

# IAM final

Daniela Pico Arredondo

27/8/2021

La siguiente base de datos proporciona información recolectada de un estudio sobre la eficacia del control de infecciones nosocomiales, cuyo objetivo principal fue determinar si los programas de vigilancia y control de infecciones han reducido las tasas de infección nosocomial (adquiridas en los hospitales) en hospitales de Estados Unidos. Estos datos consisten de una muestra aleatoria de  $n = 33$  hospitales, respectivamente, seleccionados de los 338 hospitales originales investigados. Los datos presentados corresponden al periodo de estudio 1975-76. Las 12 variables son descritas en la Tabla 1.

# variable	Código	Nombre	Descripción
1	ID	Número de identificación del registro	numérica
2	DPERM	Longitud de permanencia	Longitud promedio de permanencia de todos los pacientes en el hospital (en días).
3	EDAD	Edad	Edad promedio de los pacientes (en años).
4	RINF	Riesgo de infección	Probabilidad promedio estimada de adquirir infección en el hospital (en porcentaje).
5	RRC	Razón de rutina de cultivos	Razón del número de cultivos desarrollados al número de pacientes sin signos o síntomas de infección adquirida en el hospital, por 100.
6	RRX	Razón de rutina de rayos X del pecho	Razón del número de rayos X llevados a cabo al número de pacientes sin signos o síntomas de neumonía, por 100.
7	NCAMAS	Número de camas	Número promedio de camas en el hospital durante el periodo de estudio.
8	AEM	Afiliación a escuela de medicina	1=SÍ, 2=NO
9	PDP	Censo promedio diario	Número promedio de pacientes en el hospital por día durante el periodo de estudio.
10	NENFERM	Número de enfermeras	Número promedio de tiempos completos equivalentes registrados y enfermeras de práctica licenciadas durante el periodo de estudio (número de tiempos completos + 1/2 del número de tiempo parcial).
11	FSD	Facilidades y servicios disponibles	Porcentaje de 35 facilidades potenciales y servicios que son proporcionados por el hospital.
12	REGION	Región	Región geográfica, donde 1=NE, 2=NC, 3=S, 4=W.

En el siguiente informe se realizará el análisis de componentes para las variables numéricas, DPERM, EDAD, RINF, RRC, RRX, NENFERM, FSD; recordemos que tiene sentido realizar un análisis de componentes principales cuando hay presencia de varias variables ya que estas pueden estar altamente correlacionadas y terminan midiendo lo mismo por lo cual se puede simplificar, existen factores o componentes que puedan representar ese conjunto de variables sin tener que recurrir a la totalidad de variables de esta manera el ACP ayuda a simplificar el estudio de los objetos, de esta manera, se realizará un análisis de la correlación de los datos

## Resumen de Correlaciones mas altas por pares

Table 1: Coorrelaciones mayores en pares de vbles

	First.Variable	Second.Variable	Correlation
60	NCAMAS	PDP	0.9748026
69	NCAMAS	NENFERM	0.9377819
70	PDP	NENFERM	0.8951219
80	NENFERM	FSD	0.8784788
78	NCAMAS	FSD	0.8410032
79	PDP	FSD	0.8135074
19	DPERM	RINF	0.5426557
55	DPERM	PDP	0.5301933
46	DPERM	NCAMAS	0.4593552
48	RINF	NCAMAS	0.4530611
39	RINF	RRX	0.4305867
30	RINF	RRC	0.4271066
57	RINF	PDP	0.4239785
37	DPERM	RRX	0.3966850
66	RINF	NENFERM	0.3932925

## Gráfico de Matriz de Correlaciones

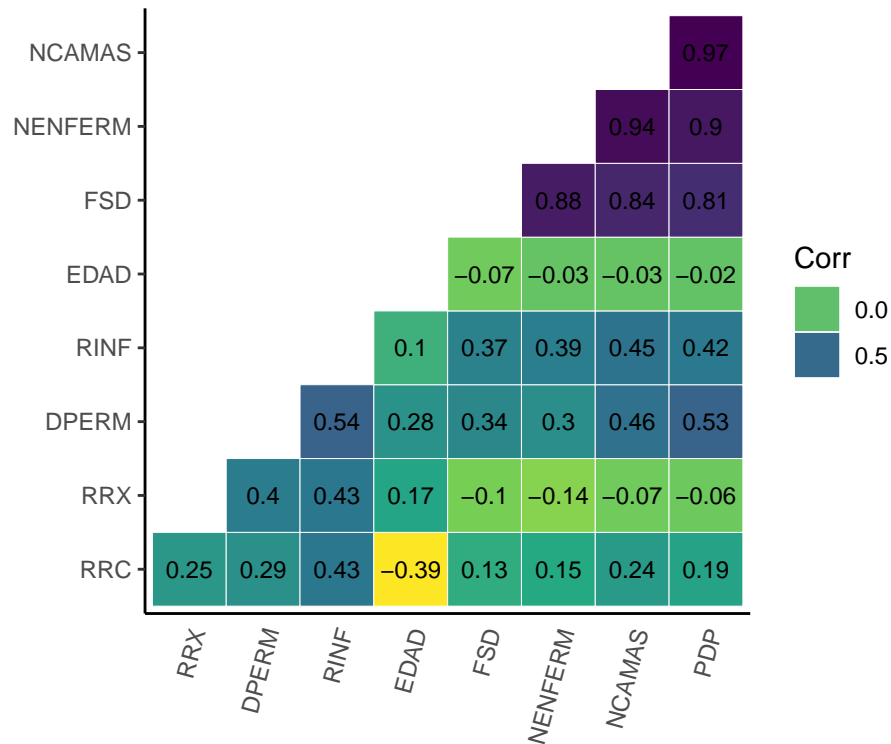


Figure 1: Matriz de Correlación

El resumen de correlaciones indica que hay una alta dependencia lineal entre las variables Numero de camas y Censo promedio diario, igualmente entre las variables Numero de camas y Numero de enfermeras. En general se observan correlaciones altas entre varios pares de variables, por lo cual es pertinente realizar un análisis de componentes principales.

## ACP con la función: princomp del paquete stats

Se tiene la siguiente matriz de varianzas-covarianzas

##

## Matriz de var-cov s:

	DPERM	EDAD	RINF	RRC	RRX	NCAMAS	PDP	NENFERM
DPERM	7.163438	3.5748246	1.9097056	8.996932	20.52782	272.66526	255.01613	118.899526
EDAD	3.574825	22.1015726	0.6284677	-21.006391	15.71107	-31.24395	-15.05806	-22.8356855
RINF	1.909706	0.6284677	1.7288710	6.449234	10.94657	132.11694	100.18387	76.0520161
RRC	8.996932	-21.0063911	6.4492339	131.880635	56.12380	601.13367	392.83871	260.128528
RRX	20.527818	15.7110685	10.9465726	56.123800	373.82854	-317.92560	-199.32581	-408.45837
NCAMAS	272.665262	-31.2439516	132.1169355	601.133669	-317.92560	49185.93145	38851.67742	30586.94153
PDP	255.016129	-15.0580645	100.1838710	392.838710	-199.32581	38851.67742	32295.74194	23657.48387
NENFERM	118.899526	-22.8356855	76.0520161	260.128528	-408.45837	30586.94153	23657.48387	21628.51628
FSD	13.155565	-5.0061290	7.0712903	22.151452	-28.49919	2692.43226	2110.38065	1864.9061290

## Warning: package 'factoextra' was built under R version 4.0.5

Table 2: Correlaciones de las vbles sobre ejes

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
DPERM	0.4570625	-0.4764253	-0.0392629	0.3618891	0.0448995
EDAD	-0.0274878	-0.0366075	-0.0553505	0.1390918	-0.5591028
RINF	0.4391091	-0.0673594	0.2235157	0.4558869	0.1213852
RRC	0.2086444	-0.1018300	0.3862472	0.2822313	0.8429608
RRX	-0.0847175	-0.2306870	0.1544724	0.9520346	-0.0947987
NCAMAS	0.9961290	-0.0228651	0.0842615	-0.0089428	-0.0049061
PDP	0.9828986	-0.1586643	-0.0932494	0.0043515	0.0046119
NENFERM	0.9531609	0.2967276	-0.0563817	0.0153016	0.0032969
FSD	0.8548051	0.2062621	-0.0665616	0.0309506	-0.0042795

Recordemos que las correlaciones son los cosenos del angulo que forma la variable con la componente, La columna de correlaciones representa la correlación entre las componentes principales y la variables, estas correlaciones se usan para mirar la calidad de cada variable sobre los ejes, pero con las correlaciones al cuadrado que hacen referencia a los cosenos cuadrados y se presentan en la siguiente tabla

Table 3: Coordneadas, cosenos cuadrados y contribuciones de las variables sobre los ejes 1,2,3

	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
DPERM	1.204	-1.255	0.209	0.227	0.001	0.059
EDAD	-0.127	-0.169	0.001	0.001	0.000	0.001
RINF	0.568	-0.087	0.193	0.005	0.000	0.000
RRC	2.358	-1.151	0.044	0.010	0.006	0.049
RRX	-1.612	-4.390	0.007	0.053	0.003	0.717
NCAMAS	217.441	-4.991	0.992	0.001	48.894	0.927
PDP	173.855	-28.065	0.966	0.025	31.257	29.300
NENFERM	137.970	42.951	0.909	0.088	19.686	68.628
FSD	12.145	2.931	0.731	0.043	0.153	0.319

Esta tabla muestra la información de las coordenadas, cosenos cuadrados y contribuciones respectivamente, en dos dimensiones cada una, las columnas de los cosenos cuadrados representan la calidad de cada variable sobre los ejes, por ejemplo, la variable que mejor contribuyen de manera significativa a la formación de la componente 1 son NCAMAS y PDP ya que tienen valores de 0.921 y 0.897

Las contribuciones hacen referencia a la proyección de la variables, entre mas grande es la magnitud de la proyección más grande es la contribución de la variable, en la tabla se representan en porcentajes, cada componente explica un porcentaje de la variación total, es decir, la columna 5 explica en cada entrada cierto porcentaje de la varianza del componente 1.

### Vectores Propios

Table 4: Vectores Propios (Loadings)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
DPERM	0.0039	0.0242	0.0039	0.0510	0.0117
EDAD	-0.0004	0.0033	0.0096	0.0344	-0.2551
RINF	0.0018	0.0017	-0.0109	0.0316	0.0155
RRC	0.0076	0.0222	-0.1643	0.1706	0.9397
RRX	-0.0052	0.0847	-0.1107	0.9690	-0.1779
NCAMAS	0.6992	0.0963	-0.6924	-0.1044	-0.1056
PDP	0.5591	0.5413	0.6209	0.0412	0.0805
NENFERM	0.4437	-0.8284	0.3072	0.1185	0.0471
FSD	0.0391	-0.0565	0.0356	0.0235	-0.0060

### Valores Propios

99818.86	2774.865	728.4845	360.8657	106.1301	46.87424	16.19193	2.237263	0.7973104
----------	----------	----------	----------	----------	----------	----------	----------	-----------

Las respectivas primeras componentes principales son:

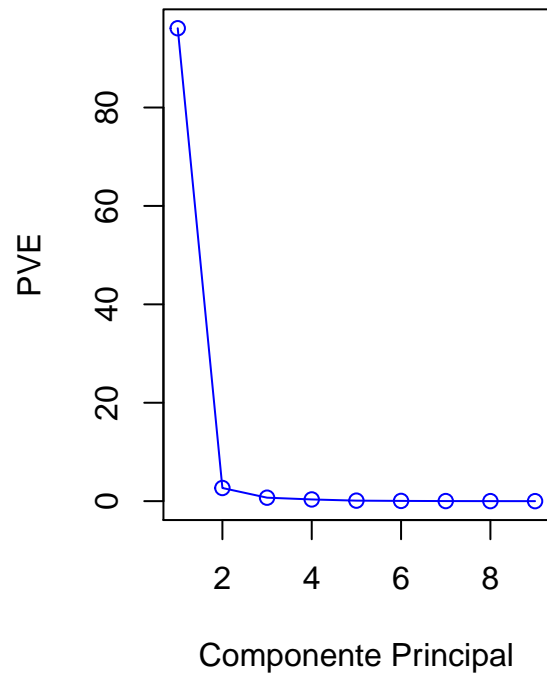
Table 5: Porcentajes de Var Explicada

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
96.11339	2.671857	0.7014417	0.3474696	0.1021904	0.0451342	0.0155909	0.0021542	0.0007677

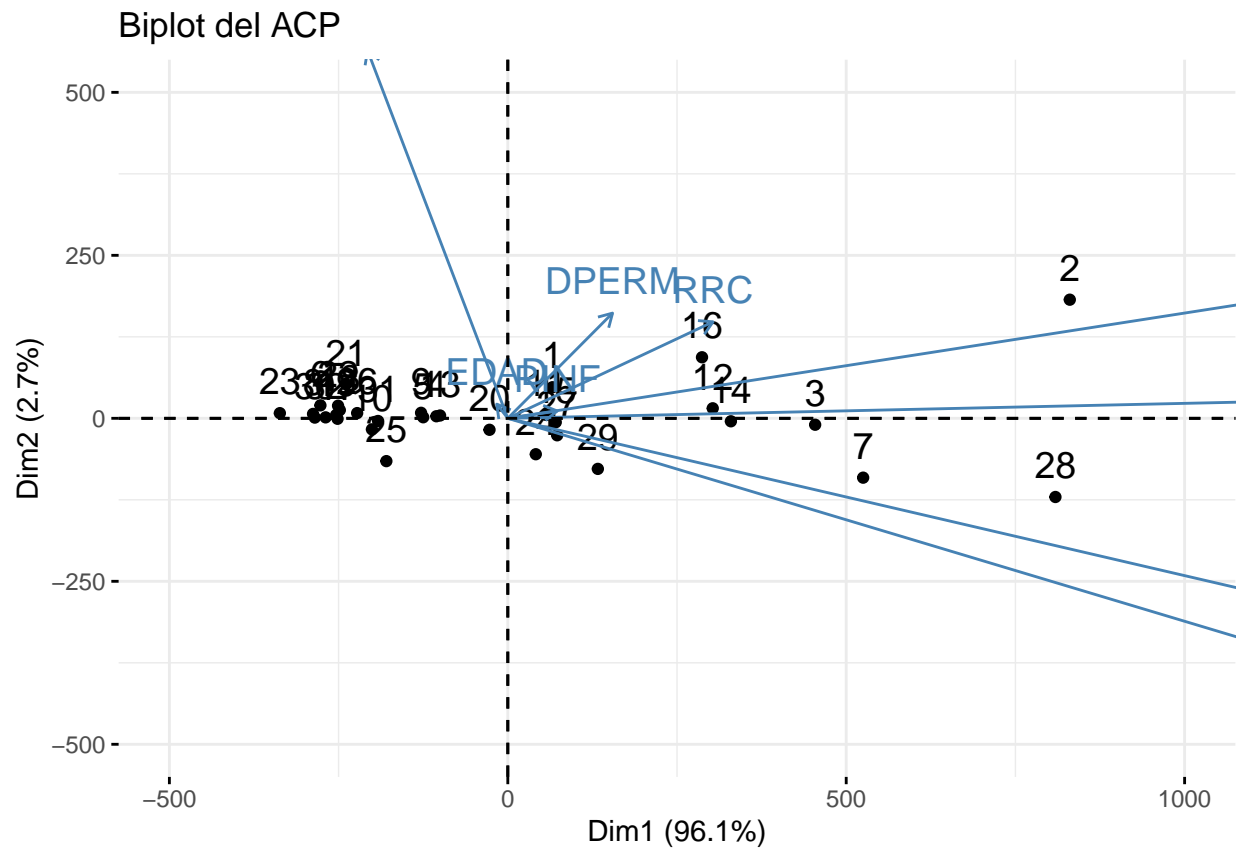
De la tabla anterior se observa que la componente 1 explica el 96.11339 de la variabilidad total, y las dos primeras componentes principales explican el 98.785247% de dicha variabilidad.

A continuación se presentaran algunas tablas que representan lo analizado en las tablas anteriores

Gráfico de Valores Propios



En la Gráfica se observa que después del número 2 en el eje X la tendencia de los puntos se empieza a estabilizar, esto indica que las dos primeras componentes pueden explicar gran parte de la variabilidad total.



### Gráfico de Coeficientes

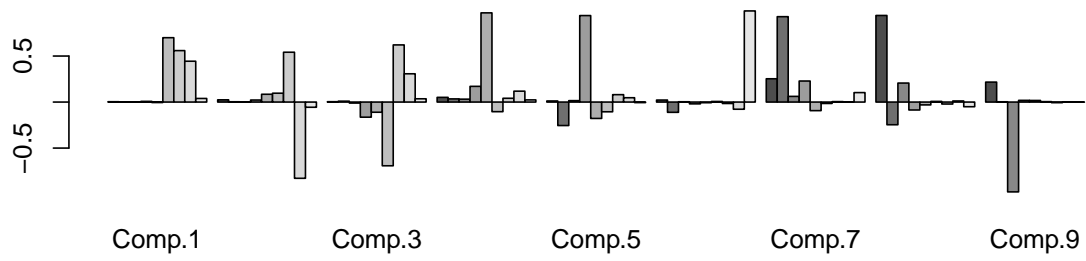


Figure 2: Gráfico de Coeficientes