Distribucion de Porbabilidad a partir de una funcion dada

Generate data with errors,
OLS vs MLE

In[ ]:=
```
IntegerPart@AbsoluteTime[]
parte entera tiempo desde 1900
```

Out[ ]= 3 939 872 133

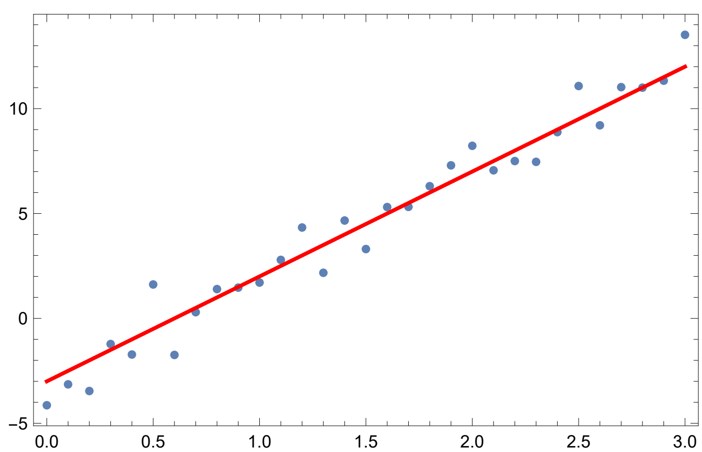
In[ ]:=
```
SeedRandom[IntegerPart@AbsoluteTime[]];
semilla aleato· parte entera tiempo desde 1900
```

In[ ]:=
```
error = 1.;
dat = Table[{x, 5. x - 3. + RandomVariate[NormalDistribution[0, error]]}, {x, 0, 3, .1}];
tabla variable aleatoria distribución normal
```

In[ ]:=
```
Length@dat
longitud
```

Out[ ]= 31

In[ ]:=
```
Show[ListPlot[dat, Axes → False],
mue·· representación d·· ejes falso
 Plot[5. x - 3., {x, 0, 3}, PlotStyle → Red, Axes → False], Frame → True]
 representación gráfica estilo de repr··· rojo ejes falso marco verdade
```

Out[ ]=


In[ ]:=
```
Dimensions@dat
dimensiones
```

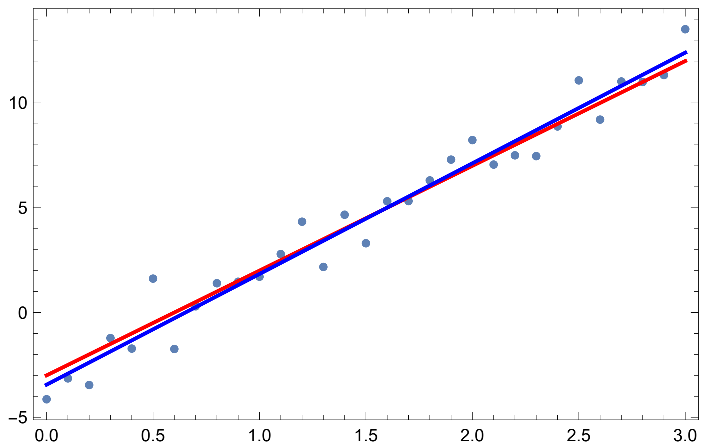
Out[ ]= {31, 2}

OLS (Ordinary Least Squares)

```
In[•]:= Clear[modelo]
        borra

In[•]:= modelo[x_] = a x + b;

In[•]:= mincuad = Sum[(dat[[k, 2]] - modelo[dat[[k, 1]]])^2, {k, Length@dat}]
                  suma                                          longitud

Out[•]= (-4.13966 - b)^2 + (13.5187 - 3. a - b)^2 + (11.3315 - 2.9 a - b)^2 + (10.999 - 2.8 a - b)^2 +
  (11.0241 - 2.7 a - b)^2 + (9.20466 - 2.6 a - b)^2 + (11.0773 - 2.5 a - b)^2 + (8.8785 - 2.4 a - b)^2 +
  (7.46298 - 2.3 a - b)^2 + (7.50112 - 2.2 a - b)^2 + (7.05963 - 2.1 a - b)^2 + (8.22786 - 2. a - b)^2 +
  (7.29565 - 1.9 a - b)^2 + (6.30384 - 1.8 a - b)^2 + (5.31756 - 1.7 a - b)^2 + (5.30774 - 1.6 a - b)^2 +
  (3.30557 - 1.5 a - b)^2 + (4.66586 - 1.4 a - b)^2 + (2.17456 - 1.3 a - b)^2 + (4.33295 - 1.2 a - b)^2 +
  (2.78906 - 1.1 a - b)^2 + (1.708 - 1. a - b)^2 + (1.46789 - 0.9 a - b)^2 + (1.39728 - 0.8 a - b)^2 +
  (0.296811 - 0.7 a - b)^2 + (-1.74102 - 0.6 a - b)^2 + (1.61727 - 0.5 a - b)^2 +
  (-1.72162 - 0.4 a - b)^2 + (-1.22101 - 0.3 a - b)^2 + (-3.46088 - 0.2 a - b)^2 + (-3.14374 - 0.1 a - b)^2

In[•]:= mincuad // FullSimplify
                   simplifica completamente

Out[•]= 31. (43.1946 + 3.05 a^2 + (-8.95727 + b) b + a (-21.8868 + 3. b))

In[•]:= AbsoluteTiming[solOLS = NMinimize[mincuad, {a, b}]]
        duración absoluta        minimiza aproximadamente

Out[•]= {0.003667, {25.373, {a → 5.28181, b → -3.44408}}}

In[•]:= AbsoluteTiming[solOLS = FindMinimum[mincuad, {a, b}, Method → "LevenbergMarquardt"]]
        duración absoluta        encuentra mínimo              método

Out[•]= {0.004057, {25.373, {a → 5.28181, b → -3.44408}}}

In[•]:= AbsoluteTiming[otraSol = Fit[dat, {1, x}, x]]
        duración absoluta          ajusta

Out[•]= {0.002227, -3.44408 + 5.28181 x}

In[•]:= eca = D[mincuad, a];
              deriva
        ecb = D[mincuad, b];
              deriva

In[•]:= Solve[{eca == 0, ecb == 0}, {a, b}]
        resuelve

Out[•]= {{a → 5.28181, b → -3.44408}}

In[•]:= solOLS[[2]]

Out[•]= {a → 5.28181, b → -3.44408}

In[•]:= modelo[x] /. solOLS[[2]]

Out[•]= -3.44408 + 5.28181 x
```

In[•]:=
```
Show[ListPlot[dat, Axes → False], Plot[5. x - 3., {x, 0, 3}, PlotStyle → Red, Axes → False],
 Plot[modelo[x] /. solOLS[[2]], {x, 0, 3}, PlotStyle → Blue, Axes → False], Frame → True]
```
mue·· representación d·· ejes falso representación gráfica estilo de repr··· rojo ejes falso
representación gráfica estilo de repr··· azul ejes falso marco verdadero

Out[•]=



MLE (Maximum Likelihood Estimator)

In[•]:=
```
Clear[modelo, dist]
```
borra

In[•]:=
```
modelo[x_] = a x + b;
```

In[•]:=
```
dist[y_, x_] = PDF[NormalDistribution[modelo[x], σ], y]
```
fun·· distribución normal

Out[•]=
$$\frac{e^{-\frac{(-b-a\,x+y)^2}{2\,\sigma^2}}}{\sqrt{2\,\pi}\,\sigma}$$

In[•]:=
```
Product[dist[dat[[k, 2]], dat[[k, 1]]], {k, Length@dat}]
```
producto longitud

Out[•]=
$$\frac{1}{32768\,\sqrt{2}\,\pi^{31/2}\,\sigma^{31}}\,e^{-\frac{(-4.13966-b)^2}{2\,\sigma^2}-\frac{(13.5187-3.\,a-b)^2}{2\,\sigma^2}-\frac{(11.3315-2.9\,a-b)^2}{2\,\sigma^2}-\frac{(10.999-2.8\,a-b)^2}{2\,\sigma^2}-\frac{(11.0241-2.7\,a-b)^2}{2\,\sigma^2}-\frac{(9.20466-2.6\,a-b)^2}{2\,\sigma^2}-\frac{(11.0773-2.5\,a-b)^2}{2\,\sigma^2}-\frac{(8.8785-2.4\,a-b)^2}{2\,\sigma^2}-\frac{(7.46298-2.3\,a-b)^2}{2\,\sigma^2}-}$$

```
In[•]:= Log[Product[dist[dat[[k, 2]], dat[[k, 1]]], {k, Length@dat}]] // PowerExpand
```
lo··· producto longitud expande potencias

Out[•]=
$$-\frac{(-4.13966-b)^2}{2\sigma^2}-\frac{(13.5187-3.\,a-b)^2}{2\sigma^2}-\frac{(11.3315-2.9\,a-b)^2}{2\sigma^2}-\frac{(10.999-2.8\,a-b)^2}{2\sigma^2}-\frac{(11.0241-2.7\,a-b)^2}{2\sigma^2}-\frac{(9.20466-2.6\,a-b)^2}{2\sigma^2}-\frac{(11.0773-2.5\,a-b)^2}{2\sigma^2}-\frac{(8.8785-2.4\,a-b)^2}{2\sigma^2}-\frac{(7.46298-2.3\,a-b)^2}{2\sigma^2}-\frac{(7.50112-2.2\,a-b)^2}{2\sigma^2}-\frac{(7.05963-2.1\,a-b)^2}{2\sigma^2}-\frac{(8.22786-2.\,a-b)^2}{2\sigma^2}-\frac{(7.29565-1.9\,a-b)^2}{2\sigma^2}-\frac{(6.30384-1.8\,a-b)^2}{2\sigma^2}-\frac{(5.31756-1.7\,a-b)^2}{2\sigma^2}-\frac{(5.30774-1.6\,a-b)^2}{2\sigma^2}-\frac{(3.30557-1.5\,a-b)^2}{2\sigma^2}-\frac{(4.66586-1.4\,a-b)^2}{2\sigma^2}-\frac{(2.17456-1.3\,a-b)^2}{2\sigma^2}-\frac{(4.33295-1.2\,a-b)^2}{2\sigma^2}-\frac{(2.78906-1.1\,a-b)^2}{2\sigma^2}-\frac{(1.708-1.\,a-b)^2}{2\sigma^2}-\frac{(1.46789-0.9\,a-b)^2}{2\sigma^2}-\frac{(1.39728-0.8\,a-b)^2}{2\sigma^2}-\frac{(0.296811-0.7\,a-b)^2}{2\sigma^2}-\frac{(-1.74102-0.6\,a-b)^2}{2\sigma^2}-\frac{(1.61727-0.5\,a-b)^2}{2\sigma^2}-\frac{(-1.72162-0.4\,a-b)^2}{2\sigma^2}-\frac{(-1.22101-0.3\,a-b)^2}{2\sigma^2}-\frac{(-3.46088-0.2\,a-b)^2}{2\sigma^2}-\frac{(-3.14374-0.1\,a-b)^2}{2\sigma^2}-\frac{31\,\mathrm{Log}[2]}{2}-\frac{31\,\mathrm{Log}[\pi]}{2}-31\,\mathrm{Log}[\sigma]$$

```
In[•]:= Log[Product[dist[dat[[k, 2]], dat[[k, 1]]], {k, Length@dat}]] // PowerExpand // FullSimplify
```
lo··· producto longitud expande potencias simplifica completamen

Out[•]=
$$\frac{-669.517-47.275\,a^2+a\,(339.245-46.5\,b)+(138.838-15.5\,b)\,b-28.4871\,\sigma^2-31.\,\sigma^2\,\mathrm{Log}[\sigma]}{\sigma^2}$$

```
In[•]:= lmle1 =
  PowerExpand[Log[Product[dist[dat[[k, 2]], dat[[k, 1]]], {k, Length@dat}]]] // FullSimplify
```
expande poten··· lo··· producto longitud simplifica complet

Out[•]=
$$\frac{-669.517-47.275\,a^2+a\,(339.245-46.5\,b)+(138.838-15.5\,b)\,b-28.4871\,\sigma^2-31.\,\sigma^2\,\mathrm{Log}[\sigma]}{\sigma^2}$$

```
In[•]:= AbsoluteTiming[solMLE = NMaximize[{lmle1, σ > 0}, {a, b, σ}]]
```
duración absoluta maximiza aproximadamente

```
Out[•]= {0.064981, {-40.8824, {a → 5.28181, b → -3.44408, σ → 0.904702}}}
```

```
In[•]:= (a /. solOLS[[2]]) - (a /. solMLE[[2]])
```

Out[•]= -2.66454×10^{-15}

```
In[•]:= Chop[%]
```
cambia números pequeños por 0

```
Out[•]= 0
```

In[ ]:= (b /. solOLS⟦2⟧) - (b /. solMLE⟦2⟧)

Out[ ]= 5.77316×10^{-15}

In[ ]:= Chop[%]
cambia números pequeños por 0

Out[ ]= 0

Otra forma de la distribución de errores

In[ ]:= Clear[a, b, modelo]
borra

In[ ]:= modelo[x_] = a x + b;

In[ ]:= dist[y_, x_] = PDF[StudentTDistribution[modelo[x], σ, Length@dat - 2], y]
fun·· distribución t de Student longitud

Out[ ]= $$\frac{9\,984\,397\,484\,360\,291\,143\,009\,697\,792\,\sqrt{29}}{5\,014\,575\,\pi\left(29+\frac{(-b-a\,x+y)^2}{\sigma^2}\right)^{15}\sigma}$$

In[ ]:= lmle1 =
PowerExpand[Log[Product[dist[dat⟦k, 2⟧, dat⟦k, 1⟧], {k, Length@dat}]]] // FullSimplify;
expande poten·· lo··· producto longitud simplifica completa

In[ ]:= AbsoluteTiming[solMLE = NMaximize[{lmle1, a > 0 && σ > 0}, {a, b, σ}]]
duración absoluta maximiza aproximadamente

Out[ ]= {1.53054, {-40.9283, {a → 5.29403, b → -3.47605, σ → 0.87698}}}

In[ ]:= (a /. solOLS⟦2⟧) - (a /. solMLE⟦2⟧)

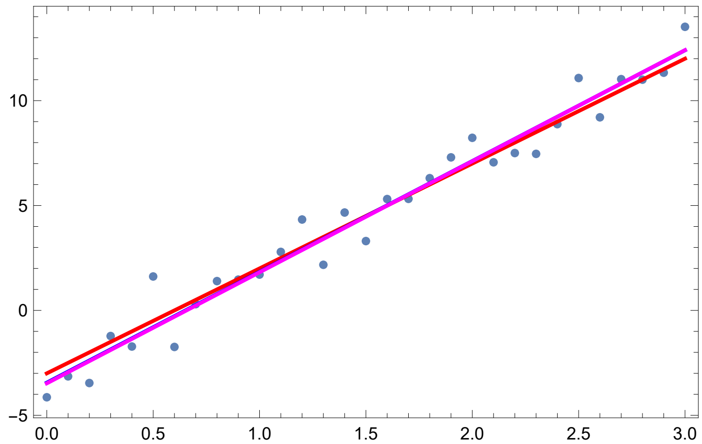
Out[ ]= -0.0122253

In[ ]:= (b /. solOLS⟦2⟧) - (b /. solMLE⟦2⟧)

Out[ ]= 0.0319717

In[ ]:=
```
Show[ListPlot[dat, Axes → False], Plot[5. x - 3., {x, 0, 3}, PlotStyle → Red, Axes → False],
  Plot[modelo[x] /. solOLS⟦2⟧, {x, 0, 3}, PlotStyle → Blue, Axes → False],
  Plot[modelo[x] /. solMLE⟦2⟧, {x, 0, 3}, PlotStyle → Magenta, Axes → False], Frame → True]
```

Out[ ]=



Kernel Density Estimation (https://en.wikipedia.org/wiki/Kernel_density_estimation)

Kernel Normal

In[ ]:=
```
Clear@f
```

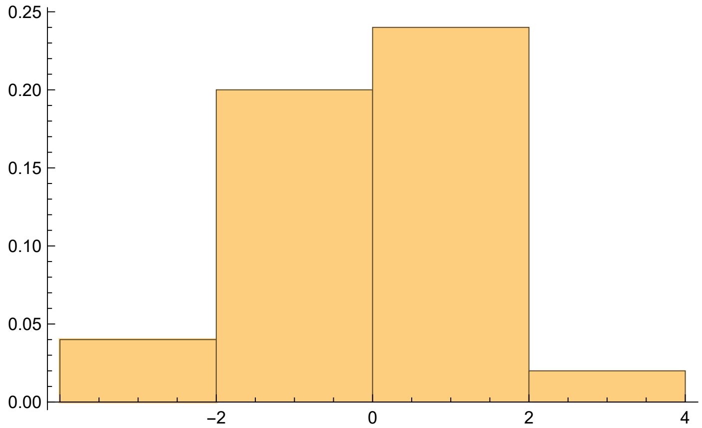
In[ ]:=
$$f[x\_, xi\_, h\_, \sigma\_, n\_] := \frac{1}{n\,h\,\sigma}\,\frac{1}{\sqrt{2\pi}}\,\mathrm{Sum}\left[\mathrm{Exp}\left[-\frac{(x - xi[[k]])^2}{2\,h^2\,\sigma^2}\right], \{k, n\}\right]$$

In[ ]:=
```
n = 25;
```

In[ ]:=
```
dat = RandomVariate[StudentTDistribution[5], n];
```

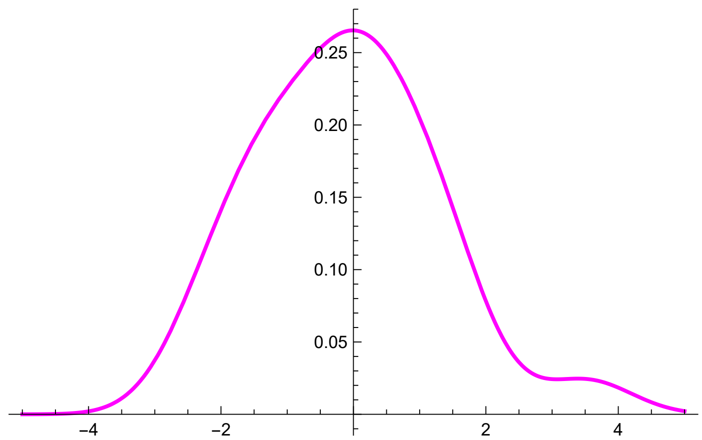
```
In[ ]:= grh = Histogram[dat, Automatic, "PDF"]
               histograma       automático   función
```

Out[ ]=



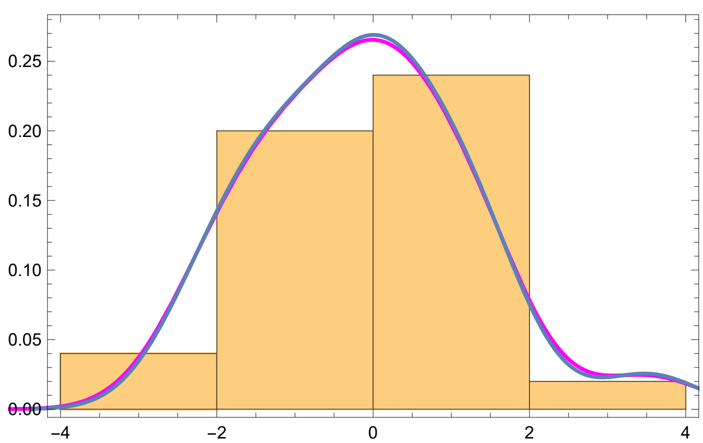
```
In[ ]:= grkde =
         Plot[f[x, dat, .5, StandardDeviation[dat], Length@dat], {x, -5, 5}, PlotStyle → Magenta]
         representación gráfica  desviación estándar  longitud  estilo de repr⋯ magenta
```

Out[ ]=



```
In[ ]:= Show[grh, grkde, SmoothHistogram[dat], Frame → True]
        muestra           histograma suave        marco   verdade
```

Out[ ]=

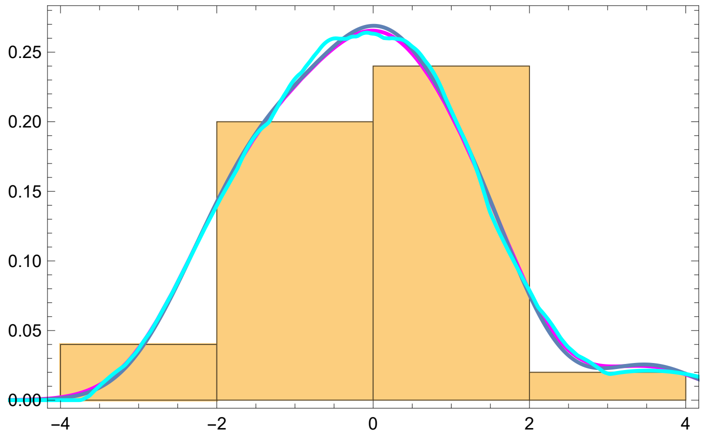


In[ ]:=
```
kde = SmoothKernelDistribution[dat, Automatic, "Epanechnikov"]
```
distribución de núcleo suave — automático

Out[ ]= DataDistribution[Type: SmoothKernel Data points: 25]

In[ ]:=
```
Show[grh, grkde, SmoothHistogram[dat],
 Plot[PDF[kde, x], {x, -5, 5}, PlotStyle → Cyan], Frame → True]
```
muestra — histograma suave — repr··· — función de densidad de probabi·· — estilo de repr··· — cian — marco — verdade

Out[ ]=



In[ ]:=
```
FindDistribution[dat]
```
encuentra distribución

Out[ ]= NormalDistribution[-0.0565974, 1.51965]

In[ ]:=
```
FindDistribution[dat, TargetFunctions → {StudentTDistribution}]
```
encuentra distribución — funciones objetivo — distribución t de Student

Out[ ]= StudentTDistribution[-0.134879, 1.26363, 19.996]

In[ ]:=
```
FindDistribution[dat, 5]
```
encuentra distribución

Out[ ]=
```
{NormalDistribution[-0.0565974, 1.51965], UniformDistribution[{-2.32081, 3.5216}],
 WeibullDistribution[1.66059, 2.82029, -2.58729],
 LogisticDistribution[-0.138034, 0.858585],
 StudentTDistribution[-0.0847648, 1.46812, 29.5562]}
```

Exoplanetas

In[ ]:=
```
Directory[]
```
directorio

Out[ ]= /Users/michel

In[ ]:=
```
NotebookDirectory[]
```
directorio de cuaderno

Out[ ]=
```
/Users/michel/Library/CloudStorage/OneDrive-uv.cl/cursos/pregrado/Estadisticas/2024/
  Clases/
```

In[ ]:=
```
SetDirectory[NotebookDirectory[]]
```
establece direct··· directorio de cuaderno

Out[ ]=
```
/Users/michel/Library/CloudStorage/OneDrive-uv.cl/cursos/pregrado/Estadisticas/2024/
  Clases
```

In[ ]:=
```
Directory[]
```
directorio

Out[ ]=
```
/Users/michel/Library/CloudStorage/OneDrive-uv.cl/cursos/pregrado/Estadisticas/2024/
  Clases
```

In[ ]:=
```
exopl = Import["exoplanetas-11-2024.csv"];
```
importa

In[ ]:=
```
Length@exopl
```
longitud

Out[ ]=
```
7345
```

In[ ]:=
```
header = exopl[[1]]
```

Out[ ]=
```
{name, planet_status, mass, mass_error_min, mass_error_max, mass_sini,
 mass_sini_error_min, mass_sini_error_max, radius, radius_error_min, radius_error_max,
 orbital_period, orbital_period_error_min, orbital_period_error_max,
 semi_major_axis, semi_major_axis_error_min, semi_major_axis_error_max,
 eccentricity, eccentricity_error_min, eccentricity_error_max, inclination,
 inclination_error_min, inclination_error_max, angular_distance, discovered,
 updated, omega, omega_error_min, omega_error_max, tperi, tperi_error_min,
 tperi_error_max, tconj, tconj_error_min, tconj_error_max, tzero_tr,
 tzero_tr_error_min, tzero_tr_error_max, tzero_tr_sec, tzero_tr_sec_error_min,
 tzero_tr_sec_error_max, lambda_angle, lambda_angle_error_min, lambda_angle_error_max,
 impact_parameter, impact_parameter_error_min, impact_parameter_error_max,
 tzero_vr, tzero_vr_error_min, tzero_vr_error_max, k, k_error_min, k_error_max,
 temp_calculated, temp_calculated_error_min, temp_calculated_error_max,
 temp_measured, hot_point_lon, geometric_albedo, geometric_albedo_error_min,
 geometric_albedo_error_max, log_g, publication, detection_type, mass_measurement_type,
 radius_measurement_type, alternate_names, molecules, star_name, ra, dec,
 mag_v, mag_i, mag_j, mag_h, mag_k, star_distance, star_distance_error_min,
 star_distance_error_max, star_metallicity, star_metallicity_error_min,
 star_metallicity_error_max, star_mass, star_mass_error_min, star_mass_error_max,
 star_radius, star_radius_error_min, star_radius_error_max, star_sp_type,
 star_age, star_age_error_min, star_age_error_max, star_teff, star_teff_error_min,
 star_teff_error_max, star_detected_disc, star_magnetic_field, star_alternate_names}
```

```
In[ ]:= Counts@exopl[[All, 2]]
         conteos        todo

Out[ ]= <|planet_status → 1, Confirmed → 7344|>

In[ ]:= masas = Select[exopl[[All, 3]], NumberQ];
                selecciona     todo      ¿número?

In[ ]:= Length@masas
         longitud

Out[ ]= 4386

In[ ]:= distmasas = SmoothKernelDistribution[Log10@masas]
                    distribución de núcleo suave  logaritmo en base

Out[ ]= DataDistribution[ Type: SmoothKernel
                          Data points: 4386 ]

In[ ]:= 10^Mean@distmasas
            media

Out[ ]= 1.30675

In[ ]:= Log10@Min[masas]
         logari… mínimo

Out[ ]= -4.20066
```

In[∘]:= 10^(-4.20066) jupiter mass in earth mass

Input interpretation:

convert $\frac{1}{10^{4.20066}} M_{♃}$ (Jupiter masses) to Earth masses

Result:

$0.02002\, M_{\oplus}$ (Earth masses)

Step-by-step solution

Additional conversions:

$\approx 1.2 \times 10^{23}$ kg (kilograms)

$\approx 1.2 \times 10^{26}$ grams

Comparisons:

$\approx 0.36 \times$ Mercury mass (3.30104×10^{23} kg)

$\approx 1.6 \times$ Earth's moon mass ($\approx 7.35 \times 10^{22}$ kg)

$\approx 7.2 \times$ Eris mass ($\approx 1.7 \times 10^{22}$ kg)

Interpretation:

mass

Corresponding quantities:

Relativistic energy E from $E = mc^2$:

1.075×10^{40} J (joules)

6.708×10^{58} eV (electronvolts)

Weight w of a body from $w = mg$:

1.173×10^{24} N (newtons)

1.173×10^{29} dynes

1.196×10^{26} ponds

2.636×10^{23} lbf (pounds-force)

WolframAlpha

```
In[∘]:= Sort[masas][[1 ;; 10]]
        ordena
Out[∘]= {0.000063, 0.00007, 0.00019, 0.00021, 0.00076, 0.00082, 0.0009, 0.00091, 0.001041, 0.00107}

In[∘]:= Reverse[Sort[masas]][[1 ;; 10]]
        invierte ⋯ ordena
Out[∘]= {74.6, 73.4, 72.07, 72., 72., 71.7, 71.64, 71.12, 71.03, 70.69}

In[∘]:= Max[masas]
        máximo
Out[∘]= 74.6
```

In[ ]:=
```
Min[masas]
mínimo
```
Out[ ]= 0.000063

In[ ]:=
```
Log10@Max[masas]
logari··· máximo
```
Out[ ]= 1.87274

In[ ]:=
```
Log10@Min[masas]
logari··· mínimo
```
Out[ ]= -4.20066

In[ ]:= **Mass Moon / mass Earth**

Input interpretation:

$$\frac{\text{Moon} \mid \text{mass}}{\text{Earth} \mid \text{mass}}$$

Result: Show details

0.0123 (1.23%)

WolframAlpha

In[ ]:=
```
grma = Plot[PDF[distmasas, m], {m, -5.68, 2.2},
       repr··· función de densidad de probabilidad
  Frame → True, Axes → False, PlotStyle → Magenta]
  marco  verd··· ejes  falso  estilo de repr··· magenta
```

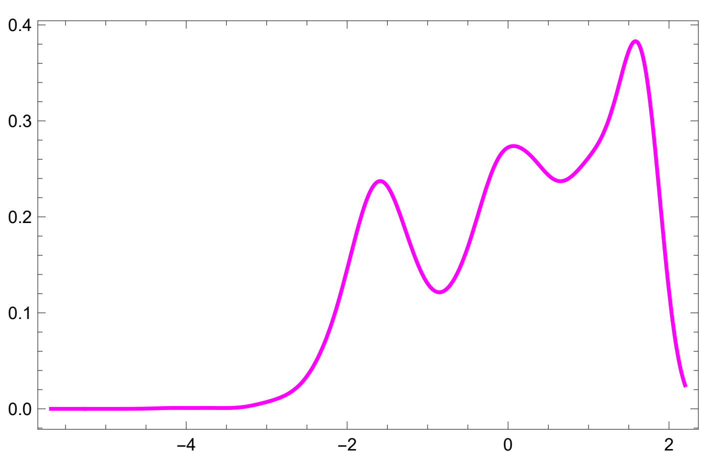
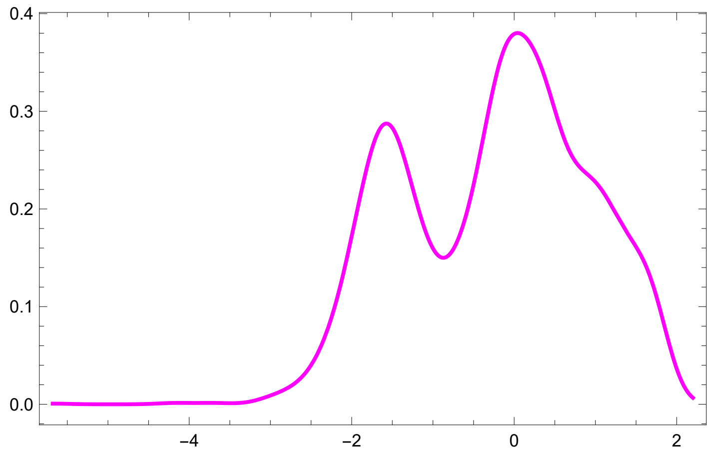
Out[ ]=


grafico con datos del año pasado



In[◦]:=
```
FindDistribution[Log10@masas]
```
encuentra distribución   logaritmo en base

Out[◦]=
```
MixtureDistribution[{0.266876, 0.464191, 0.268933},
 {NormalDistribution[-1.54588, 0.53526],
  NormalDistribution[0.317156, 0.541194], NormalDistribution[1.52709, 0.216679]}]
```

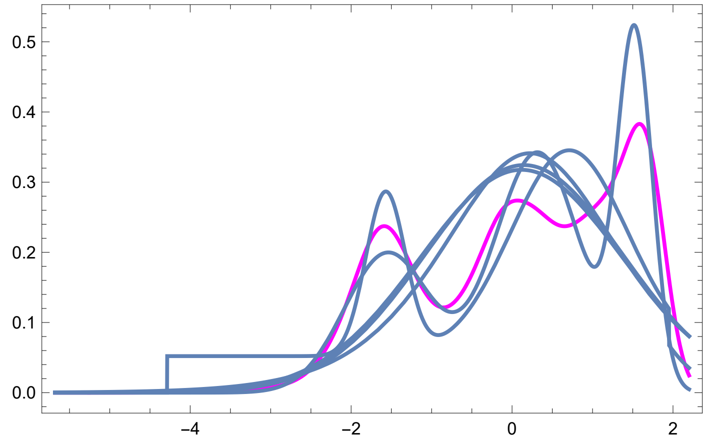
In[◦]:=
```
fdist = FindDistribution[Log10@masas, 5]
```
encuentra distribución   logaritmo en base 10

Out[◦]=
```
{MixtureDistribution[{0.266876, 0.464191, 0.268933},
  {NormalDistribution[-1.54588, 0.53526], NormalDistribution[0.317156, 0.541194],
   NormalDistribution[1.52709, 0.216679]}], MixtureDistribution[
  {0.143155, 0.532003, 0.324841}, {NormalDistribution[-1.5713, 0.245506],
   NormalDistribution[0.710644, 0.723507], UniformDistribution[{-4.28793, 1.95038}]}],
 NormalDistribution[0.117122, 1.25613], StudentTDistribution[0.146324, 1.22999, 4771.72],
 LogisticDistribution[0.223105, 0.732639]}
```

In[◦]:=
```
% // TableForm
```
forma de tabla

Out[◦]//TableForm=
```
MixtureDistribution[{0.266876, 0.464191, 0.268933}, {NormalDistribution[-1.54588, 0.53526],
MixtureDistribution[{0.143155, 0.532003, 0.324841}, {NormalDistribution[-1.5713, 0.245506],
NormalDistribution[0.117122, 1.25613]
StudentTDistribution[0.146324, 1.22999, 4771.72]
LogisticDistribution[0.223105, 0.732639]
```

In[ ]:=
```
Show[grma, Plot[Table[PDF[fdist[[k]], lgm], {k, 5}], {lgm, -5.68, 2.2}], PlotRange → All]
```
muestra repr··· tabla función de densidad de probabilidad rango de rep··· todo

Out[ ]=


In[ ]:=
```
fdist[[1]]
```

Out[ ]=
```
MixtureDistribution[{0.266876, 0.464191, 0.268933},
 {NormalDistribution[-1.54588, 0.53526],
  NormalDistribution[0.317156, 0.541194], NormalDistribution[1.52709, 0.216679]}]
```

In[ ]:=
```
Show[grma, Plot[PDF[fdist[[1]], lgm], {lgm, -5.68, 2.2}], PlotRange → All]
```
muestra repr··· función de densidad de probabilidad rango de rep··· todo

Out[ ]=