Experiment 1

1. Absolute to Pure Coordinates

given

$Y = \{5, 3, 2\}$

calculation

$$mean(Y) = \frac{10}{3}$$

$$sd(y) = \sqrt{\frac{\sum_{i=1}^{3}\left(y_i - mean(Y)\right)^2}{n-1}} = \sqrt{\frac{\left(5-\frac{10}{3}\right)^2 + \left(3-\frac{10}{3}\right)^2 + \left(2-\frac{10}{3}\right)^2}{2}} \approx 1.527525$$

$$\frac{5-\frac{10}{3}}{1.527525} \approx 1.09109$$

$$\frac{3-\frac{10}{3}}{1.527525} \approx -0.2182179$$

$$\frac{2-\frac{10}{3}}{1.527525} \approx -0.8728717$$

Hint: R-Console was used for calculation.

2. Mean Distance

given

$Y^a = \{2, 3, 5\}$

$Y^b = \{3, 1, 2\}$

calculation

$$dist_{avg}^{y}\left(P^a, P^b\right) = \frac{1}{n} \cdot \sum_{i=1}^{n}\left|y_i^a - y_i^b\right| = \frac{|2-3| + |3-1| + |5-2|}{3} = \frac{1+2+3}{3} = 2$$

3. ROC-Plots

given

$ranks$	1	1	1	2	3	5	6	1	1	4
$scores$	0.04	0.08	0.11	0.08	0.22	0.31	0.12	0.15	0.04	0.2
$scores_{sort}$	0.04	0.04	0.08	0.08	0.11	0.12	0.15	0.2	0.22	0.31

$t = 3$ (num of top scores)

Hint: $scores$ is a multiset!

calculation

$$recall(0.04) = \frac{|\emptyset \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = 0$$

$$fallout(0.04) = \frac{|\emptyset \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = 0$$

~ (same and irrelevant since plotted at the same point)

$$recall(0.08) = \frac{|\{0.04, 0.04\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{2}{7}$$

$$fallout(0.08) = \frac{|\{0.04, 0.04\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = 0$$

~

$$recall(0.11) = \frac{|\{0.04, 0.04, 0.08, 0.08\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{4}{7}$$

$$fallout(0.11) = \frac{|\{0.04, 0.04, 0.08, 0.08\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = 0$$

$$recall(0.12) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{5}{7}$$

$$fallout(0.12) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = 0$$

$$recall(0.15) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{5}{7}$$

$$fallout(0.15) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = \frac{1}{3}$$

$$recall(0.2) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{6}{7}$$

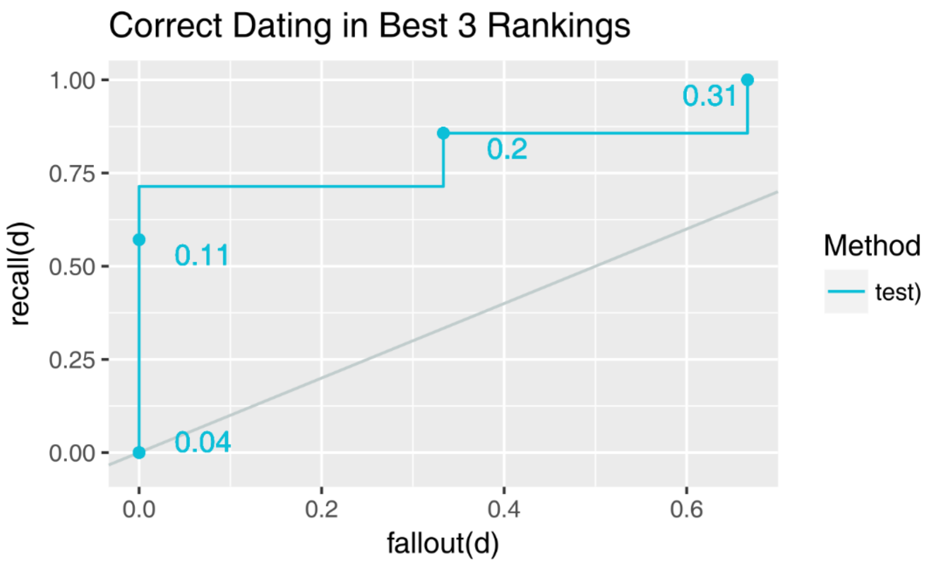
$$fallout(0.2) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = \frac{1}{3}$$

$$recall(0.22) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15, 0.2\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = \frac{6}{7}$$

$$fallout(0.22) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15, 0.2\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = \frac{2}{3}$$

$$recall(0.31) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15, 0.2, 0.22\} \cap \{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|}{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.15, 0.22\}|} = 1$$

$$fallout(0.31) = \frac{|\{0.04, 0.04, 0.08, 0.08, 0.11, 0.12, 0.15, 0.2, 0.22\} \cap \{0.31, 0.12, 0.2\}|}{|\{0.31, 0.12, 0.2\}|} = \frac{2}{3}$$

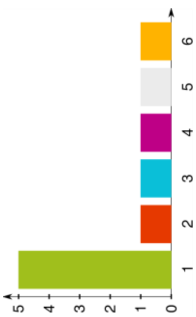


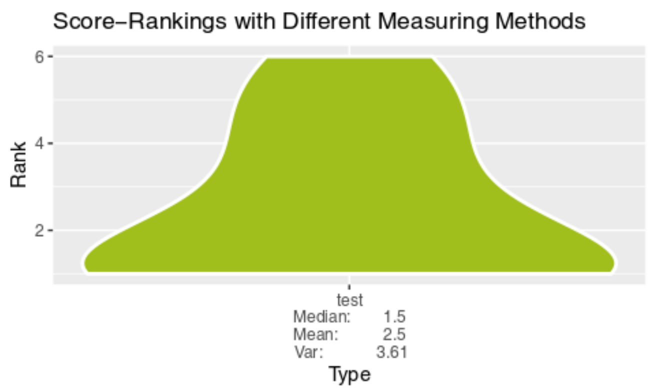
The plot is correct!
As said in the Thesis it is drawn for $d_S < d$, so there is no fallout or recall value at the end, since it is not necessary if the amount of values is high enough.
And this is a common behavior due to the definition:
http://sachsmc.github.io/plotROC/ (here: it is looked only on values bigger c)

4. Violine-Plots

given

$ranks$	1	1	1	2	3	5	6	1	1	4





calculation

Median:

1	1	1	1	$\boxed{1}$	$\boxed{2}$	3	4	5	6

$$\Rightarrow \tilde{r} = \frac{1}{2} \cdot \left(r_{\frac{u}{2}} + r_{\frac{u}{2}+1} \right) = \frac{1}{2} \cdot (1 + 2) = 1.5$$

Mean:

$$\frac{1+\cdots+1+2+\cdots+6}{10}=\frac{25}{10}=2.5$$

Var:

$$\frac{\sum_{i=1}^{10}(v_i-mean)^2}{n-1}$$

$$=\frac{(1-2.5)^2}{9}\cdot 5+\frac{(2-2.5)^2}{9}+\cdots+\frac{(6-2.5)^2}{9}$$

$$=1.25+\frac{1}{36}+\frac{1}{36}+0.25+\frac{25}{36}+\frac{49}{36}=\frac{47}{18}\approx 3.61$$

5. ROC-Plots & Violine Plots Ranking Function

given

$scores_{correct}$	0.02	0.02	0.04	0.03		
$scores_{sample\ 1}$	0.05	0.03	0.04	0.02	0.02	0.12
$scores_{sample\ 2}$	0.15	0.02	0.03	0.05	0.04	0.08
$scores_{sample\ 3}$	0.07	0.18	0.04	0.02	0.05	0.12
$scores_{sample\ 4}$	0.06	0.03	0.04	0.17	0.02	0.12

calculation

$scores_{sample\ 1}^{sort}$	0.02	0.02	0.03	0.04	0.05	0.12
$scores_{sample\ 2}^{sort}$	0.02	0.03	0.04	0.05	0.08	0.15
$scores_{sample\ 3}^{sort}$	0.02	0.04	0.05	0.07	0.12	0.18
$scores_{sample\ 4}^{sort}$	0.02	0.03	0.04	0.06	0.12	0.17
$ranks(scores_{correct})$	2	1	2	2		

6. Fishing Correct Score for a Curve

given

Test-Curve 1:	*Test-Curve 2:*	*Test-Scores 1:*	*Test-Scores 2:*
"year","GD"	"year","GD"	*"year","score"*	*"year","score"*
1964,0.25	1973,0.25	*"1960",0.83*	*"1960",0.83*
1964,0.253	1973,0.253	*"1961",0.85*	*"1961",0.85*
1964,0.26	1973,0.26	*"1962",0.88*	*"1962",0.88*
1964,0.268	1973,0.268	*"1963",1.04*	*"1963",0.83*
1964,0.27	1973,0.27	*"1964",0.96*	*"1964",0.85*
1965,0.271	1974,0.271	*"1965",1.08*	*"1965",0.88*
1965,0.259	1974,0.259	*"1966",1.06*	*"1966",1.04*
1965,0.253	1974,0.253	*"1967",1.07*	*"1967",0.96*
1965,0.27	1974,0.27	*"1968",1.08*	*"1968",1.08*
1965,0.273	1974,0.273	*"1969",1.05*	*"1969",1.06*
		"1970",1.07	*"1970",1.07*
		"1971",1.08	*"1971",1.08*
		"1972",1.04	*"1972",1.04*
		"1973",1.02	*"1973",1.05*

Test-Curve 3:	"year","score"
"year","GD"	"1960",0.82
1960,0.35	"1961",0.85
1960,0.351	"1962",0.88
1960,0.36	"1963",0.83
1960,0.268	"1964",0.85
1960,0.28	"1965",0.88
1961,0.271	"1966",1.04
1961,0.279	"1967",0.96
1961,0.253	"1968",1.08
1961,0.27	"1969",1.06
1961,0.273	"1970",1.07
	"1971",1.08
	"1972",1.04
	"1973",1.05

calculation

Test-Curve 1:	0.96
Test-Curve 2:	1.05
Test-Curve 3:	0.82