MULTIPLE COMPARISONS IN QOE ANALYSIS

Kjell Brunnström

Mittuniversitetet
MID SWEDEN UNIVERSITY

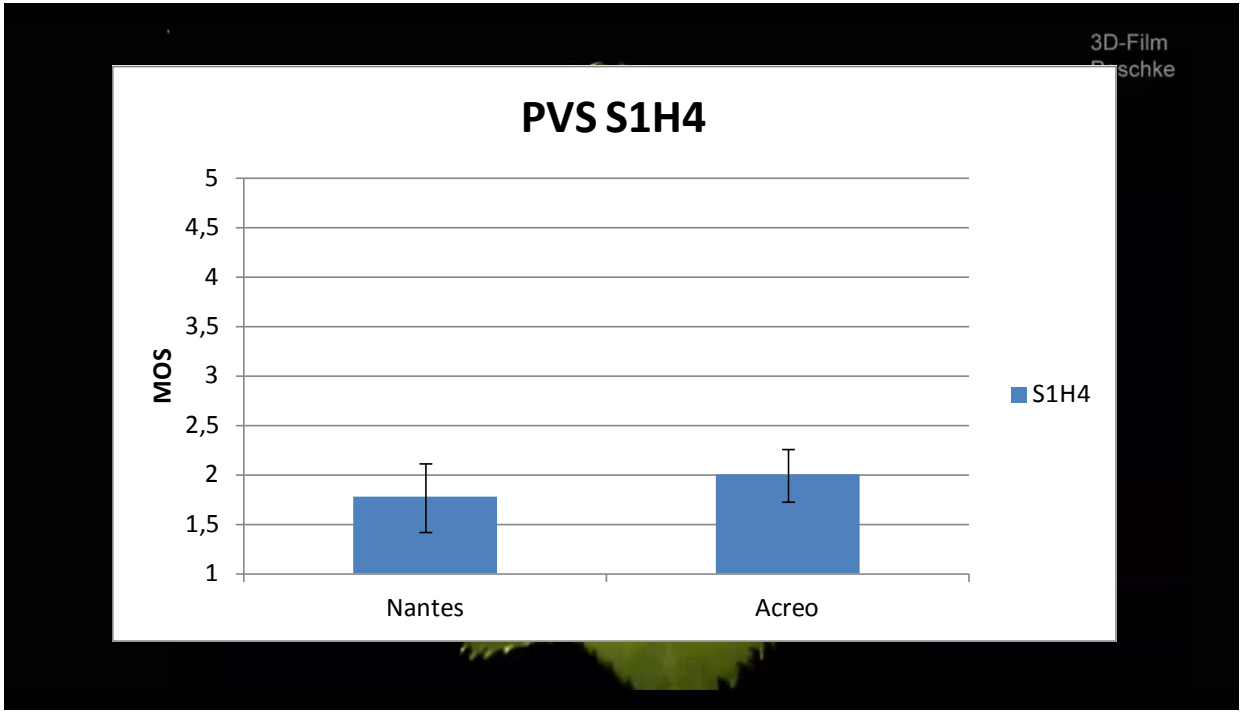
SWEDISH ICT ACREO

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

- Data is collected, then what?
- Mean and standard deviation is usually easily computed
- Common question – are two means the same or different?

EXAMPLE

- Example: One video with an error. A number of people in France and an equal number in Sweden rates the quality on a five graded scale



STATISTICAL TEST OR HYPOTHESIS TEST

- Null hypothesis (H_0): One possible arrangement
  - H_0: $\mu_1 = \mu_2$ (in this case)
- Alternative hypothesis (H_1): all other arrangement
  - H_1: $\mu_1 \neq \mu_2$ (in this case)
- Student T-test
  - If $t_{obs} \geq t_{critical}$ reject the null hypothesis
  - For example (unequal sample sizes, unequal variance)

$$t_{obs} = \frac{\mu_1 - \mu_2}{s_{diff}}; s_{diff} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; df = (n_1 - 1) + (n_2 - 1)$$

- Our example:
  - $\mu_1 = 1.77$; $\mu_2 = 2.00$; $s_1 = 0.869$; $s_2 = 0.667$; $n_1 = 22$; $n_2 = 19$;
  - t = -0.927; df = 39 + Table => t(40) = 2.02 at 0.05 significance level
  - Abs(t) < t(40)
  - Statistics packages gives p = 0.359

MORE THAN TWO MEAN?

- What if there are more than two means?
- Are the mean the same?
- If we test all the pairs with a t-test, then we know they are the same?
- Unfortunately, there is an increased risk of type-I errors
  - Reject the null hypothesis, although it is true
  - At each pairwise test there is small risk
  - Eg significance level $\alpha = 0.05$ gives 5% risk one comparison and about $n*\alpha$ for n comparisons
- Thus we need to handle this

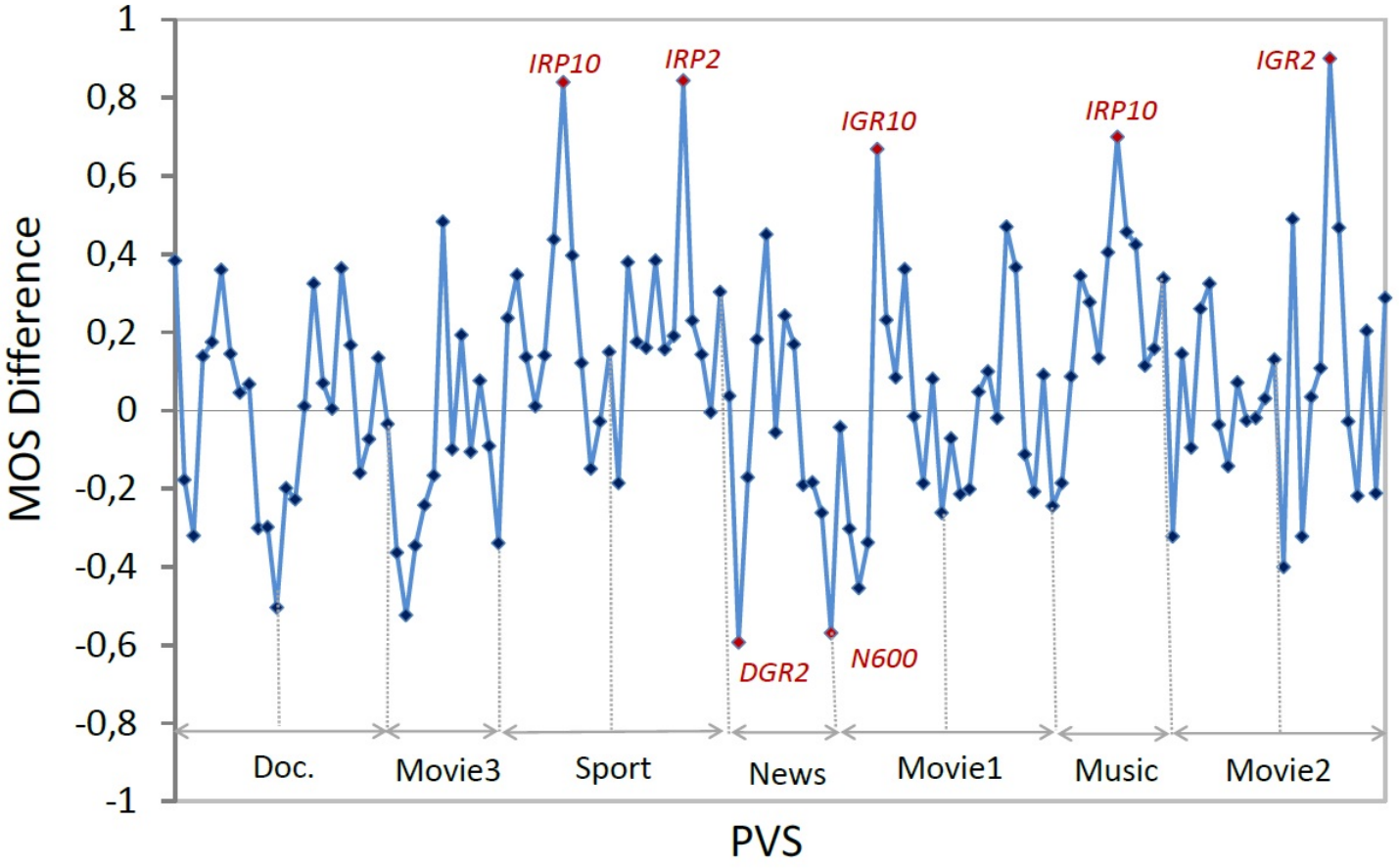
BONFERRONI CORRECTION METHOD

- the signicance level (α) should be divided by the number of comparisons N
- α_comp = α_total/N
- 10 different mean => 10*9/2 comparison 45
- α_comp =0.05/45 = 0.0011

EXAMPLE

- Compare exp 1 and exp 2
  - Same PVSs used in both exp
  - N_PVS = 100
  - Interesting comparison
    - PVS_exp1 (i) with PVS_exp2 (i)
  - If pre-planned gives 100 comparisons
  - If post-hoc gives 100*99/2 = 4950

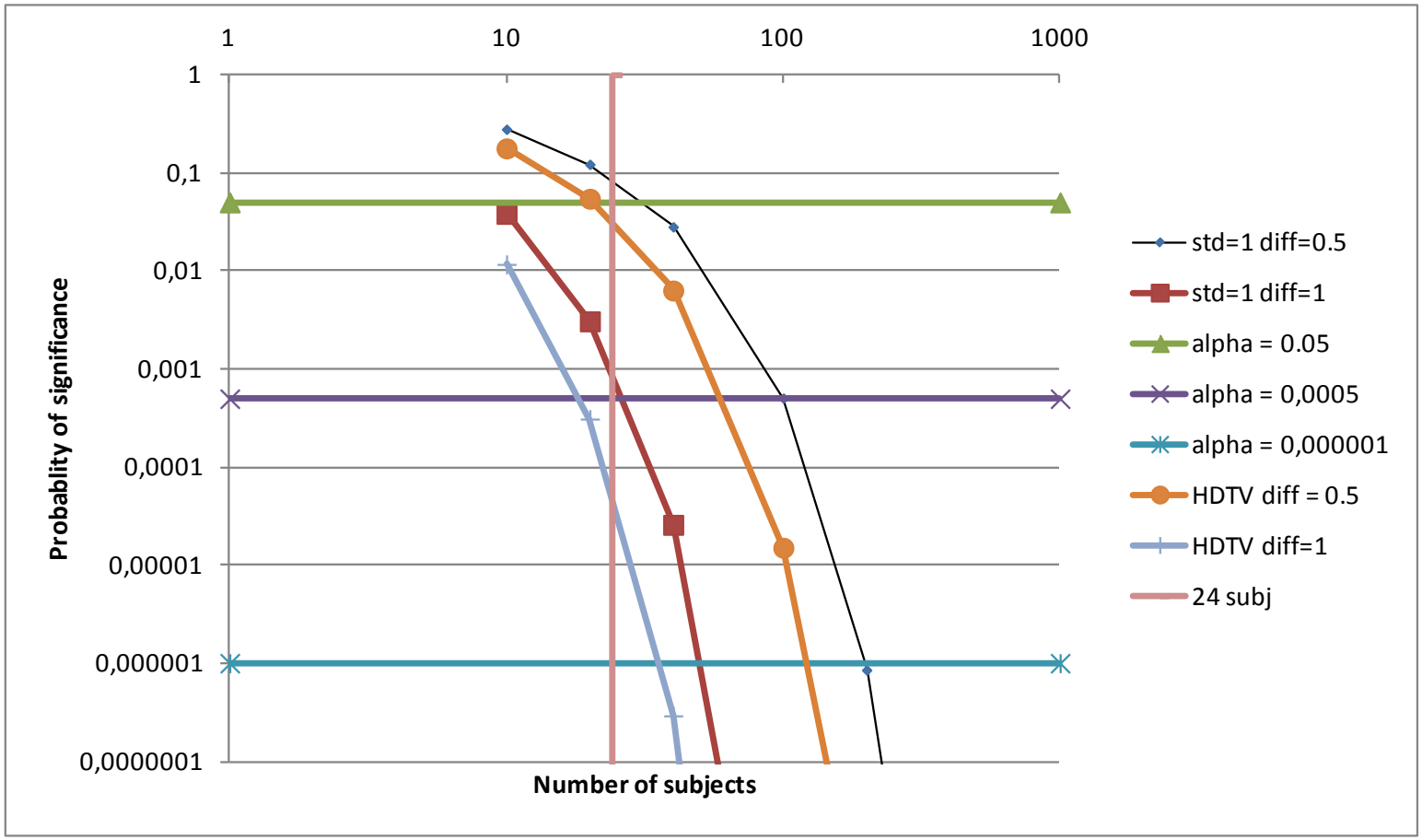
RECENT RESULT



Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

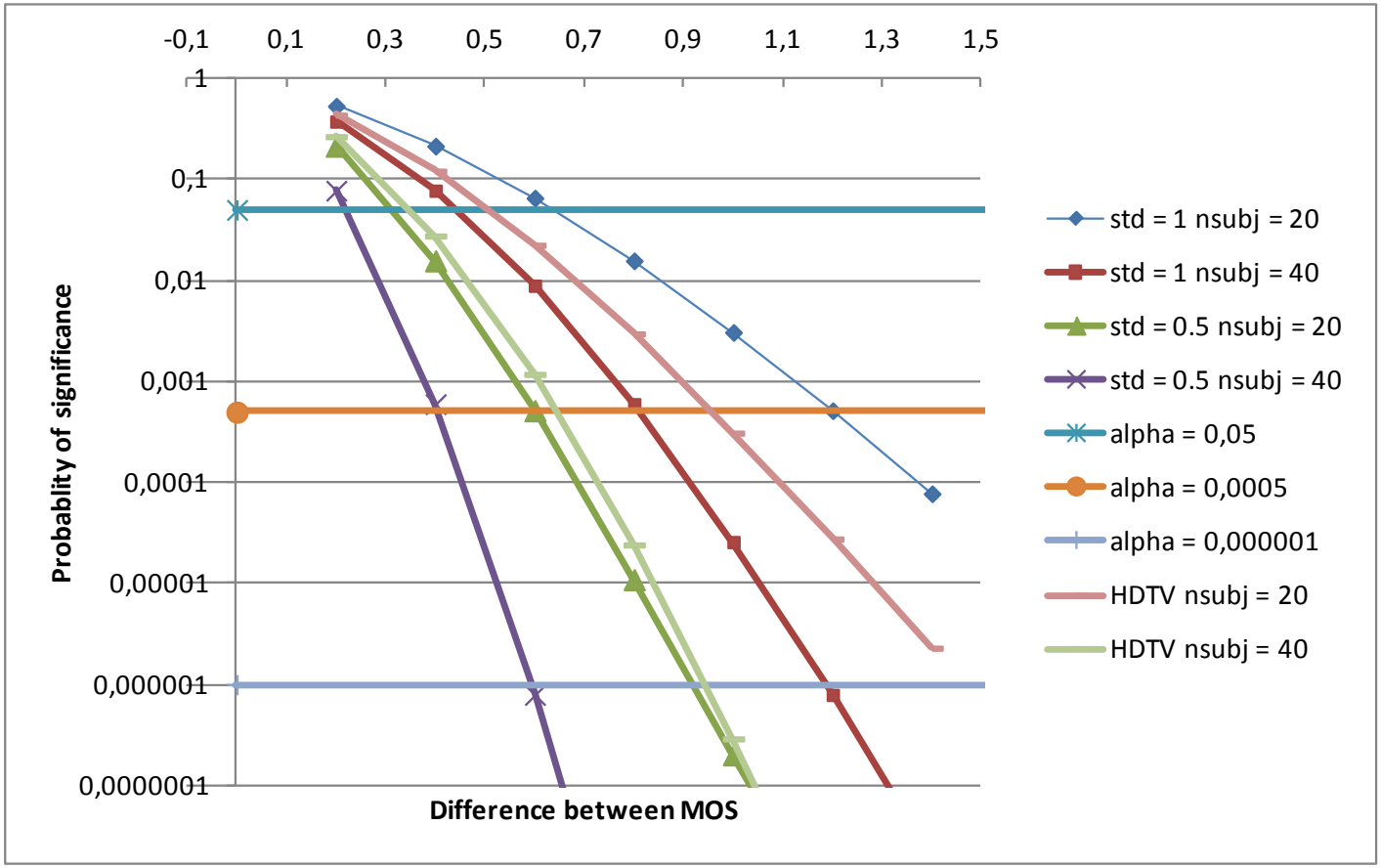
INFLUENCE OF THE NUMBER OF SUBJECTS



Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

INFLUENCE OF DIFFERENCE



Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

CORRELATION SIGNIFICANCE TESTING

- Fisher z-statistics
- Compare with two-tailed Student statistics
- If $Z_N \geq t_{critical}$ reject the null hypothesis

$$Z_N = \frac{z1 - z2 - \mu_{(z1-z2)}}{\sigma_{(z1-z2)}}$$

$$\mu_{(z1-z2)} = 0$$

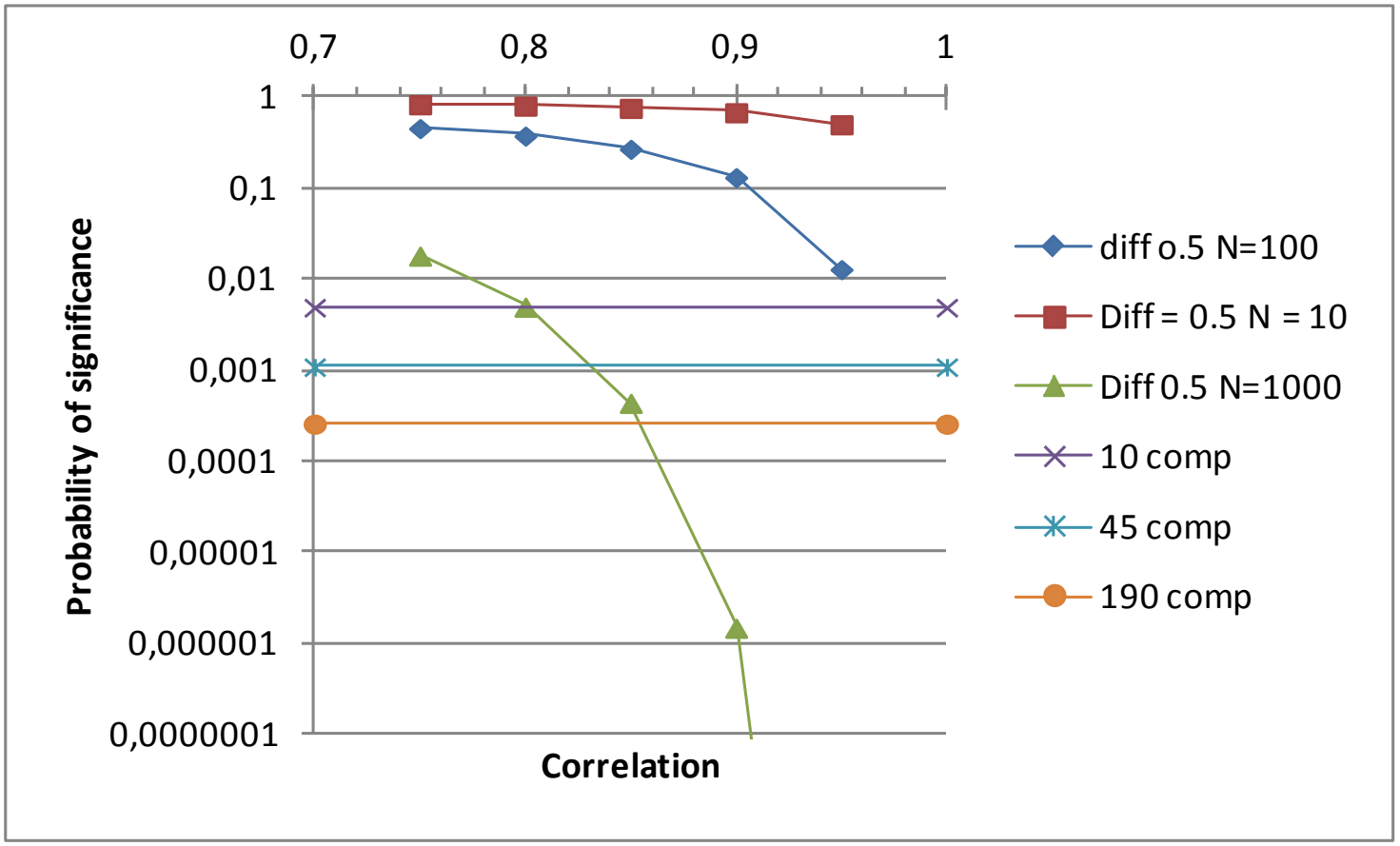
$$\sigma_{(z1-z2)} = \sqrt{\sigma_{z1}^2 + \sigma_{z2}^2}$$

$$\sigma_z = \sqrt{\frac{1}{N-3}}$$

Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

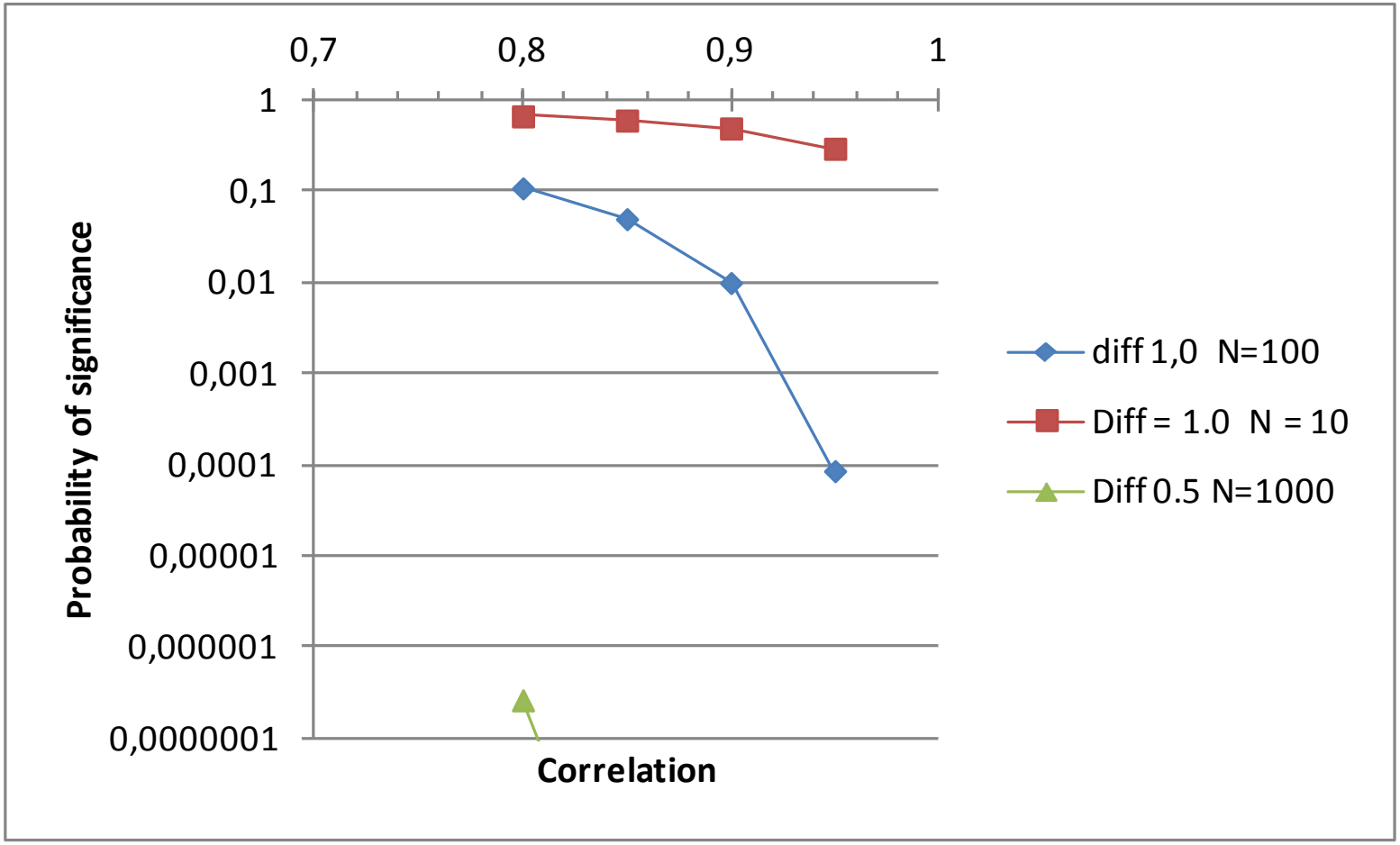
INFLUENCE ON CORRELATION



Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

INFLUENCE ON CORRELATION



Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

INFLUENCE OF RECENT TEST

TABLE II. TEST OF SIGNIFICANT DIFFERENCE FOR SROCC.

	PSNR	SSIM	MS-SSIM	MARZ	VAR	VQM	VQM-VFD	PEVQ	V-BLIINDS
PSNR			-		+		-		+
SSIM				+	+			+	+
MS-SSIM	+			+	+			+	+
MARZ		-	-		-	-	-		
VAR	-	-	-	+		-	-	+	
VQM				+	+			+	+
VQM-VFD	+			+	+			+	+
PEVQ		-	-		-	-	-		
V-BLIINDS	-	-	-			-	-		

Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

INFLUENCE OF RECENT TEST

TABLE II. TEST OF SIGNIFICANT DIFFERENCE FOR SROCC.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PSNR			-		-		+
SSIM						+	+
MS-SSIM	+					+	+
VQM						+	+
VQM-VFD	+					+	+
PEVQ		-	-	-	-		
V-BLIINDS	-	-	-	-	-		

Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO

DISCUSSION

- Not correcting for Type-I error may see effect that are not there.
- Correcting will lower efficeincy
- What can be done:
  - More test subjects may be needed
  - Reduce variance
  - Plan comparisons ahead

WWW.ACREO.SE

Mittuniversitetet
MID SWEDEN UNIVERSITY

SWEDISH ICT ACREO