Riesgo en Vanedela

Derik Castillo

May 18, 2017

## Exploración de los datos

Me interesan las columnas 5 a 7, que corresponden a los datos de las tres fases del comportamiento "CD1.1A", "CD1.1B", "CD1.1C".

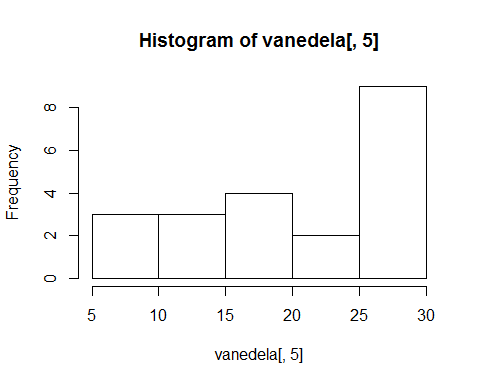
En principio, se puede definir el tiempo al que se puede decir que existe riesgo, como el percentil del 90%. La adición que se puede hacer es calcular los intervalos de confianza del 90% para el percentil. Para calcular el intervalo de confianza se utiliza un método de bootstrap, en el que se muestrean aleatoriamente una fracción de los datos, y se calcula el percentil. Se repite este procedimiento 1000 veces y con los 1000 percentiles se hace el intervalo de confianza. Este procedimiento funciona aún cuando la distribución de los datos se aleja de la normal.

En primer lugar se hace una exploración general de los datos, tomando sólo el primer comportamiento.

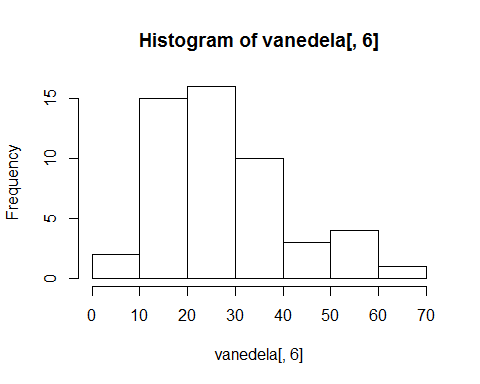
El número de niños en cada etapa en el orden A, B, C.

## CD1.1A CD1.1B CD1.1C   
## 21 51 65

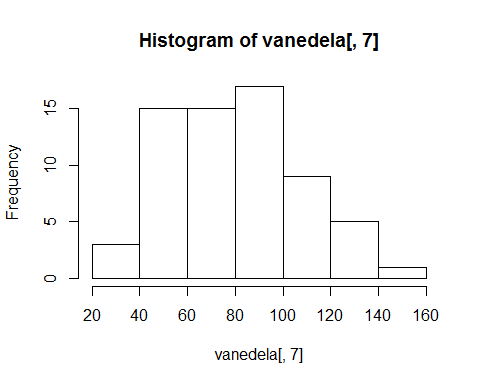
Ahora los histogramas para cada una de las fases. Si el histograma es simétrico, se puede decir que su distribución se parece a la normal. Primero fase A



Ahora la fase B

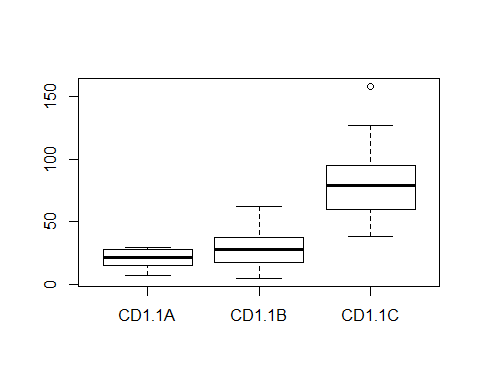


Finalmente la fase C



Se puede ver que sólo B y C tienen distribuciones que se aproximan a la normal.

Una gráfica de cajas puede ayudar a comparar las medianas, y los percentiles en las tres fases. En esta representación la línea gruesa en el medio representa la mediana. Los lados de la caja el primer y tercer cuartil.



Se puede ver que la mediana aumenta. Sabiendo que la fase C es como una meta, es decir que no hay algo más allá, entonces se puede pensar en esto como una especie de asíntota.

## Medianas y percentiles 90%

Las medianas de las tres fases son

## CD1.1A CD1.1B CD1.1C   
## 22 28 79

Los percentiles de 90% para cada una de las fases son

## CD1.1A CD1.1B CD1.1C   
## 30.0 49.0 119.6

Se pueden calcular los intervalos de confianza para los percentiles 90. La idea sería que se puede cortar en la fecha más tardía del intervalo de confianza para la fase C. Esto definiría un extremo de la ventana de no-riesgo.

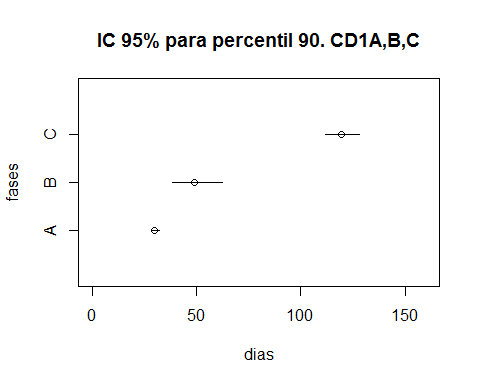
## Loading required package: boot

Para la fase A, el percentil de 90% es 30. Para esta misma fase, los extremos del intervalo de confianza al 95% para el percentil de 90% son 28.7310189, 32.4429811.

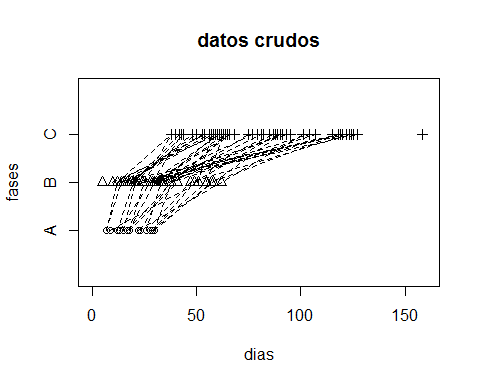
Para la fase B, el percentil de 90% es 49. Para esta misma fase, los extremos del intervalo de confianza al 95% para el percentil de 90% son 38.4551475, 62.2320525.

Para la fase C, el percentil de 90% es 119.6. Para esta misma fase, los extremos del intervalo de confianza al 95% para el percentil de 90% son 111.7017227, 127.8168773

Una representación gráfica de los percentiles ayuda a ubicarlos con respecto a las fases y el tiempo. En la siguiente gráfica, los dias se muestran en el eje x, en el eje y se muestran las tres fases. La bolita es el percentil 90% para una fase y la raya indica el intervalo de confianza.



La gráfica de los percentiles 90 y los intervalos de confianza es un resumen, pero la gráfica con todos los datos puede dar una idea más completa de cómo se están comportando. Cada niño está representado por un símbolo. La bolita es corresponde a la fase A, el triángulo a la fase B y la cruz a la fase C. Las líneas punteadas conectan mediciones para el mismo niño. Los símbolos que no están conectados, indican mediciones aisladas.



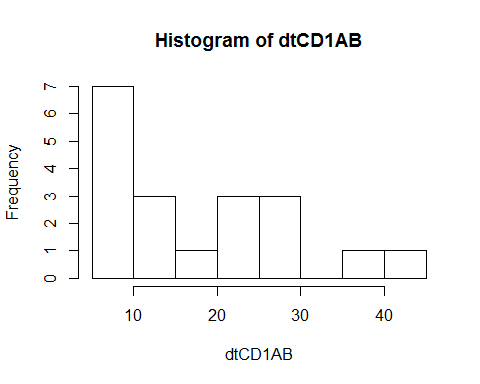
## Intervalo de tiempo entre fases

Otra forma de caracterizar las fases, es con el retraso o intervalo de tiempo entre dos de ellas. Esto solo se puede hacer con los niños que tienen al menos dos mediciones.

# Intervalo entre fases A y B

Si consideramos el intervalo de tiempo para las fases A y B, lo primero seria contar cuantos casos tenemos de niños que tienen mediciones en ambas columnas.

El número de niños que tienen ambas mediciones es 19. El histograma de los 19 intervalos es



El histograma indica que se tiene una distribución muy sesgada, tipo exponencial. Ahora se puede ver el percentil 90% de los intervalos junto con su intervalo de confianza.

El percentil del 90% es 31.2, los extremos del intervalo de confianza son 17.9541853, 42.0178147.

Como ejemplo, usando los datos de intervalo de tiempo, se puede decir que a partir de que un niño presenta el comportamiento A, entonces se debe esperar entre 17.9541853 y 42.0178147 dias para que presente el comportamiento B. Si se tarda mas esta en riesgo.

# Intervalo entre fases B y C

Se hace lo mismo para el otro intervalo.

Hay 39 niños con ambas mediciones. El percentil 90 de este intervalo es 89.4 y los extremos del intevalo de confianza son 71.5321071, 115.7462929.

A manera de ejemplo, se puede decir que un niño presenta riesgo si una vez que presentó el comportamiento B tarda mas de 115.7462929 dias en presentar el comportamiento C.

### Nota

En el artículo de WHO se usan failure time models. Falta analizar si se puede ocupar este modelo para estos datos.