

Julia言語を使った統計学の勉強の仕方

- 黒木玄
- 2022-05-28~2022-05-28

このノートでは[Julia言語 \(https://julialang.org/\)](https://julialang.org/)を使用している:

- [Julia言語のインストールの仕方の一例 \(https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb\)](https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb)

自明な誤りを見つけたら、自分で訂正して読んで欲しい。大文字と小文字の混同や書き直しが不完全な場合や符号のミスは非常によくある。

このノートに書いてある式を文字通りにそのまま読んで正しいと思ってしまうとひどい目に合う可能性が高い。しかし、数が使われている文献には大抵の場合に文字通りに読むと間違っている式や主張が書いてあるので、内容を理解した上で訂正しながら読んで利用しなければいけない。実践的に数学を使う状況では他人が書いた式をそのまま信じていけない。

このノートの内容よりもさらに詳しいノートを自分で作ると勉強になるだろう。膨大な時間を取られることになるが、このノートの内容に関係することで飯を食っていく可能性がある人にはそのためにかけた時間は無駄にならないと思

目次

- ▼ [1 Julia言語の使い方](#)
 - [1.1 Juliaのインストールと実行](#)
 - [1.2 グラフの描き方](#)
 - [1.3 標準正規分布乱数のプロット](#)
 - [1.4 確率分布の扱い方](#)
 - [1.5 正規分布の確率密度関数のプロット](#)
- ▼ [2 Anscombeの例のプロット](#)
 - [2.1 RDatasets.jlパッケージのインストール](#)
 - [2.2 データのプロットの仕方](#)
- ▼ [3 Datasaurusの散布図のプロット](#)
 - [3.1 データの取得](#)
 - [3.2 散布図の作成](#)
- ▼ [4 中心極限定理のプロット](#)
 - [4.1 素朴なワークフロー](#)
 - [4.2 Revise.jlを使うワークフロー](#)

1 Julia言語の使い方

1.1 Juliaのインストールと実行

- (1) [JuliaLangについて検索 \(https://www.google.com/search?q=JuliaLang\)](https://www.google.com/search?q=JuliaLang) する。
- (2) [Julia言語のウェブサイト \(https://julialang.org/\)](https://julialang.org/) に移動する。
- (3) Downloadボタンを押して、[Download Julia \(https://julialang.org/downloads/\)](https://julialang.org/downloads/) のページに移動する。
- (4) 自分のパソコンに合わせて、current stable release のインストーラをダウンロード
- (5) それを実行してJulia言語の公式バイナリを自分のパソコンにインストールする。
- (6) juliaを実行すると以下のように表示される。

```
 _ _ _ _ _ _ _ _ _ _ | Documentation: https://docs.julialang.org
( ) _ _ _ _ _ _ _ _ _ |
 _ _ _ _ _ _ _ _ _ _ | Type "?" for help, "]"? for Pkg help.
| | | | | | | | | | |
| | | | | | | | | | | Version 1.7.3 (2022-05-06)
- / | \ _ _ _ _ _ _ _ _ _ | Official https://julialang.org/ release
| _ _ /
```

julia>

- julia> プロンプトで ? を入力するとヘルプモードに移る.
- julia> プロンプトで] を入力するとパッケージ管理モードに移る.
- それらのモードから julia> プロンプトに戻るにはバックスペースキーを押す.
- 至るところでタブキーによる補完を使える.
- julia> プロンプトでJulia言語のコードを入力して終わりに ; を付けて実行すると計算結果が表示されない.

問題: 以下を自分がインストールしたjuliaで実行してみよ.

```
In [1]: 1 1 + 1
```

```
Out[1]: 2
```

```
In [2]: 1 sin(pi/6)
```

```
Out[2]: 0.49999999999999994
```

```
In [3]: 1 sinpi(1/6)
```

```
Out[3]: 0.5
```

1.2 グラフの描き方

(7) 統計学対応のグラフ作画パッケージを入れるためには

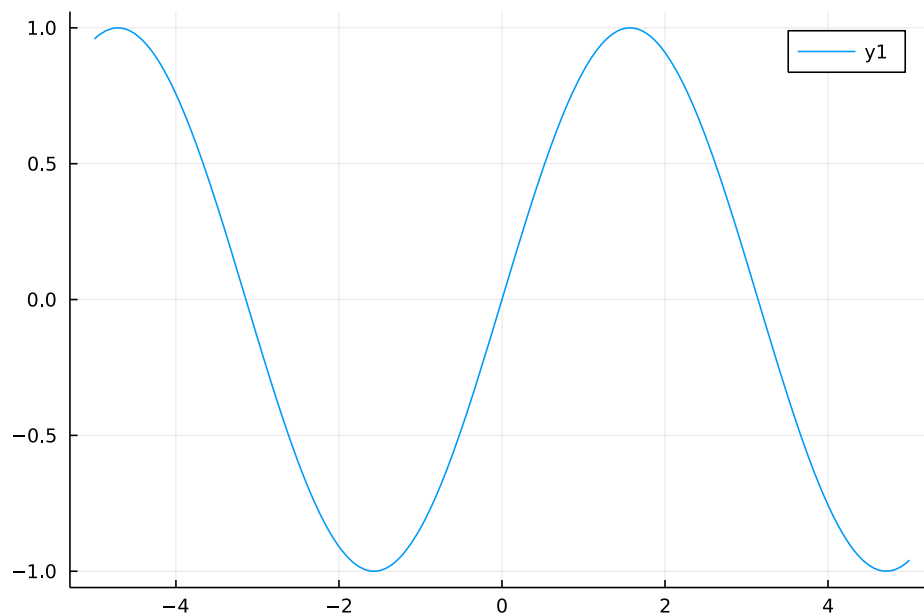
```
julia> using StatsPlots
```

を実行するだけでよい. インストールするかどうかをjuliaが聞いて来た場合には y と答える.

問題: 自分のパソコンで using StatsPlots の後に以下を実行してグラフを作画してみよ.

```
In [4]: 1 using StatsPlots
        2 plot(sin)
```

```
Out[4]:
```

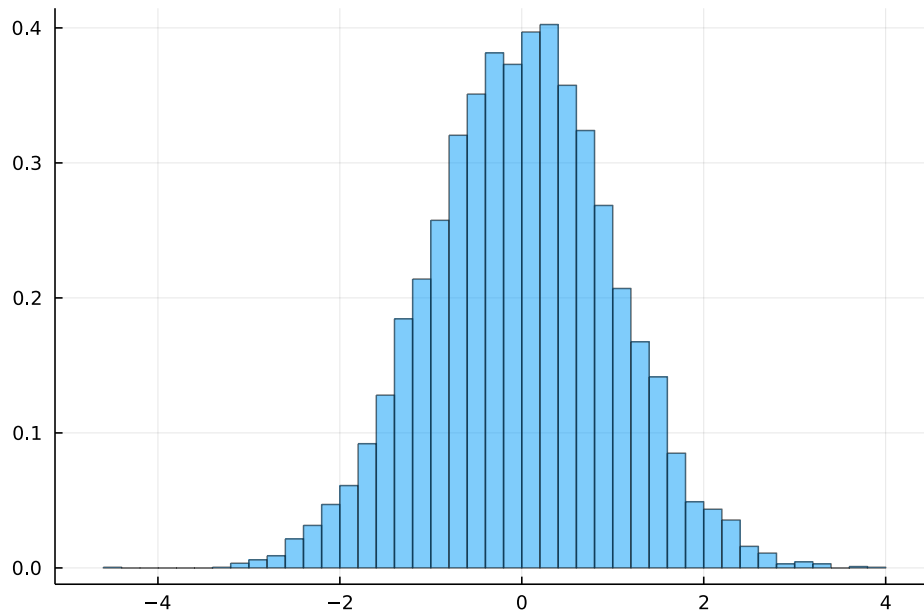


1.3 標準正規分布乱数のプロット

```
In [5]: 1 Z = randn(10^4);
```

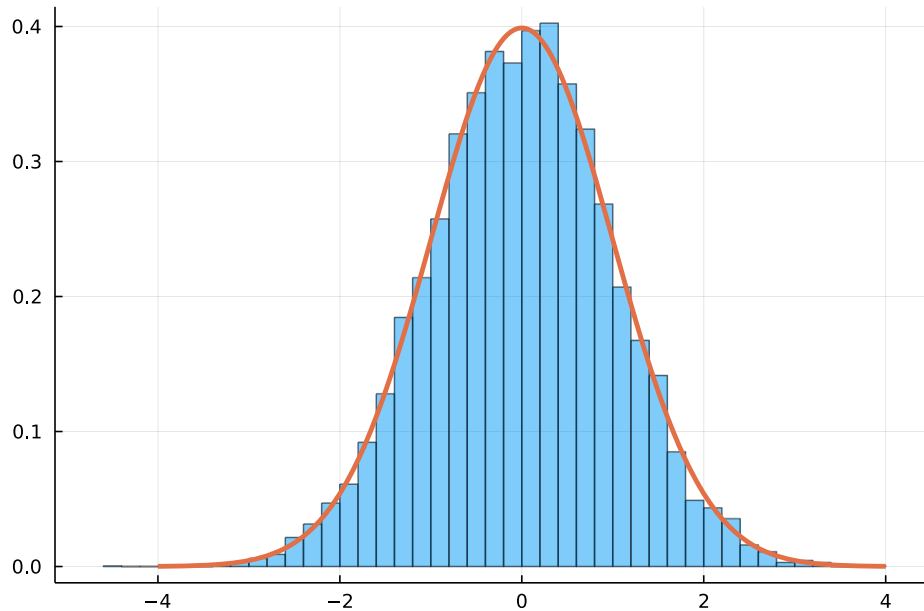
```
In [6]: 1 histogram(Z; norm=true, alpha=0.5, label="")
```

Out[6]:



```
In [7]: 1 plot!(x → exp(-x^2/2)/sqrt(2pi), -4, 4; label="", lw=3)
```

Out[7]:



1.4 確率分布の扱い方

(7) 確率分布を扱うためのパッケージを入れるためには

```
julia> using Distributions
```

を実行するだけでよい。インストールするかどうかをjuliaが聞いて来た場合には y と答える。

しかし、この機会にパッケージモードでのパッケージの入れ方についてもこの機会に学んでおいた方がよい。

```
julia> ]
```

と] を入力すると、

```
pkg>
```

と表示されてパッケージ管理モードに移行できる(戻るには pkg> プロンプトの直後にバックスペースを押せばよい)。

以下を実行すると、Distributions.jl パッケージをjuliaの環境にインストールできる。

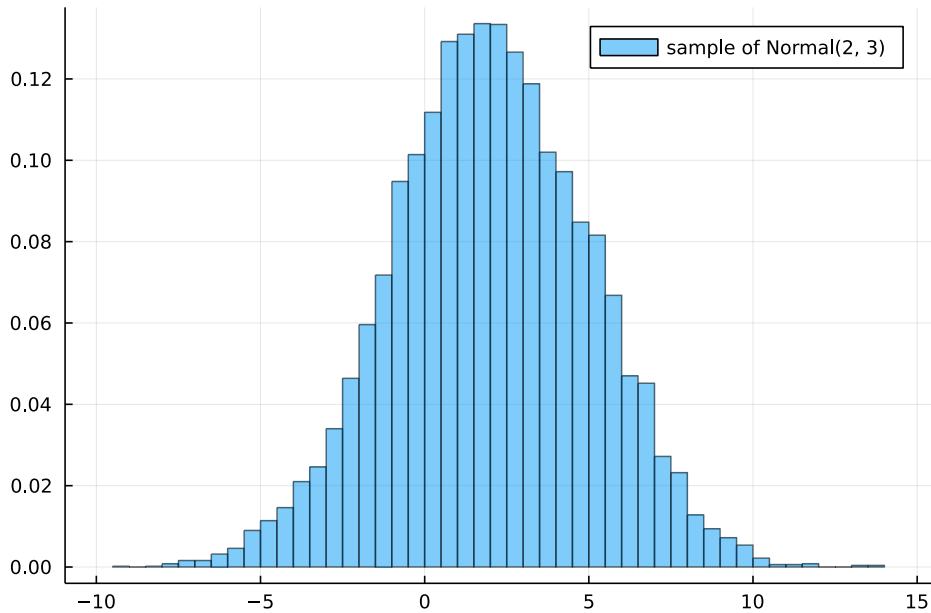
```
pkg> add Distributions
```

1.5 正規分布の確率密度関数のプロット

```
In [8]: 1 using Distributions
        2 X = rand(Normal(2, 3), 10^4);
```

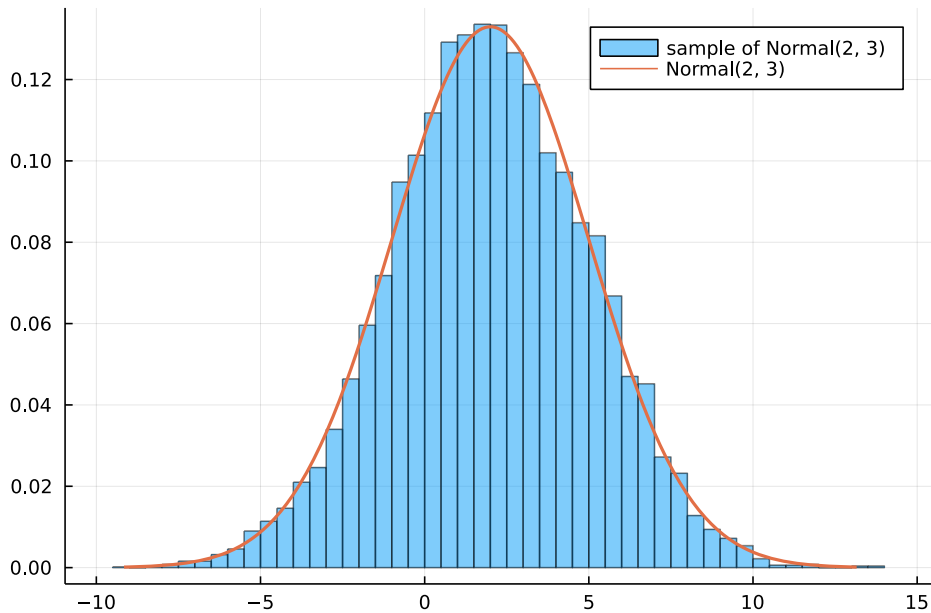
```
In [9]: 1 histogram(X; norm=true, alpha=0.5, label="sample of Normal(2, 3)")
```

Out[9]:



```
In [10]: 1 plot!(Normal(2, 3); label="Normal(2, 3)", lw=2)
```

Out[10]:



2 Anscombeの例のプロット

2.1 RDatasets.jlパッケージのインストール

確率分布を扱うためのパッケージを入れるためには

```
julia> using RDatasets
```

を実行するだけでよい。インストールするかどうかをjuliaが聞いて来た場合には y と答える。もしくはパッケージ管理モードで

```
pkg> add RDatasets
```

を実行してもよい。その場合には julia> プロンプトで改めて using RDatasets を実行する必要がある。

```
In [11]: 1 using RDatasets
        2 anscombe = dataset("datasets", "anscombe")
```

Out[11]: 11 rows × 8 columns

	X1	X2	X3	X4	Y1	Y2	Y3	Y4
	Int64	Int64	Int64	Int64	Float64	Float64	Float64	Float64
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.1	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.1	5.39	12.5
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89

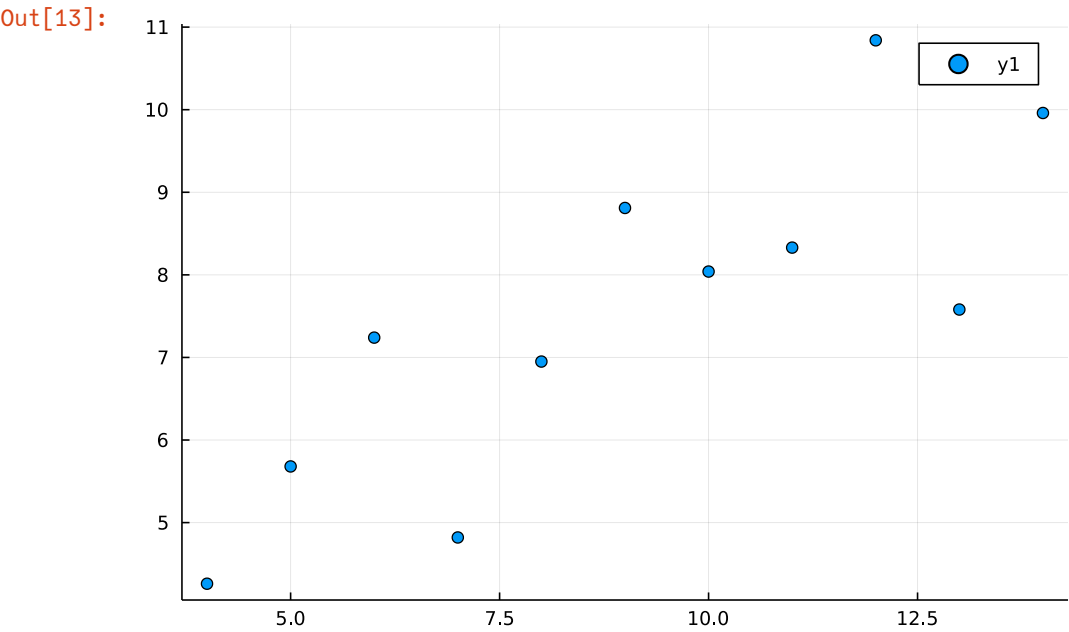
2.2 データのプロットの仕方

以下ではデータ1の場合のプロットの仕方を説明しよう.

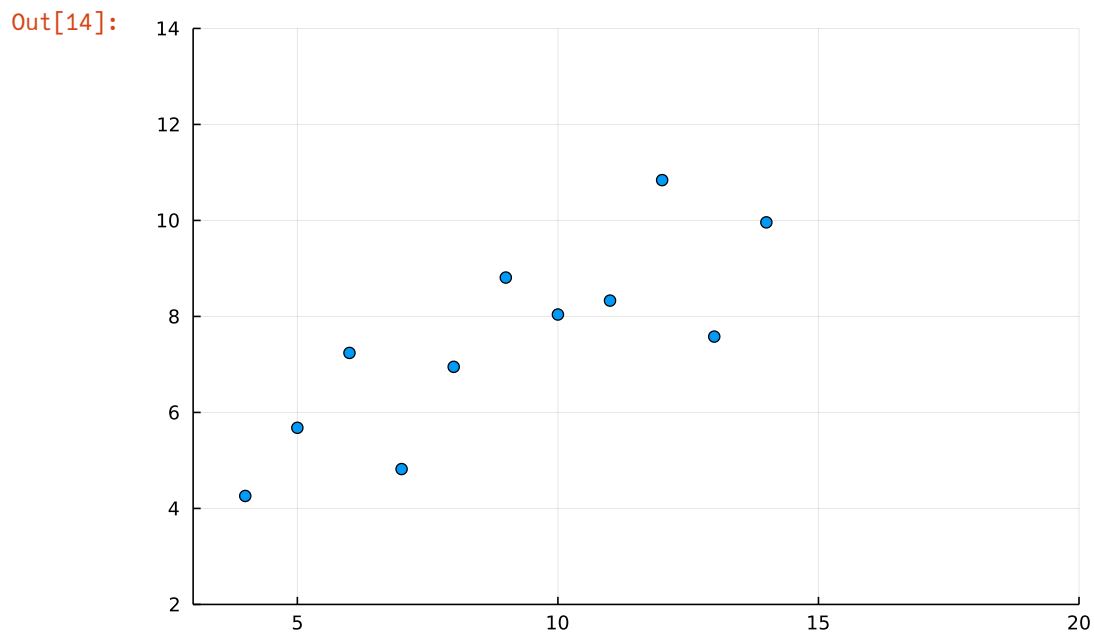
```
In [12]: 1 # x, y にデータを入れる
        2 x, y = anscombe.X1, anscombe.Y1
```

Out[12]: ([10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5], [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68])

```
In [13]: 1 # 散布図を描いてみる
        2 using StatsPlots
        3 scatter(x, y)
```



```
In [14]: 1 # xlim, ylimなどを追加
        2 scatter(x, y; label="", xlim=(3, 20), ylim=(2, 14))
```



```
In [15]: 1 # データの標本平均や不偏分散・不偏共分散を計算
        2 xbar = mean(x)
```

Out[15]: 9.0

```
In [16]: 1 ybar = mean(y)
```

Out[16]: 7.500909090909093

```
In [17]: 1 sx2 = var(x)
```

Out[17]: 11.0

```
In [18]: 1 sy2 = var(y)
```

Out[18]: 4.127269090909091

```
In [19]: 1 sxy = cov(x, y)
```

Out[19]: 5.501

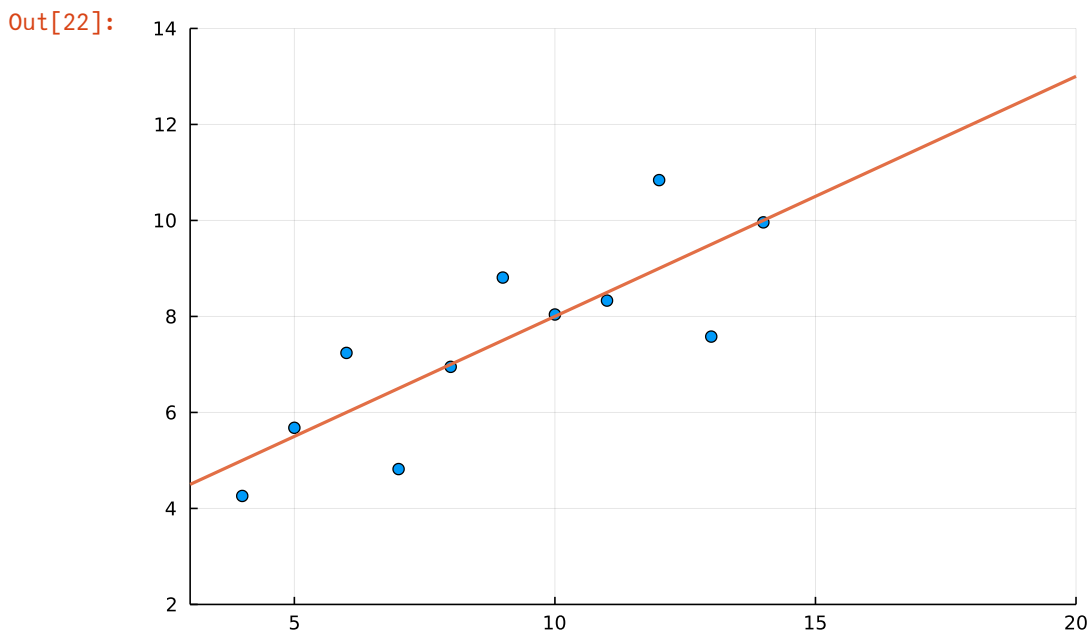
```
In [20]: 1 betahat = sxy/sx2
```

Out[20]: 0.5000909090909091

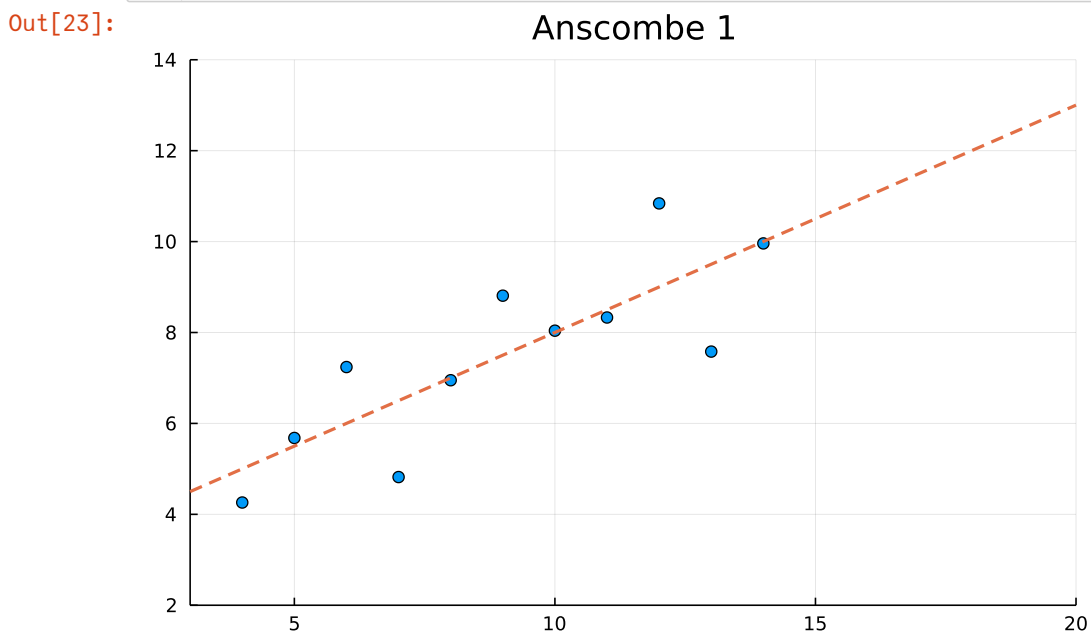
```
In [21]: 1 alphahat = ybar - betahat*xbar
```

Out[21]: 3.0000909090909103

```
In [22]: 1 scatter(x, y; label="", xlim=(3, 20), ylim=(2, 14))
2 plot!(x → alphahat + betahat*x, 3, 20; label="", lw=2)
```



```
In [23]: 1 scatter(x, y; label="", xlim=(3, 20), ylim=(2, 14), title="Anscombe 1")
2 plot!(x → alphahat + betahat*x, 3, 20; label="", lw=2, ls=:dash)
```



```
In [24]: 1 # design matrix
2 X = x .^ (0:1)'
```

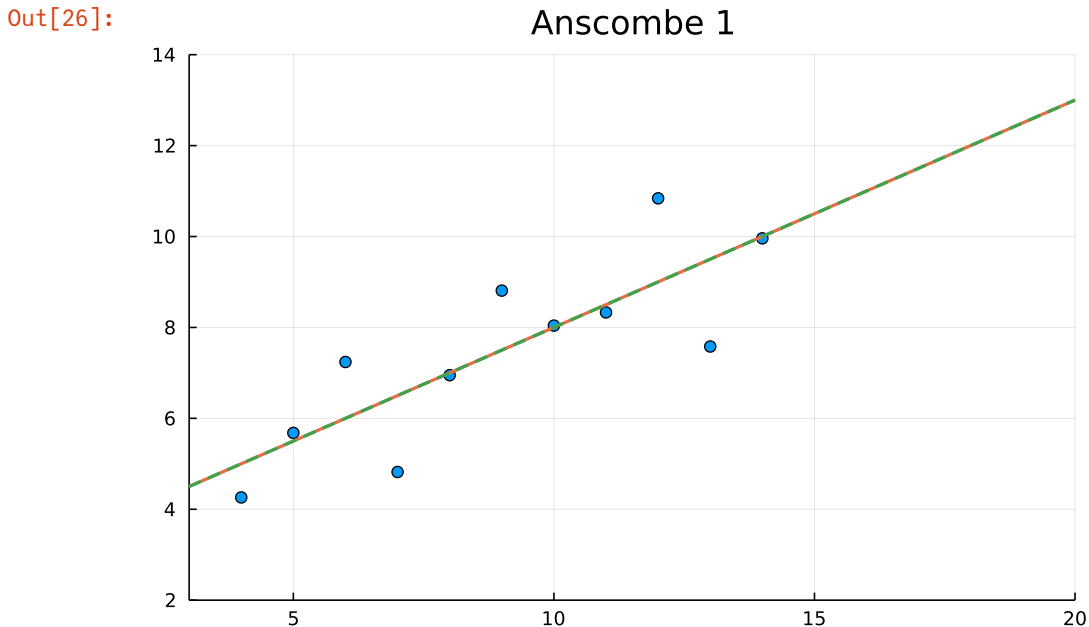
Out[24]: 11×2 Matrix{Int64}:

```
1 10
1  8
1 13
1  9
1 11
1 14
1  6
1  4
1 12
1  7
1  5
```

```
In [25]: 1 # 最小二乗法を一発実現 (計画行列の一般逆行列をyにかける)
        2 alphahat2, betahat2 = X \ y
```

```
Out[25]: 2-element Vector{Float64}:
          3.000090909090908
          0.5000909090909091
```

```
In [26]: 1 # 2つの直線はぴったり重なり合う。
        2 scatter(x, y; label="", xlim=(3, 20), ylim=(2, 14), title="Anscombe 1")
        3 plot!(x → alphahat + betahat*x, 3, 20; label="", lw=2)
        4 plot!(x → alphahat2 + betahat2*x, 3, 20; label="", lw=2, ls=:dash)
```



問題: 他のアンスコムของデータについて同様のグラフを作成せよ.

3 Datasaurusの散布図のプロット

以下のデータは「[条件付き確率分布, 尤度, 推定, 記述統計](#)

<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estin>からのコピー＆ペースト.



3.1 データの取得

- <http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html>
(<http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html>)
 - <https://www.dropbox.com/sh/xaxpz3pm5r5awes/AADUbGVagF9i4RmM9JkPtviEa?dl=0>
(<https://www.dropbox.com/sh/xaxpz3pm5r5awes/AADUbGVagF9i4RmM9JkPtviEa?dl=0>)
- <https://visualizing.jp/the-datasaurus-dozen/> (<https://visualizing.jp/the-datasaurus-dozen/>)

In [27]:

```
1 datasaurus = [  
2     55.3846 97.1795  
3     51.5385 96.0256  
4     46.1538 94.4872  
5     42.8205 91.4103  
6     40.7692 88.3333  
7     38.7179 84.8718  
8     35.6410 79.8718  
9     33.0769 77.5641  
10    28.9744 74.4872  
11    26.1538 71.4103  
12    23.0769 66.4103  
13    22.3077 61.7949  
14    22.3077 57.1795  
15    23.3333 52.9487  
16    25.8974 51.0256  
17    29.4872 51.0256  
18    32.8205 51.0256  
19    35.3846 51.4103  
20    40.2564 51.4103  
21    44.1026 52.9487  
22    46.6667 54.1026  
23    50.0000 55.2564  
24    53.0769 55.6410  
25    56.6667 56.0256  
26    59.2308 57.9487  
27    61.2821 62.1795  
28    61.5385 66.4103  
29    61.7949 69.1026  
30    57.4359 55.2564  
31    54.8718 49.8718  
32    52.5641 46.0256  
33    48.2051 38.3333  
34    49.4872 42.1795  
35    51.0256 44.1026  
36    45.3846 36.4103  
37    42.8205 32.5641  
38    38.7179 31.4103  
39    35.1282 30.2564  
40    32.5641 32.1795  
41    30.0000 36.7949  
42    33.5897 41.4103  
43    36.6667 45.6410  
44    38.2051 49.1026  
45    29.7436 36.0256  
46    29.7436 32.1795  
47    30.0000 29.1026  
48    32.0513 26.7949  
49    35.8974 25.2564  
50    41.0256 25.2564  
51    44.1026 25.6410  
52    47.1795 28.7180  
53    49.4872 31.4103  
54    51.5385 34.8718  
55    53.5897 37.5641  
56    55.1282 40.6410  
57    56.6667 42.1795  
58    59.2308 44.4872  
59    62.3077 46.0256  
60    64.8718 46.7949  
61    67.9487 47.9487  
62    70.5128 53.7180  
63    71.5385 60.6410  
64    71.5385 64.4872  
65    69.4872 69.4872  
66    46.9231 79.8718  
67    48.2051 84.1026  
68    50.0000 85.2564  
69    53.0769 85.2564  
70    55.3846 86.0256  
71    56.6667 86.0256  
72    56.1538 82.9487  
73    53.8462 80.6410  
74    51.2821 78.7180  
75    50.0000 78.7180  
76    47.9487 77.5641  
77    29.7436 59.8718
```

```

78      29.7436 62.1795
79      31.2821 62.5641
80      57.9487 99.4872
81      61.7949 99.1026
82      64.8718 97.5641
83      68.4615 94.1026
84      70.7692 91.0256
85      72.0513 86.4103
86      73.8462 83.3333
87      75.1282 79.1026
88      76.6667 75.2564
89      77.6923 71.4103
90      79.7436 66.7949
91      81.7949 60.2564
92      83.3333 55.2564
93      85.1282 51.4103
94      86.4103 47.5641
95      87.9487 46.0256
96      89.4872 42.5641
97      93.3333 39.8718
98      95.3846 36.7949
99      98.2051 33.7180
100     56.6667 40.6410
101     59.2308 38.3333
102     60.7692 33.7180
103     63.0769 29.1026
104     64.1026 25.2564
105     64.3590 24.1026
106     74.3590 22.9487
107     71.2821 22.9487
108     67.9487 22.1795
109     65.8974 20.2564
110     63.0769 19.1026
111     61.2821 19.1026
112     58.7179 18.3333
113     55.1282 18.3333
114     52.3077 18.3333
115     49.7436 17.5641
116     47.4359 16.0256
117     44.8718 13.7180
118     48.7179 14.8718
119     51.2821 14.8718
120     54.1026 14.8718
121     56.1538 14.1026
122     52.0513 12.5641
123     48.7179 11.0256
124     47.1795  9.8718
125     46.1538  6.0256
126     50.5128  9.4872
127     53.8462 10.2564
128     57.4359 10.2564
129     60.0000 10.6410
130     64.1026 10.6410
131     66.9231 10.6410
132     71.2821 10.6410
133     74.3590 10.6410
134     78.2051 10.6410
135     67.9487  8.7180
136     68.4615  5.2564
137     68.2051  2.9487
138     37.6923 25.7692
139     39.4872 25.3846
140     91.2821 41.5385
141     50.0000 95.7692
142     47.9487 95.0000
143     44.1026 92.6923
144 ];

```

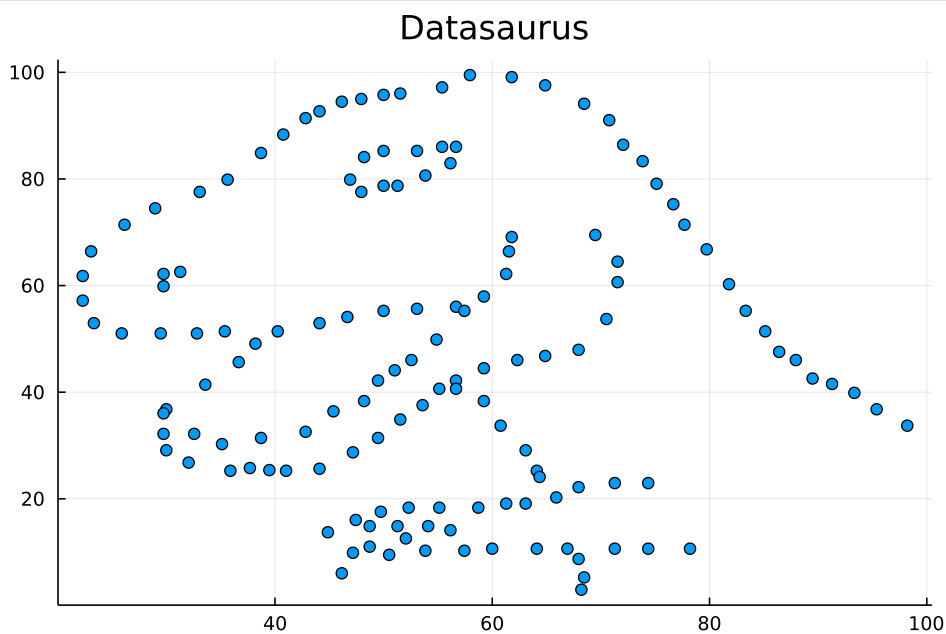
3.2 散布図の作成

```
In [28]: 1 # 行列Aの第j列はA[:,j]
        2 @show datasaurus[:,1];
```

```
datasaurus[:, 1] = [55.3846, 51.5385, 46.1538, 42.8205, 40.7692, 38.7179, 35.641, 33.0769, 28.9744, 26.1538, 23.0769, 22.3077, 22.3077, 23.3333, 25.8974, 29.4872, 32.8205, 35.3846, 40.2564, 44.1026, 46.6667, 50.0, 53.0769, 56.6667, 59.2308, 61.2821, 61.5385, 61.7949, 57.4359, 54.8718, 52.5641, 48.2051, 49.4872, 51.0256, 45.3846, 42.8205, 38.7179, 35.1282, 32.5641, 30.0, 33.5897, 36.6667, 38.2051, 29.7436, 29.7436, 30.0, 32.0513, 35.8974, 41.0256, 44.1026, 47.1795, 49.4872, 51.5385, 53.5897, 55.1282, 56.6667, 59.2308, 62.3077, 64.8718, 67.9487, 70.5128, 71.5385, 71.5385, 69.4872, 46.9231, 48.2051, 50.0, 53.0769, 55.3846, 56.6667, 56.1538, 53.8462, 51.2821, 50.0, 47.9487, 29.7436, 29.7436, 31.2821, 57.9487, 61.7949, 64.8718, 68.4615, 70.7692, 72.0513, 73.8462, 75.1282, 76.6667, 77.6923, 79.7436, 81.7949, 83.3333, 85.1282, 86.4103, 87.9487, 89.4872, 93.3333, 95.3846, 98.2051, 56.6667, 59.2308, 60.7692, 63.0769, 64.1026, 64.359, 74.359, 71.2821, 67.9487, 65.8974, 63.0769, 61.2821, 58.7179, 55.1282, 52.3077, 49.7436, 47.4359, 44.8718, 48.7179, 51.2821, 54.1026, 56.1538, 52.0513, 48.7179, 47.1795, 46.1538, 50.5128, 53.8462, 57.4359, 60.0, 64.1026, 66.9231, 71.2821, 74.359, 78.2051, 67.9487, 68.4615, 68.2051, 37.6923, 39.4872, 91.2821, 50.0, 47.9487, 44.1026]
```

```
In [29]: 1 using StatsPlots
        2 scatter(datasaurus[:,1], datasaurus[:,2]; label="", title="Datasaurus")
```

Out[29]:



問題: [Datasaurusについて検索 \(https://www.google.com/search?q=Datasaurus\)](https://www.google.com/search?q=Datasaurus) して見つけた解説を読み。

4 中心極限定理のプロット

4.1 素朴なワークフロー

```
In [30]: 1 using StatsPlots, Distributions
```

```
In [31]: 1 # 確率分布を dist と書く.
        2 dist = Gamma(2, 3)
```

Out[31]: Gamma{Float64}{α=2.0, θ=3.0}

```
In [32]: 1 # 確率分布 dist のサイズ n のサンプルを L 個生成
2 n = 10
3 L = 10^4
4 Xs = [rand(dist, n) for _ in 1:L]
```

```
Out[32]: 10000-element Vector{Vector{Float64}}:
 [2.635522985081041, 0.9622722381233564, 3.5738291251461263, 0.21932438876644544, 7.751688469946
 901, 0.9637638427242243, 10.847338678100066, 26.562798194507465, 12.50287321503098, 8.2277393033
 12017]
 [8.876744459276773, 10.172269240841013, 7.08818041686744, 2.6362749040534648, 3.186617065796940
 4, 5.734861388858096, 2.671493527764686, 0.6220595247340548, 7.1161239463474715, 10.887243036283
 943]
 [13.601542797928643, 2.6084439639961854, 26.656076788709935, 8.67810296714426, 8.0937007293061
 3, 13.86589240784676, 2.090466178759308, 11.53700424968032, 0.379263087555967, 2.00899260251785
 4]
 [1.6775323174685992, 1.3441534450369752, 15.119281829495774, 1.914934502256298, 3.5301851328134
 01, 7.075089832221566, 3.4048861302843223, 10.438000208123263, 1.0568737652978486, 0.59471588773
 62813]
 [4.666043919770652, 3.8613213192471125, 4.188269282690952, 3.6140996208898124, 4.26473256200404
 55, 3.1858763753400137, 3.646091305224089, 5.269327014004344, 3.8978674658394077, 4.908973983758
 463]
 [4.9150965826734225, 1.6356136527533214, 7.365650095843915, 3.3197495183256427, 6.6165013405022
 26, 3.0900322804206244, 12.590976661785037, 6.819410094562741, 10.047401966192925, 2.72092125108
 8169]
 [4.649099093021919, 15.617752905155864, 2.657022989328737, 2.270826774437134, 12.4678783988036
 2, 5.766110708171157, 6.459146280816039, 4.045462629682636, 5.417845915283674, 17.37550407257361
 6]
 [6.113216098645418, 4.8258673520655275, 2.4444269477312774, 16.244502981814332, 6.4488859504075
 275, 4.977376520247162, 2.8639248804504263, 12.160311817502649, 2.485147918581819, 1.22915263065
 66785]
 [4.300358215944927, 1.3309383706019464, 1.9758808728268358, 9.272231090331937, 20.6751643406904
 7, 10.980382475014505, 3.557686362572265, 5.097032284227044, 3.018208845957548, 5.77224564234784
 5]
 [0.8390910927717168, 2.203666553831285, 8.80242040133952, 17.4266509172475, 5.869291451072286,
 6.830167670949732, 5.13882141091262, 1.9384345006952224, 23.252548292062006, 2.7892406681422677]
 [2.3610238973206323, 3.616696953870841, 3.251292725970478, 5.849390657493353, 4.86821233166904
 6, 3.900239669339899, 0.8443142417767185, 4.7761189056539095, 6.889543051447067, 2.4287068863957
 257]
 [3.2224715436517988, 3.0033627457697656, 5.547896341515606, 8.283541753661872, 3.0446283809264
 2, 4.345030637882381, 2.7081859307855947, 5.375304642331395, 2.5122002969662063, 4.2487114214977
 32]
 [3.5471129125172016, 5.225903457665368, 8.53993351283934, 0.539506301822759, 10.45102295880514,
 1.1771322928374441, 6.840263620404789, 20.41932474271385, 6.540120134451808, 3.8069609257632333]
 ⋮
 [13.009159988426457, 3.556269234358348, 9.797586002281372, 2.7119851716893613, 6.53735969091318
 8, 8.14096139055462, 19.490049987206625, 2.325337856422863, 3.7532907820249717, 2.40038589460639
 83]
 [1.817989651448665, 2.6414941269947096, 3.5098988772102047, 12.063175403473396, 6.2526853416055
 59, 5.0347309071717845, 1.4723995399138587, 6.094213252035036, 2.7720880139179194, 9.32116599096
 858]
 [3.8082483736682358, 2.5685535528980434, 12.51150731132481, 5.545818553141018, 0.44370916521910
 46, 7.202521796566505, 5.072440944211424, 3.4626737081391443, 5.121829411377483, 6.2024312109375
 27]
 [17.795956061401537, 31.146995610517628, 5.783226772792608, 8.390421466989544, 3.4170951558673
 5, 12.190330523264787, 1.8366357530355784, 11.406632990857426, 5.5253728326473786, 3.33985163731
 7272]
 [3.326342485768449, 1.304335237930767, 6.275737393886105, 2.619295588676364, 9.651904375708838,
 2.951088575404821, 0.7899213114176036, 1.8036084911958676, 1.9693513682220471, 6.405736676583086
 5]
 [7.8489999584813415, 7.741305350920602, 3.475801119949569, 2.745254761589453, 5.50332275681481
 2, 12.430904030787374, 8.706168977533476, 16.681748488207962, 15.544687903130555, 1.046240791446
 1607]
 [10.096165874356464, 5.5203293781323675, 12.73071937705351, 7.6205328884235115, 7.0867356755884
 98, 1.7792149812276317, 6.183054412507923, 6.8332358687926, 3.0519461626647684, 8.1847137910106
 67]
 [4.949132313819705, 8.896846170314591, 5.9436521488783995, 2.930579613964694, 5.730015087208041
 5, 7.526276984044081, 2.0982709189985336, 1.7464567580027777, 3.2362011620782427, 2.279723195717
 854]
 [9.363795636652016, 4.437918147456635, 14.581643301896534, 10.780484279711924, 5.61535015918286
 7, 3.6015530212172244, 4.425760322849657, 2.6441451108551135, 6.085925625467905, 6.5570305307824
 79]
 [2.729589913535599, 2.980883626731754, 6.041774874807794, 3.8276778258349466, 4.08544711579297
 9, 4.93850819956958, 3.378306158690931, 7.528773340782513, 11.71877414772664, 1.373617757554562
 3]
 [4.561559895196093, 2.370956204147278, 3.770299973212631, 7.654742646613231, 4.020797433307321
 5, 2.512538418565086, 1.7415272827491948, 17.329706751167016, 4.696589586857167, 5.7978730672332
 11]
```

```
[4.461865516416442, 11.987602191139581, 2.273978859983986, 3.4352784152852935, 4.125372130213126, 7.311477065605714, 7.010946375951115, 1.5059718008167007, 3.2083213922876417, 4.580927243999986]
```

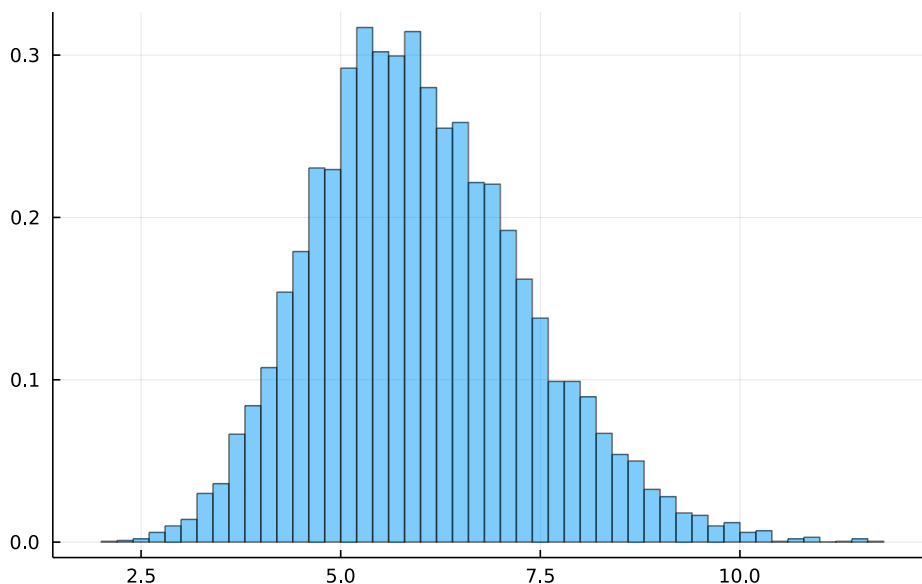
```
In [33]: 1 # L個のサイズnのサンプルの各々の標本平均を計算
        2 Xbars = mean.(Xs)
```

```
Out[33]: 10000-element Vector{Float64}:
```

```
7.424715044073862
5.8991867510823885
8.951948577344536
4.615565305073433
4.1502602848768895
5.912135344414803
7.6726649767274395
5.979281309810282
6.598012850051532
7.509033295902417
3.8785539320937668
4.229133369498877
6.708728085982093
⋮
7.172238599848422
5.097984110473972
5.193973402748329
10.08325188046911
3.7097321504793945
8.172443413886132
6.90867361278446
4.5337154353026925
6.809360613607235
4.86033529610273
5.4456591259048235
4.990174099169959
```

```
In [34]: 1 # Xbarのヒストグラムを表示
        2 histogram(Xbars; norm=true, alpha=0.5, label="", title="$dist, n=$n")
```

```
Out[34]: Gamma{Float64}(α=2.0, θ=3.0), n=10
```

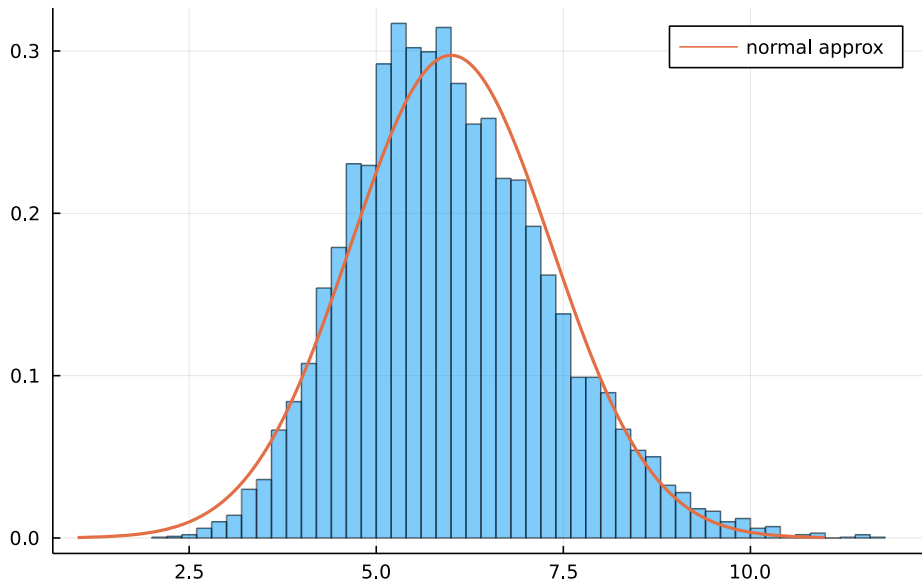


```
In [35]: 1 # 中心極限定理による正規分布近似を設定
        2 mu = mean(dist)
        3 sigma = std(dist)
        4 normal_approx = Normal(mu, sigma/sqrt(n))
```

```
Out[35]: Normal{Float64}(μ=6.0, σ=1.3416407864998736)
```

```
In [36]: 1 # 上のグラフに重ねて正規分布をプロット
        2 plot!(normal_approx; label="normal approx", lw=2)
```

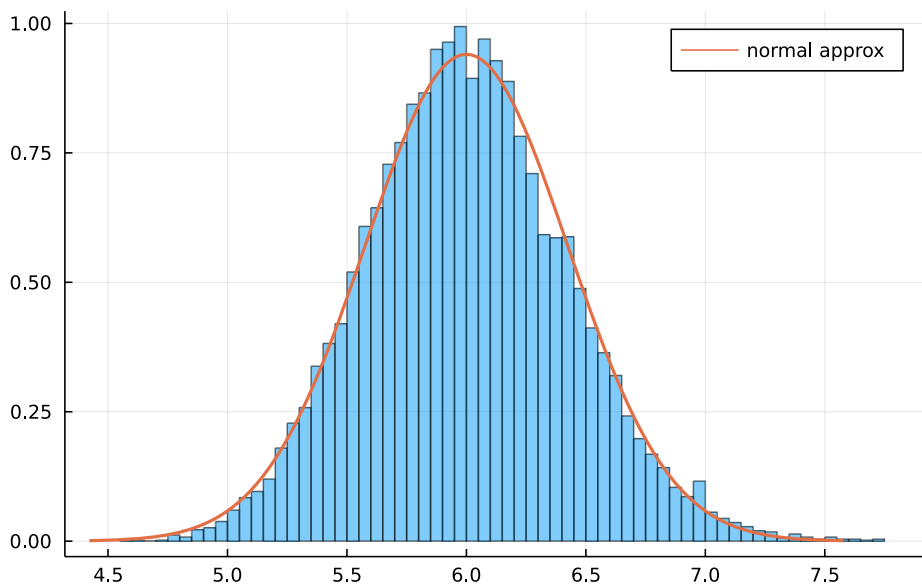
Out[36]: **Gamma{Float64}($\alpha=2.0$, $\theta=3.0$), $n=10$**



$n = 10$ が小さすぎてずれが大きい.

```
In [37]: 1 # nを大きくしてやり直してみる.
        2 n = 100
        3 L = 10^4
        4 Xs = [rand(dist, n) for _ in 1:L]
        5 Xbars = mean.(Xs)
        6 histogram(Xbars; norm=true, alpha=0.5, label="", title="$dist, n=$n")
        7 mu = mean(dist)
        8 sigma = std(dist)
        9 normal_approx = Normal(mu, sigma/sqrt(n))
        10 plot!(normal_approx; label="normal approx", lw=2)
```

Out[37]: **Gamma{Float64}($\alpha=2.0$, $\theta=3.0$), $n=100$**



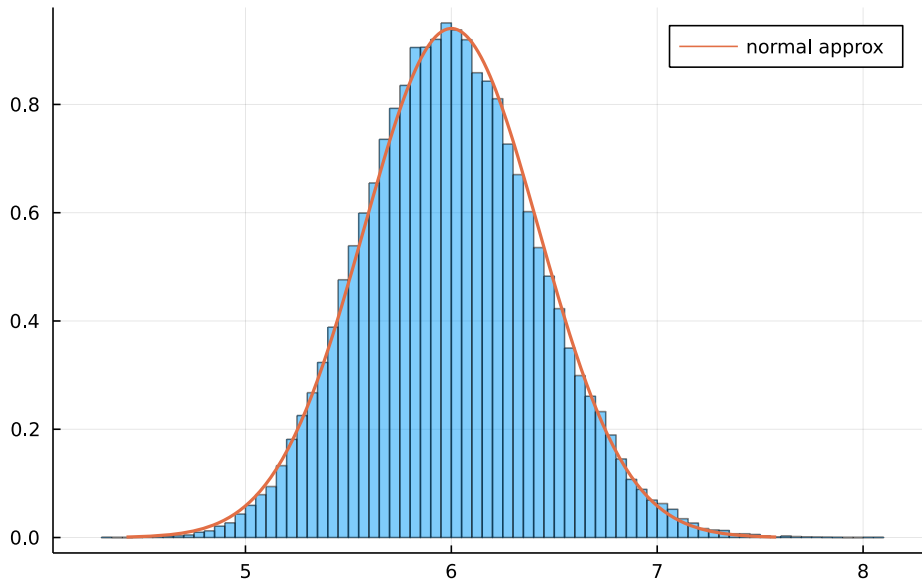
$n = 100$ にしたら, 正規分布とよく一致するようになった.

```

In [38]: 1 # Lも大きくしてやり直してみる.
2 n = 100
3 L = 10^5
4 Xs = [rand(dist, n) for _ in 1:L]
5 Xbars = mean.(Xs)
6 histogram(Xbars; norm=true, alpha=0.5, label="", title="$dist, n=$n")
7 mu = mean(dist)
8 sigma = std(dist)
9 normal_approx = Normal(mu, sigma/sqrt(n))
10 plot!(normal_approx; label="normal approx", lw=2)

```

Out[38]: Gamma{Float64}($\alpha=2.0$, $\theta=3.0$), n=100

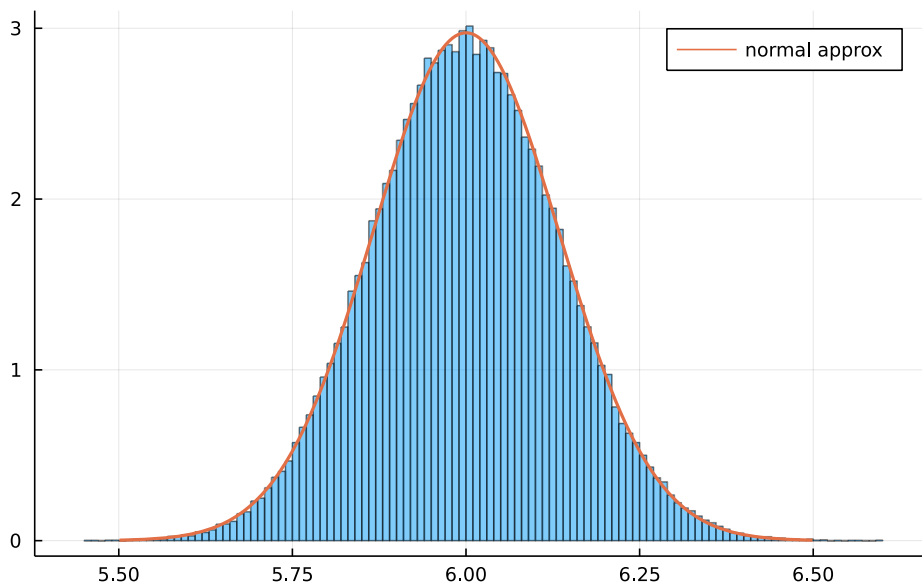


```

In [39]: 1 # Lも大きくしてやり直してみる.
2 n = 1000
3 L = 10^5
4 Xs = [rand(dist, n) for _ in 1:L]
5 Xbars = mean.(Xs)
6 histogram(Xbars; norm=true, alpha=0.5, label="", title="$dist, n=$n")
7 mu = mean(dist)
8 sigma = std(dist)
9 normal_approx = Normal(mu, sigma/sqrt(n))
10 plot!(normal_approx; label="normal approx", lw=2)

```

Out[39]: Gamma{Float64}($\alpha=2.0$, $\theta=3.0$), n=1000



4.2 Revise.jlを使うワークフロー

のように素朴に毎回コードを入力することは非常に面倒である。

似た仕事は関数化して1行の入力で実行できるようにしておく方がよい.

しかし, 関数の定義を `julia>` プロンプトに直接入力すると, 試行錯誤で関数の定義を何度も変える作業が非常に面倒になる.

もしも, 関数の定義をファイルに書いておき, ファイル内の関数の定義を書き換えると, 自動的に `julia>` プロンプトの側に関数の定義の変更が反映されるようにできれば非常に便利である. それを実現するのが [Revise.jl](https://github.com/timholly/Revise.jl) (<https://github.com/timholly/Revise.jl>) パッケージである. `Revise.jl` パッケージは

```
pkg> add Revise
```

でインストールできる.

問題: `using Revise` を実際に使ってみよ.

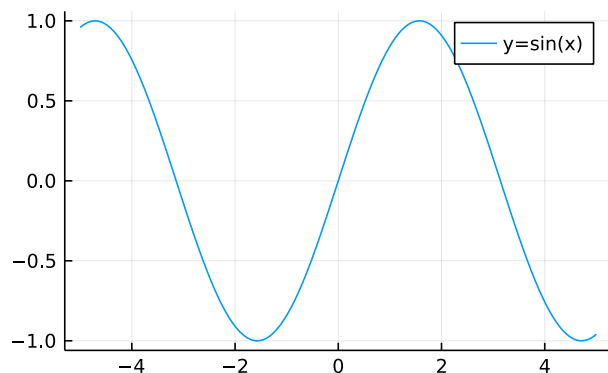
```
In [40]: 1 using StatsPlots, Distributions
2 default(size=(400, 250), titlefontsize=10)
3
4 function hello_sine()
5     println("Hello, Sine!")
6     plot(sin; label="y=sin(x)")
7 end
8
9 function plot_central_limit_theorem(dist, n; L=10^4, bin=:auto)
10     distname = mydistname(dist)
11     mu = mean(dist)
12     sigma = std(dist)
13     Xs = [rand(dist, n) for _ in 1:L]
14     Xbars = mean.(Xs)
15     normal_approx = Normal(mu, sigma/sqrt(n))
16
17     if dist isa DiscreteUnivariateDistribution
18         mu = mean(dist)
19         sigma = std(dist)
20         a = round(n*mu - 4.5sqrt(n)*sigma)
21         b = round(n*mu + 4.5sqrt(n)*sigma)
22         ran = a-0.5:b+0.5
23         bin = ran / n
24     end
25
26     histogram(Xbars; bin, norm=true, alpha=0.5, label="Xbars")
27     plot!(normal_approx; lw=2, label="normal approx")
28     title!("$distname, n=$n")
29 end
30
31 mypdf(dist, x) = pdf(dist, x)
32 mypdf(dist::DiscreteUnivariateDistribution, x) = pdf(dist, round{Int}(x))
33 mydistname(dist) = replace(string(dist), r"{{^}}*" => "")
34
35 function plot_dist(dist; xlim0=nothing)
36     distname = mydistname(dist)
37     if isnothing(xlim0)
38         mu = mean(dist)
39         sigma = std(dist)
40         a = max(minimum(dist), mu - 4.5sigma)
41         b = min(maximum(dist), mu + 4.5sigma)
42         if dist isa DiscreteUnivariateDistribution
43             a, b = a-1, b+1
44         else
45             a, b = a-0.025(b-a), b+0.025(b-a)
46         end
47         xlim0 = (a, b)
48     end
49     plot(x -> mypdf(dist, x), xlim0...; label="", title="$distname")
50 end
51
52 function plot_dist_clt(dist, n; L=10^4, xlim0=nothing)
53     P0 = plot_dist(dist; xlim0)
54     P1 = plot_central_limit_theorem(dist, n; L)
55     plot(P0, P1; size=(800, 250), layout=(1, 2))
56 end
```

Out[40]: `plot_dist_clt` (generic function with 1 method)


```
In [41]: 1 hello_sine()
```

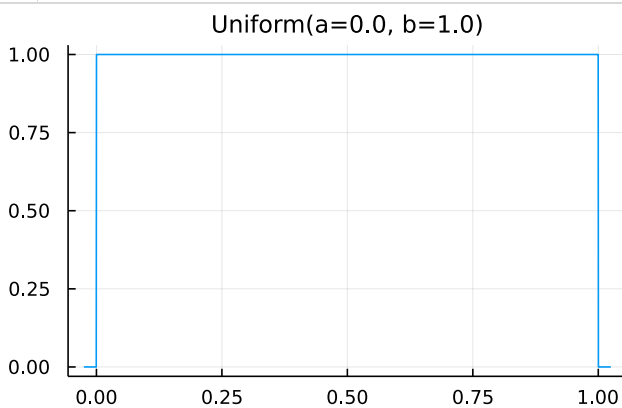
Hello, Sine!

Out[41]:

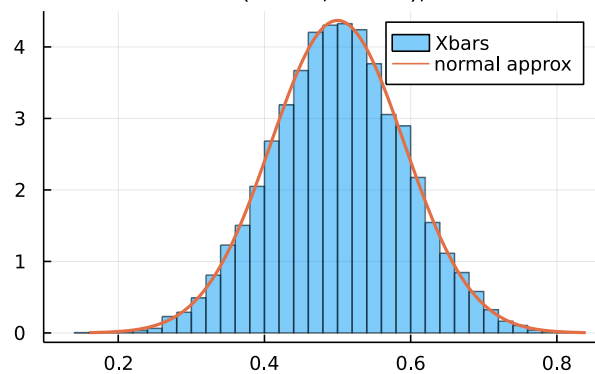


```
In [42]: 1 plot_dist_clt(Uniform(), 10)
```

Out[42]:

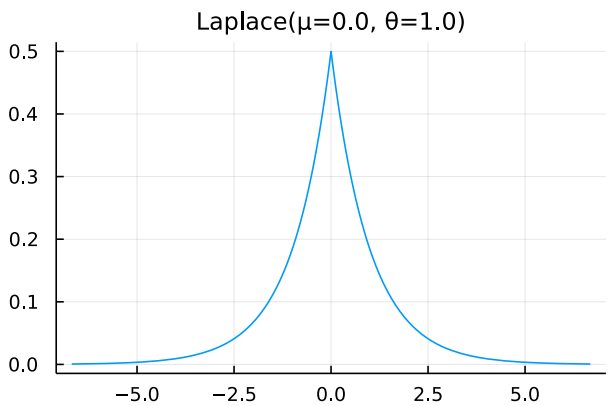


Uniform($a=0.0$, $b=1.0$), $n=10$

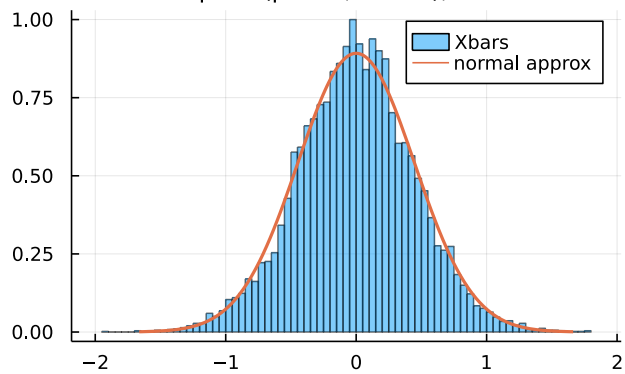


```
In [43]: 1 plot_dist_clt(Laplace(), 10)
```

Out[43]:

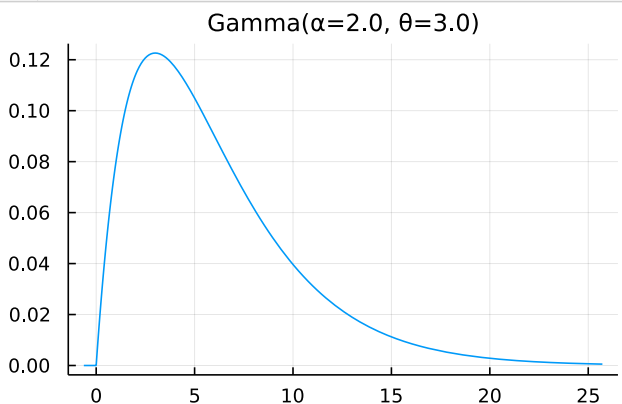


Laplace($\mu=0.0$, $\theta=1.0$), $n=10$

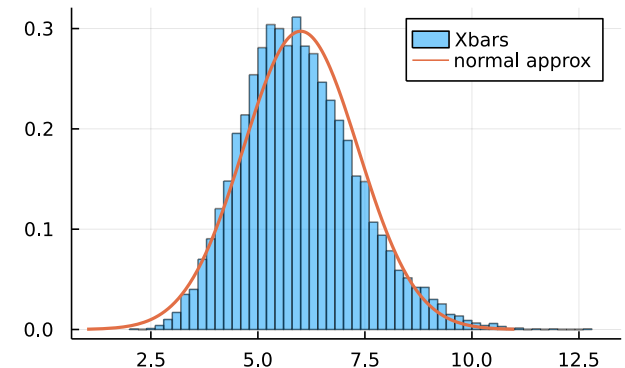


```
In [44]: 1 plot_dist_clt(Gamma(2, 3), 10)
```

Out[44]:

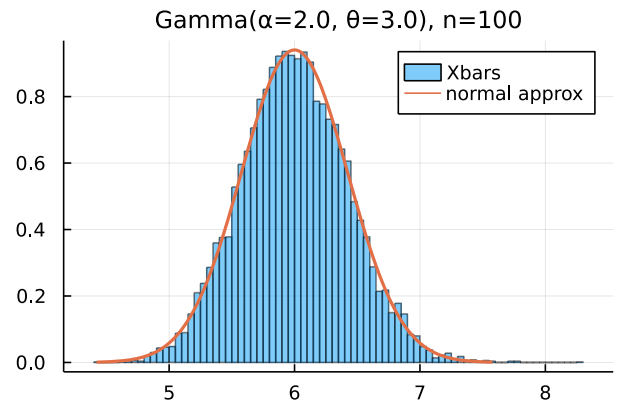
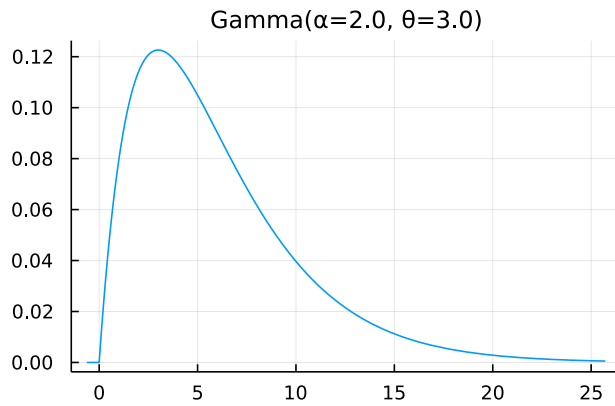


Gamma($\alpha=2.0$, $\theta=3.0$), $n=10$



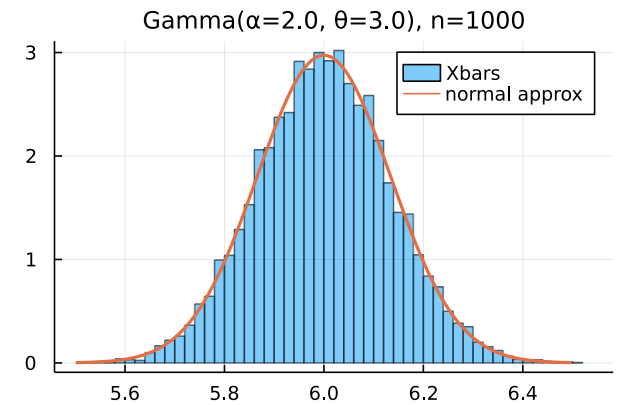
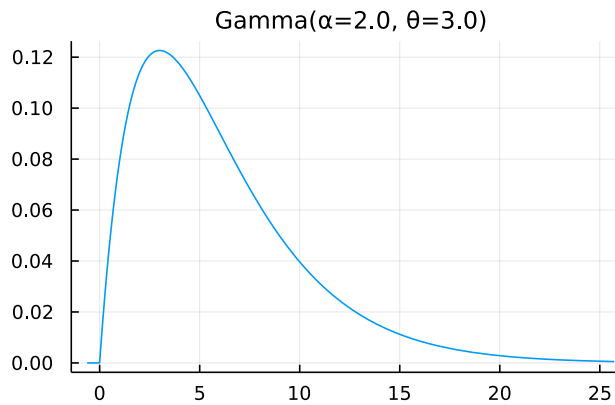
In [45]: 1 plot_dist_clt(Gamma(2, 3), 100)

Out[45]:



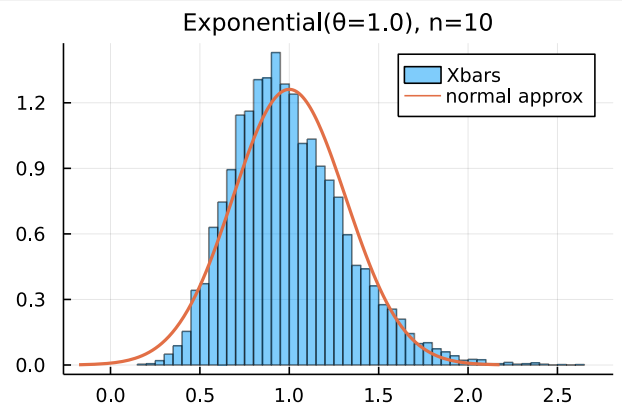
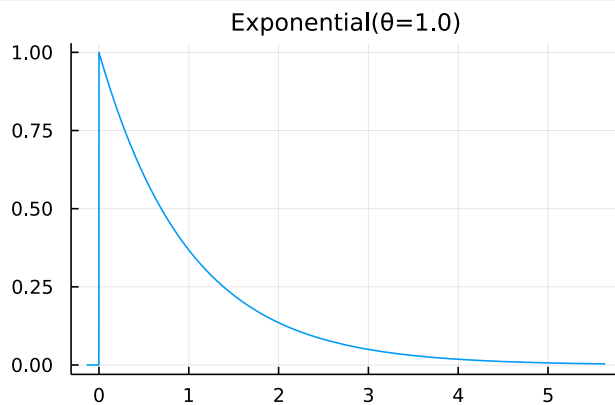
In [46]: 1 plot_dist_clt(Gamma(2, 3), 1000)

Out[46]:



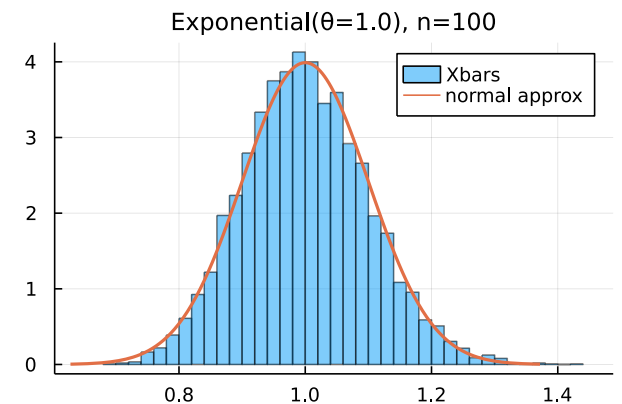
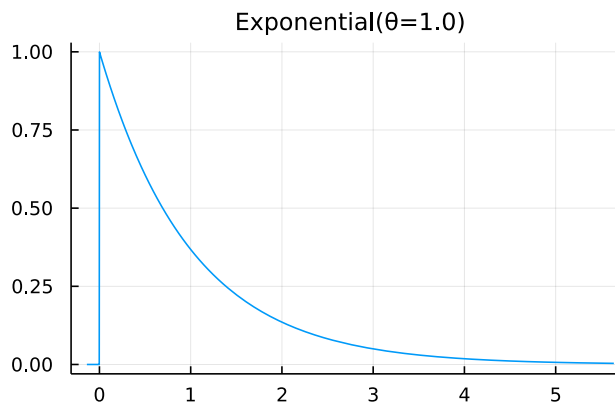
In [47]: 1 plot_dist_clt(Exponential(), 10)

Out[47]:



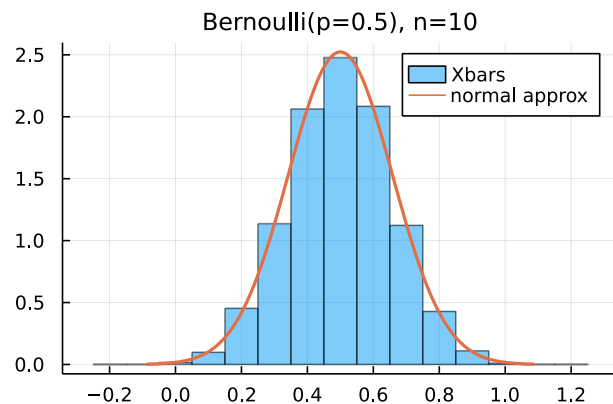
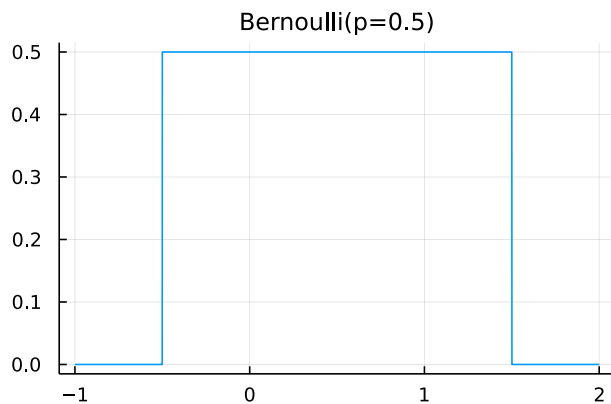
In [48]: 1 plot_dist_clt(Exponential(), 100)

Out[48]:



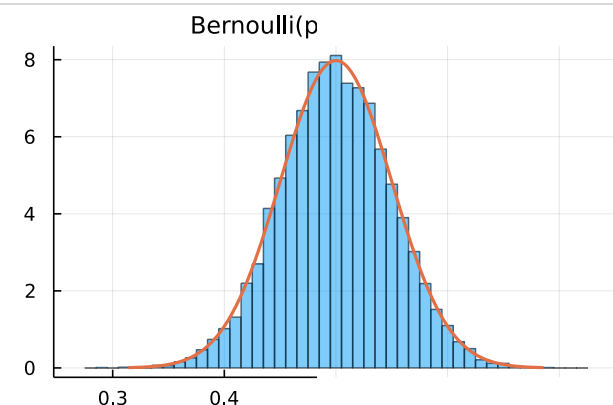
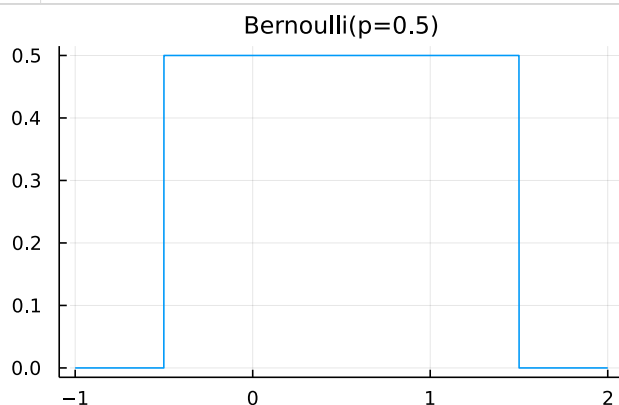
In [49]: 1 plot_dist_clt(Bernoulli(), 10)

Out[49]:



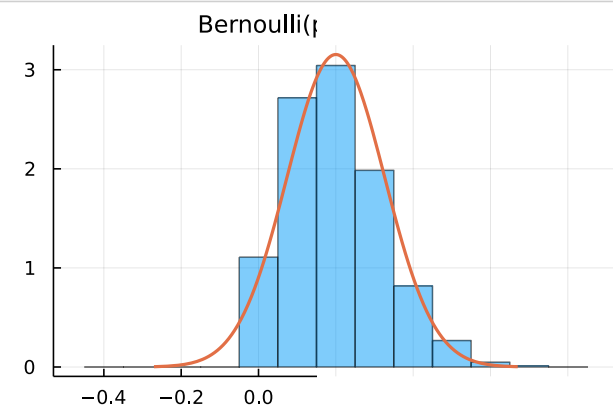
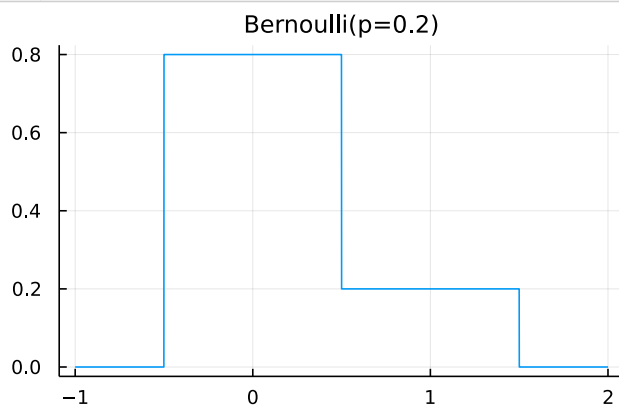
In [50]: 1 plot_dist_clt(Bernoulli(), 100)

Out[50]:



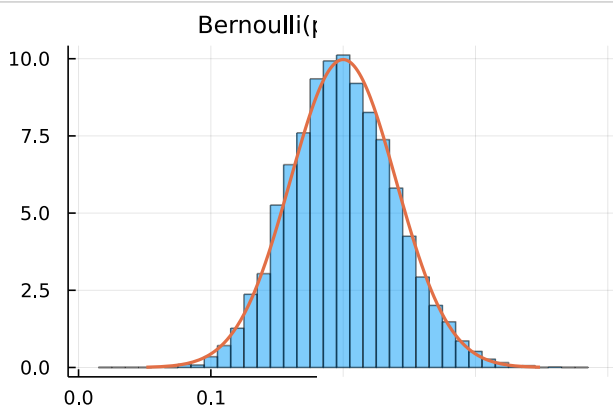
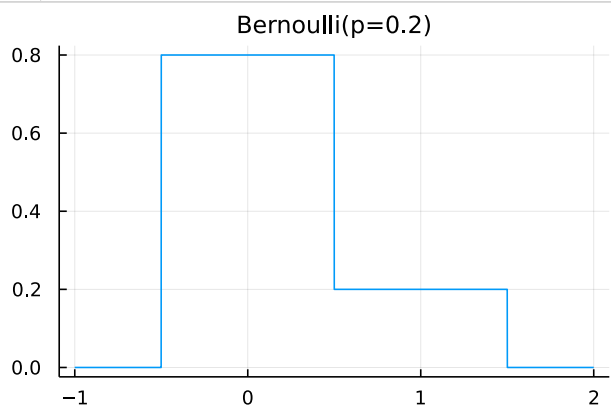
In [51]: 1 plot_dist_clt(Bernoulli(0.2), 10)

Out[51]:

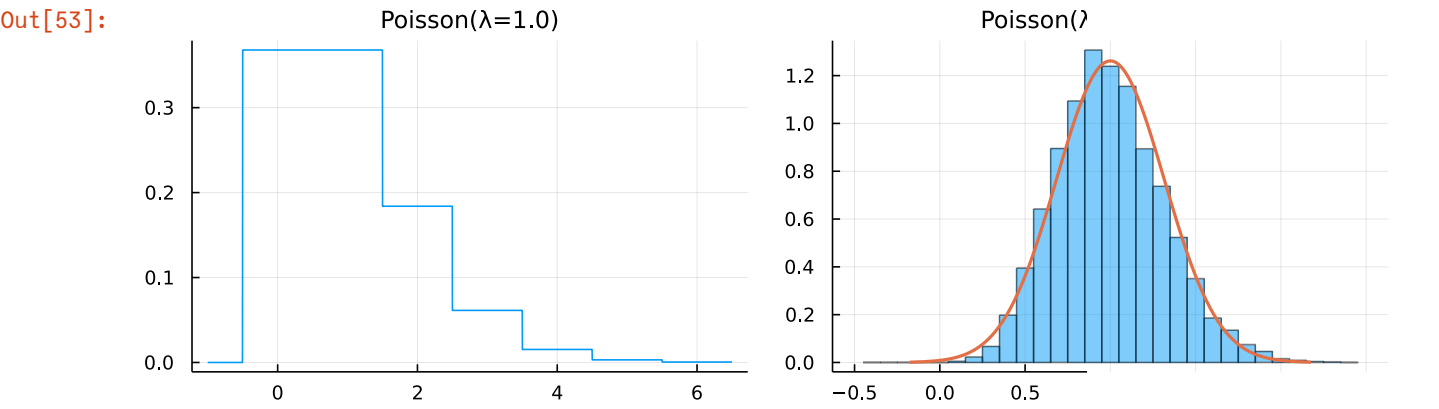


In [52]: 1 plot_dist_clt(Bernoulli(0.2), 100)

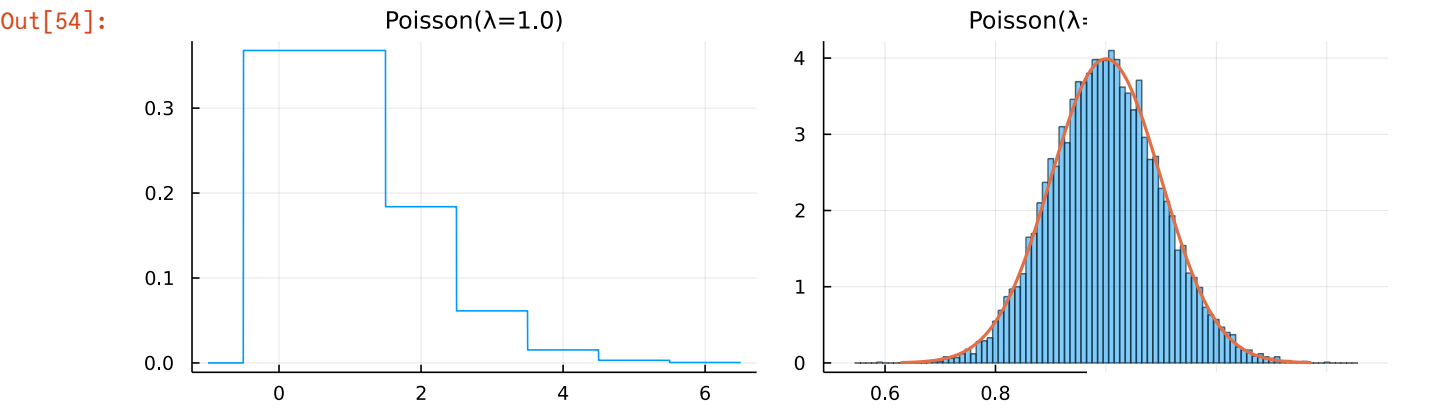
Out[52]:



```
In [53]: 1 plot_dist_clt(Poisson(), 10)
```



```
In [54]: 1 plot_dist_clt(Poisson(), 100)
```



```
In [ ]: 1
```