

تحلیل پوششی داده ها و تحلیل های آماری

بهروز دانشیان^۱، صادق پورجعفر، آذرنوش کافی، پریناز دهقان نژاد

گروه ریاضی دانشگاه آزاد اسلامی واحد لاهیجان

چکیده

در این تحقیق تلاش شده است تا آزمون های آماری در تحلیل پوششی داده ها برای تحلیل دقیق تر مدل های تحلیل پوششی داده ها مورد استفاده قرار گیرد. همچنین با به کار گیری تحلیل خوشه ای در تحلیل پوششی داده ها و به کار گیری آن در مورد سی و پنج سرپرستی شعبه های یک بانک تجاری عملکرد آنها را در ماه های مختلف سال بررسی کرده ایم.

کلمات کلیدی: آزمون آماری، تحلیل خوشه ای، توزیع کارایی، تحلیل پوششی داده ها، سرپرستی بانک.

۱ مقدمه

تحلیل پوششی داده ها برای ارزیابی کارایی با توجه به تحقیقات گسترده در این زمینه روشی شناخته شده است. بسیاری از مطالعات در این زمینه با قطعی فرض کردن داده ها مرز کارایی مشخصی را تعیین و بر اساس آن کارایی را می سنجند. این اندازه گیری ها به خطاهای داده ها و نمونه ها توجهی ندارند و یا در حالت کلی به ساختارهای آماری و احتمالی بی توجه هستند. اکنون معلوم شده است که توجه بیشتر به ویژگی های آماری در تحلیل پوششی داده ها می تواند در تبیین بیشتر عملکرد کمک شایانی باشد.

به عنوان نمونه بنکر (۱۹۹۳) سازگاری برآوردگرهای کارایی *DEA* را تحت مفروضات معینی را نشان داد و آزمون های آماری را معرفی کرد. *Korostelev, Simar, Tsybakov (1995a, 1995b)* همگرایی برآوردگرهای کارایی غیر پارامتری را نشان دادند. *Kneip, Park, Simar (1996)* این نتایج را در حالت کلی تعمیم دادند. *Simar, Wilson (1995)* روش *BootStrap* را برای برآورد تخمینی از اریب و فاصله اطمینان پیشنهاد کردند. هدف این مقاله استفاده از آزمون های آماری جدید برای نتایج مدل های تحلیل پوششی داده ها و همچنین استفاده از خوشه بندی برای کارایی است. نتایج کارایی ۳۵ سرپرستی یک بانک تجاری (دیواندری و همکاران) در این مطالعه مورد بررسی قرار گرفته است. ساختار این مقاله به این صورت است که در بخش اول روش تحلیل پوششی داده ها و خوشه بندی به صورت اجمالی معرفی می شود. در بخش دوم استفاده از آزمون های آماری برای نتایج کارایی حاصل از مدل های تحلیل پوششی داده ها جهت مقایسه معنی دار بودن اختلاف واریانس آنها

و ساختار احتمال توزیع کارایی خواهد بود. در بخش آخر کارایی ۳۵ سرپرستی بانک در ۱۲ ماه بررسی و نتایج حاصل از خوشه بندی و آزمون ANOVA ارائه خواهد شد. نمونه خوشه بندی برای یک ماه نیز به عنوان نمونه ارائه خواهد شد همچنین توزیع کارایی سرپرستی ها نیز در این بخش ارائه می شود. بخش آخر به نتیجه گیری از بررسی انجام شده خواهد پرداخت.

۱ تحلیل پوششی داده ها و ابزارهای تحلیل آماری

نتیجه تحقیقات چارنر کوپر و رودز^۲ برای ارائه مدلی ریاضی برای ارزیابی عملکرد واحد های تصمیم گیری متجانس که ورودی های از یک نوع را برای تولید خروجی های همگون مورد استفاده قرار می دهند بر اساس تحقیقات فارل با ارائه مدل CCR به نتیجه رسید. با این مدل تحلیل پوششی داده ها به عنوان روشی مبتنی بر برنامه ریزی خطی مسیری جدیدی را برای ارزیابی عملکرد و محاسبه کارایی گشود. با این فرض که n واحد تصمیم گیری DMU_j $j=1, \dots, n$ ، m ورودی برای تولید s خروجی استفاده می کنند کارایی واحد تصمیم گیری DMU_o ، $o=1, \dots, n$ که با صرف ورودی X_{io} ، $i=1, \dots, m$ به تولید Y_{ro} ، $r=1, \dots, s$ دست یافته است چنین است:

Min θ

$$s.t \quad \sum \lambda_j X_j + S^- = \theta X_o \quad (1)$$

$$\sum \lambda_j Y_j - S^+ = Y_o$$

$$S^-, S^+, \lambda_j \geq 0 \quad j=1, \dots, n$$

θ میزان کارایی واحد تصمیم گیری DMU_o که $o=1, \dots, n$ را به دست می دهد.

تحلیل خوشه ای یک روش قدیمی است که در آن هیچ فرضی در مورد گروه ها یا ساختمان آنها در نظر گرفته نمی شود و دسته بندی کردن بر اساس مشابهت ها یا فواصل (عدم شباهت ها) انجام می شود. در تحلیل خوشه ای متغیر ها بر اساس ضرایب همبستگی با اندازه های مشابه دسته بندی می شوند. فاصله اقلیدسی برای تعیین فاصله معمولاً ترجیح داده می شود. هنگامی که همبستگی های نمونه به عنوان اندازه های مشابهت استفاده می کنیم متغیرهایی که همبستگی های منفی بزرگی دارند را غیر متشابه و متغیرهایی که همبستگی های مثبت بزرگی دارند را مشابه تلقی می کنیم. در این حالت فاصله بین خوشه ها را به صورت کمترین مشابهت بین اعضای خوشه های متناظر اندازه گیری می کنند.

² Charnes, Cooper and Rhodes
Decision Making Unit ³

۲ آزمون آماری و تحلیل خوشه ای در ارزیابی سرپرستی شعبه های بانک

آزمون مجموع رتبه ویلکسون - من - ویتنی^۴ برای کارایی قبلا مورد استفاده قرار گرفته است. در اینجا قصد داریم آزمون نسبت واریانس ها را برای کارایی ۳۵ سرپرستی مربوط به یک بانک تجاری که کارایی آنها توسط حسین زاده و همکاران مورد سنجش قرار گرفته است به کار بگیریم. نتیجه این بررسی در جدول ۱۱ آمده است.

فرض صفر بر این تاکید دارد که واریانس ۲ جامعه مساوی است. برای نمونه دو ماه اول و هفتم در بررسی اول و دو ماه اول و دوازدهم را در بررسی دوم مورد آزمون قرار داده ایم. برای ماه اول به ترتیب برآورد میانگین و واریانس عبارت است از: $\bar{x} = 0.44, S_1^2 = 0.0608$ و برای ماه دوازدهم $\bar{y} = 0.43, S_2^2 = 0.0713$ خواهیم داشت: $F = \frac{S_1^2}{S_2^2} = 1.1727$ و با توجه به این که $F(34,34)_{\alpha=0.05} = 1.69$ در سطح اطمینان $\alpha = 0.05$ فرض صفر رد نمی شود. و این یعنی این که بین واریانس کارایی شعبه ها در دو ماه از سال اختلاف معنی داری وجود دارد. نتیجه برای ماه دوازدهم عبارت است از $\bar{y} = 0.48, S_2^2 = 0.0521$ و در نتیجه داریم: $F = 1.1670$ مشابه ها در اینجا نیز فرض صفر در همان سطح اطمینان آماری رد نمی شود. لذا در ماه های بررسی شده کارایی ها هم واریانس هستند. جدول زیر اطلاعات بیشتری در مورد توزیع کارایی بدست می دهد:

	کشیدگی چندگی	ضریب کشیدگی چندگی	چولگی چندگی
ماه اول	0.095	-0.259	0.05
ماه دوم	0.095	-0.259	0.05
ماه هفتم	0.07	-0.260	0
دوازدهم	0.08	-0.260	0.02

جدول ۱

نتایج جدول ۱ نشان می دهد که توزیع کارایی در هر چهار ماه مورد بررسی دارای چوله به راست بوده و میزان انحراف توزیع کارایی از توزیع نرمال استاندارد بیشتر و شکل بشقاب گونه دارد. در ادامه چندین طرح خوشه بندی با استفاده از نرم افزار SPSS آمده است: در طرح اول ماه های سال خوشه بندی شده است تا مدیریت از بعد زمانی اطلاعاتی در مورد هم سانی عملکرد سرپرستی ها داشته باشد:

⁴ Rank-Sum Test (Wilcoxon-Mann-Whitney)

Case Number	Cluster
1	1
2	2
3	1
4	3
5	3
6	2
7	3
8	3
9	2
10	2
11	2
12	2

جدول ۲: اعضای خوشه

این نتیجه نشان می دهد که سرپرستی ها در ۳ خوشه خوشه بندی می شوند. هم چنین معلوم شده است که کدام ماه های سال در یک خوشه قرار گرفته اند. به عنوان نمونه معلوم شده است که سرپرستی ها در چهار برج آخر سال مثل هم رفتار می کنند و در ماه های میانی تیر، مرداد، مهر و آبان مانند هم بوده اند. فروردین و خرداد در یک خوشه قرار گرفته اند. در جداول زیر به ترتیب فاصله بین مراکز خوشه بندی و جدول تحلیل واریانس را ملاحظه می کنید:

Cluster	1	2	3
1		.289	.231
2	.289		.270
3	.231	.270	

جدول ۳: فاصله نهایی بین مراکز خوشه

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VAR00001	.004	2	.001	9	3.663	.069
VAR00002	.006	2	.000	9	18.454	.001
VAR00003	.002	2	.004	9	.521	.611
VAR00004	.027	2	.008	9	3.248	.087
VAR00005	.000	2	.000	9	.	.
VAR00006	.000	2	.000	9	1.000	.405
VAR00007	.000	2	.000	9	.	.
VAR00008	.001	2	.001	9	1.715	.234
VAR00009	.008	2	.001	9	6.436	.018
VAR00010	.003	2	.001	9	6.783	.016
VAR00011	.002	2	.001	9	3.298	.084
VAR00012	.012	2	.001	9	8.888	.007
VAR00013	.002	2	.001	9	3.407	.079
VAR00014	.002	2	.000	9	9.581	.006
VAR00015	.003	2	.000	9	10.442	.005

VAR00016	.000	2	.000	9	.974	.414
VAR00017	.007	2	.001	9	6.964	.015
VAR00019	.004	2	.001	9	7.218	.013
VAR00018	.002	2	.001	9	3.131	.093
VAR00020	.002	2	.000	9	5.449	.028
VAR00021	.001	2	.000	9	4.120	.054
VAR00022	.002	2	.000	9	5.829	.024
VAR00023	.002	2	.000	9	8.230	.009
VAR00024	.001	2	.000	9	5.977	.022
VAR00025	.003	2	.001	9	2.387	.147
VAR00026	.004	2	.000	9	8.572	.008
VAR00027	.005	2	.001	9	7.017	.015
VAR00028	.001	2	.000	9	4.193	.052
VAR00029	.005	2	.003	9	1.599	.255
VAR00030	.001	2	.000	9	2.935	.104
VAR00031	.001	2	.000	9	3.174	.091
VAR00032	.015	2	.002	9	6.904	.015
VAR00033	.001	2	.001	9	.912	.436
VAR00034	.001	2	.000	9	1.234	.336
VAR00035	.001	2	.000	9	2.787	.114

جدول ۴: نتایج آزمون ANOVA

نتایج جدول نشان دهنده مربع اختلافات درون خوشه ای و بین خوشه ای است. ستون آخر نشان می دهد که واقع شدن ماه های سال در خوشه های متفاوت درست بوده است و سه خوشه تعداد خوشه مناسب برای خوشه بندی است. نتیجه نشان می دهد که ۲ ماه در خوشه اول ۶ ماه در خوشه دوم و ۴ ماه در خوشه سوم قرار دارد.

Cluster	1	2.000
	2	6.000
	3	4.000
Valid		12.000
Missing		.000

جدول ۵: تعداد اعضای هر خوشه

در طرح دوم سرپرستی ها در یک ماه معین از نظر کارایی خوشه بندی شده اند. به عنوان نمونه نتایج حاصل از خوشه بندی کارایی سرپرستی ها در ماه شهریور به صورت زیر است:

2	2
---	---

3	2
4	3
5	2
6	2
7	2
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	3
30	1
31	1
32	1
33	1
34	1
35	1

جدول ۶: اعضای هر خوشه

جدول فوق به وضوح نشان می دهد که کدام سرپرستی ها در ماه شهریور از نظر کارایی کمترین اختلاف را داشته اند به گونه ای در تحلیل خوشه ای در یک خوشه قرار گرفته اند. جداول زیر به ترتیب مرکز خوشه ها فاصله بین آنها و آزمون تحلیل واریانس را نشان می دهد.

VAR00006	.35	.97	.60
----------	-----	-----	-----

جدول ۷: مراکز نهایی خوشه ها

Cluster	1	2	3
1		.629	.251
2	.629		.377
3	.251	.377	

جدول ۸: فاصله نهایی بین مراکز خوشه

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VAR00006	.866	2	.004	32	235.998	.000

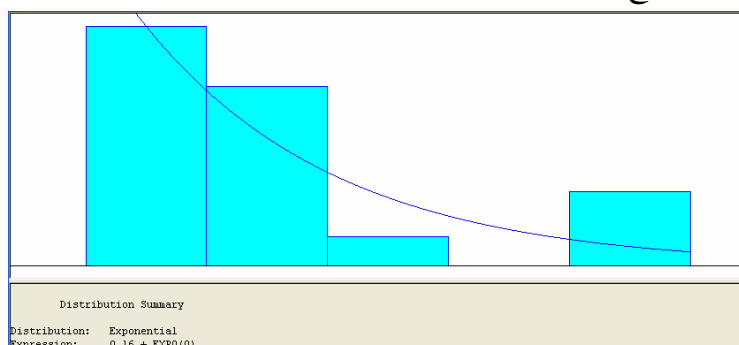
جدول ۹: نتایج آزمون ANOVA

Cluster	1	27.000
	2	5.000
	3	3.000
Valid		35.000
Missing		.000

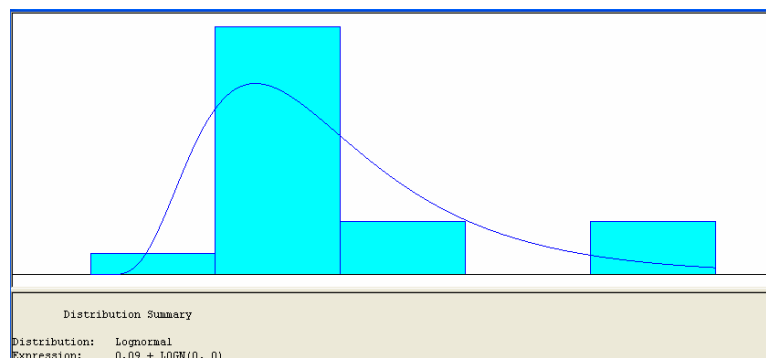
جدول ۱۰: تعداد اعضای هر خوشه

در ماه شهریور ۲۷ سرپرستی مانند هم ۵ سرپرستی مثل هم و ۳ سرپرستی نیز نزدیک به یکدیگر عمل کرده اند. آزمون معنی دار بودن سه خوشه را نشان می دهد.

تعیین و بررسی توزیع کارایی عملکرد سرپرستی ها نشان می دهد که ۱۰ ماه از سال بجز ماه های پنجم و دوازدهم همگی از توزیع لاگ نرمال تبعیت می کنند. البته پارامترهای آنها با یکدیگر تفاوت دارند. توزیع کارایی در دو ماه پنجم و دوازدهم نمایی است. برای تعیین توزیع ما از نرم افزار Arena استفاده کرده ایم. علت تغییر توزیع کارایی را شاید در شدت رقابت بین سرپرستی ها برای اخذ امتیاز بیشتر در این ماه ها دانست. شکل های ۱ و ۲ توزیع و پارامتر توزیع ها را مشخص می کند.



شکل ۱: توزیع کارایی مربوط به ماه پنجم



شکل ۲: توزیع کارایی مربوط به ماه سوم

۳ نتیجه گیری

در این مطالعه تلاش شد تا به کارگیری آزمون های آماری تحلیل روشن تری از توزیع کارایی داشته باشیم لذا با توجه به این مطلب چندین آماره و آزمون نسبت واریانس ها را در مورد کارایی ۳۵ سرپرستی یک بانک انجام دادیم. مشخص شد که سرپرستی ها در چندین ماه از سال اختلاف معنی داری در واریانس کارایی دارند. همچنین خوشه بندی را در دو طرح سالیانه برای تعیین ماه های نزدیک به هم از نظر کارایی و همچنین خوشه بندی در یک ماه از سال را انجام دادیم. تعیین توزیع کارایی نشان می دهد که عملکرد سرپرستی ها در این سال دو بار دچار تغییر توزیع شده است. این امر می تواند از جنبه مدیریتی برای یافتن دلیل این تغییر توزیع مورد توجه قرار گیرد.

تقدیر و تشکر

نویسندگان مقاله مراتب تشکر و امتنان خود را از راهنمایی های آقای دکتر بهروز فتحی و آقای دکتر حسین جباری خامنه ای و همچنین آقای امیر زینل ابراز می دارند.

	mah 1	mah 2	mah 3	mah 4	mah 5	mah 6	mah 7	mah 8	mah 9	mah 10	mah 11	mah 12
S 1	0.47	0.57	0.57	0.52	0.54	0.56	0.48	0.52	0.58	0.53	0.57	0.58
S 2	1	0.96	0.95	0.92	0.92	0.98	0.92	0.93	1	1	1	1
S 3	0.93	0.88	0.86	0.87	0.85	0.89	1	1	0.91	0.97	0.86	0.84
S 4	0.63	0.58	0.55	0.59	0.57	0.68	0.47	0.73	0.75	0.75	0.75	0.84
S 5	1	1	1	1	1	1	1	1	1	1	1	1
S 6	1	1	1	1	1	1	0.99	1	1	1	1	1
S 7	1	1	1	1	1	1	1	1	1	1	1	1
S 8	0.36	0.46	0.45	0.4	0.41	0.43	0.37	0.4	0.43	0.39	0.43	0.43
S 9	0.35	0.45	0.43	0.37	0.37	0.39	0.35	0.41	0.48	0.44	0.48	0.47
S 10	0.41	0.51	0.48	0.43	0.44	0.47	0.41	0.44	0.48	0.46	0.48	0.49
S 11	0.26	0.33	0.31	0.27	0.26	0.27	0.23	0.25	0.29	0.26	0.3	0.3
S 12	0.22	0.28	0.18	0.29	0.25	0.28	0.24	0.27	0.31	0.32	0.4	0.34
S 13	0.28	0.38	0.35	0.3	0.31	0.33	0.29	0.3	0.34	0.31	0.34	0.34
S 14	0.24	0.31	0.29	0.26	0.27	0.3	0.25	0.27	0.31	0.28	0.31	0.31
S 15	0.36	0.42	0.4	0.36	0.38	0.4	0.36	0.37	0.43	0.39	0.42	0.41

S 16	0.3	0.32	0.32	0.29	0.28	0.3	0.3	0.32	0.33	0.28	0.32	0.34
S 17	0.56	0.48	0.46	0.4	0.41	0.44	0.42	0.45	0.5	0.48	0.5	0.5
S 18	0.33	0.42	0.42	0.38	0.39	0.41	0.37	0.38	0.42	0.38	0.43	0.44
S 19	0.35	0.41	0.38	0.34	0.35	0.4	0.39	0.39	0.43	0.39	0.44	0.45
S 20	0.35	0.36	0.38	0.32	0.32	0.35	0.34	0.35	0.38	0.35	0.39	0.39
S 21	0.31	0.32	0.31	0.28	0.31	0.34	0.32	0.32	0.36	0.31	0.35	0.35
S 22	0.27	0.33	0.31	0.26	0.26	0.28	0.24	0.26	0.29	0.27	0.31	0.31
S 23	0.29	0.34	0.33	0.29	0.3	0.34	0.3	0.31	0.35	0.31	0.36	0.35
S 24	0.28	0.32	0.3	0.27	0.28	0.31	0.28	0.29	0.31	0.28	0.33	0.33
S 25	0.46	0.43	0.43	0.36	0.35	0.35	0.38	0.42	0.42	0.32	0.4	0.4
S 26	0.3	0.36	0.35	0.31	0.32	0.36	0.33	0.35	0.4	0.36	0.4	0.39
S 27	0.39	0.47	0.43	0.39	0.39	0.41	0.36	0.36	0.43	0.4	0.45	0.47
S 28	0.22	0.26	0.25	0.22	0.24	0.26	0.24	0.26	0.28	0.24	0.28	0.29
S 29	0.34	0.39	0.36	0.32	0.4	0.55	0.44	0.42	0.44	0.37	0.41	0.42
S 30	0.34	0.32	0.3	0.27	0.27	0.28	0.29	0.29	0.3	0.26	0.29	0.3
S 31	0.28	0.3	0.28	0.25	0.26	0.27	0.25	0.25	0.27	0.24	0.28	0.28
S 32	0.63	0.54	0.51	0.44	0.42	0.42	0.42	0.43	0.43	0.43	0.4	0.44
S 33	0.27	0.35	0.34	0.29	0.3	0.31	0.28	0.28	0.3	0.27	0.31	0.32
S 34	0.35	0.36	0.34	0.33	0.34	0.35	0.3	0.31	0.35	0.29	0.34	0.34
S 35	0.29	0.31	0.29	0.26	0.26	0.27	0.25	0.26	0.28	0.24	0.28	0.29

شکل ۱۱: کارایی های ۳۵ سرپرستی در ۱۲ ماه

منابع

- [1] Banker, R. D., 1993, Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273
- [2] Banker, R.D., 1996, Hypothesis Tests Using Data Envelopment Analysis, *Journal of Productivity Analysis* 7, 139-159.
- [3] Charnes, A., W.W. Cooper and E. Rhodes, 1978, Measuring the efficiency of Decision Making Units, *European Journal of Operational Research* 2, 429-444.
- [4] Kneip, A., B.U. Park and L. Simar, 1996, A note on the convergence of nonparametric DEA efficiency measures, Discussion Paper 9603, Institut de Statistique, Universite Catholique de Louvain.
- [5] Koostelev, A. P., L. Simar and A. B. Tsybakov, 1995a, Efficient estimation of monotone boundries, *Annals of Statistics* 23(2), 476-489.
- [6] Koostelev, A. P., L. Simar and A. B. Tsybakov, 1995b, On estimation of monotone and convex boundries, *Publication de l'Institut de statistique de l'Universite de Paris* 39:, 3-18.
- [7] Simar, L. and P.W. Wilson, 1995, Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science*

[۸] دیواندری علی مهناز محمودی فرهاد حسین زاده لطفی غلامرضا جهانشاهلو و ولی... طالقانی برنکی ارزیابی سرپرستیهای بانک ملت ایران در مقاطع زمانی مختلف کاربردی از تحلیل پوششی داده ها در حال چاپ مجله علمی - پژوهشی ریاضیات کاربردی دانشگاه آزاد اسلامی واحد لایهجان.

[۹] جانسون. ریچارد ا. و دین. دبلیو. ویچرن "تحلیل آماری چند متغیری کاربردی" ترجمه حسینعلی نیرومند انتشارات دانشگاه فردوسی مشهد - چاپ دوم ۱۳۸۴.

