Unit67_HW

2020 年 6 月 2 日

1 Unit6-7 HW

```python
[2]:        import numpy as np
            import matplotlib.pyplot as plt
            import seaborn as sns
            import astropy
            sns.set_style('darkgrid')
            import scipy.stats as stats
            from IPython.core.interactiveshell import InteractiveShell
            InteractiveShell.ast_node_interactivity = 'all'
```

1.1 HW-U6-1

(a) Compute the sample mean direction

```python
[8]:        theta=np.array([0,45,90,135,180,225])*astropy.units.deg
            freq=np.array([2,4,3,1,1,1])
            theta
            freq
            astropy.stats.circmean(theta,weights=freq)
```

[8]: $[0, \ 45, \ 90, \ 135, \ 180, \ 225]\,^{\circ}$

[8]:   array([2, 4, 3, 1, 1, 1])

[8]: $67.5\,^{\circ}$

方向的均值是 67.5°

(b) Compute the angular deviation for the data

[15]:
```
n=sum(freq)
Y=sum(freq*np.sin(theta))/n
X=sum(freq*np.cos(theta))/n
r=np.sqrt(X**2+Y**2)
r
1-r
2*(1-r)
np.log(1-r)
```

[15]:
0.52572034

[15]:
0.47427966

[15]:
0.94855932

[15]:
−0.74595813

(c) Determine 95% confidence limits for the population mean

[19]:
```
def meand_ci(r,n,alpha):
R=r*n
chi2=stats.chi2.isf(alpha, 1, loc=0, scale=1)
if r>=0.9:
return np.arccos(np.sqrt(n**2-(n**2-R**2)*np.e**(chi2/n))/R)
else:
return np.arccos(np.sqrt(2*n*(2*R**2-n*chi2)/(4*n-chi2))/R)
```

[31]:
```
data_mean=astropy.stats.circmean(theta,weights=freq)
ci=(data_mean-meand_ci(r,n,0.05),data_mean+meand_ci(r,n,0.05))
ci
```

[31]:
12

[31]:
(<Quantity 20.06103815 deg>, <Quantity 114.93896185 deg>)

均值的 95% 置信区间是 (20.06°,114.94°)

(d) Determine the sample median direction

[32]:
```python
def sample_median(theta,freq):
    n=sum(freq)
    tmp=np.cumsum(freq)
    if n%2:
        for i in range(len(tmp)):
            if tmp[i]>=(n-1)/2:
                break
        return theta[i]
    else :
        for i in range(len(tmp)):
            if tmp[i]>=n/2-1:
                break
        left=i
        for i in range(len(tmp)):
            if tmp[i]>=n/2+1:
                break
        right=i
        return (theta[left]+theta[right])/2
```

[33]:
```python
sample_median(theta,freq)
```
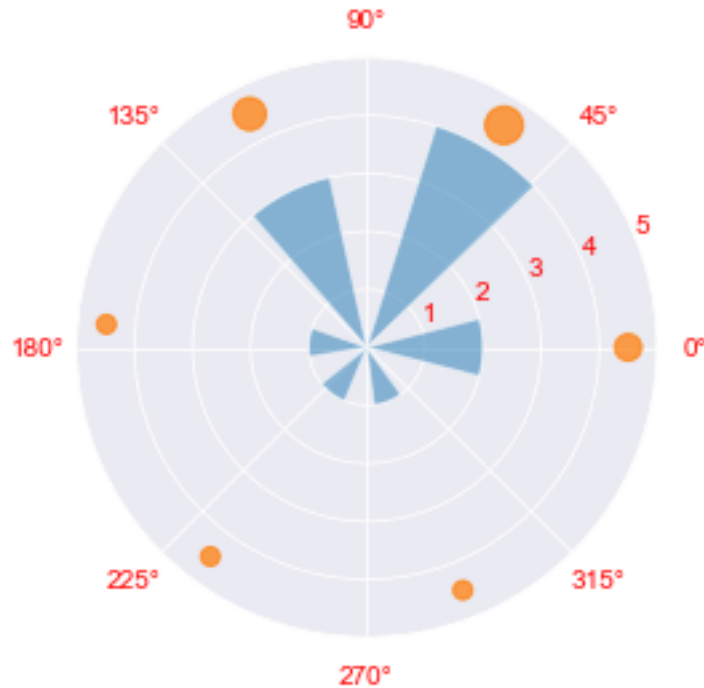
[33]:
$67.5\,°$

1.2 HW-U6-2

Visualize the data in Problem HW-6-1 in polar coodinates

[65]:
```python
width=0.5
data=np.array([0,45,90,135,180,225])
ax = plt.subplot(111, projection='polar')
ax.bar(data, freq, width=width, bottom=0.0, alpha=0.5)
ax.scatter(data,4.5*np.ones(len(theta)),marker='o', s=50*freq,alpha=0.
↪75)
ax.tick_params(colors='red')
ax.set_rlim(0,5)
plt.show()
```

[65]:
```
<BarContainer object of 6 artists>
```

[65]:　　　　<matplotlib.collections.PathCollection at 0x1aa4609a9c8>

[65]:　　　　(0, 5)

1.3　HW-U7-1

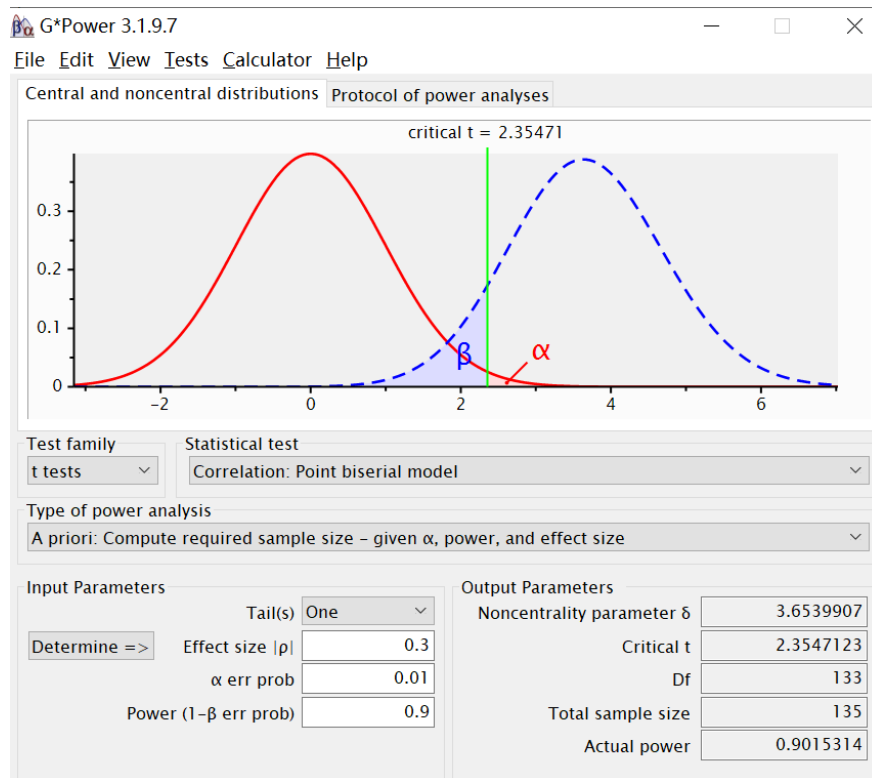如果在某一个地区 covid-19 的感染率在 10%，从该地区人口中随机取样 N 个样本进行实验研究，要使得样本感染率误差在 5% 以内（95% confidence level），N 至少应该要采样多少？

[66]:
```python
def qualitative_n(z,p,d):
    return z**2*p*(1-p)/d**2
```

[67]:
```python
qualitative_n(2,0.1,0.05)
```

[67]:　　　　144.0

N 至少要采样 144。

1.4 HW-U7-2

如果空腹血糖平均水平在 90 mg/dl，初步实验表明某种饮料在半小时内会提高血糖 3 mg/dl；假设血糖的标准差在 10 mg/dl；研究同一批人喝饮料前后的血糖水平增加量，采用 paired -t 检验，要求 NHST 显著水平达到 99%，统计 Power 达到 0.9 ，估算需要多少大的样本量？ 根据 G-Power 的计算结果可知，需要样本量为 135

[ ]: