

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 hemos visto 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

1. 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.
2. 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

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 paramétrica 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 "EP12-respuesta-grupo-i", donde i es el número de grupo asignado. Las respuestas deben subirse antes de las 23:30 del sábado 17 de junio.

PREGUNTAS

Grupo 1:

1. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de las mismas unidades (cada una de un árbol diferente) de la variedad Winter Nelly durante las semanas 15 y 20 de crecimiento. Desea saber si existe diferencia en el peso de las peras en este periodo de tiempo. ¿Qué puede concluir?

Semana 15	Semana 20
67,0975	68,53608
65,20517	68,70861
66,46108	67,6798
69,68791	71,75897
67,25581	71,24894
76,00934	77,59157
66,03579	67,70297
75,89793	77,45975
69,54054	71,33677
64,58422	66,71235
66,01125	69,30179

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de algunas unidades (cada una de un árbol diferente) de dos variedades distintas, Winter Nelly y Golden Bosc, durante la semana 15 de crecimiento. Desea saber si el peso de ambas variedades en esta etapa de su desarrollo es el mismo. ¿Qué puede concluir?

Winter Nelly	Golden Bosc
150,6948	130,2934
144,935	192,9736
321,2211	227,2896
153,4385	160,5874
122,8188	132,915
171,2131	118,7482
157,9422	149,1814
219,7431	236,0405
204,1157	191,3631
	133,928
	367,991
	144,3353

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de las mismas unidades (cada una de un árbol diferente) de la variedad Winter Nelly durante las semanas 15 y 20 de crecimiento. Desea saber si existe diferencia en el peso de las peras en este periodo de tiempo. ¿Qué puede concluir?

Semana 15	Semana 20
66,94086	73,18442
64,45227	66,37679
73,04155	74,8069
68,0122	70,09922
67,9556	69,64732
68,83456	71,1403
67,82337	70,15615
76,05828	76,87465
73,8764	75,33465
68,07721	69,8663
73,92357	76,55669

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de algunas unidades (cada una de un árbol diferente) de dos variedades distintas, Winter Nelly y Golden Bosc, durante la semana 15 de crecimiento. Desea saber si el peso de ambas variedades en esta etapa de su desarrollo es el mismo. ¿Qué puede concluir?

Winter Nelly	Golden Bosc
213,41	170,8754
185,9876	144,3119
221,9353	166,2316
320,1673	114,3064
201,3828	112,9957
180,1779	126,6177
168,4547	117,5733
222,9346	128,7313
183,4629	143,4207
	210,4415
	138,7852
	272,3915

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de las mismas unidades (cada una de un árbol diferente) de la variedad Winter Nelly durante las semanas 15 y 20 de crecimiento. Desea saber si existe diferencia en el peso de las peras en este periodo de tiempo. ¿Qué puede concluir?

Semana 15	Semana 20
72,6196	75,71317
65,43683	72,9827
69,57628	71,07957
67,69712	70,93275
74,08071	76,35463
68,65335	70,99375
73,43684	74,77972
65,17816	66,22893
68,57111	70,06971
69,34316	71,02664
66,382	68,64397

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de algunas unidades (cada una de un árbol diferente) de dos variedades distintas, Winter Nelly y Golden Bosc, durante la semana 15 de crecimiento. Desea saber si el peso de ambas variedades en esta etapa de su desarrollo es el mismo. ¿Qué puede concluir?

Winter Nelly	Golden Bosc
106,1104	161,4966
311,9078	176,9673
305,8252	173,6628
114,1533	120,6394
165,3824	238,9576
113,5277	321,0362
358,7241	145,1212
150,6982	206,1193
147,674	163,5642
	141,4695
	147,4953
	167,8039

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de las mismas unidades (cada una de un árbol diferente) de la variedad Winter Nelly durante las semanas 15 y 20 de crecimiento. Desea saber si existe diferencia en el peso de las peras en este periodo de tiempo. ¿Qué puede concluir?

Semana 15	Semana 20
74,46436	76,10206
65,48316	67,53608
70,85975	72,46521
66,60465	71,78476
65,61149	67,81848
67,43604	69,01986
70,70847	71,8967
75,31325	77,47667
75,90031	77,78669
73,29737	75,00866
73,43978	76,31516

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de algunas unidades (cada una de un árbol diferente) de dos variedades distintas, Winter Nelly y Golden Bosc, durante la semana 15 de crecimiento. Desea saber si el peso de ambas variedades en esta etapa de su desarrollo es el mismo. ¿Qué puede concluir?

Winter Nelly	Golden Bosc
135,4449	280,424
144,9288	159,7303
200,543	221,4824
306,6218	399,6019
210,0958	183,6954
158,0067	125,9711
468,6698	145,1197
233,7936	270,5207
226,3744	147,9605
	130,0475
	209,3887
	150,7068

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de las mismas unidades (cada una de un árbol diferente) de la variedad Winter Nelly durante las semanas 15 y 20 de crecimiento. Desea saber si existe diferencia en el peso de las peras en este periodo de tiempo. ¿Qué puede concluir?

Semana 15	Semana 20
64,68956	66,0537
64,86934	69,8352
69,18784	70,19561
70,91504	72,72097
64,31397	66,65992
66,70795	68,18177
64,08226	66,90458
72,01905	73,69559
68,11424	69,84229
73,80995	75,28577
69,60985	70,79695

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. Tras el éxito obtenido recientemente por un equipo de agrónomos en la mejora del cultivo de manzanas de exportación, con la ayuda de algunos modelos estadísticos, un colega dedicado a la producción de peras desea estudiar algunas características de sus productos. Para ello, ha registrado los pesos (en gramos) de algunas unidades (cada una de un árbol diferente) de dos variedades distintas, Winter Nelly y Golden Bosc, durante la semana 15 de crecimiento. Desea saber si el peso de ambas variedades en esta etapa de su desarrollo es el mismo. ¿Qué puede concluir?

Winter Nelly	Golden Bosc
137,3024	265,6924
147,0478	238,226
184,1737	176,9622
228,9383	162,8725
243,3874	257,0427
131,7825	138,5463
156,725	199,1035
151,9297	205,7466
135,5634	141,2139
	174,1134
	393,7728
	142,6864

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 1:

- Obtienen correctamente los datos necesarios para realizar la prueba solicitada, en un formato pertinente.
- Argumentan convincentemente el incumplimiento de todas las condiciones requeridas para aplicar una prueba robusta adecuada.
- Seleccionan, argumentando con gráficos u otras herramientas, una transformación apropiada para poder aplicar una prueba paramétrica a los datos.
- Formulan con claridad y explícitamente hipótesis nulas y alternativas adecuadas para responder la pregunta planteada, tanto en lenguaje natural como en lenguaje matemático.
- Realizan de forma completa y correcta una prueba paramétrica adecuada para la pregunta planteada en el enunciado, utilizando los datos correctos.
- 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.
- 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.

- Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema.

Pregunta 2:

- Proponen una pregunta de investigación, interesante y novedosa, que involucra la comparación de las medias de dos grupos independientes de personas encuestadas en la Encuesta Casen 2017.
- 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.
- 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.
- Basándose en el análisis anterior, proponen explícitamente una prueba robusta que permite docimar las hipótesis propuestas, justificando su elección apropiadamente.
- 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.
- Entrega una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema.
- 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.
- Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema.

Pregunta 3:

- 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 Encuesta Casen 2017.
- 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.
- 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.
- Basándose en el análisis anterior, proponen explícitamente una prueba robusta que permite docimar las hipótesis propuestas, justificando su elección apropiadamente.
- 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.
- 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.
- Entrega una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema.

- 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.
- Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema.