

# Diferencias en el tiempo de supervivencia de pacientes de cáncer con respecto al órgano en el cual se desarrolla el tumor

Transformaciones para homogeneizar la varianza en el modelo de análisis de varianza con un criterio de clasificación

Se aplicó un tratamiento con ascorbato a 64 pacientes con cáncer avanzado de estómago, bronquios, colon, ovario y pecho. El propósito del estudio era determinar si el tiempo de supervivencia del paciente difería de acuerdo con el órgano afectado por el cáncer. (Referencia: Cameron, E. and Pauling, L. (1978) Supplemental ascorbate in the supportive treatment of cancer: re-evaluation of prolongation of survival times in terminal human cancer. Proceedings of the National Academy of Science USA, 75, 4538-4542.)

Los datos se muestran en la tabla 1, y corresponden al tiempo de supervivencia (en días) de cada paciente tratado con ascorbato.

En la figura 1 se muestran los diagramas de cajas correspondientes a estos datos. En ellos puede verse que el tiempo de supervivencia de las pacientes con cáncer de seno es mayor que el de aquellos pacientes con cánceres localizados en otros órganos. Sin embargo, también se observa que la dispersión parece aumentar con el tiempo de supervivencia, lo cual nos hace sospechar que puede haber heterocedasticidad en los datos.

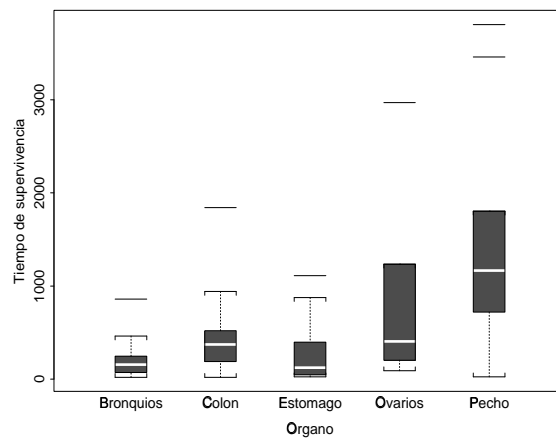


Figura 1: Gráficos de caja para los tiempos de supervivencia de acuerdo con el órgano en el cual se encuentra el tumor

Cuando ajustamos un modelo de análisis de varianza con un criterio de clasificación, obtenemos la siguiente tabla de análisis de varianza:

Analysis of Variance Table

Estómago	Bronquios	Colon	Ovario	Pecho
124	81	248	1234	1235
42	461	377	89	24
25	20	189	201	1581
45	450	1843	356	1166
412	246	180	2970	40
51	166	537	456	727
1112	63	519		3808
46	64	455		791
103	155	406		1804
876	859	365		3460
146	151	942		719
340	166	776		
396	37	372		
	223	163		
	138	101		
	72	20		
	245	283		

Tabla 1: Tiempos de supervivencia de pacientes con cancer con respecto al órgano en el cual se encuentra el tumor

Response: Survival

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Organ	4	11535761	2883940	6.433437	0.0002294532
Residuals	59	26448144	448274		
Total	63	37983905			

De acuerdo con esta tabla, existe una diferencia significativa entre los tiempos promedio de supervivencia para los diferentes órganos (rechazamos  $H_0 : \mu_1 = \mu_2 = \dots = \mu_6$ ).

Analicemos los residuos de este modelo, los cuales se muestran en la figura 2. Aquí se observa un claro patrón de apertura, lo cual nos indica que es necesario realizar una transformación de los datos para que un análisis por medio de un modelo lineal sea válido.

Con el fin de estimar empíricamente la transformación que debe ser empleada para estabilizar la varianza, graficamos los logaritmos de los promedios de cada grupo versus los logaritmos de sus desviaciones standard (ver figura 3). Nuevamente es claro un patrón de crecimiento. Cuando ajustamos un modelo de regresión a estos puntos, obtenemos una pendiente de 0.9604954 . Esto sugiere el uso de la transformación logaritmo.

Usando los logaritmos de los datos, obtenemos la tabla que se muestra a continuación

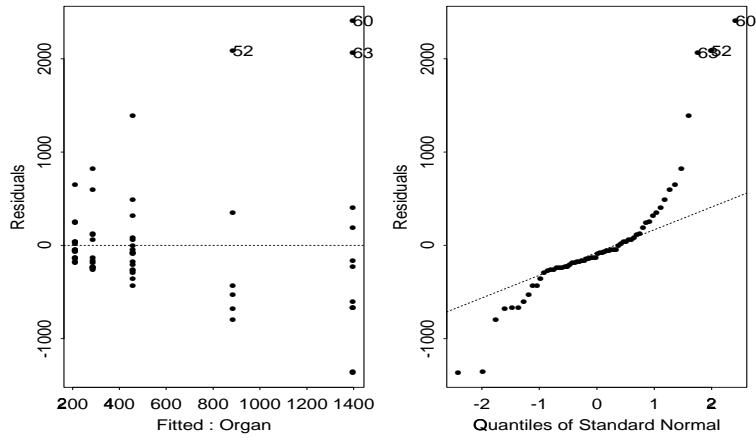


Figura 2: Gráficos de residuos para el modelo de análisis de varianza con un criterio de clasificación

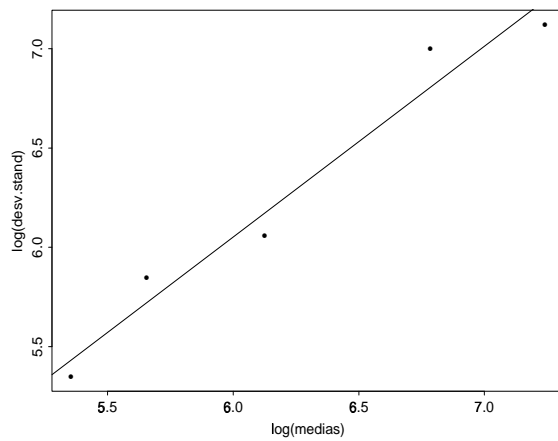


Figura 3: Gráfico del logaritmo de las medias de los grupos vs. el logaritmo de las desviaciones standard.

## Analysis of Variance Table

Response: log(Survival)

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Organ	4	24.48656	6.121639	4.285967	0.004122
Residuals	59	84.26959	1.428298		
Total	63	108.7561			

Nuevamente, rechazamos la hipótesis de igualdad entre las medias de los grupos, lo cual nos indica que existen diferencias en el tiempo de supervivencia de acuerdo con el órgano en el cual se desarrolla el tumor.

El nuevo gráfico de residuos se muestra en la figura 4. En ella podemos ver que desapareció el patrón de apertura, y que el gráfico normal de residuos tiene un comportamiento sustancialmente mejor. Sin embargo, hay tres observaciones que parece muy alejadas del resto, y que deben ser analizadas con mayor detenimiento para determinar si se trata de observaciones atípicas.

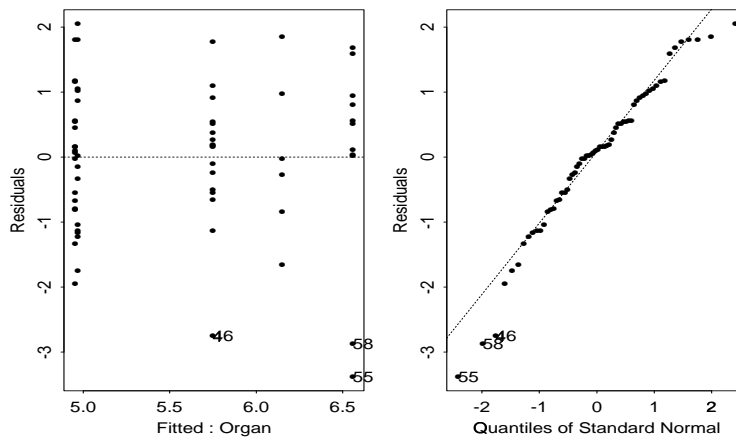


Figura 4: Gráficos de residuos para el modelo de análisis de varianza con un criterio de clasificación