# PRÁCTICA 3 ESTADÍSTICA APLICADA

**1. Genera dos muestras de tamaño 20 de una población normal con media 0 y varianza 1 y aplica el contraste t adecuado para contrastar si las medias de ambos grupos son la misma. Utiliza como semilla del generador de números aleatorios la denominación del grupo quitando la barra baja (_). Luego, aplica un procedimiento de simulación para aproximar el p-valor de este contraste. Comenta los resultados que obtienes.**

```
        Two Sample t-test

data:  muestra1 and muestra2
t = -1.5492, df = 38, p-value = 0.1296
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9893323  0.1315703
sample estimates:
  mean of x   mean of y
-0.33120908  0.09767192
```

Como el p-valor obtenido (p = 0.1296) es mayor que el nivel de significación (α= 0.05), no existe suficiente evidencia a favor de la hipótesis alternativa, en otras palabras, no rechazaremos que las medias sean iguales.

Mediante el procedimiento de simulación obtenemos un p-valor realmente próximo al obtenido con el contraste t (con 10.000 repeticiones el p-valor obtenido ronda entre los 0.126 y los 0.136), por lo que llegamos a la misma conclusión.

**2. Genera otras dos muestras de tamaño 20 de una población uniforme con mínimo 0 y máximo 10 y aplica el contraste adecuado para contrastar que ambos grupos provienen de la misma población. Utiliza como semilla del generador de números aleatorios la denominación del grupo quitando la barra baja (_). Luego, aplica un procedimiento de simulación para aproximar el p-valor de este contraste. Comenta los resultados que obtienes.**

De nuevo, el p-valor obtenido (0.07627) es superior al nivel de significación (0.05), por lo que no rechazaremos que las muestras provengan de la misma distribución.

Mediante el procedimiento de simulación obtenemos un p-valor muy similar, siempre por encima del nivel de significación cuando el número de repeticiones es alto.

Cabe destacar que, pese a que el p-valor es superior al nivel de significación, lo es por relativamente poco, lo que podría dejar dudas cuando el nivel de repeticiones es bajo. También es destacable que, si quisiéramos un contraste con un nivel de significación mayor, ya no podríamos asegurar si las muestras provienen de la misma distribución.

**3. Obtén el tamaño muestral mínimo necesario para, usando una muestra de ese tamaño con α = 0.05, poder detectar una diferencia de una unidad entre la media real de una población normal con desviación típica 1.5 y la media de contraste con una probabilidad de al menos 0.8.**

Fijando los datos del enunciado en la función “*power.t.test*” obtenemos que, para ser capaces de detectar la diferencia, el tamaño muestral mínimo es *n=20*.

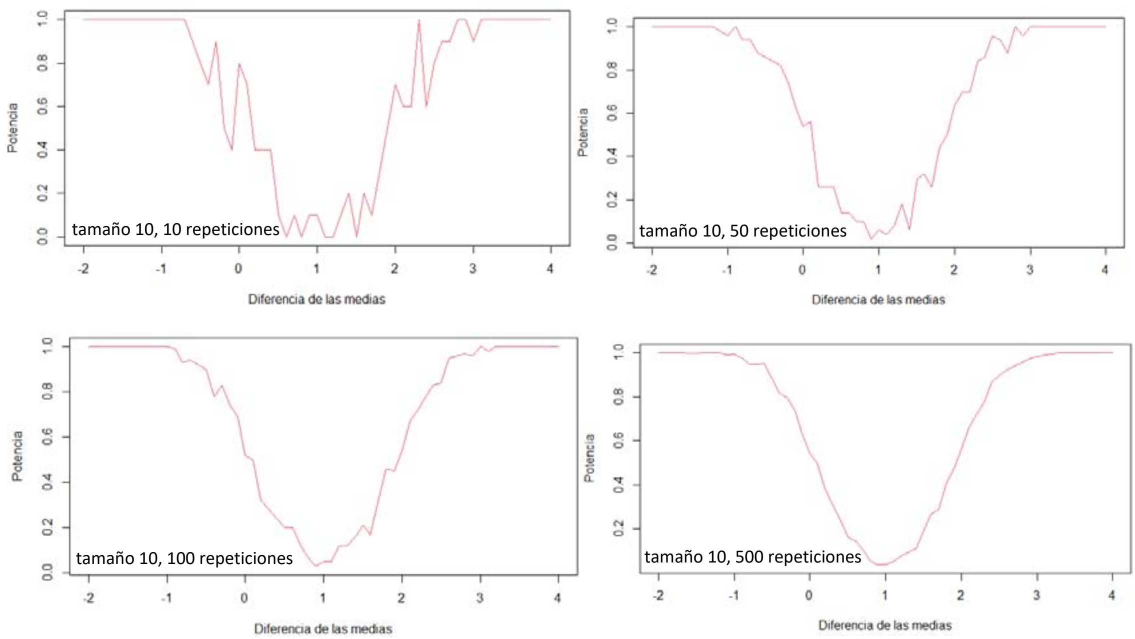
```
     One-sample t test power calculation

              n = 19.66697
          delta = 1
             sd = 1.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

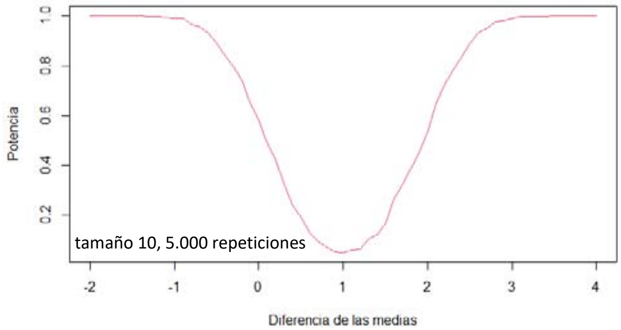
**4. Estima mediante simulación el tamaño muestral mínimo para, usando una muestra de ese tamaño con α = 0.05, poder detectar una diferencia de una unidad entre la media real de una población normal con desviación típica 1.5 (pero que asumiremos desconocida) y la media de contraste con una probabilidad de al menos 0.8.**

Utilizando la función “*estimar_prob_rechazar_H0*” y un bucle *while*, hemos simulando normales de tamaño cada vez mayor, obteniendo que el primer *n* con una probabilidad de al menos 0.8 es *n=19*.

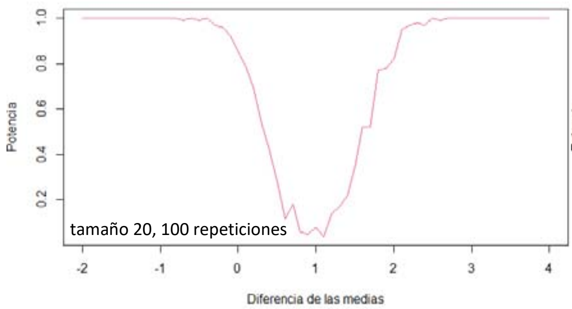
**5. Construye una función que aproxime la curva de potencia del contraste t para la comparación de la media de dos poblaciones normales con la misma varianza.**

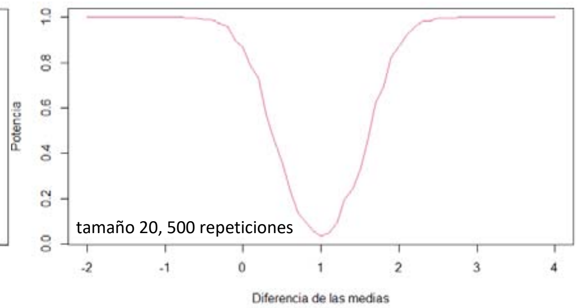


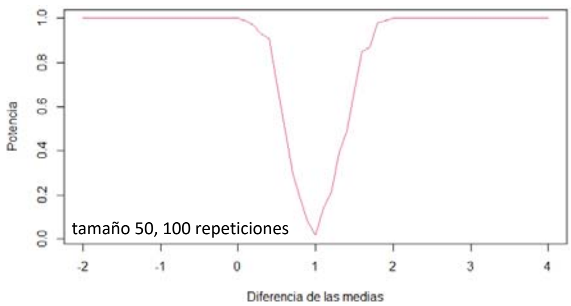
Si aproximamos la curva mediante simulación, al aumentar la cantidad de muestras tomadas obtendremos una curva más exacta.

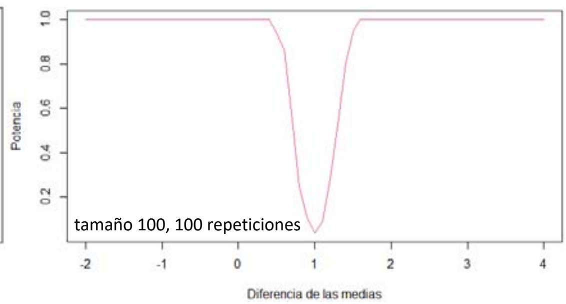



Veamos qué pasa si aumentamos el tamaño de la muestra y no el de las repeticiones. La varianza (como en los casos anteriores) siempre será 1.




Cuanto más grandes son nuestras muestras, más fina se vuelve la curva, hasta el punto que, con muestras de gran tamaño, apenas se vislumbra la apertura. Esto pasa porque los valores de nuestra función potencia son de la forma n/10.

Salta a la vista entonces que con esta función de potencia es mejor aumentar el número de muestras que el tamaño de la muestra.