

Taller de Consultoria - TP3

Gonzalo Barrera Borla

23/09/2019

Setup

Ejercicio 1

Los datos siguientes corresponden a un experimento realizado por Charles Darwin en 1876. En cada maceta se plantan dos brotes de maíz, uno producido por fertilización cruzada, y el otro por auto-fertilización. El objetivo era mostrar las ventajas de la fertilización cruzada. Los datos son las altura finales de las plantas después de un período de tiempo. ¿Alguno de los dos tipos de maíz es demostrablemente mejor?. Si es así, ¿cómo se puede describir la diferencia?.

Sea Ω_0 la familia de distribuciones tal que

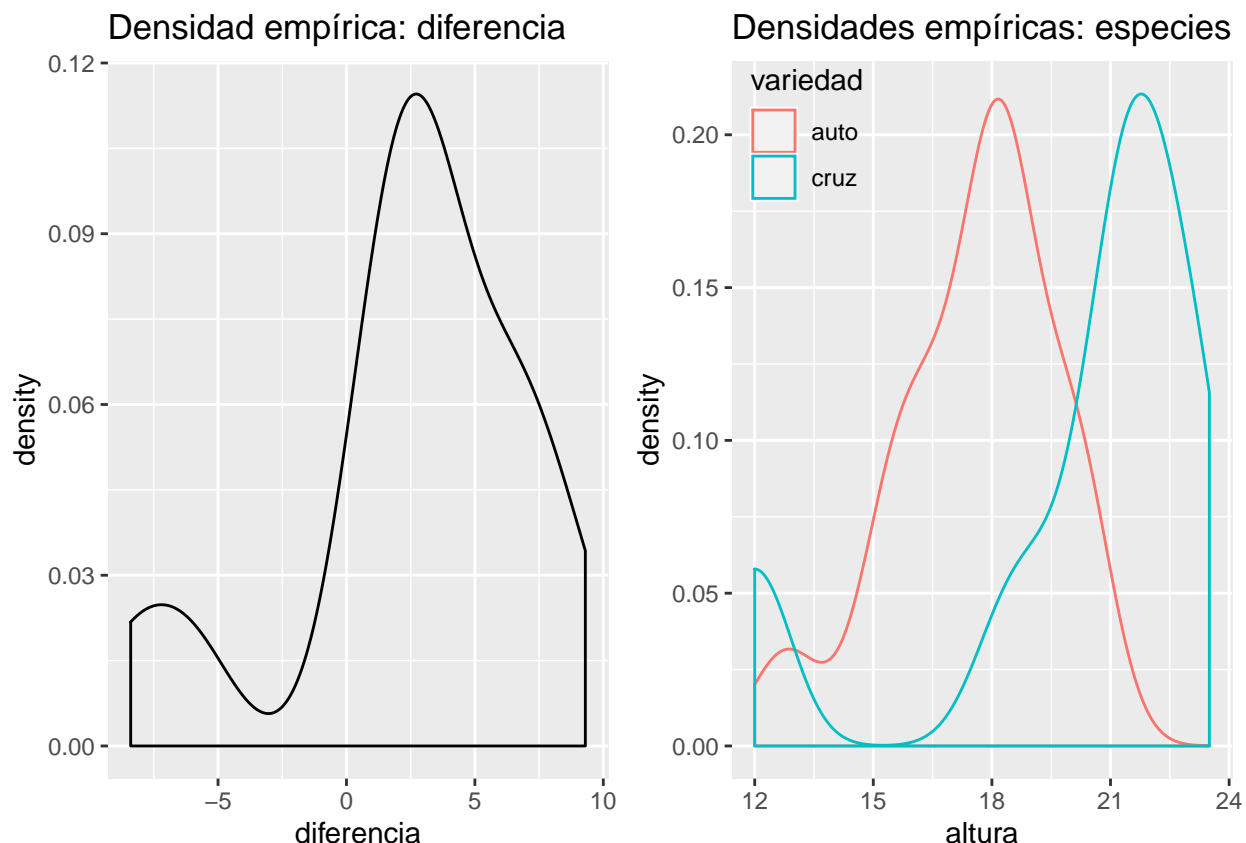
$$\Omega_0 = \{F : F \text{ es absolutamente continua con única mediana en } 0\}$$

Sea Y_i la altura del brote de maíz producido por fertilización cruzada de la maceta $i \in \{1, \dots, n\}$, $n = 15$, y X_i la altura del producido por auto-fertilización. Supondremos además, razonablemente, que $Y_i \stackrel{iid}{\sim} F_Y(t) = F(t - \theta_Y)$, $F \in \Omega_0$ (es decir que F_Y es absolutamente continua con única mediana en θ_Y). Análogamente, $X_i \stackrel{iid}{\sim} F_X(t) = F(t - \theta_X)$. Como “*el objetivo era mostrar las ventajas de la fertilización cruzada*”, una forma razonable de plantear este test será:

$$H_0 := \theta_Y = \theta_X \quad \text{vs.} \quad H_1 := \theta_Y > \theta_X$$

Como los brotes están naturalmente apareados en sus respectivas macetas, es razonable realizar un test de muestras apareadas. Para determinar qué clase de test realizar, nos resta contestar: ¿Es normal la distribución de las alturas de ambos tipos de brotes? Si la respuesta es “sí”, podemos usar un “test t”, mientras que si no será conveniente recurrir a alguna alternativa no paramétrica, como el test de Wilcoxon de rangos signados, que sólo requiere que la distribución bajo la hipótesis nula sea simétrica.

Para testear la normalidad de los datos, comensamos por graficar la estimación empírica de la densidad de cada tipo de maíz, y de la diferencia de alturas por maceta:



Las densidades individuales *podrían* llegar a ser normales, aunque tienen colas a izquierda un tanto pesadas, lo cual se nota aún más en la densidad de las diferencias. Para cuantificar esta primera impresión, recurrimos al test de Shapiro univariado. Para la diferencia $D_i = Y_i - X_i$, el p-valor es de 0.093, que dependiendo del nivel de significación utilizado puede alcanzar o no para rechazar la normalidad. Si realizamos dos tests por separado, para la muestra (X_1, \dots, X_n) de brotes autofertilizados obtenemos un p-valor 0.38 con lo cual es razonable asumir su normalidad, pero al aplicar el test a la muestra de fertilización cruzada, obtenemos un p-valor de 9.7×10^{-4} que nos lleva a rechazar su normalidad. Resulta difícil suponer que por casualidad las diferencias de altura resulten normales si cada muestra no tiene una distribución normal subyacente, así que por seguridad convendrá recurrir a un test no paramétrico.

Nótese que bajo $H_0 := \theta_Y = \theta_X \Rightarrow F_X = F_Y$, de manera que la distribución de $Y_i - X_i$ será simétrica, y podemos utilizar un test de Wilcoxon para datos apareados con confianza.

Este test arroja un p-valor de 0.021, de manera que con el tradicional criterio de significación $\alpha = 0.05$, podemos rechazar la hipótesis nula y concluir que la diferencia entre las medianas de los brotes de fertilización cruzada y los autofertilizados es positiva. Aprovechando el hecho de que el estadístico T^+ es un estimador de Hodges-Lehmann, podemos encontrar también el intervalo de confianza de nivel 95% correspondiente al test, que resulta ser $[1, \infty)$.

Por último, y a modo ilustrativo, incluimos en la siguiente tabla los p-valores e intervalos de confianza de nivel 95% correspondientes a cuatro alternativas de test razonables en este problema:

Muestras	Test	p-valor	Lím. inf. IC
apareadas	Wilcoxon	0.0206	1.0000
2 muestras	Wilcoxon	0.0013	2.0000
apareadas	T de Student	0.0251	0.4634
2 muestras	T de Student	0.0118	0.7666

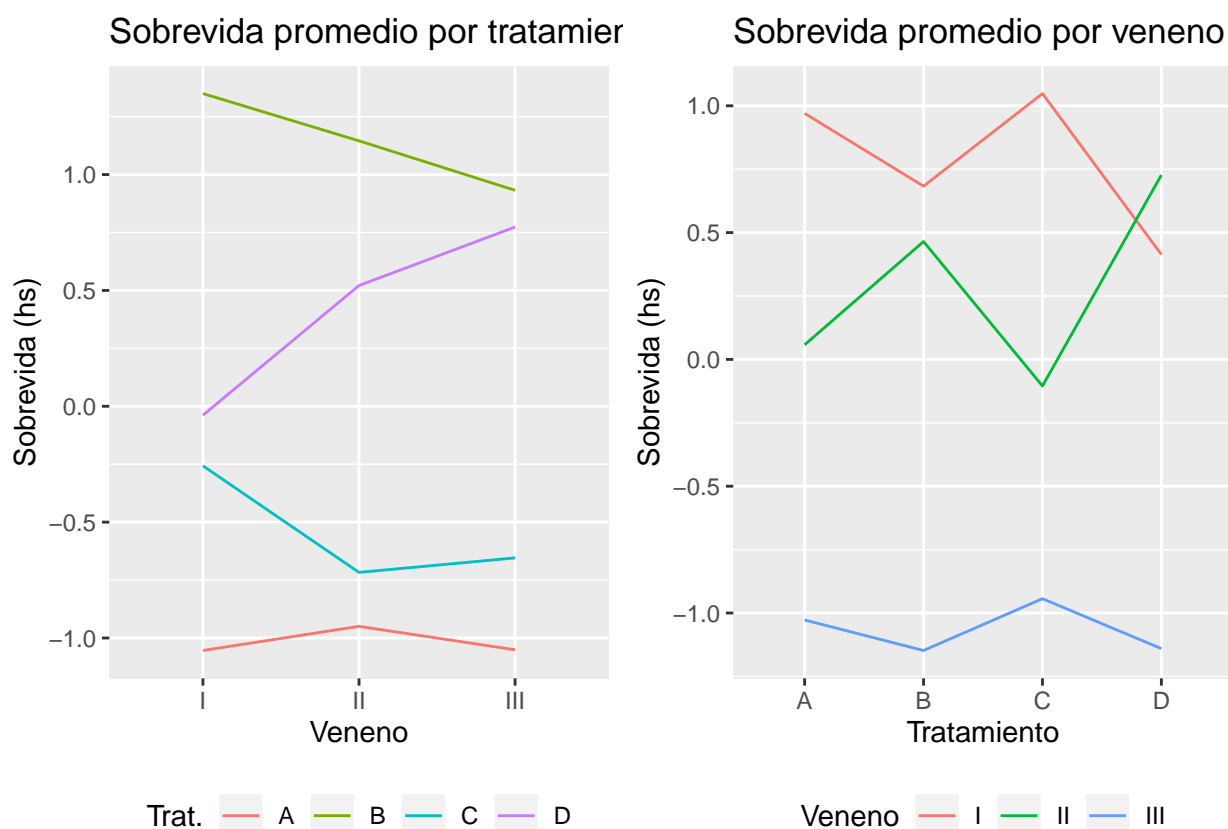
Es decir, que en todos los casos hubiésemos rechazado la hipótesis nula, pero la estimación de la diferencia en las medianas/medias hubiese sido bastante distinta.

Ejercicio 2

Se dan a continuación los tiempos de sobrevida (en unidades de 10 horas) de animales, sometidos a 3 tipos de veneno, y 4 tratamientos antitóxicos. Cada combinación veneno-tratamiento se prueba con 4 animales. Describir la influencia de ambos factores en la sobrevida. ¿Hay algún tratamiento demostrablemente mejor?

Para hacer más intuitiva la información, cambiamos las unidades de **sobrevida** de modo que esté en horas. Sea y_{ijk} la sobrevida del k -ésimo animal, sometida al tratamiento j para el veneno i , en general planteamos $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, donde $i \in \{I, II, III\}$, $j \in \{A, B, C, D\}$, $k \in \{1, 2, 3, 4\}$ y $\epsilon_{ijk} \sim N(0, \sigma^2)$.

Para darnos una idea de la relación entre venenos y tratamientos, comenzamos graficando los perfiles:



A primera vista, no se observa paralelismo en ninguno de los dos gráficos. Sin embargo, salvo un ligero entrecruzamiento en los perfiles de los venenos II y III para los tratamientos C y D, los perfiles no se cruzan en ningún otro punto, por lo que tal vez sea razonable ignorar las interacciones y considerar un modelo de efectos únicamente aditivos.

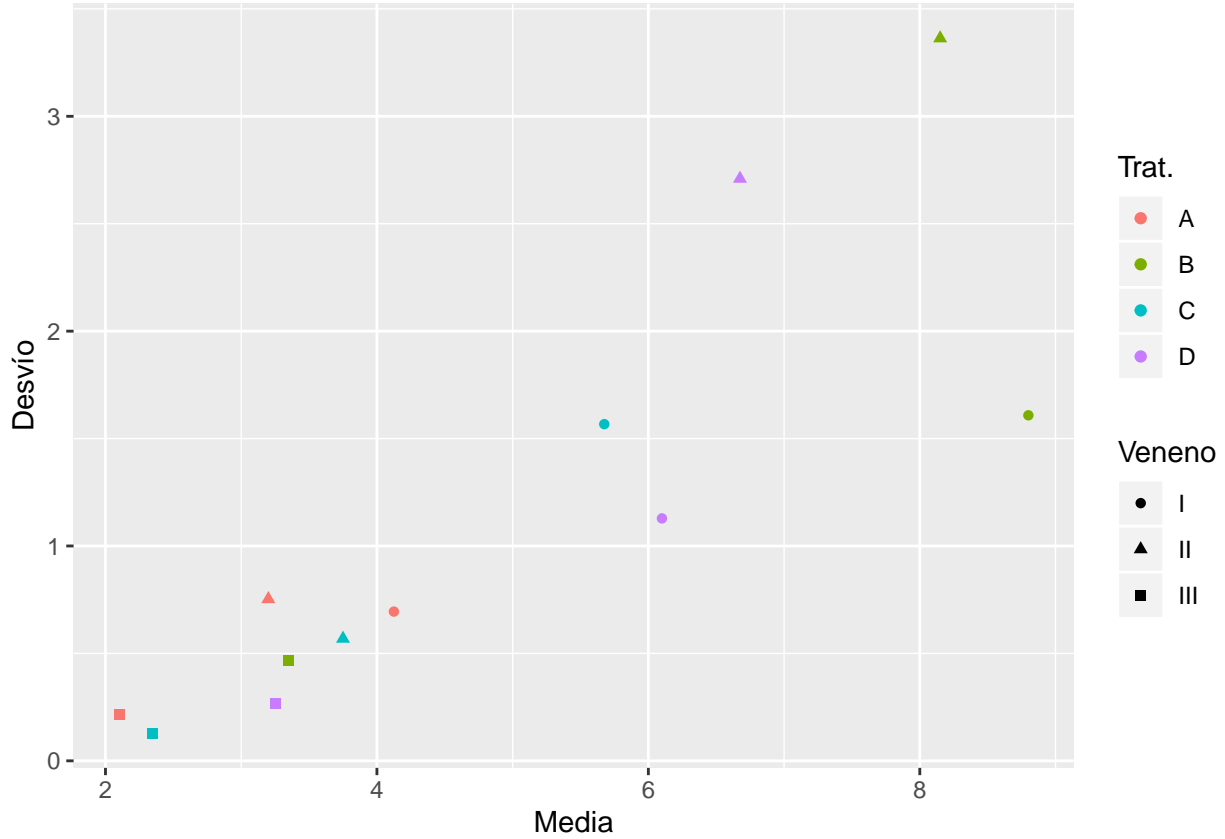
Por otra parte, debemos corroborar el fortísimo supuesto de *homocedasticidad* de los ϵ_{ijk} que hicimos inicialmente. Seber [1977, p. 195] siguiendo a Scheffé menciona que cuando se cuenta con un diseño balanceado como aquí, la desigualdad en las varianzas para cada combinación de bloque y tratamiento no hace demasiado “daño”, siempre y cuando la razón entre el mayor y el menor desvío estándar es menor a 2. Lamentablemente, nuestros datos son “bochornosos” en tal sentido:

Table 2: La tasa entre los desvios de II:B y III:C es mayor a 26

veneno	trt	mean	sd
II	B	8.150	3.363
II	D	6.675	2.710
I	B	8.800	1.608
I	C	5.675	1.567
I	D	6.100	1.128
II	A	3.200	0.753
I	A	4.125	0.695
II	C	3.750	0.569
III	B	3.350	0.465
III	D	3.250	0.265
III	A	2.100	0.216
III	C	2.350	0.129

Antes de proseguir, nos convendrá entonces considerar una transformación [estabilizadora de la varianza](#) que al menos morigere esta situación.

Si graficamos el desvío en función de la media de cada grupo, observamos una estructura razonablemente lineal en la media:

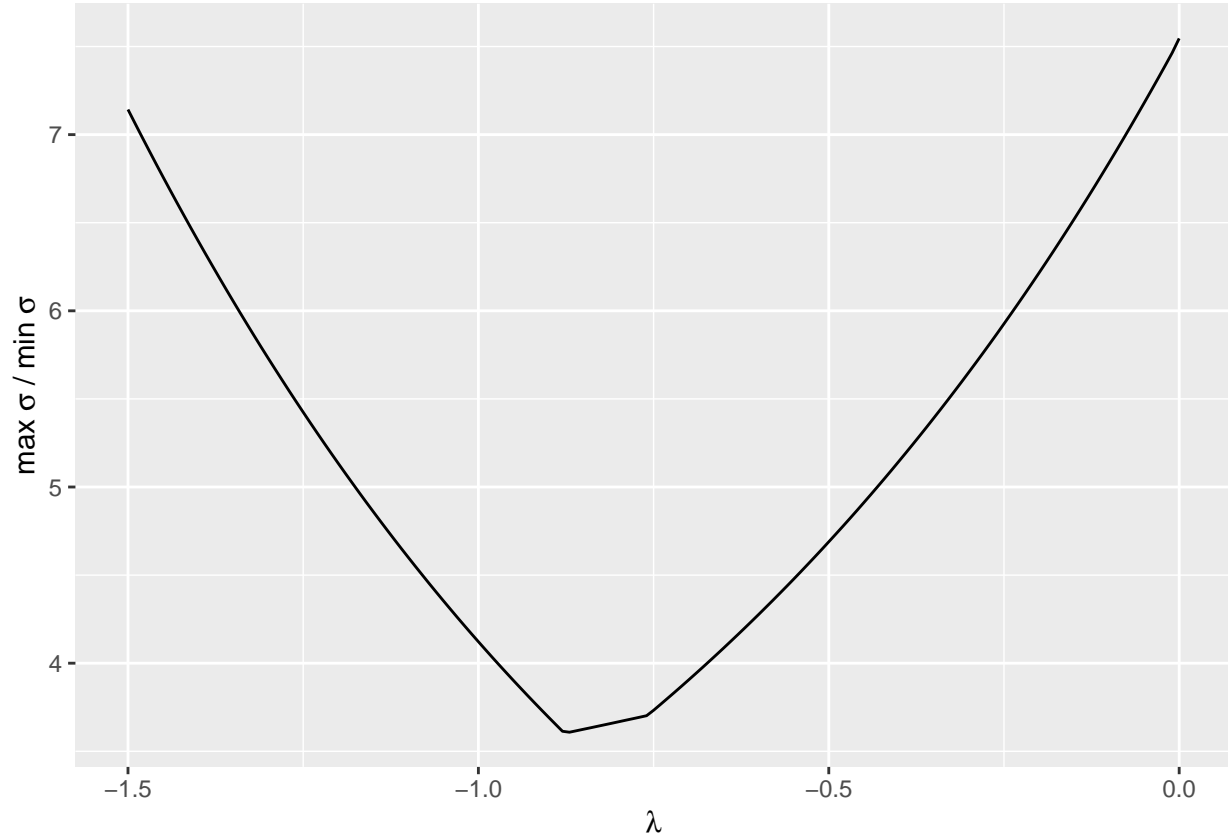


En este caso, donde $\sigma \approx k \times \mu$, la transformación sugerida es el logaritmo. Esto es razonable, ya que al hablar por ejemplo del logaritmo base 2 de la *sobrevivida*, un tratamiento que duplica la supervivencia, aumentará $\log_2(y)$ en una unidad. Aplicando esta transformación, el la razón entre el máximo y mínimo desvíos anteriormente mencionada mejora bastante, pero sigue estando por encima de 7, muy lejos del corte de 2 que propone Seber.

Haremos un intento más, y en lugar de considerar la transformación logarítmica, buscaremos la transformación “óptima” según este criterio de optimalidad, de entre la familia de transformaciones [Box-Cox](#) de un parámetro, que incluye el logaritmo natural como un caso especial cuando $\lambda = 0$:

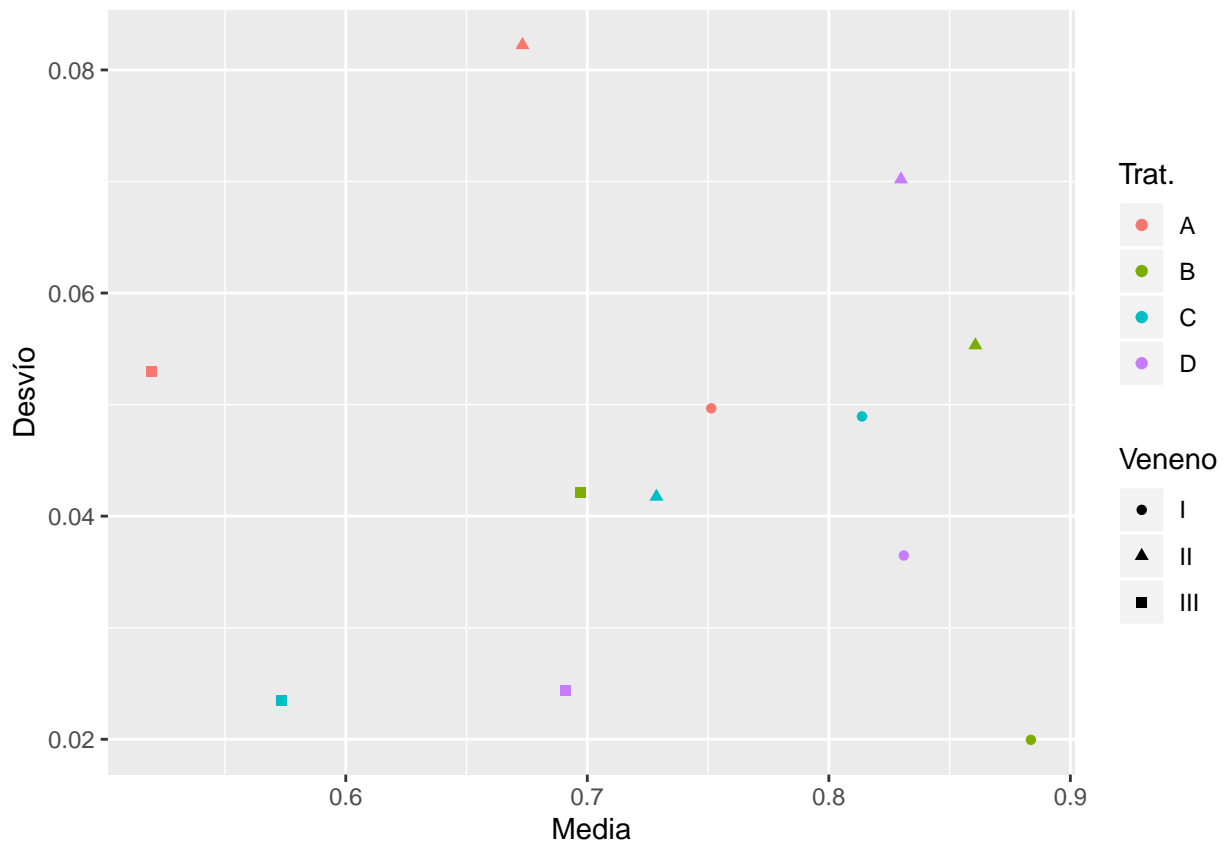
$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0, \end{cases}$$

Si graficamos la relación entre el mayor y menor desvío estándar por bloque y tratamiento para distintos λ , observamos que el mínimo se encuentra alrededor de $\lambda = -0.87$, que lleva la tasa de desvíos a ≈ 3.6 :

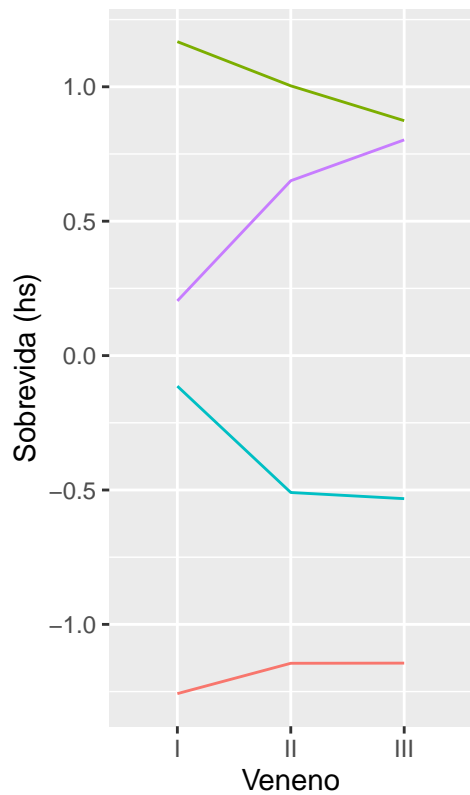


Reducir la tasa de desvíos a 3.6 sigue sin ser suficiente para el criterio originalmente autoimpuesto, pero tampoco está demasiado lejos, y es una enorme mejoría respecto a la tasa original de más de 26, así que aplicaremos la transformación de Box-Cox y calificaremos momentáneamente de “tolerable” la heterocedasticidad restante. Lo que sí, para ayudar a la comprensión, en lugar de tomar el óptimo $\lambda^* = -0.87$, tomaremos el mucho más interpretable $\lambda = -1$, que consigue una tasa de desvíos similar (4.12), pero provee una transformación más clara: $g(y) = 1 - \frac{1}{y}$, que cuando $y \in (1, +\infty) \Rightarrow g(y) \in (0, 1)$.

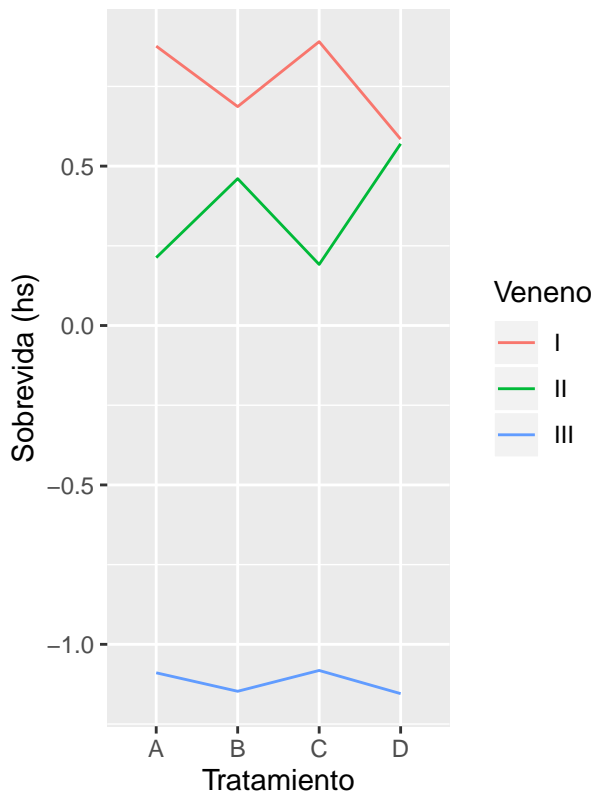
Aplicamos la transformación, y volvemos a graficar los perfiles.



Sobrevida promedio por tratamier



Sobrevida promedio por veneno



Las varianzas observadas para cada tratamiento y veneno ahora ya no presentan ninguna tendencia obvia, y

aunque no sean realmente paralelos los perfiles, al menos no se observan entrecruzamientos. Comenzamos planteando un modelo con interacciones, $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, y comenzamos testeando la hipótesis de “no-interacción” $H_0 : \gamma_{ij} = 0 \forall i \in \{II, III\}, j \in \{B, C, D\}$:

Coefs	GL	C. Medios	F obs.	P-valor
veneno	2	72.6	72.6	2.31e-13
trt	3	28.3	28.3	1.38e-09
veneno:trt	6	1.09	1.09	0.387
Residuals	36	NA	NA	NA

Efectivamente, no sólo el p-valor del estadístico F correspondiente a las interacciones es altísimo, sino que la razón entre los “cuadrados medios” de las interacciones y cualquiera de los tratamiento es realmente muy pequeña, así que podemos despreocuparnos de las interacciones. A continuación consideramos un modelo aditivo únicamente, y repetimos los tests de significatividad para los efectos principales, tanto de los tratamientos como de los venenos. Mantenemos la transformación de la variable respuesta con $\lambda = -1$, ya que la elección del parámetro se realizó independientemente del modelo a ajustar.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \text{con} \quad \epsilon_{ijk} \sim N(0, \sigma^2)$$

Los principales descriptores del modelo muestran que

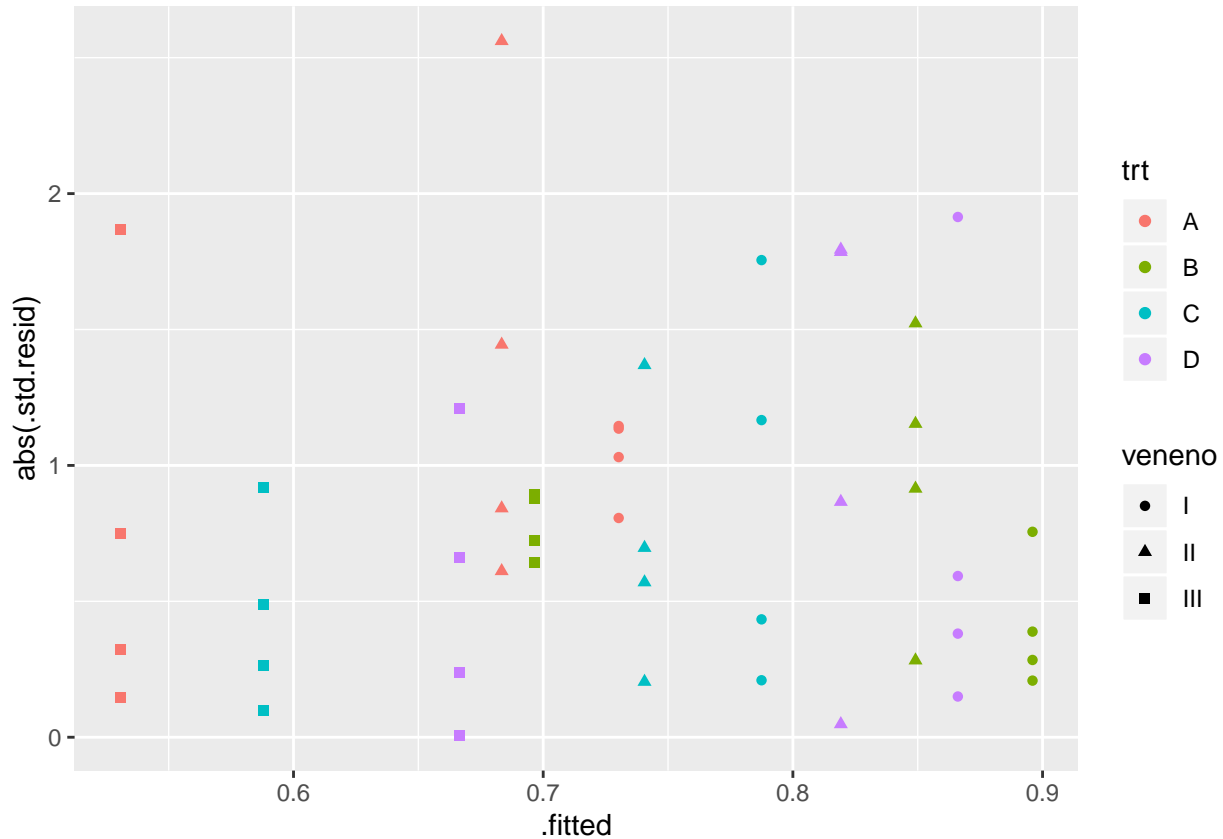
- es globalmente significativo,
- tanto los efectos principales de los venenos como de los tratamientos son significativos y
- todos los coeficientes son significativamente distintos de 0 a nivel 0.01 o menor:

R ² aj.	F obs.	P-valor
0.826	45.5	6.97e-16

Coefs	GL	C. Medios	F obs.	P-valor
veneno	2	71.7	71.7	2.86e-14
trt	3	28	28	4.19e-10
Residuals	42	NA	NA	NA

Coef.	Estimado	P-valor
(Intercept)	0.73	6.91e-36
venenoII	-0.0469	0.0103
venenoIII	-0.2	1.69e-14
trtB	0.166	2.66e-10
trtC	0.0572	0.00689
trtD	0.136	3.35e-08

Graficamos el módulo de los residuos estudentizados contra los valores ajustados $\widehat{g(y)}$, y nos tranquiliza la ausencia de estructura evidente. Hay una observación con un residuo > 2.5 que puede merecer mayor consideración, pero resulta razonable dejarla así, ya que de quitarla el diseño deja de ser balanceado, y ya bastantes transformaciones hemos tenido que hacer para obtener el modelo actual:



Hasta ahora, hemos concluido que bajo una transformación g de los datos, - las observaciones son aproximadamente homocedásticas, y - un modelo multiplicativo no aporta más que el aditivo para el análisis.

Nos resta *describir* la influencia de los venenos y los tratamientos en la sobrevida, con la complicación de que estamos trabajando sobre datos transformados. Como la transformación es estrictamente creciente en la sobrevida, el signo de los efectos se mantiene. En principio en orden *creciente* de sobrevida, - los venenos se ordenan $III < II < I$, y - los tratamientos se ordenan $A < C < D < B$.

Usando el método de “Diferencias Significativas Honestas” de Tukey, podemos comparar de a pares los venenos y tratamientos. En las unidades transformadas $g(y)$ que venimos trabajando, todos los venenos y tratamientos son distintos entre sí, salvo los tratamientos D y B, para los cuales el IC de su diferencia de medias incluye al 0.

Coef.	Comp.	Dif. medias	Lím. inf. IC	Lím. sup. IC	P-valor
veneno	II-I	-0.0469	-0.0892	-0.00451	0.0272
veneno	III-I	-0.2	-0.242	-0.157	1.12e-12
veneno	III-II	-0.153	-0.195	-0.11	1.49e-10
trt	B-A	0.166	0.112	0.22	1.58e-09
trt	C-A	0.0572	0.00336	0.111	0.0335
trt	D-A	0.136	0.082	0.19	1.98e-07
trt	C-B	-0.109	-0.162	-0.0547	1.72e-05
trt	D-B	-0.0299	-0.0838	0.0239	0.455
trt	D-C	0.0786	0.0248	0.132	0.00184

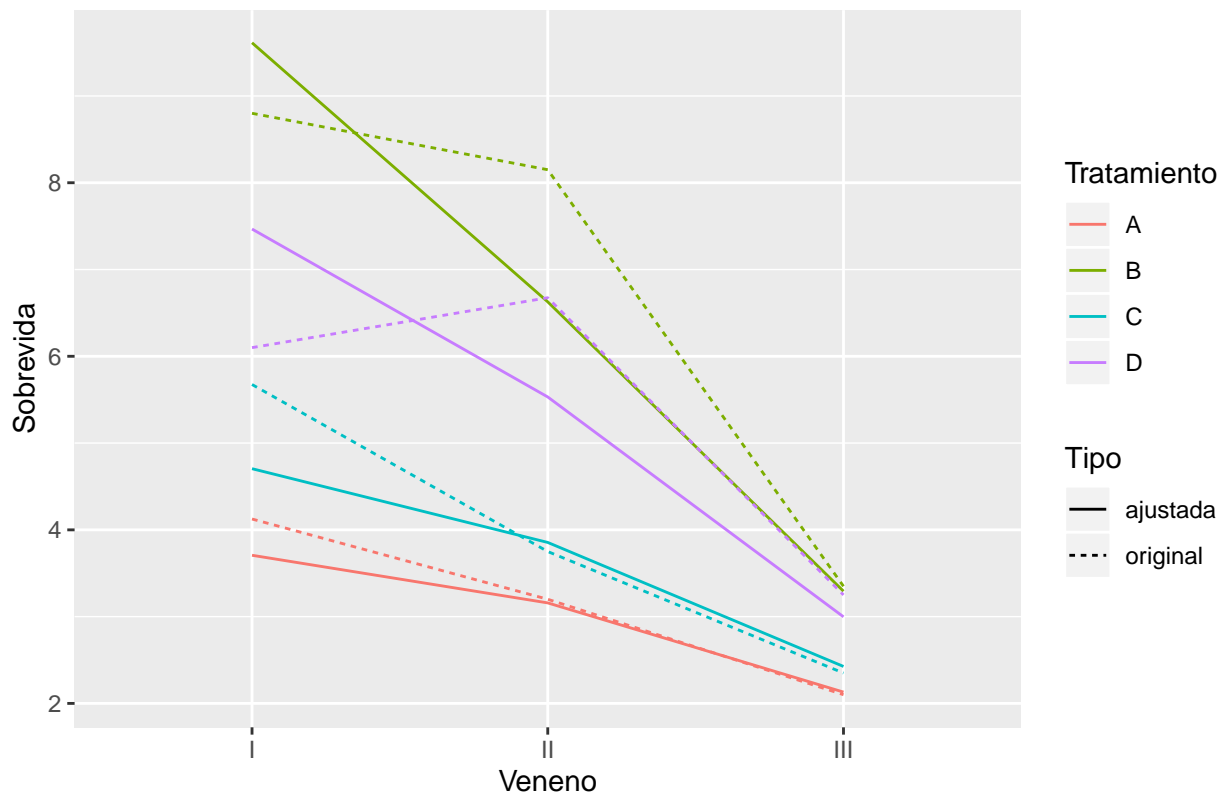
La inferencia mas allá de estas conclusiones es algo limitada. En primer lugar, al no contar con un control, ni siquiera podemos argumentar que alguno de los tratamientos más efectivo que un simple placebo. Por otra parte, tampoco tenemos una forma directa de responder *por cuánto* mejora la sobrevida en términos

absolutos en la escala original. Lo que sí podemos dar, es tomar $g^{-1}(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$ como una aproximación de la sobrevida media para cada combinación de venenos y tratamientos. Ordenadas por g^{-1} , tenemos:

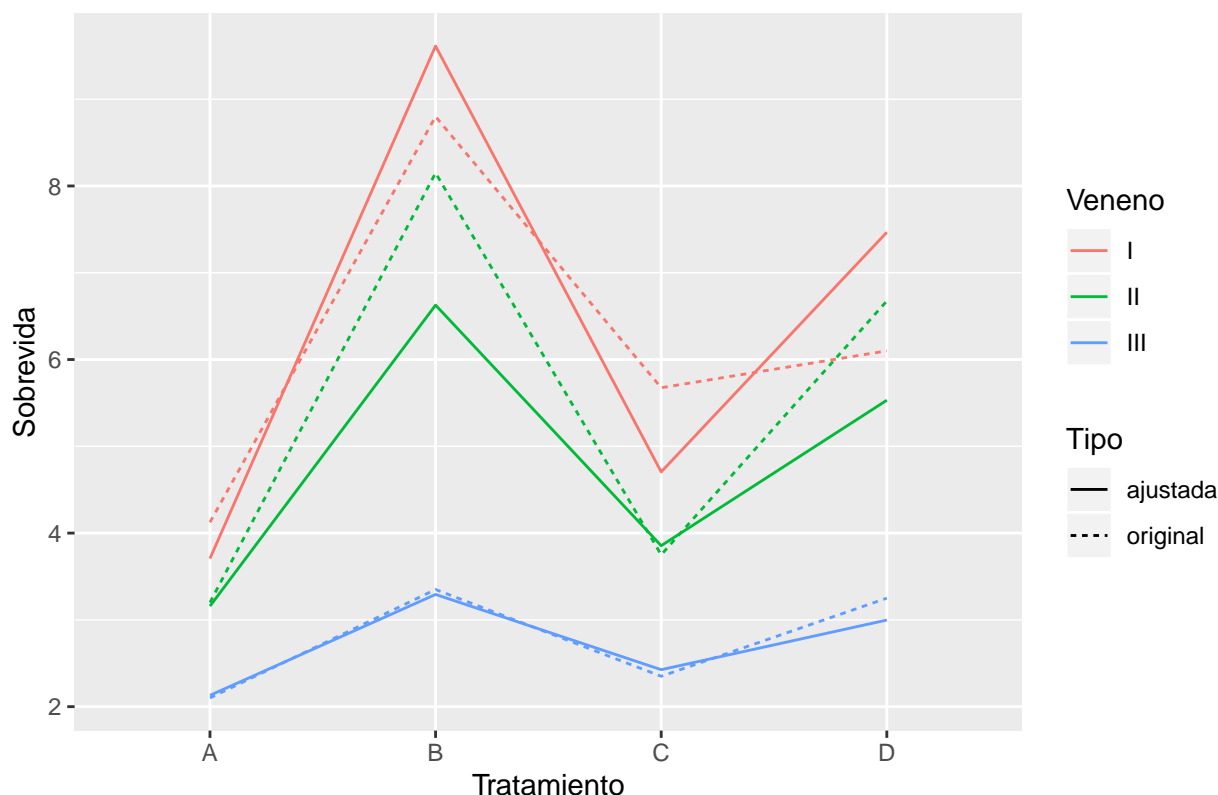
Veneno	Tratamiento	Original	Ajustada
I	B	8.8	9.61
I	D	6.1	7.47
II	B	8.15	6.63
II	D	6.68	5.53
I	C	5.68	4.7
II	C	3.75	3.85
I	A	4.12	3.71
III	B	3.35	3.29
II	A	3.2	3.16
III	D	3.25	3
III	C	2.35	2.43
III	A	2.1	2.13

O si lo preferimos, podemos volver a graficar los perfiles de sobrevida por veneno y tratamiento, con los valores originales y los ajustados. Esta vez, no restamos la media, para darnos una mejor idea de las sobrevidas comparadas:

Perfiles de sobrevida (original y ajustada) para cada tratamiento



Perfiles de sobrevida (original y ajustada) para cada veneno



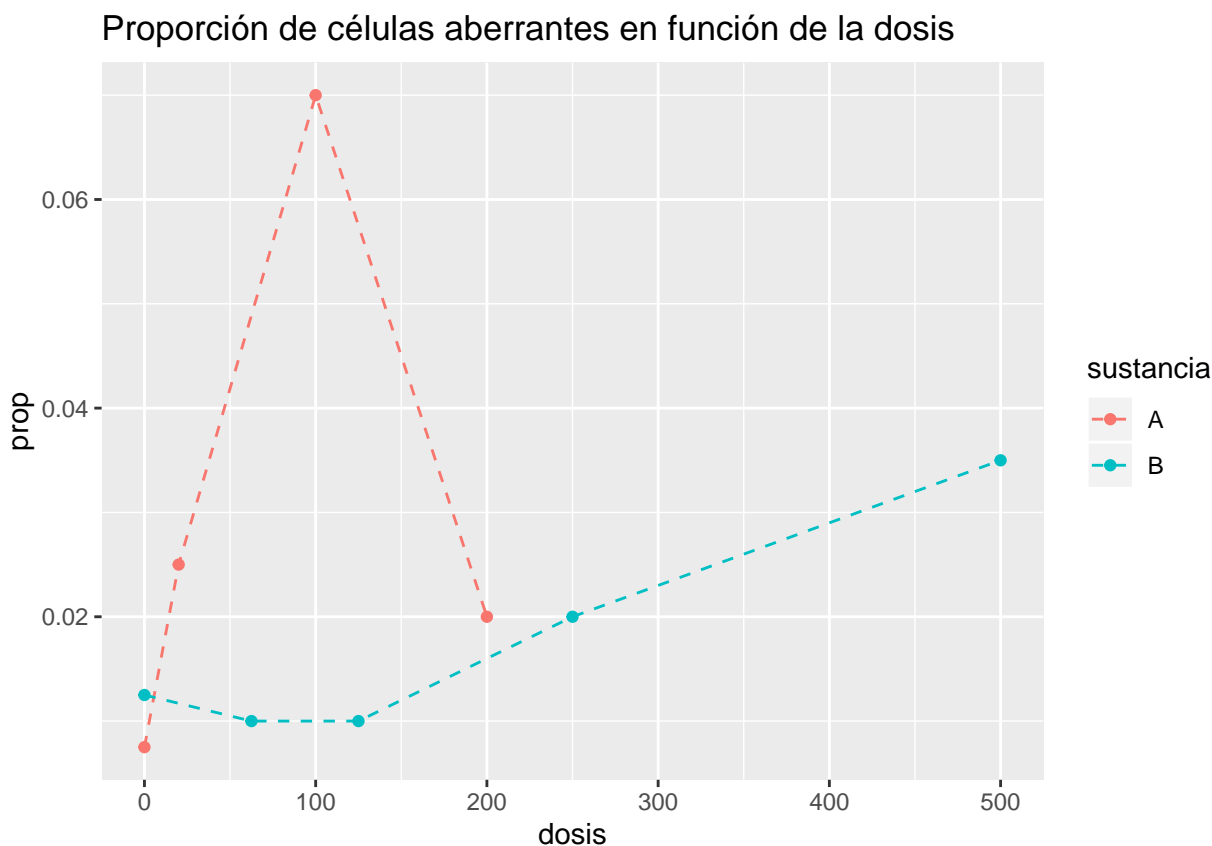
Ejercicio 3

Un ensayo de aberración cromosómica sirve para determinar si una sustancia produce cambios estructurales en los cromosomas. Los datos que siguen muestran los resultados de la aplicación de dos sustancias (“A” y “B”), cada una en varias dosis (la primera “0” es un control). Para cada dosis se muestrea una cantidad de células, y mediante un análisis microscópico se determina la cantidad de células con aberraciones. Se desea determinar si hay diferencias en los efectos de las dos sustancias; y en caso afirmativo, describirlas. Los datos son: sustancia, dosis (en mg./ml.), cantidad total de células muestreadas, y cantidad de células aberrantes. Tener en cuenta que “Dosis 0 de A” es lo mismo que “Dosis 0 de B”.

Obs.	Sustancia	Dosis	Muestreadas	Aberrantes	Prop.
1	A	0.0	400	3	0.0075
2	A	20.0	200	5	0.0250
3	A	100.0	200	14	0.0700
4	A	200.0	200	4	0.0200
5	B	0.0	400	5	0.0125
6	B	62.5	200	2	0.0100
7	B	125.0	200	2	0.0100
8	B	250.0	200	4	0.0200
9	B	500.0	200	7	0.0350

Inmediatamente observamos que la cantidad de células muestreadas no es la misma en todas las repeticiones. Para hacer una comparación más justa, no intentaremos inferir directamente la cantidad de células aberrantes en cada muestra, sino la *proporción* de ellas. A continuación, realizamos un primer gráfico de esta proporción

en función de la dosis de cada sustancia:



La dosis de la sustancia B parece afectar la proporción de células aberrantes de manera aproximadamente lineal, pero el comportamiento para la sustancia A es realmente misterioso. Uno podría suponer que la observación #3 es un *outlier* y considerar eliminarla, pero contando con sólo tres observaciones con dosis positivas de la sustancia A, descartar una equivale a descartar un tercio de los datos disponibles.

A continuación, hemos de elegir un modelo a ajustar. Una primera condición necesaria, será que estime idénticamente la proporción de células aberrantes cuando la dosis de la sustancia aplicada es nula. En segundo lugar, ya que no contamos con observaciones con dosis positivas de ambas sustancias, no tenemos otra opción más que considerar efectos aditivos, aunque químicamente resulte ingenuo tal planteo. Por último, considerando el gráfico anterior, plantearemos una relación dosis-respuesta cuadrática en la sustancia A. En un principio, tal idea puede resultar un mero artilugio matemático para “ajustar mejor”, pero una somera búsqueda por la internet indica que existe tal cosa como una relación [dosis-respuesta cuadrática](#). En este trabajo por ejemplo, los investigadores encuentran que un modelo cuadrático en la dosis de cafeína estima mucho mejor que otro lineal la actividad cerebral registrada en un electroencefalograma (EEG). Sean, para cada observación, p, d_A, d_B, n la proporción de células aberrantes, la dosis de las sustancias A y B, y en número de células muestreadas, respectivamente. Además, llamemos π a la proporción de células aberrantes “basal”, en ausencia de sustancia alguna. Nuestro modelo será:

$$p_i = \pi + \beta_1 d_{A,i} + \beta_2 (d_{A,i})^2 + \beta_3 d_{B,i} + \beta_4 d_B + \beta_5 n_i + \epsilon_i \quad \text{con } \epsilon_i \sim N(0, \sigma^2)$$

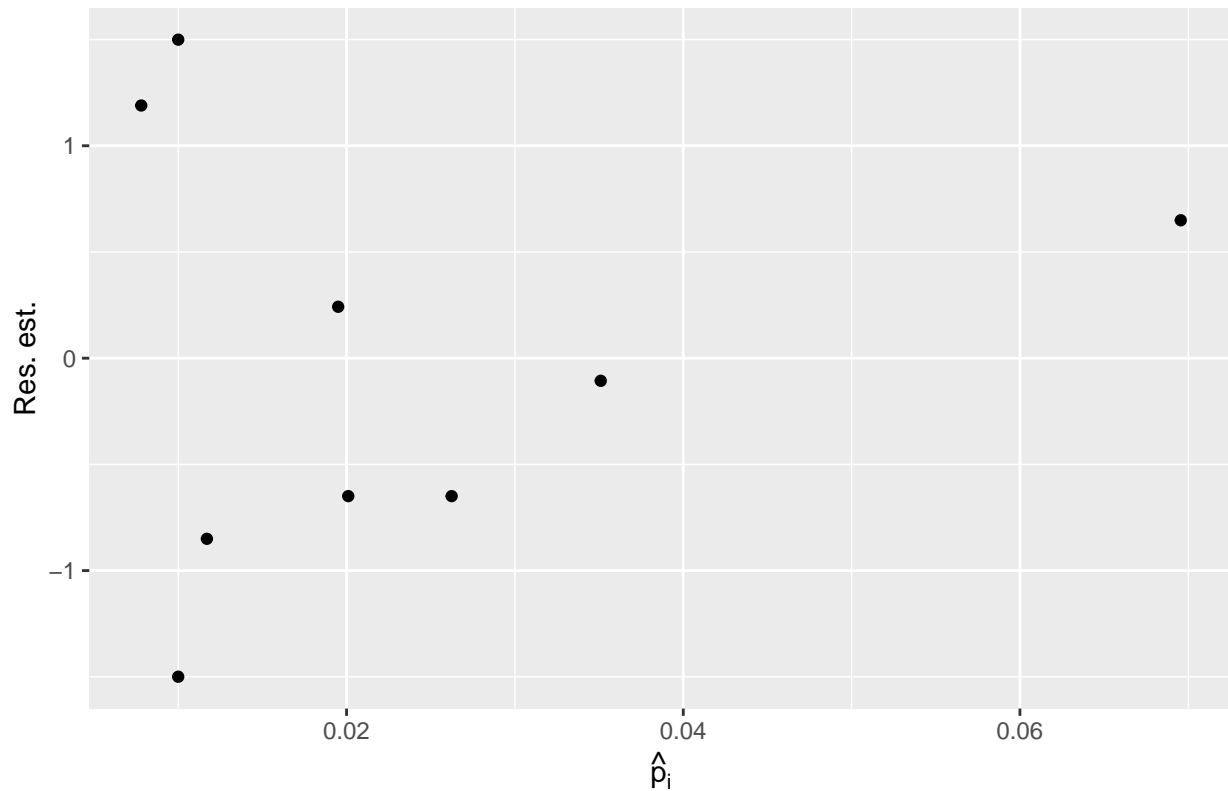
El ajuste es *sorprendentemente* bueno, para lo descorazonadores que resultaban los datos originalmente. A continuación presentamos un resumen de los estadísticos principales, incluido el R^2 ajustado, el p-valor del test F de significatividad global, y los coeficientes estimados:

R ² aj.	F obs.	P-valor
0.985	137	0.000157

Coef.	Estimado	P-valor
(Intercept)	0.0159	0.00655
poly(sustanciaA, 2)1	0.0331	0.000345
poly(sustanciaA, 2)2	-0.0526	3.37e-05
poly(sustanciaB, 1)	0.0302	0.000628
muestreadas	3.05e-05	0.0654

Tenemos muy pocos puntos como para observar tendencias en los residuos estudentizados en función de las proporciones predichas, pero lo que sí vemos, es que son todos realmente pequeños:

Residuos estudentizados en función de la proporción predicha



Los coeficientes de π ((Intercept)) y n_i (muestreadas) son significativos a nivel ≈ 0.05 , mientras que los de d_A, d_B son significativos a nivel < 0.001 . Como la regresión fue hecha sobre covariables generadas usando la función `poly` que computa polinomios *ortogonales* para las dosis de sustancia, ninguno de los coeficientes se presta a una interpretación directa, ni siquiera la ordenada π . Lo que sí podemos hacer con las herramientas disponibles, es elegir una grilla de valores de n, d_A, d_B , realizar predicciones con el modelo ajustado y graficarlas. Así, se observa la principal limitación del enfoque elegido: para dosis por encima de 200 mg/ml (por fuera del soporte de los datos), el efecto de la sustancia A da predicciones *negativas* para la proporción de células aberrantes:

Predicción de la roporción de células aberrantes para tamaños de muestra

