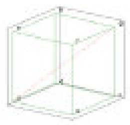


ch08 sec8.7 分布函数

```
md"""
# ch08  sec8.7 分布函数
"""
```

Table of Contents


```
begin
        using PlutoUI   , Plots   ,DataFrames   ,HypertextLiteral   ,LaTeXStrings
,Symbolics   ,StatsBase
        gr()
        theme(:bright)
        @htl("""<script
src="https://cdn.bootcdn.net/ajax/libs/mathjax/3.2.0/es5/tex-svg-full.min.js">
</script>
        """)
        PlutoUI.TableOfContents()

end

```

datacollection = Dict()

```
datacollection=Dict()
```

美国人口的年龄分布

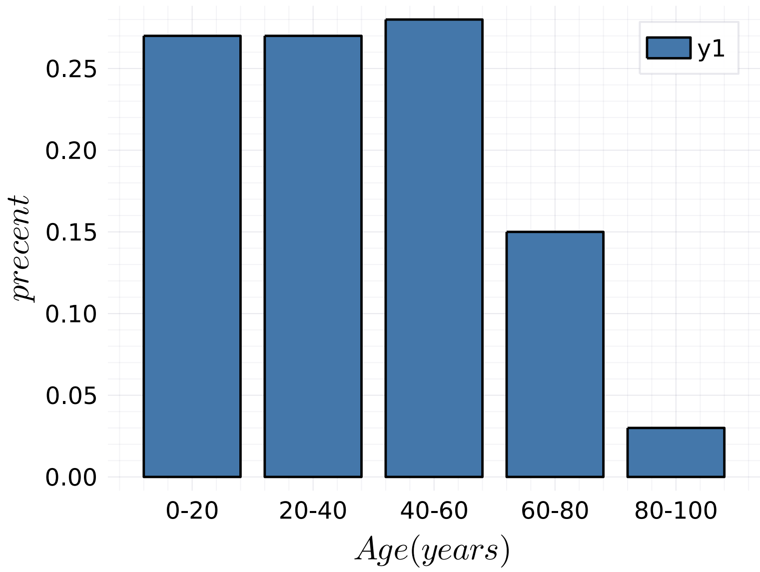
下图为 2012年的人口分布图.

每一个年龄组都是一个矩形, 所有矩形的面积和为1,现在看看单个年龄组$0-20$, 这个年龄组占总人口的27,每个年龄的人占总人口的比例为$\frac{27}{20}=0.0135$

```
md"""
## 美国人口的年龄分布

下图为 2012年的人口分布图.

每一个年龄组都是一个矩形, 所有矩形的面积和为$1$,现在看看单个年龄组$0-20$, 这个年龄组占总人口的$27%$,每个年龄的人占总人口的比例为$\frac{27%}{20}=0.0135$
"""
```

```
Dict("agedistribution1" ⇒ Dict("agegap" ⇒ 20, "agegroup" ⇒ 0:20:100, "percent" ⇒ [0.
```

```
let
    agegap=20
    agegroup=0:agegap:100
    getstr(t)="$(t)-$(t+agegap)"
    agegroupstr=[ getstr(t) for t in 0:agegap:100]
    precent=[0.27,0.27,0.28,0.15,0.03]
    data1=Dict(
       "agegap"=>20,
       "agegroup"=>agegroup,
        "agegroupstr"=>agegroupstr,
        "percent"=>precent,
    )

   merge!(datacollection,Dict("agedistribution1"=>data1))
end
```




```
let
    data=datacollection["agedistribution1"]
    xs,ys=data["agegroupstr"],data["percent"]
    plot(bar(xs,ys),label=false,xlabel=L"Age(years)",ylabel=L"precent",size=
    (400,300))
end
```

Example

example 1

- (a).估计一下$20-60$岁的人口的比例
- (b).十岁一下人口的比例
- (c). $75-80$岁人口的比例
- (d). $80-85$岁之间的人口比例

计算如下:

1. $20-60$是$20-40$,$40-60$ 两个gap 的数据和,$percent_{at}20-40+percent_{at}40-60$

2. 小于十岁,由于处于$0-20$区间中, 近似认为在$0-20$区间内分布是均匀的, 所以$0-10$岁一半

3. $75-80$岁之间, 占了$60-80$区间的$\frac{80-75}{80-60}$ 比例
4. $80-85$ 就算与 3 相同

```
md"""
!!! example

    example 1

    - (a).估计一下$20-60$岁的人口的比例
    - (b).十岁一下人口的比例
    - (c). $75-80$岁人口的比例
    - (d). $80-85$岁之间的人口比例

计算如下:
1.  $20-60$是$20-40$,$40-60$ 两个gap 的数据和,$percent_{at}20-40+percent_{at}40-60$


2. 小于十岁,由于处于$0-20$区间中, 近似认为在$0-20$区间内分布是均匀的, 所以$0-10$岁一半


3.  $75-80$岁之间, 占了$60-80$区间的$\frac{80-75}{80-60}$ 比例

4.  $80-85$ 就算与 3 相同
"""
```

0.0075

```julia
let
    data=datacollection["agedistribution1"]  #获取数据

    #1. 20-60 是 20-40,40-60两个gap 的数据和
    @show fraction20to40=data["percent"][2]+data["percent"][3]

    #2. 小于十岁,由于处于$0-20$区间中, 近似认为在$0-20$区间内分布是均匀的, 所以$0-10$岁一半
    @show fraction0to10=data["percent"][1]/2

    #3
     ratio1=(80-75)/data["agegap"]
     @show   fraction75to80=data["percent"][4]*ratio1

    #4 计算与 3相同
    ratio2=(85-80)/data["agegap"]
     @show   fraction80to85=data["percent"][5]*ratio2
end
```

```
fraction20to40 = (data["percent"])[2] + (data["percent"])[3] = 0.55
fraction0to10 = (data["percent"])[1] / 2 = 0.135
fraction75to80 = (data["percent"])[4] * ratio1 = 0.0375
fraction80to85 = (data["percent"])[5] * ratio2 = 0.0075
```

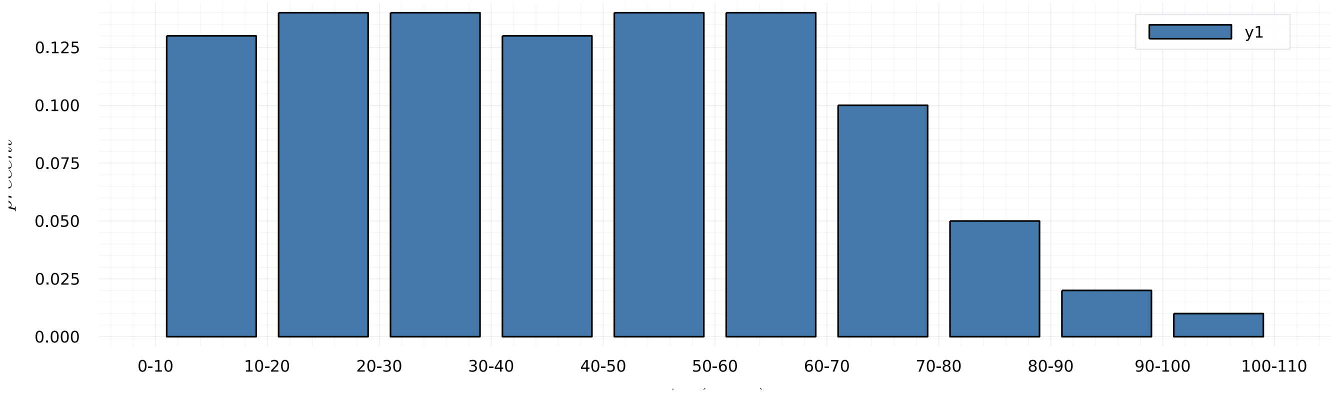
平滑柱形图

在上面的例子中, 20岁是一个比较大的跨度, 中间每个年龄的人数是不均匀的, 如果统计的区间变得更小一点, 估计的结果会更准确. 比如我们取10 岁为区间

```julia
md"""

### 平滑柱形图

在上面的例子中, $20$岁是一个比较大的跨度, 中间每个年龄的人数是不均匀的, 如果统计的区间变得更小一点, 估计的结果会更准确. 比如我们取$10岁为区间$
"""
```

```
Dict("agedistribution1" ⇒ Dict("agegap" ⇒ 10, "agegroup" ⇒ 0:10:100, "percent" ⇒ [0.
```

```julia
let
    agegap=10
    agegroup=0:agegap:100
    getstr(t)="$(t)-$(t+agegap)"
    agegroupstr=[ getstr(t) for t in 0:10:100]
    precent=[0.13,0.14,0.14,0.13,0.14,0.14,0.10,0.05,0.02,0.01]
    data2=Dict(
       "agegap"=>agegap,
       "agegroup"=>agegroup,
        "agegroupstr"=>agegroupstr,
        "percent"=>precent,
    )

   merge!(datacollection,Dict("agedistribution2"=>data2))
end
```




```
let
    data=datacollection["agedistribution2"]
    xs,ys=data["agegroupstr"],data["percent"]
    plot(bar(xs,ys),label=false,xlabel=L"Age(years)",ylabel=L"precent",size=
    (1000,300))
end
```

在人口分布上的变化是以时间为变量的函数, 和前面的积分计算一样, 要研究人口分布问题, 我们沿着年龄变化的方向把年龄段分割成小的区间(这里用agegap 表示), 当$\Deta t \to 0$时, 区间内的人口占比近似为同一个值. 如果定义人口占比为时间的函数 $p(t)$,那么在一个时间区间内的人口占比可以求定积分获得

$$fraction_{at}a-b= \int_{a}^{b}p(t)dt$$

如果是全部人口占比,就有:

$$fraction_{at}0-100= \int_{0}^{100}p(t)dt=1$$

在前面的积分应用中,都会定义一个函数来表示沿着变量表示的变化, 那么 $p(t)$ 是如何定义的呢?

这就是下面要谈到的问题.

```
md"""

在人口分布上的变化是以时间为变量的函数, 和前面的积分计算一样, 要研究人口分布问题, 我们沿着年龄变化的方向把年龄段分割成小的区间(这里用agegap 表示), 当$\Deta t \to 0$ 时, 区间内的人口占比近似为同一个值. 如果定义人口占比为时间的函数 $p(t)$,那么在一个时间区间内的人口占比可以求定积分获得

$fraction_{at}a-b= \int_{a}^{b}p(t)dt$

如果是全部人口占比,就有:

$fraction_{at}0-100= \int_{0}^{100}p(t)dt=1$


在前面的积分应用中,都会定义一个函数来表示沿着变量表示的变化, 那么 $p(t)$ 是如何定义的呢?

这就是下面要谈到的问题.

"""
```

概率密度函数

在前面积分物理应用里, 我们首先讲到了一个跷跷板质心的问题, 如果跷跷板内部各处的密度不均一, 那么质心的位置可能会变化.

在这里人口问题中, 我们也可以把人口 $0-100$的区间看做是一根长杆, 长杆内部的密度各处不一,解决问题, 我们首先要通过测量获得一个密度和长度之间的函数关系.对于人口问题是类似的, 我们需要一个人口比例和年龄关系的函数. 然后按照长杆一样的处理进行积分.

人口分布的特殊点在于, 所有年龄的人口占比中和为 1, 所以在整个区间内积分的值为 1:

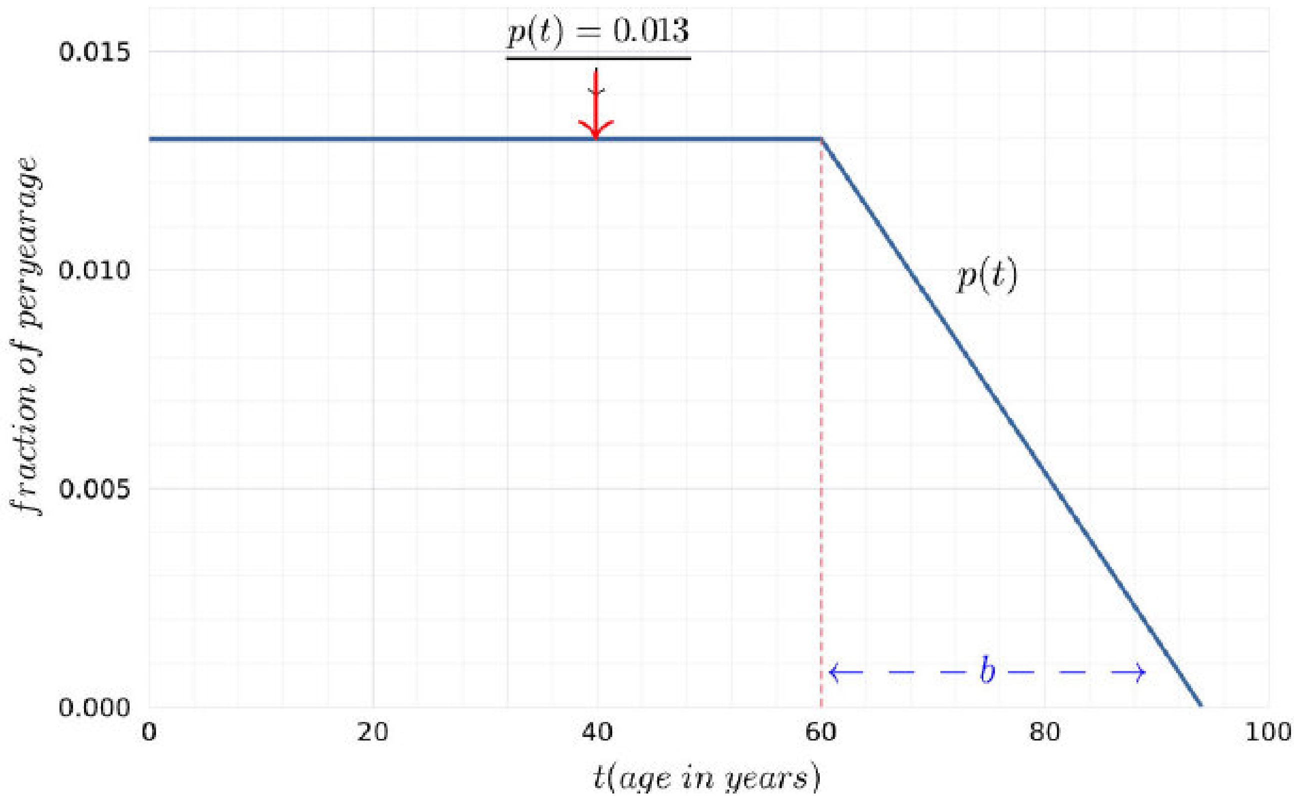
$$fraction_{at}0-100= \int_{0}^{100}p(t)dt=1$$

```
md"""
## 概率密度函数

在前面积分物理应用里，我们首先讲到了一个跷跷板质心的问题，如果跷跷板内部各处的密度不均一，那么质心的位置可能会变化.

在这里人口问题中，我们也可以把人口 $0-100$的区间看做是一根长杆，长杆内部的密度各处不一,解决问题，我们首先要通过测量获得一个密度和长度之间的函数关系.对于人口问题是类似的，我们需要一个人口比例和年龄关系的函数.  然后按照长杆一样的处理进行积分.

人口分布的特殊点在于，所有年龄的人口占比中和为 1，所以在整个区间内积分的值为 1:

$fraction_{at}0-100= \int_{0}^{100}p(t)dt=1$
"""
```

Example

example 2

根据美国人口分布构建一个连续函数,反映人口占比随年龄的变化问题,为简化问题,
$0-60$岁之间的密度函数为:$p(t)=0.013, t\in(0,60]$



$0-60$之间的人口占比变化率与时间的函数为$p(t)=0.013$,所以 $0-60$之间的人口所占比例就是图中矩形的面积:

$$\int_0^{60} p(t)dt = 60(0.013)$$

从 $60-$ 接近 100 岁时, 所占比例是图中的三角形的面积:

$$\int_{60}^{100} p(t)dt = \frac{1}{2}(0.013)b$$

两个合计为总人口数量 1:

$$\frac{1}{2}(0.013)b + 60(0.013) = 1$$

由此求出$b=33.85$ 含义是 93.85 岁以上的人口占比就可以忽略不计了

在三角形内计算出从$60\rightarrow 93.85$区间的变化, 即图中斜线的斜率,可得$m=-0.00038$

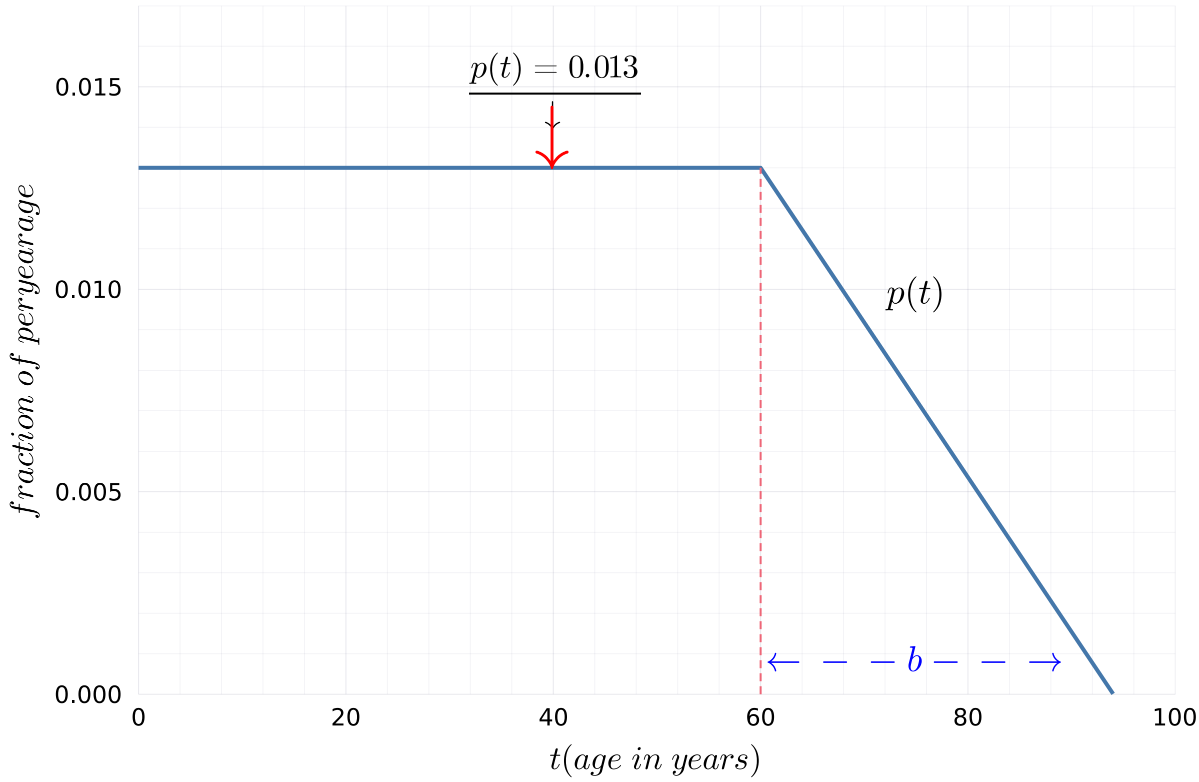
所以在$60\rightarrow 93.85$区间内变化的仿射直线方程为:

$$p(t)=0.0358-0.00038t\ ,t\in(60,93.85]$$

所以整个人口的分布密度函数就是:

$$\left\{\begin{matrix} p(t)=0.013\ ,t\in(0,60]\\ p(t)=0.0358-0.00038t\ ,t\in(60,93.85] \end{matrix}\right.$$

```
md"""

!!! example
    example 2

    根据美国人口分布构建一个连续函数,反映人口占比随年龄的变化问题,为简化问题, $0-60$岁之间的密度函数为:$p(t)=0.013, t \in (0,60]$

    ![](https://tva1.sinaimg.cn/mw690/e6c9d24egy1h37zzdiw9oj20xg0k8wg4.jpg)

 $0-60$之间的人口占比变化率与时间的函数为$p(t)=0.013$,所以 $0-60$之间的人口所占比例就是图中矩形的面积:

$\int_{0}^{60}p(t) dt=60(0.013)$

从 $60- 接近 100 岁$时, 所占比例是图中的三角形的面积:

$\int_{60}^{100}p(t)dt=\frac{1}{2}(0.013)b$

两个合计为总人口数量 $1$:

$\frac{1}{2}(0.013)b+60(0.013)=1$

由此求出$b=33.85 \ 含义是 93.85 岁以上的人口占比就可以忽略不计了$

在三角形内计算出从$60 \to 93.85 $区间的变化, 即图中斜线的斜率,可得$m=-0.00038$

所以在$60 \to 93.85$区间内变化的仿射直线方程为:

$p(t)=0.0358-0.00038t\ , t \in (60,93.85]$


所以整个人口的分布密度函数就是:



$\left\{\begin{matrix}
p(t)=0.013  \ , t \in (0,60]\\
p(t)=0.0358-0.00038t \ , t \in (60,93.85]
\end{matrix}\right.$

"""
```




```
let
    ann=[
        (40,0.015,text(L" \frac{p(t)=0.013}{\downarrow}",pointsize=11)),
        (40,0.014,text(L"\downarrow",pointsize=24,color=:red)),
        (75,0.010,text(L"p(t)",pointsize=12,color=:black)),
         (75,0.001,text(L"\leftarrow-- b-- \rightarrow",pointsize=12,color=:blue))
    ]
    plot([0,60,94],[0.013,0.013,0.00001],label=false,lw=2,xlims=(0,100),
        ylims=(0,0.017),xlabel=L"t(age \ in \ years)",ylabel=L"fraction \ of\  per
      year age",ann=ann)
    plot!([60,60],[0,0.013],lw=1,ls=:dash, colo=:green,label=false)
end
```

```
let
@htl("""

<script>hljs.highlightAll();</script>
<script src="https://cdn.bootcdn.net/ajax/libs/mathjax/3.2.0/es5/tex-svg-
full.min.js">
<script src="http://127.0.0.1:8080/tex-svg-full.min.js"></script>
""")
end
```