

2 Pregunta 2

Se sabe que la duración de las conexiones a un sitio de ventas de calzado tiene una distribución normal con una media de 30 minutos y una desviación estándar de 14 minutos. Debido a un rediseño de la pagina web se espera que los clientes pasen el triple de tiempo en el sitio y se ha agregado una consulta de uso de cookies al inicio de la sesión que toma 3 minutos contestarla. Los clientes no pueden ingresar al sitio si no contestan la consulta.

2.1

Cree una muestra aleatoria de 400 datos que representen las conexiones al sitio actualmente. Luego transforme los datos para que representen como cambiará la muestra con la nueva página web.

Solución

Previo al rediseño de la página web, los datos tenían los siguientes parámetros estadísticos,

$$\begin{aligned} d &= \text{normal} \\ \mu_1 &= 30\text{min} \\ \sigma_1 &= 14\text{min} \end{aligned}$$

Por lo que una muestra aleatoria que represente estas conexiones se consigue con los siguientes comandos:

```
# 2.1
con = rnorm(400, mean=30, sd=14)
con
```

Figure 9: código 2.1

```
> con
[1] 27.7071479 24.2903409 17.4122317 22.3147858 3.4968619 41.2978357 31.9155406 33.9988827 51.9317896 41.4563213 28.0734510 32.9483330 28.7510139
[14] 38.5579576 22.2776835 47.4490511 27.6487499 47.9361898 43.3999107 56.6556796 18.1606211 40.1413772 42.4205873 29.8960007 35.7537669 52.3146047
[27] 22.6981557 33.1394034 20.9668022 21.8159558 38.3545849 27.6648241 48.5398936 33.7807497 36.2639145 20.9316173 1.6721195 17.7136283 41.6902773
[40] 33.6995713 45.2630940 44.5725207 16.5605602 58.6957585 24.8812725 13.5949603 9.5636469 16.5010004 34.0928983 45.1682147 34.5685632 28.3115912
[53] 40.1604352 38.1372093 25.0998634 21.8228125 57.2806000 15.9120735 57.3720245 34.1190436 18.2781617 23.6030949 55.0178156 8.6492081 50.5516214
[66] 61.3918209 26.8615916 10.6666631 42.6464334 28.5263334 39.7371478 33.3577182 39.1509771 34.0592753 4.6332742 31.1109563 11.4656386 44.2490080
[79] 12.3670474 10.3753404 47.8543374 13.4085136 35.1750750 4.0127305 23.2569665 26.9048398 38.0357458 28.1299035 26.5476051 42.6483390 25.1406943
[92] 32.8446813 29.7468953 13.1230251 50.2077366 18.6071531 23.0395770 55.7729883 31.3626177 15.7304259 36.1840640 42.6480104 45.8582838 24.0827120
[105] 32.2519494 36.6156926 24.4598763 21.9944155 27.8207546 35.2831017 40.4342173 34.6136465 40.6068411 64.8828397 59.7623700 44.1328735 49.5096684
[118] 29.8930650 35.8780074 20.4122903 20.7427842 3.4300891 47.5010956 49.2824615 -2.0619917 8.0966408 4.9484609 6.8471655 18.4383175 34.7753335
[131] 66.0320472 19.6570708 26.1681242 32.7369196 56.6680203 50.5033691 20.1751600 37.0448767 19.2667035 36.4155696 22.0289163 19.9387772 17.2292030
[144] 42.9457651 37.4143586 15.5224300 36.0320799 28.6234677 31.8845960 37.1868296 14.7686267 5.5526979 27.1776032 30.2015633 17.3504313 65.3598344
[157] 32.4708673 22.9775038 31.5695757 21.1167544 51.1600345 17.3651717 20.1679322 28.9438607 34.6043255 35.0798218 44.0038014 51.8533537 16.4337879
[170] 20.1745187 17.4837670 28.6028907 29.3849577 48.5024317 30.9188809 57.5573112 25.9984012 14.9964326 27.8682021 71.2176616 34.2214519 27.3843872
[183] 45.1380286 3.1581465 30.3022460 -3.1462436 38.0473563 49.9795692 50.3045616 -6.6727070 17.2846076 34.3879948 19.6131073 36.9563869 44.7984682
[196] 14.6054143 22.0022142 36.7090993 38.8624678 8.8779698 46.3256712 24.9491185 38.4060627 41.7482887 33.2529271 31.2275052 26.1134820 15.9253893
[209] -5.6201324 46.0539696 7.2686768 26.8549512 38.0578491 31.2521407 38.7628656 51.7702251 20.5565199 43.7905662 39.5162783 43.1194666 6.5230979
[222] 23.0671452 32.4620316 54.8382028 34.5354528 16.5759804 14.7839918 56.4582824 33.7030062 33.3942511 27.4768854 38.0777311 22.7306317 31.1392935
[235] 21.2201643 41.4810757 0.8788632 21.5309121 26.8930665 30.0354437 39.4535944 29.8873145 39.9477907 23.8258748 24.9419445 21.1019993 14.7775888
[248] 14.8225964 48.8225964 26.3229865 58.0157932 42.1482109 39.1973740 22.2928735 41.6647070 27.3635695 28.7079043 10.9898961 6.1314366 46.4299418
[261] 8.9589175 49.8861121 30.9957297 31.4481644 18.3322282 56.2163358 38.2146500 36.6480744 55.8359909 13.1532064 39.3213055 18.2258578 29.6959780
[274] 15.3210014 7.9727969 34.2990431 32.5465073 32.3982042 18.1260317 8.6815794 28.2814430 45.5767654 6.0779652 41.6028479 31.5811574 10.8725583
[287] 49.2955887 23.2982449 43.1504375 18.0746790 24.4337570 38.4156668 30.3927771 34.3425072 43.8688141 14.5687787 50.1614905 21.7768246 36.0045695
[300] 29.8226147 27.8128584 1.5280352 29.6737785 66.6446653 39.6073614 32.9583694 36.0238191 14.1920160 53.1812508 14.2387554 21.5562323 15.9959138
[313] 29.9385104 21.8719683 57.0937057 41.5839601 29.8399889 42.7316230 28.3417395 31.6662946 3.6248356 -10.9105625 49.7794724 24.1204868 46.5003451
[326] 27.1228711 20.8835334 7.0652691 34.6609634 34.2779882 39.4383767 52.4335954 43.1640639 44.4518401 25.4182124 10.6289169 32.1807800 35.2057463
[339] 18.7148755 33.6625134 29.527381 7.4603466 22.6348004 3.0212097 19.2736449 30.5490324 38.8425924 53.3669967 50.5738784 31.4866663 39.6344128
[352] 28.6870845 37.5964712 15.8274263 40.3968362 34.8004202 11.0006217 25.9500866 32.3348347 26.5137136 31.5964400 44.7444474 42.7599725 15.8880282
[365] 43.5555506 35.6115636 25.7730474 23.5070185 28.4509289 31.7718434 25.5769641 34.2582063 36.1025359 46.7554384 39.2750051 15.3609417 24.0620212
[378] 33.3314836 52.6130659 46.6079589 31.1726435 53.2918896 39.4137627 23.6550034 22.6945072 35.4868513 32.2202635 -8.2146500 40.2183326 21.5679450
[391] 46.7415211 10.7556438 23.8151882 7.9556903 35.3421521 44.4943599 11.0685565 13.8084994 22.7087280 36.6244773
```

Figure 10: resultados 2.1

Después del rediseño de la página web, cada uno de las conexiones al sitio x_i sufriría la siguiente transformación hacia y_i cuando se dice ahora que *pasan el triple de tiempo en el sitio y se ha agregado una consulta de 3 min antes de ingresar al sitio*:

$$y_i = 3x_i + 3 \quad (1)$$

Por lo tanto, si transformamos los datos para que representen cómo cambiará la muestra con la nueva página web, obtenemos los estadísticos:

x_i	$y_i = 3x_i + 3$
$\mu_1 = 30$	$\mu_2 = 30 \times 3 + 3 = 93$
$\sigma_1 = 14$	$\sigma_2 = 14 \times 3 = 42$

y si transformamos cada uno de los datos de la muestra 1 con la Eq. (1) obtenemos la nueva muestra:

`con2 = con*3 + 3`

Figure 11: código 2.1 para transformar los datos de la muestra.

```
> con2
[1] 86.121444 75.871023 55.236695 69.944357 13.490586 126.893507 98.746622 104.996648 158.795369 127.368964 87.220353 101.844999 89.253042 118.673873
[15] 69.833050 145.347153 85.946250 146.808570 133.199732 172.967039 57.481863 123.424132 130.261762 92.688002 110.261301 159.943814 71.094467 102.418210
[29] 65.900407 68.447867 118.063755 85.994472 148.619681 104.342249 111.791744 65.794852 8.016358 56.140885 128.070832 104.098714 138.789282 136.717562
[43] 52.681681 179.087275 77.643818 43.784881 31.690941 52.503001 105.278695 138.504644 106.705690 87.934773 123.481305 117.411628 78.299590 68.468437
[57] 174.841800 50.736221 175.116073 105.357131 57.834485 73.809285 168.053447 28.947624 154.654864 187.175463 83.584775 34.999989 130.939300 88.579000
[71] 122.211443 103.073155 120.452931 105.177826 16.899822 96.332869 37.396916 135.747024 40.101142 34.126021 146.563012 43.225541 108.525225 15.038192
[85] 72.770899 83.714519 117.107237 87.389711 82.642815 130.945017 78.422083 101.534044 92.240686 42.369075 153.623210 58.821459 72.118731 170.318965
[99] 97.087853 50.191278 111.552192 130.944031 140.574851 75.248136 99.755848 112.847078 76.379629 68.983247 86.462264 108.849305 124.302652 106.840940
[113] 124.820523 197.648519 182.287110 135.398621 151.529005 92.679195 110.634022 64.236871 65.228352 13.290267 145.503287 150.847384 -3.185975 27.289922
[127] 17.845383 23.541496 58.314952 107.326001 201.096141 61.971212 81.504373 101.210759 173.004061 154.510107 63.525480 114.134630 60.800110 112.246709
[141] 69.086749 62.816331 54.687609 131.837295 115.243076 49.567290 111.096240 88.870403 98.653788 114.560489 47.305880 19.658094 84.532810 93.604690
[155] 55.051294 199.079503 100.412602 71.932512 97.708727 66.350263 156.480104 55.095515 63.503797 89.831582 106.812976 108.239465 135.011404 158.560061
[169] 52.301364 63.523556 55.451301 88.808672 91.154873 148.507295 95.756643 175.671934 80.995204 47.989298 86.604606 216.652985 105.664356 85.153162
[183] 138.414086 12.474440 93.906738 -6.438731 117.142069 152.938708 153.913685 -17.018121 54.853823 106.163984 61.839322 113.869161 137.395404 46.816243
[197] 69.006643 113.129728 119.587403 29.633909 141.977014 77.847355 118.218188 128.244866 102.758781 96.682516 81.340446 50.776168 -13.860397 141.161909
[211] 24.806030 83.564854 117.173547 96.756422 119.288597 158.310675 64.669560 134.371699 121.548835 132.358400 22.569294 72.201436 100.386095 167.514609
[225] 106.606358 52.727941 47.351975 172.374846 104.109019 103.182753 85.430656 117.233193 71.191895 96.417881 66.660493 127.443227 5.636590 67.592736
[239] 83.679199 93.106331 121.360783 92.661944 122.843372 74.477624 77.825833 66.305998 47.332766 45.978386 149.467789 81.968960 177.047380 129.444633
[253] 120.592122 69.878620 127.994121 85.090709 89.123713 35.969688 21.394310 142.289825 29.876752 152.658336 95.987189 97.344493 57.996685 171.649007
[267] 117.643950 112.944223 170.507973 42.459619 120.963917 57.677574 92.087934 48.963004 26.918391 105.897129 100.639522 100.194613 57.378095 29.044738
[281] 87.844329 139.730296 21.233896 127.808625 97.743472 35.617675 150.886766 72.894735 132.451312 57.224037 76.301271 118.247000 94.178331 106.027522
[295] 134.606442 46.706336 153.484471 68.330474 111.013708 92.467844 86.438575 7.584105 92.021335 202.933996 121.822084 101.875108 111.071457 45.576048
[309] 162.543752 45.716266 67.668697 50.987742 92.815311 68.615905 174.281117 127.751880 92.519967 131.194869 88.025219 97.998884 13.874507 -29.731687
[323] 152.338417 75.361460 142.501035 84.368613 65.650600 24.195807 106.982890 105.833965 121.315130 160.300786 132.492192 136.355520 79.254637 34.886751
[337] 99.542340 108.617239 59.144626 103.987540 91.581714 25.381040 70.904401 12.063629 60.820935 94.647097 119.527777 163.100990 154.721635 97.459999
[351] 121.903239 89.061254 115.789414 50.482279 124.190509 107.401261 36.001865 80.850260 100.004504 82.541141 97.789320 137.233343 131.279917 50.664085
[365] 133.666652 109.834691 80.319142 73.521055 88.352787 98.315530 79.730892 105.774619 111.307608 143.266315 120.825015 49.082825 75.186064 102.994451
[379] 160.839198 142.823877 96.517930 162.875669 121.241288 73.965010 71.083522 109.460554 99.666191 -21.643950 123.654998 67.703835 143.224563 35.266931
[393] 74.445565 26.867071 109.026456 136.483080 36.205669 44.425498 71.126184 112.873432
```

Figure 12: resultados 2.1 de la nueva muestra.

2.2

Calcule la media, mediana, rango de datos, IQR y sd, de ambas muestras y compárelas.

Solución

x_i (dado)	x_i (muestra 1)	$y_i = 3x_i + 3$ (calculado)	$y_i = 3x_i + 3$ (muestra 2)
$\mu_1 = 30$	$\mu_2 = 30.53929$	$\mu_2 = 30 \times 3 + 3 = 93$	$\mu_2 = 94.61787$
-	$med_1 = 31.15597$	$med_2 = med_1 \times 3 + 3 = 96.46791$	$med_2 = 96.46791$
-	$range_1 = 82.12822$	$range_2 = range_1 \times 3 = 246.3847$	$range_2 = 246.3847$
-	$IQR_1 = 18.72161$	$IQR_2 = IQR_1 \times 3 = 54.16483$	$IQR_2 = 56.16483$
$\sigma_1 = 14$	$\sigma_1 = 14.40793$	$\sigma_2 = 14 \times 3 = 42$	$\sigma_2 = 43.2238$

Al comparar las medias, medianas, rangos, IQR's y desviaciones estándar, se puede ver que, teóricamente (calculándolo a mano), la transformación de los datos con $y_i = 3x_i + 3$ afecta las medidas de la siguiente manera: la media y mediana se multiplican por 3 y se les suma 3, mientras que el nuevo rango, IQR

```

# 2.1
con = rnorm(400, mean=30, sd=14)
mean1 = mean(con)
median1 = median(con)
range1 = range(con)
iqr1 = IQR(con)
sd1 = sd(con)
mean1
median1
range1
iqr1
sd1

con2 = con*3 + 3
mean2 = mean(con2)
median2 = median(con2)
range2 = range(con2)
iqr2 = IQR(con2)
sd2 = sd(con2)
mean2
median2
range2
iqr2
sd2

```

Figure 13: código 2.2 para calcular medidas.

> mean1	> mean2
[1] 30.53929	[1] 94.61787
> median1	> median2
[1] 31.15597	[1] 96.46791
> range1	> range2
[1] 82.12822	[1] 246.3847
> iqr1	> iqr2
[1] 18.72161	[1] 56.16483
> sd1	> sd2
[1] 14.40793	[1] 43.2238

Figure 14: resultados 2.2

y desviación estándar sólo deben multiplicarse por 3. En la práctica, es decir, al extraer estas medidas directamente de la muestra 2 en la Figura 12, se comprueba que estos cálculos son correctos, ya que los resultados son prácticamente los mismos, salvo por algunas unidades de diferencia debido a que la muestra fue tomada al azar.

2.3

Explique cada uno de los cambios observados.

Solución

Si se dice que ahora tardan el triple de tiempo en el sitio (x_i), la distribución de los datos cambia de forma, ya que se está ampliando el espacio entre los valores de los datos, pero la cuenta de cada una de las barras del histograma siguen siendo las mismas, lo que hace que la distribución se vea como extendida en x. Si se le agregan 3 minutos por la encuesta de cookies, la distribución se traslada 3 unidades en el eje x.

- La media cambia de 30 a 93 con la transformación $y_i = 3x_i + 3$ ya que al alargar la distancia multiplicando por 3 los datos en el set, ahora la forma, y por lo tanto, la campana en la distribución se mueve de sitio, por lo que la media es afectada al ser la medida en el centro de dicha campana.

Luego, al trasladar 3 unidades la distribución por la encuesta, la posición de la campana y de la media es afectada también.

- La mediana cambia de 31 a 96 por las mismas razones que la media, ya que en una distribución normal la media, la mediana y la moda coinciden en el mismo punto, y las razones por las que la mediana cambia son entonces las mismas por las que la media cambia.
- El rango cambia de 82 a 246, es decir, sólo se le debe multiplicar por 3 y no sumar los 3 minutos, ya que el rango de valores sólo se afecta cuando la campana cambia de forma por alargarse (en este caso) o encogerse, pues las colas o los últimos valores se mueven con este cambio y por lo tanto el máximo y mínimo valor también, lo que hace que el rango (diferencia de valores) cambie. En cambio, los 3 minutos de corrimiento no afectan al rango ya que si se mueve toda la distribución a lo largo del eje x por 3 unidades, la distancia entre el máximo y el mínimo valor es la misma y no debe cambiar.
- La desviación estándar cambia de 14 a 42 (43 a veces), porque la desviación estándar describe el ancho de la campana de distribución, por lo que el ensanchamiento al multiplicar por tres el tiempo en el sitio hace que la desviación anterior se triplique, pero la traslación por los 3 minutos no afecta el ancho de la campana, por lo que no mueve la desviación a otros valores.

2.4

Si la compañía recibe 5 centavos por el despliegue de un anuncio por minuto en su página, y el sitio recibe en promedio 50 visitas al día, calcule cuanto ingresa en promedio actualmente por la publicación de dos anuncios en la pagina web y en cuanto incrementará el monto del ingreso con la nueva página web.

Solución

Conexiones:

sitio 1	sitio 2
d = normal	d = normal
$\mu_1 = 30min$	$\mu_2 = 93min$
$\sigma_1 = 14min$	$\sigma_2 = 42min$

Ingresos:

¢5 anuncio / min
2 anuncios: ¢10 / min
50 visitas / día
\$ promedio (?)

Si se tienen 50 visitas por día a un sitio que gana ¢10 por minuto (2 anuncios), ¿cuánto duran estas visitas en promedio para poder saber el ingreso promedio al día? Estas conexiones duran un promedio $\mu_1 = 30$ minutos con el antiguo sitio y $\mu_2 = 93$ minutos con el nuevo sitio. Por lo tanto,

Actualmente se gana:

$$\text{Ingreso promedio}_1 = \text{visitas} \times \text{minutos/visita} \times \text{ingreso/minuto} = 50 \times 30 \times 10 = 15000 = \boxed{\$150} \quad (2)$$

Y con la nueva página web se ganará:

$$\text{Ingreso promedio}_2 = \text{visitas} \times \text{minutos/visita} \times \text{ingreso/minuto} = 50 \times 93 \times 10 = 46500 = \boxed{\$465} \quad (3)$$

Por lo que el incremento del ingreso con la nueva página web es:

$$\Delta \text{Ingreso} = \frac{465 - 150}{150} \times 100 - 100 = \boxed{110\%} \text{ con respecto a la ganancia actual} \quad (4)$$

2.5

Grafique la distribución de probabilidad de la muestra original y de la muestra transformada. Explique la diferencia entre las gráficas.

Solución

```
plot(x=c(min(con):max(con)), y=dnorm(c(min(con):max(con)), mean=30, 14),
     xlab="x", ylab="probability", main=c("F(x) of con"))
plot(x=c(min(con2):max(con2)), y=dnorm(c(min(con2):max(con2)), mean=93, 42),
     xlab="x", ylab="probability", main=c("F(x) of con2"))
```

Figure 15: código 2.5 para graficar la función de probabilidad de muestra original (arriba) y transformada (abajo)

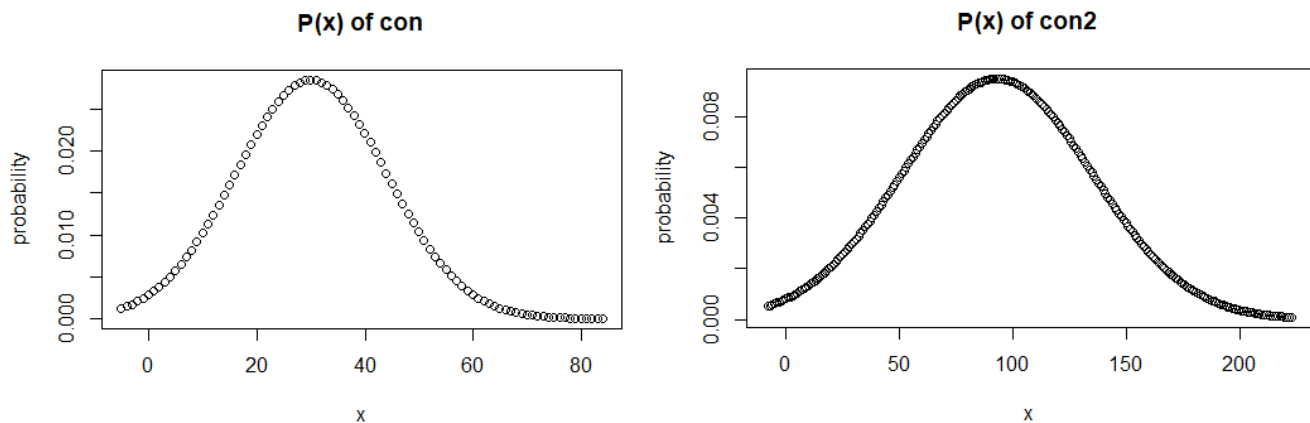


Figure 16: gráfica de función de probabilidad de la muestra original (izq) y transformada (der)

Si se analiza la diferencia entre las gráficas de la función de probabilidad, la función original está más estrecha, ya que en el eje x alcanza valores hasta 80, mientras que la transformada se extiende hasta valores alrededor de 240, lo que hace que su campana se vea más extensa y ancha, también incluso a raíz de que su desviación estándar crece al triple, por lo que se ensanchó no sólo el abarque en el eje x, sino su campana también.

```
x = c(min(con):max(con))
curve(pnorm(x, mean=30, sd=14), xlim=c(min(con),max(con)), col="blue", ylab="cumulative Distribution",
      main="Distribution Function")

x = c(min(con2):max(con2))
curve(pnorm(x, mean=93, sd=42), xlim=c(min(con2),max(con2)), col="blue", ylab="cumulative Distribution",
      main="Distribution Function")
```

Figure 17: código 2.5 para graficar la distribución de probabilidad de muestra original (arriba) y transformada (abajo)

Al analizar las dos gráficas de distribución acumulada, se nota cómo la función transformada aporta ligeramente menos elevación en y conforme avanza en x, lo cual se explica desde la gráfica de función

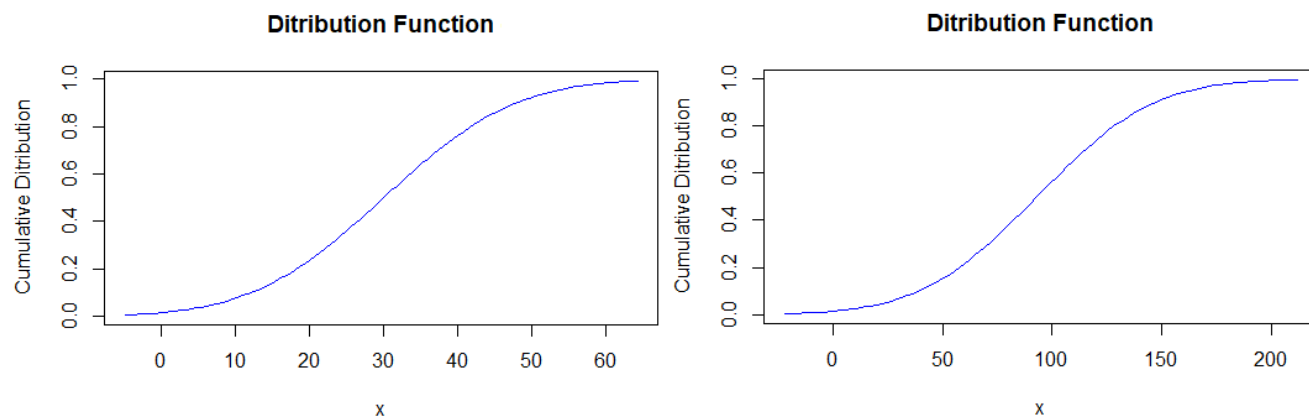


Figure 18: gráfica de distribución de probabilidad de la muestra original (izq) y transformada (der)

de probabilidad, ya que las probabilidades son mucho más pequeñas en la función transformada que en la función original, lo que hace que, al acumularlas en la función de distribución, se eleve menos por avance en x . Esto se explica ya que la forma de la distribución se alargó en x pero se aplastó en y con la transformación.