

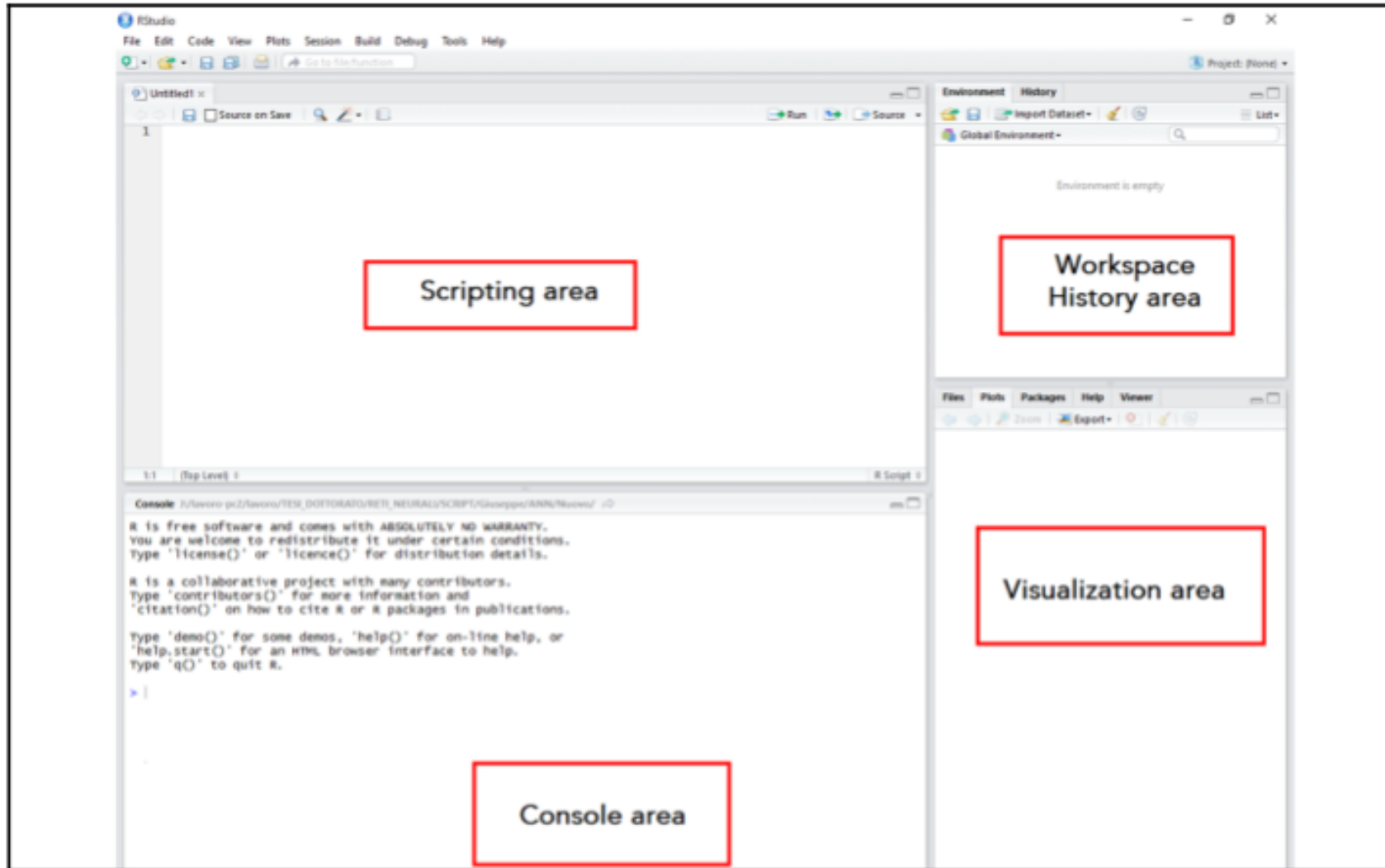
Statistical analysis for NAPPA

Yunro Chung, Ph.D.
Assistant Professor
Arizona State University

R

R/R Studio

<https://posit.co/download/rstudio-desktop/>



- **Scripting** area: We can open, create, and write out scripts
- **Console** area is a zone where commands are executed
- Workspace **History** area: you can find a list of full objects created
- **Visualization** area: You can load packages and view graphs/charts.

ROC Curve

- Cancer biomarker study:
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92
 - Assume that higher value indicate cancer

Q. Plot ROC Curve.

- We need to compute sensitivity and specificity at all possible threshold values.
- What will be the possible threshold values?
 - $(-\infty, 0.94, 1.11, 1.41, 1.53, 1.66, 1.92, 1.99, 2.40, +\infty)$
 - Often, midpoint values are used, e.g. $(0.94+1.11)/2=1.025$, $(1.11+1.41)/2=1.26$, etc.

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th = -\infty$:

	Case	Normal	Total
Value $\geq th$	4	5	9
Value $< th$	0	0	0
Total	4	5	9

- $st = 4/4 = 1$
- $sp = 0/5 = 0$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th=1.025$:

	Case	Normal	Total
Value $\geq th$	4	4	9
Value $< th$	0	1	0
Total	4	5	9

- $st=4/4=1$
- $sp=1/5=0.20$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th=1.260$:

	Case	Normal	Total
Value $\geq th$	4	3	9
Value $< th$	0	2	0
Total	4	5	9

- $st=4/4=1$
- $sp=2/5=0.40$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th=1.470$:

	Case	Normal	Total
Value \geq th	3	3	9
Value<th	1	2	0
Total	4	5	9

- $st=3/4=0.75$
- $sp=2/5=0.40$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

- At $th=1.595$:

	Case	Normal	Total
Value $\geq th$	3	2	9
Value $< th$	1	3	0
Total	4	5	9

- $st=3/4=0.75$
- $sp=3/5=0.60$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th=1.790$:

	Case	Normal	Total
Value \geq th	2	1	9
Value<th	2	4	0
Total	4	5	9

- $st=2/4=0.50$
- $sp=4/5=0.80$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

- At $th=1.955$:

	Case	Normal	Total
Value \geq th	2	0	9
Value<th	2	5	0
Total	4	5	9

- $st=2/4=0.50$
- $sp=5/5=1$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

- At $th=2.195$:

	Case	Normal	Total
Value $\geq th$	1	0	9
Value $< th$	3	5	0
Total	4	5	9

- $st=1/4=0.25$
- $sp=5/5=1$

- Possible threshold: $(-\infty, 1.025, 1.260, 1.470, 1.595, 1.790, 1.955, 2.195, +\infty)$
 - Cancer tissue: 1.41, 1.66, 1.99, 2.40
 - Normal tissue: 0.94, 1.11, 1.53, 1.66, 1.92

○ At $th = +\infty$:

	Case	Normal	Total
Value $\geq th$	0	0	9
Value $< th$	4	5	0
Total	4	5	9

- $st = 0/4 = 0$
- $sp = 5/5 = 1$

- Cancer gene expression array study.

th	St (or TPF)	Sp	1-sp (or FPF)
$-\infty$	1	0	1
1.025	1	0.2	0.8
1.260	1	0.4	0.6
1.470	0.75	0.4	0.6
1.595	0.75	0.6	0.2
1.790	0.5	0.8	0.2
1.955	0.5	1	0
2.195	0.25	1	0
∞	0	1	0

