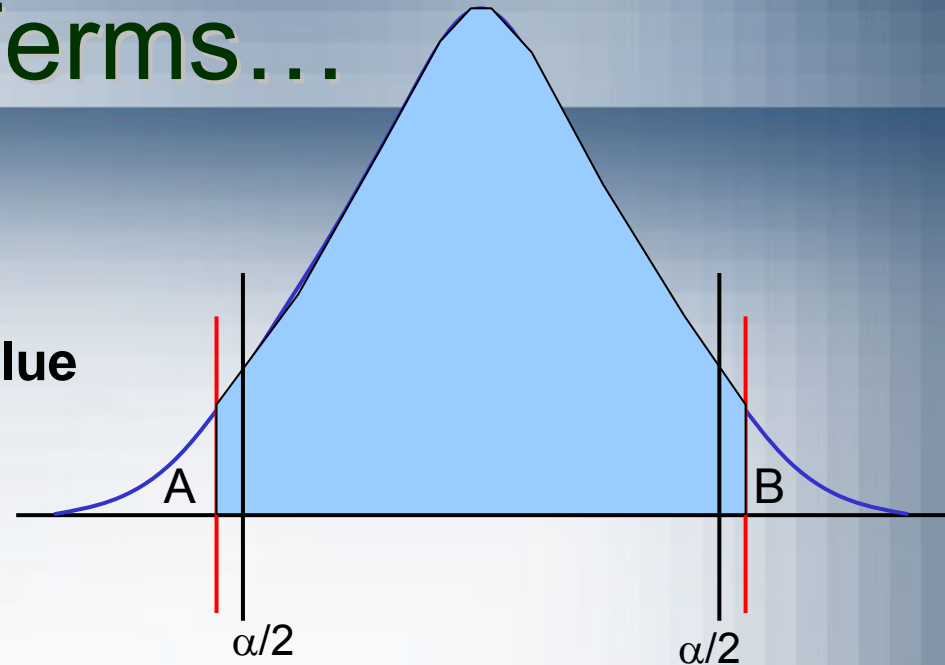null hypothesis = what we are trying to prove (for e.g., mean G_{mm} from each lab are the same)

alternative hypothesis = opposite of the null hypothesis (for e.g., mean G_{mm} from labs are different)

Null Hypothesis

$$\mu_{IN1} = \mu_{IN2} = \mu_{KS1} = \mu_{KS2} \cdots$$

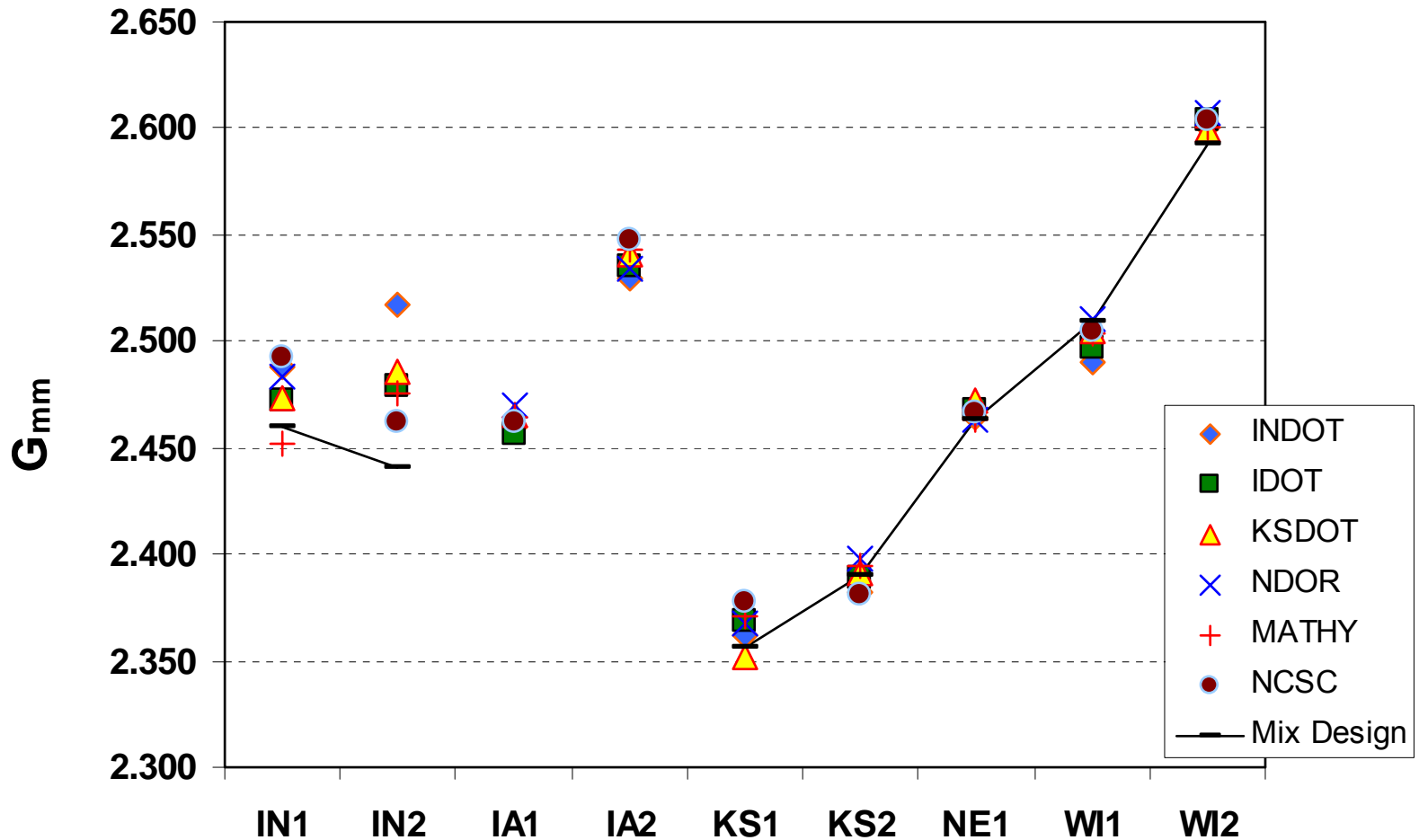# More Statistical Terms…

Area of region A + B = p-value

p-value ≥ α value $\Rightarrow$ null hypothesis is true

   i.e., mean values are equal

p-value < α value $\Rightarrow$ reject null hypothesis

   i.e., mean values are not equal; there are real differences in the test results produced

# Max. Theo. Specific Gravity, G$_{mm}$

G$_{mm}$ Statistical Inferences

- Variability between labs was not found to be significant; i.e., all participating labs produced similar results

- Variability between mixes was found to be significant; i.e., all mixes were not the same

- Only IDOT and NCSC ran replicate tests; therefore, grouping of similar/dissimilar mixes could be obtained using their data

- G$_{mm}$ of WI1, WI2, IA2, KS1 and KS2 appear to be unique

- The other mixes showed some overlap

Bulk Specific Gravity, G_{mb}, @N_{des}
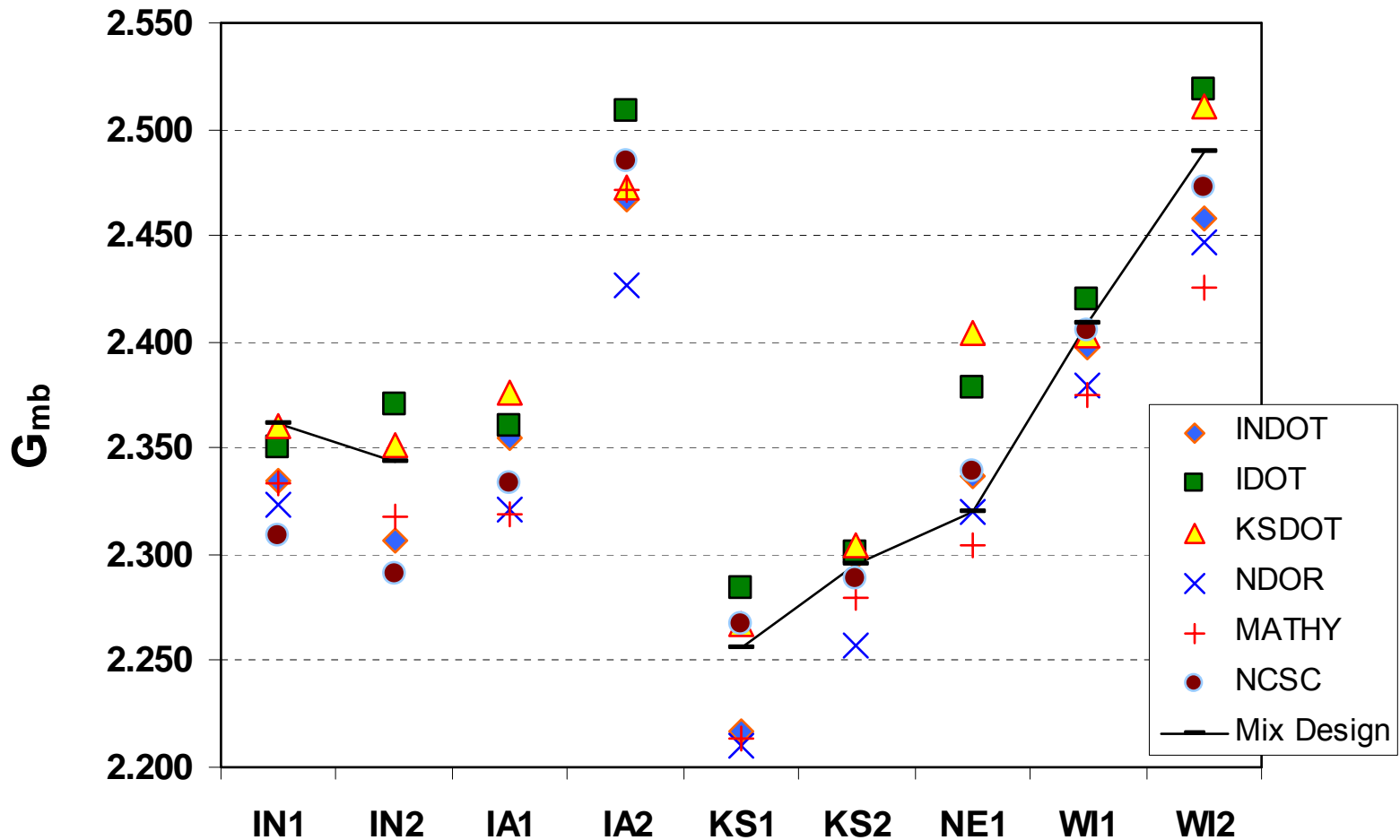
G_{mb} @N_{des} Statistical Inferences

Complete dataset…

- ❖ Variability between labs -- significant!!!
- ❖ Variability between mixes -- significant!!!

Dataset split in two for further analysis…

- ❖ Samples compacted to N_{max} (back-calculated N_{des} values)
- ❖ Samples compacted to N_{des}

G_{mb} @N_{des} Statistical Inferences

❖ Samples compacted to N_{max}

➢ KSDOT, IDOT, NDOR

- Significant variability between the three labs
- Statistical tests to determine grouping (for each mix individually) indicated that KSDOT and IDOT produced similar results.
- NDOR data appeared to be different from the other two labs!
- *However, since NDOR did not run replicate tests (lack of standard deviation data), we cannot determine whether this is a true difference due to variations in lab test procedure*
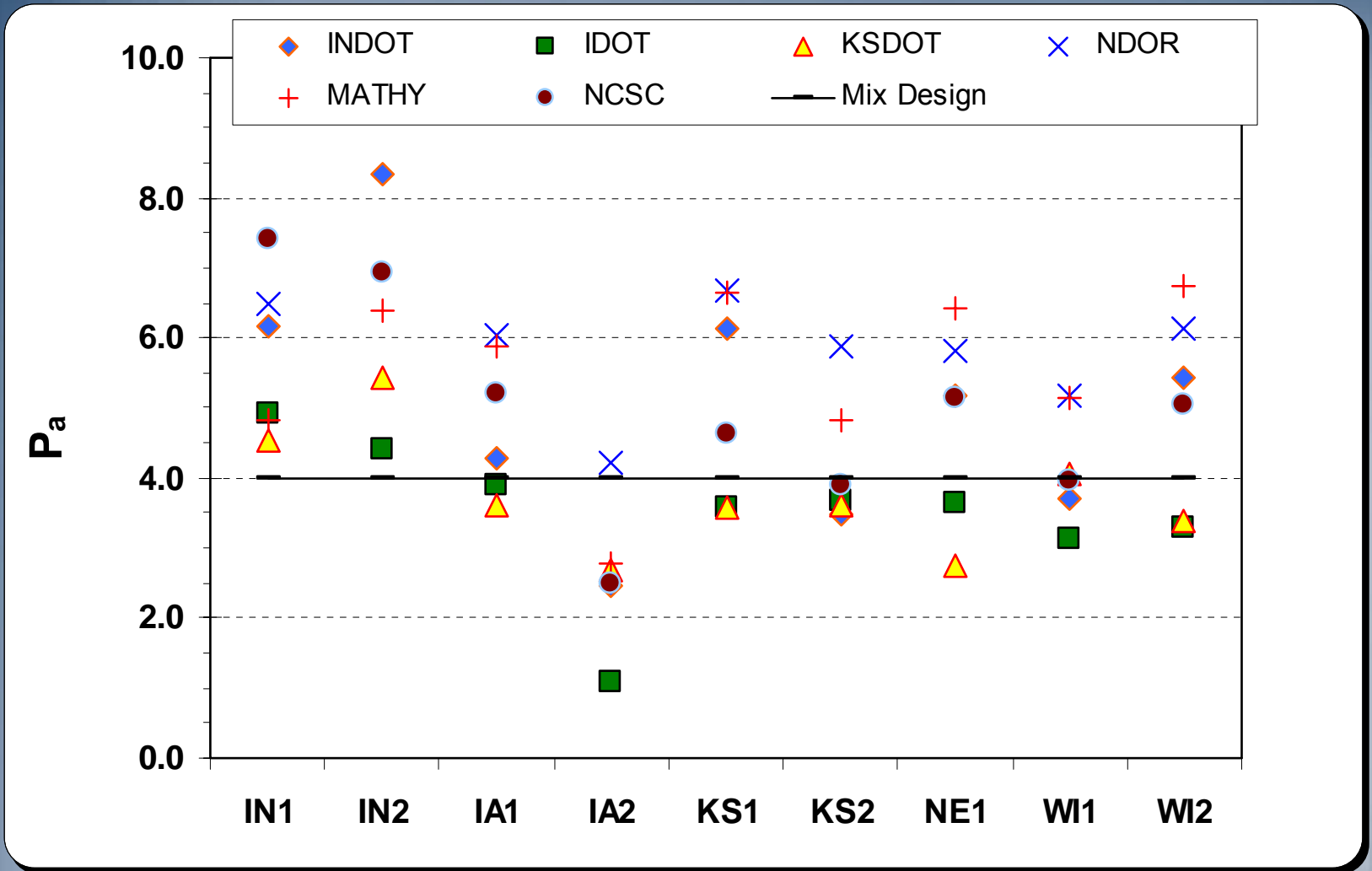
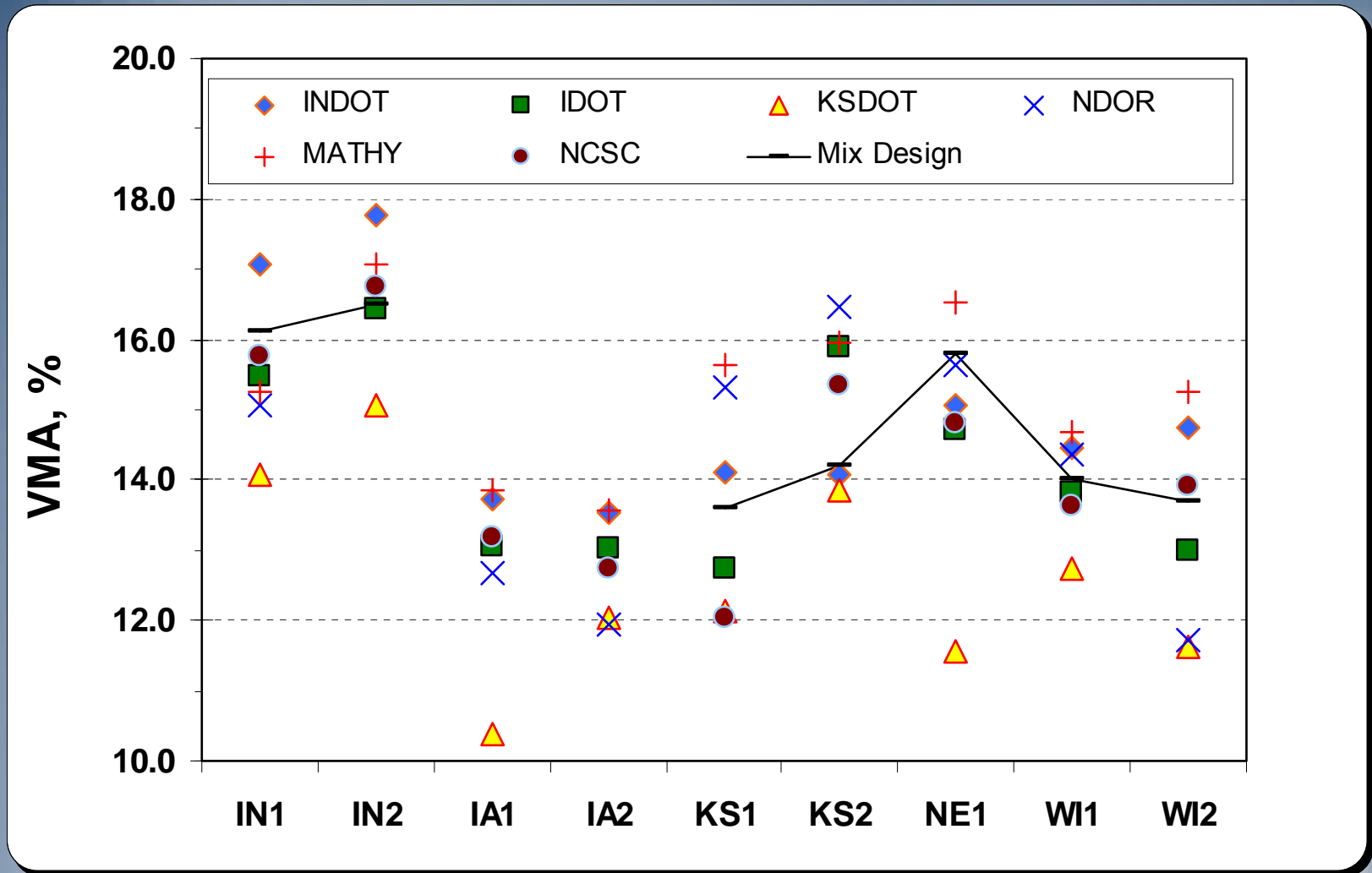# G_{mb} @N_{des} Statistical Inferences

❖ Samples compacted to N_{des}

➢ INDOT, NCSC and MATHY

- Variability between labs --- Not statistically significant; which implies that variations in lab practices did not impact the test results

- Significant variability between mixes was indicated, as in the case of G_{mm} data
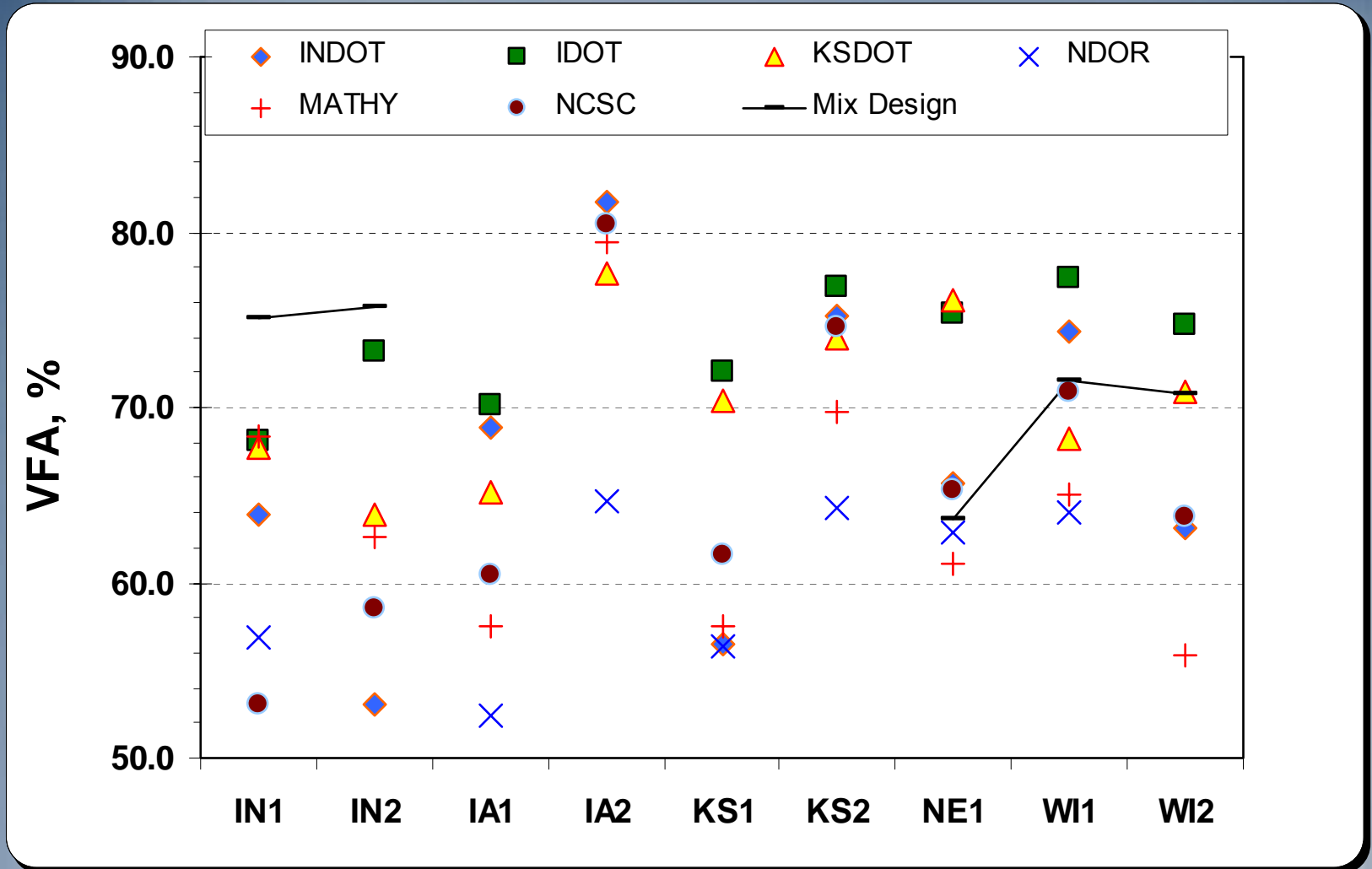
Percent Air Voids, P_a
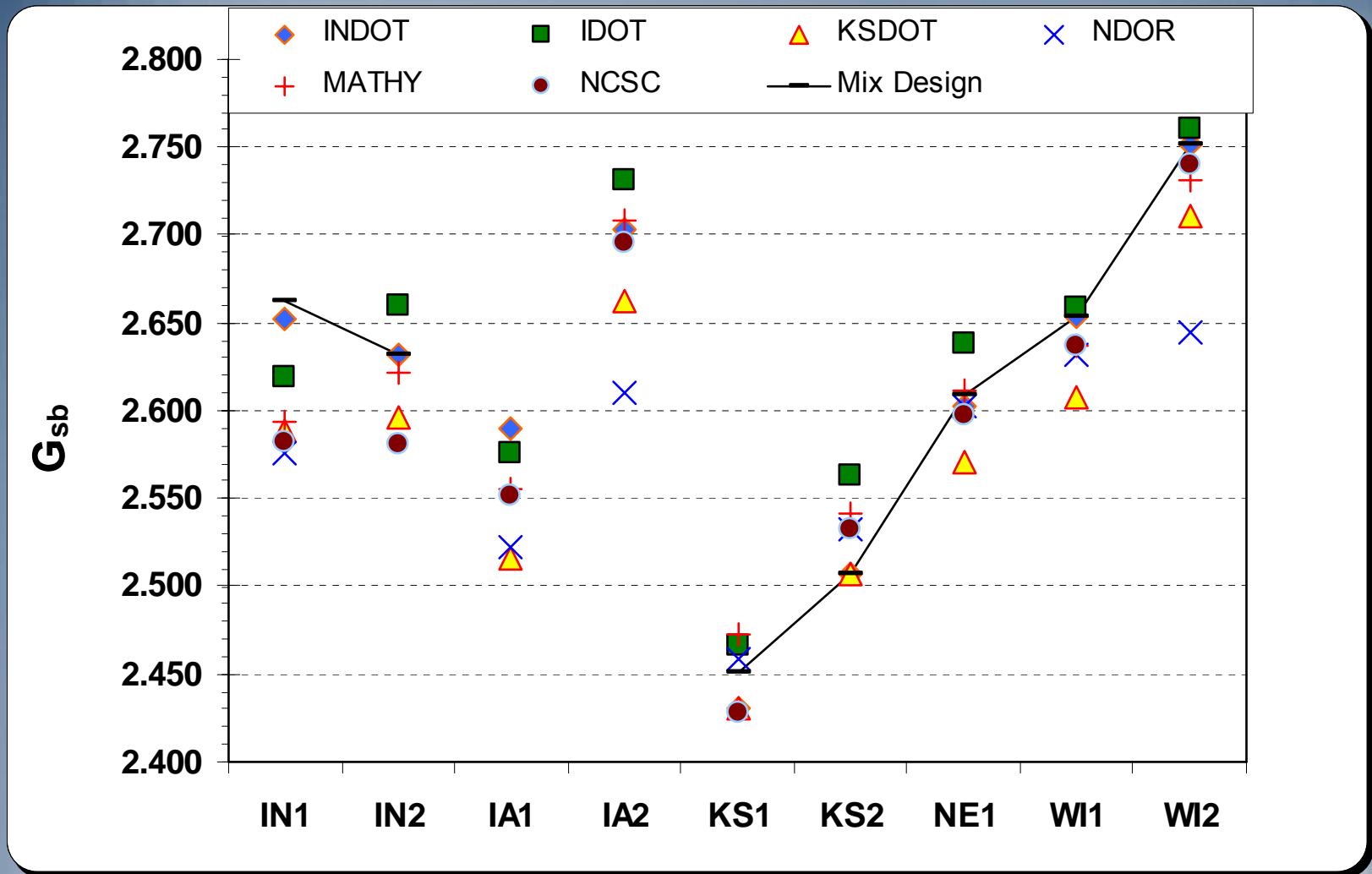
Voids in the Mineral Aggregate, VMA

Voids Filled with Aggregate, VFA

Specific Gravity of Aggregate, G_{sb}

G_{sb} Statistical Inferences

❖ Statistically significant difference between labs was observed

❖ Some labs (INDOT, KSDOT, NCSC) ran tests on individual aggregate and reported calculated G_{sb}, while other labs (IDOT, NDOR, MATHY) ran tests on blended aggregate

G$_{sb}$ Statistical Inferences

❖ Two-factor ANOVA re-run on split datasets

❖ Results indicate…

➢ Significant variability between labs and mixes, in both cases (individual and blended)

❖ Stat. tests to determine grouping was not conducted due to lack of replicate data (except for a couple of labs)

Fine Aggregate Angularity, FAA
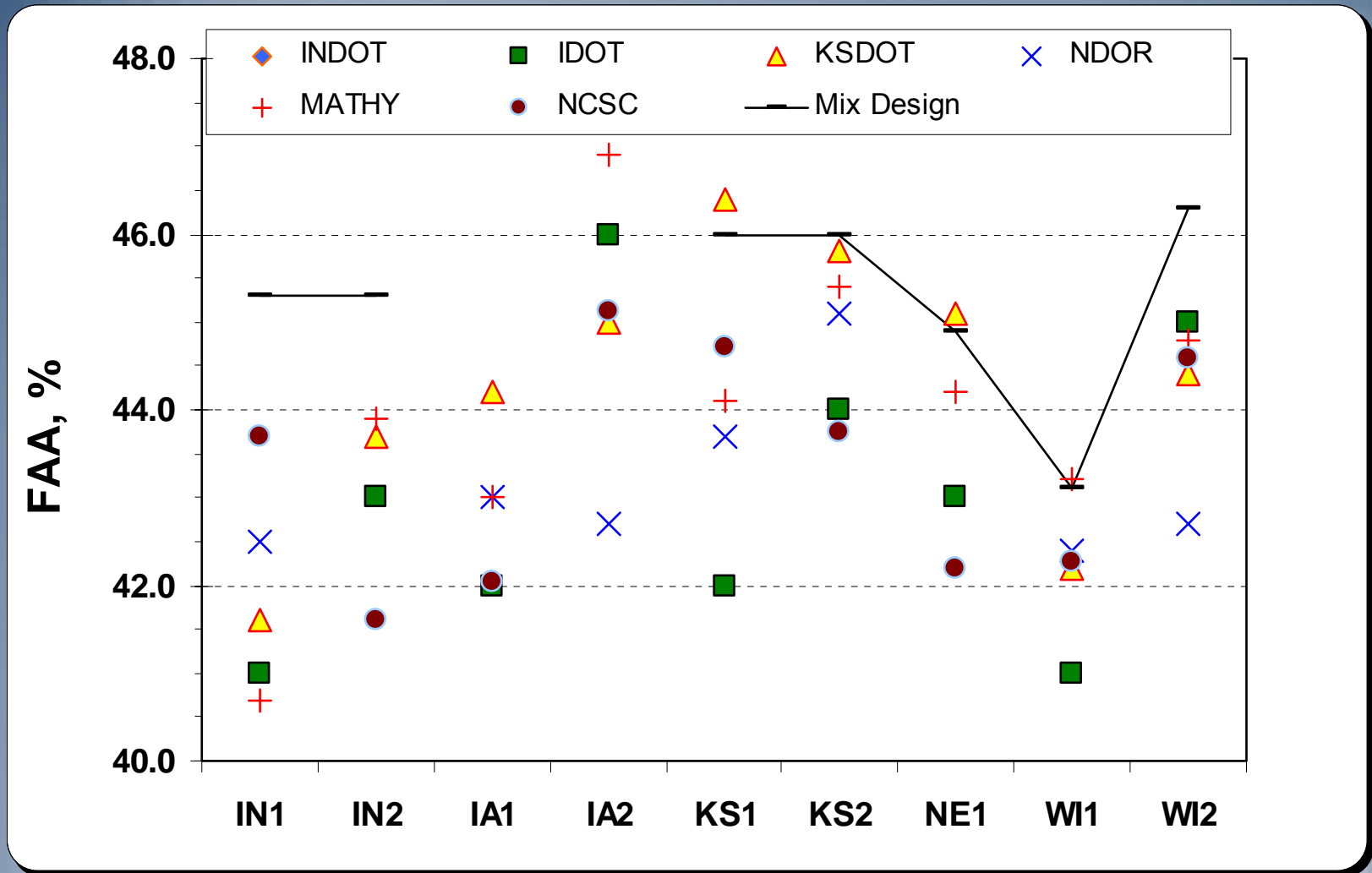
FAA Statistical Inferences

❖ No significant differences between labs

❖ Significant differences between mixes

❖ Statistical grouping tests were not run due to lack of replicate data

Summary

❖ Variability between mixes was found to be significant in all cases; expected since aggregate and binder combinations used in the nine mixes were not the same

❖ G_{mm} data -- No variability between labs

❖ FAA data -- No variability between labs

Summary

❖ G_{mb} data -- Significant variability between labs was indicated when data from N_{des} and N_{max} samples were treated as one whole dataset

❖ N_{des} dataset -- No variability between labs

❖ N_{max} dataset -- No variability between labs, except NDOR; but is this a real/true difference?

Summary

❖ G_{sb} data -- Significant variability between labs was indicated when data from all labs were used (blended and individual)

❖ Also true when treated as two separate datasets. But which labs are similar/dissimilar??? Unknown due to lack of replicate data.

❖ FAA -- No significant differences between labs

Further Study?

❖ Need more replicates!!!! Suggest at least three trials for each parameter being studied to find real differences between labs

➢ G_{mb}? G_{sb}?

❖ Test standardization?