



EJERCICIO PRÁCTICO 12: TRANSFORMACIÓN DE DATOS Y MÉTODOS ROBUSTOS

CONTEXTO

Comenzamos nuestro estudio de la inferencia estadística conociendo las pruebas paramétricas que, como hemos visto, hacen muchas suposiciones acerca de las distribuciones de los datos que se analizan. También nos hemos encontrado con que, en la práctica, no es difícil encontrarse con datos que no parecen satisfacer estas suposiciones.

Hemos ampliado nuestro conjunto de herramientas estudiando métodos no paramétricos y el remuestreo, pero existen aún más alternativas: si los datos siguen una distribución que se aleje de la normal, podemos también aplicar **transformaciones** ad-hoc a los datos, que permitan aplicar los métodos que ya conocemos cambiando las hipótesis contrastadas a unas similares (idealmente equivalentes). También podemos emplear **métodos robustos**, basados en medidas de tendencia central distintos a la media.

OBJETIVOS DE APRENDIZAJE

1. Transformar datos para aplicar métodos paramétricos con alguna validez.
2. Entender las hipótesis que se contrastan con diferentes métodos robustos.
3. Aplicar métodos robustos en diferentes situaciones utilizando el ambiente R.

ÉXITO DE LA ACTIVIDAD

- a. Transformación de datos (problema 1):
 - a. El equipo es capaz de encontrar una transformación adecuada para los datos entregados.
 - b. El equipo puede aplicar una prueba de hipótesis paramétrica sobre los datos transformados usando el entorno R.
- b. Métodos no paramétricos (problemas 2 y 3):
 - a. El equipo es capaz de mencionar una alternativa robusta para la prueba tradicional en diferentes casos de estudio.
 - b. El equipo puede aplicar la prueba de hipótesis robusta en cada caso, usando el entorno R.

ACTIVIDADES

1. Copiar el enunciado de los problemas asignados como comentarios de un script R.
2. El equipo lee los enunciados entregados y revisa los datos gráficamente.
3. Transformación de datos:
 - a. El equipo discute las posibles transformaciones y se decide por alguna(s) de ellas.
 - b. El equipo aplica las transformaciones y realiza la prueba de hipótesis correspondiente.
4. Métodos robustos:
 - a. El equipo discute qué prueba tradicional (paramétrica) correspondería en cada caso y, considerando esto, concuerda una prueba no paramétrica que ha de aplicar.
 - b. El equipo aplica la prueba de hipótesis correspondiente usando el ambiente R.
 - c. El equipo identifica las hipótesis nula y alternativa en cada caso y entrega una conclusión, interpretándola en el caso estudiado.

Fuera del horario de clases, cada equipo debe subir el script realizado UVirtual con el nombre "EP11-respuesta-grupo-i", donde i es el número de grupo asignado. Las respuestas deben subirse antes de las 23:30 del martes 17 de mayo.

PREGUNTAS

Grupo 1:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
129	1510394	134	1252837
109	402929	193	2196277
28	885722	10	120276
178	4428151	88	4629726
74	48667	142	5743260
16	834565	86	6701654
87	70599	36	6568968
108	783108	190	180141
149	210041	163	6684497
78	37449	33	35974

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 2:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
137	210041	11	2830464
19	783108	64	180141
87	834565	104	994830
150	70599	82	6684497
18	8576989	96	35974
91	251843	175	5743260
147	4428151	84	4629726
139	48667	106	48408
69	48705	121	2196277
21	885722	6	92932

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 3:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
102	842079	43	6684497
112	834565	189	180141
151	885722	140	2830464
4	48705	51	1252837
55	4428151	146	5743260
70	62764	25	6701654
98	8576989	2	1174562
135	37449	179	120276
7	48667	68	994830
183	1510394	182	48408

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 4:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
127	48667	70	6684497
106	783108	197	5743260
16	4428151	151	92932
167	842079	190	48408
175	8576989	22	994830
171	885722	148	2196277
54	834565	78	35974
29	210041	152	4629726
47	48705	53	180141
103	251843	95	1174562

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 5:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
171	70599	190	120276
27	783108	95	1174562
162	4428151	183	2196277
102	37449	198	180141
89	834565	16	35974
185	48705	151	1252837
158	842079	11	92932
64	210041	169	6701654
149	402929	187	6568968
14	62764	135	4629726

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 6:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
167	1510394	197	48408
8	251843	7	35974
65	834565	21	5743260
91	37449	195	6684497
125	48705	191	1252837
64	402929	149	6701654
196	885722	39	6568968
117	8576989	139	120276
41	62764	154	1174562
56	783108	137	2830464

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 7:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
41	48705	117	6701654
192	251843	51	1252837
138	210041	121	6568968
85	8576989	30	6684497
87	842079	113	92932
134	1510394	146	994830
8	4428151	106	120276
57	834565	153	180141
79	37449	175	5743260
17	885722	139	1174562

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 8:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
106	4428151	174	1174562
99	251843	129	48408
29	1510394	23	6684497
176	885722	139	6568968
154	37449	69	1252837
75	834565	114	35974
8	842079	137	120276
193	783108	42	5743260
66	48705	84	6701654
200	8576989	152	92932

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 9:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
127	885722	133	1174562
199	834565	70	120276
67	37449	82	6568968
29	402929	64	4629726
154	783108	122	6684497
28	62764	73	92932
130	210041	51	180141
66	48705	18	48408
179	8576989	176	994830
117	842079	39	35974

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

Grupo 10:

1. En el trabajo de título de un estudiante del DIINF se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

Instancia A	Tiempo A	Instancia B	Tiempo B
53	834565	22	5743260
109	783108	6	6684497
197	1510394	85	1174562
15	842079	186	6701654
56	70599	185	6568968
64	62764	149	4629726
145	885722	84	2196277
103	4428151	118	48408
111	402929	147	1252837
76	251843	110	994830

2. Analice la primera pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.
3. Analice la segunda pregunta abordada en el ejercicio práctico 11, con los mismos datos, utilizando un método robusto adecuado.

CRITERIOS DE EVALUACIÓN

Pregunta	Categoría	Nivel de logro	Puntos
1	Datos	Obtienen correctamente los datos necesarios para realizar la prueba solicitada, en un formato pertinente	2
	Condiciones	Argumentan convincentemente el incumplimiento de todas las condiciones requeridas para aplicar una prueba robusta adecuada	2
	Transformación	Seleccionan, argumentando con gráficos u otras herramientas, una transformación apropiada para poder aplicar una prueba paramétrica a los datos.	3
	Formulación de hipótesis	Formulan con claridad y explícitamente hipótesis nulas y alternativas adecuadas para responder la pregunta planteada, tanto en lenguaje natural como en lenguaje matemático	3
	Prueba	Realizan de forma completa y correcta una prueba paramétrica adecuada para la pregunta planteada en el enunciado, utilizando los datos correctos	4
	Conclusión	Entregan una conclusión completa a la pregunta planteada en el enunciado, en el dominio original de los datos, basándose en el resultado correcto de una prueba adecuada y el contexto del problema	4
	Código fuente	Escriben código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que realiza de forma completa y correcta la transformación y la prueba paramétrica adecuada para responder y con los datos adecuados	3
	Ortografía y redacción	Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema	3
2	Pregunta de investigación	Proponen una pregunta de investigación, interesante y novedosa, que involucra la comparación de las medias de dos grupos independientes de personas encuestadas en la Casen 2017	3
	Datos	Obtienen una muestra de datos de acuerdo a lo solicitado, revisando su comportamiento con gráficos o pruebas estadísticas y pronunciándose explícitamente sobre la necesidad de utilizar métodos para datos problemáticos	3

	Formulación de hipótesis	Formulan explícitamente hipótesis nula y alternativa correctas, que involucran la comparación de las medias de una variable numérica de dos grupos independientes, para responder la pregunta de investigación que plantean	3
	Alternativa robusta	Basándose en el análisis anterior, proponen explícitamente una prueba robusta que permite docimar las hipótesis propuestas, justificando su elección apropiadamente	3
	Prueba	Realizan, de forma completa y sin errores, una prueba robusta que permite responder la pregunta de investigación que plantean, usando una muestra de datos adecuada, obteniendo un p valor o intervalo de confianza correcto	3
	Conclusión	Entrega una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema.	3
	Código fuente	Escriben código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que realiza de forma completa y correcta la prueba seleccionada con los datos adecuados en cada caso	3
	Ortografía y redacción	Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema	3
3	Pregunta de investigación	Proponen una pregunta de investigación, interesante y novedosa, que involucra la comparación de las medias de más de dos grupos independientes de personas encuestadas en la Casen 2017	3
	Datos	Obtienen una muestra de datos de acuerdo a lo solicitado, revisando su comportamiento con gráficos o pruebas estadísticas y pronunciándose explícitamente sobre la necesidad de utilizar métodos para datos problemáticos	3
	Formulación de hipótesis	Formulan explícitamente hipótesis nula y alternativa correctas, que involucran la comparación de las medias de una variable numérica de más de dos grupos independientes, para responder la pregunta de investigación que plantean	3
	Alternativa robusta omnibus	Basándose en el análisis anterior, proponen explícitamente una prueba robusta que permite docimar las hipótesis propuestas, justificando su elección apropiadamente	3
	Prueba omnibus	Realizan, de forma completa y sin errores, una prueba robusta que permite responder la pregunta de investigación que plantean, usando una muestra de datos adecuada, obteniendo un p valor o intervalo de confianza correcto	3

	Post hoc	Aplican, de forma completa y sin errores, un procedimiento post-hoc robusto, usando una muestra de datos adecuada, obteniendo un p valor o intervalo de confianza correcto	3
	Conclusión	Entrega una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema.	3
	Código fuente	Escriben código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que realiza, de forma completa y correcta, la prueba robusta ómnibus y el procedimiento post-hoc adecuados	3
		No responden, o bien el código no realiza bootstrapping	0
	Ortografía y redacción	Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema	3
TOTAL			75
NOTA			7,0