Statistics Homework 05

84

B08705034 資管二 施芊羽

Chapter 6

In [1]:

```
import seaborn as sns
import pandas as pd
import numpy as np
import scipy.stats as stats
import statsmodels.api as sm
import statsmodels.stats.api as sms
import statsmodels.formula.api as smf
import math as math
import statistics
```

6.3

a.

$S = \{ A, B, C, D, E\}$ while A, B, C, D, E represent for the five differennt choices in the question.

b.

$\left\{ \begin{array} {}P(A) = .20 \\ P(B) = .20 \\ P(C) = .20 \\ P(D) = .20 \\ P(E) = .20 \end{array} \right.$

c.

Classical Approach is the approach I used to answer part(b).

d.

From the answer of part(b), we can know that the propability to have the right answer while taking guess on the different choices will be equal. Since $P(A) = P(B) = P(C) = P(D) = P(E) = .20$, we can conclude that the student can get the same probability to answer a question correctly with whatever choice he guess.

6.17

In [5]:

```
df_c06_17 = pd.read_excel("Xr06-17.xlsx")

df_c06_17["Probability"] = np.round(df_c06_17["Frequency"] / df_c06_17["Frequency"][7], 3)

display(df_c06_17)
```

	Language	Frequency	Probability
0	Spanish	38.4	0.619
1	Chinese	3.0	0.048
2	Tagalog	1.6	0.026
3	Vietnamese	1.4	0.023
4	French	1.3	0.021

	Language	Frequency	Probability
5	Korean	1.1	0.018
6	Other	15.2	0.245
7	Total	62.0	1.000

(*All the answer to the following parts are rounded to the three decimal space*).

a.

The probability of the individul chosen speaks Spanish is $.619$.

b.

The probability of the individul chosen speaks a language other than Spanish is $1 - .619 = .381$.

c.

The probability of the individul picked speaks Vietnamese or French is $.023 + .021 = .044$

d.

The probability of the individul chosen speaks one of the other languages is $.245$.

6.31

In [23]:

```
df_c06_31 = pd.read_excel("Xr06-31.xlsx")
display(df_c06_31)
```

	Unnamed: 0	Promoted	Not promoted
0	Female	0.03	0.12
1	Male	0.17	0.68

a.

In [19]:

```
print("The rate of promotion among female assistant professors is",
      np.round(df_c06_31["Promoted"][0]/(df_c06_31["Promoted"][0] + df_c06_31["Not promoted"][0]),
2))
```

```
The rate of promotion among female assistant professors is 0.2
```

The rate of promotion among female assistant professors is $\frac{.03}{.03 + .12} = 0.2$.

b.

In [22]:

```
print("The rate of promotion among male assistant professors is",
      np.round(df_c06_31["Promoted"][1]/(df_c06_31["Promoted"][1] + df_c06_31["Not promoted"][1]),
2))
```

```
The rate of promotion among male assistant professors is 0.2
```

The rate of promotion among male assistant professors is $\frac{.17}{.17 + .68} = 0.2$

c.

No, the promotion rate of male assistant professors and that of female assistant professors are both 0.2, which are equally the same.

6.39

In [31]:

```
df_c06_39 = pd.read_excel("Xr06-39.xlsx")
display(df_c06_39)
```

	Unnamed: 0	20-24	25-54	55-64	65 & older
0	Plant or company closed or moved	0.015	0.320	0.089	0.029
1	Insufficient work	0.014	0.180	0.034	0.011
2	Position or shift abolished	0.006	0.214	0.071	0.016

(*All the answer to the following parts are rounded to the three decimal space*).

a.

In [31]:

```
print("The probability that a 25- to 54-year-old was laid off or fired because of insufficient wor
k is"
      , np.round(df_c06_39["25-54"][1] / (df_c06_39["25-54"].sum()), 3))
```

```
The probability that a 25- to 54-year-old was laid off or fired because of insufficient work is 0.
252
```

The probability that a 25- to 54-year-old was laid off or fired because of insufficient work is $\frac{.180}{.320 + .180 + .214} = \frac{.180}{.714} \simeq 0.252$.

b.

In [32]:

```
print("The proportion of laid off or fired worker is age 65 and older is"
      , df_c06_39["65 & older"].sum())
```

```
The proportion of laid off or fired worker is age 65 and older is 0.056
```

The proportion of laid off or fired worker is age 65 and older is $.029 + .011 + .016 = .056$

c.

In [42]:

```
print("The probability that a laid-off or fired worker because the plant or company closed is 65 o
r older is"
      , np.round(df_c06_39["65 & older"][0] /
                 (df_c06_39["20-24"][0] + df_c06_39["25-54"][0] + df_c06_39["55-64"][0] + df_c06_39
"65 & older"][0]), 3))
```

```
The probability that a laid-off or fired worker because the plant or company closed is 65 or older
```

```
is 0.064
```

The probability that a laid-off or fired worker because the plant or company closed is 65 or older is $ \frac{.029}{.015 + .320 + .089 + .029} \simeq 0.064$

6.55

In [43]:

```
df_c06_55 = pd.read_excel("Xr06-55.xlsx")
display(df_c06_55)
```

	NBC News	Consistent liberal	Mostly liberal	Mixed	Mostly conservative	Consistent conservative
0	Trust	0.0896	0.1386	0.1944	0.0629	0.0144
1	Distrust	0.0096	0.0154	0.0540	0.0595	0.0558
2	Neither trust nor distrust	0.0576	0.0506	0.0864	0.0391	0.0153
3	Don't know	0.0032	0.0154	0.0252	0.0085	0.0045

a.

In [45]:

```
print("The probability that one respondent selected at random would trust NBC news is",
      df_c06_55["Consistent liberal"][0] + df_c06_55["Mostly liberal"][0] + df_c06_55["Mixed"][0] +

      df_c06_55["Mostly conservative"][0] +  df_c06_55["Consistent conservative"][0] )
```

```
The probability that one respondent selected at random would trust NBC news is 0.4999
```

The probability that one respondent selected at random would trust NBC news is $0.0896 + 0.1386 + 0.1944 + 0.0629 + 0.0144 = 0.4999$

b.

In [48]:

```
print("The probability that a consistent Conservative distruct NBC news is",
       np.round(df_c06_55["Consistent conservative"][1]/df_c06_55["Consistent conservative"].sum(),
4) )
```

```
The probability that a consistent Conservative distruct NBC news is 0.62
```

The probability that a consistent Conservative distruct NBC news is $\frac{0.0558}{0.0144 + 0.0558 + 0.0153 + 0.0045} = \frac{0.0558}{0.09} = 0.62$.

c.

In [50]:

```
print("The probability that a consistent Liberal neither trust nor distruct NBC news is",
       np.round(df_c06_55["Consistent liberal"][2]/df_c06_55["Consistent liberal"].sum(), 4) )
```

```
The probability that a consistent Liberal neither trust nor distruct NBC news is 0.36
```

The probability that a consistent Liberal neither trust nor distruct NBC news is $\frac{0.0576}{0.0896 + 0.0096 + 0.0576 + 0.0032} = \frac{0.0576}{0.16} = 0.36$.

d.

In [51]:

```
print("The probability that randomly chosen person is a consistent Liberal",
       df_c06_55["Consistent liberal"].sum())
```

```
The probability that randomly chosen person is a consistent Liberal 0.16
```

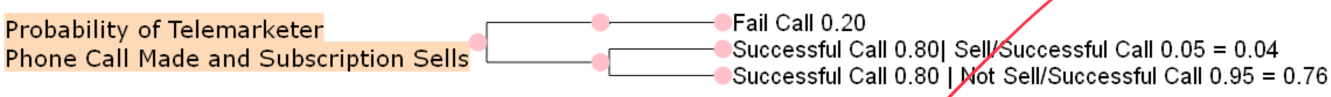
The probability that randomly chosen person is a consistent Liberal $0.0896 + 0.0096 + 0.0576 + 0.0032 = 0.16$

6.69 10

In [2]:

```
from ete3 import Tree, faces, TreeStyle, TextFace, NodeStyle
t = Tree(
    "((Fail Call 0.20), (Successful Call 0.80| Sell/Successful Call 0.05 = 0.04, Successful Call 0
.80 | Not Sell/Successful Call 0.95 = 0.76));" )
ts = TreeStyle()

Name1 = TextFace("Probability of Telemarketer")
Name2 = TextFace("Phone Call Made and Subscription Sells")
Name1.background.color = "peachpuff"
Name2.background.color = "peachpuff"
t.add_face(Name1, column=0, position = "branch-top")
t.add_face(Name2, column=0, position = "branch-bottom")

style = NodeStyle()
style["fgcolor"] = "#0f0f0e"
for n in t.traverse():
    nstyle = NodeStyle()
    nstyle["fgcolor"] = "pink"
    nstyle["size"] = 15
    n.set_style(nstyle)

ts.show_leaf_name = True
ts.show_scale = False

t.render("%%inline", tree_style =ts)
```

Out[2]:



The proportion that the seller successfully makes a sell is $0.8 \times 0.05 = 0.04$.

6.75 10

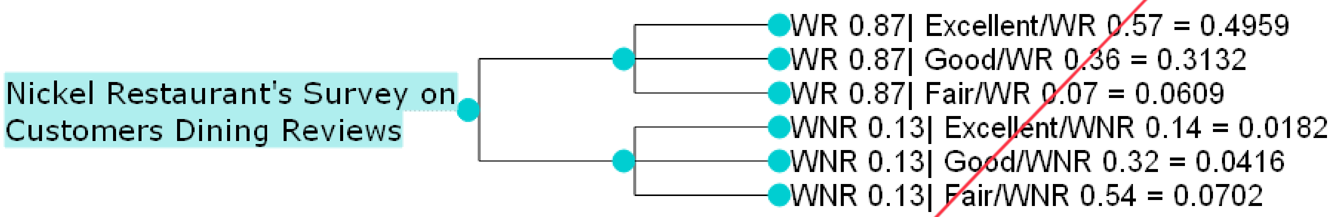
In [20]:

```
from ete3 import Tree, faces, TreeStyle, TextFace, NodeStyle
t = Tree("((WR 0.87| Excellent/WR 0.57 = 0.4959, WR 0.87| Good/WR 0.36 = 0.3132, WR 0.87| Fair/WR
0.07 = 0.0609), (WNR 0.13| Excellent/WNR 0.14 = 0.0182, WNR 0.13| Good/WNR 0.32 = 0.0416, WNR 0.13
| Fair/WNR 0.54 = 0.0702));")

Name1 = TextFace("Nickel Restaurant's Survey on")
Name2 = TextFace("Customers Dining Reviews")
Name1.background.color = "paleturquoise"
Name2.background.color = "paleturquoise"
t.add_face(Name1, column=0, position = "branch-top")
t.add_face(Name2, column=0, position = "branch-bottom")

for n in t.traverse():
    nstyle = NodeStyle()
    nstyle["fgcolor"] = "darkturquoise"
```

```
    nstyle["size"] = 15
    n.set_style(nstyle)

ts.show_leaf_name = True
ts.show_scale = False

print("WR = Customer will return ; WNR = Customer will not return")
t.render("%%inline", tree_style =ts)
```

```
WR = Customer will return ; WNR = Customer will not return
```

Out[20]:



The proportion of customers rate the restaurant as good is $0.87 \times 0.36 + (1 - 0.87) \times 0.32 = 0.3132 + 0.0416 = 0.3548$

6.83

10

In [19]:

```
df_c06_83 = pd.read_excel("Xr06-83.xlsx")

df_c06_83["P(Good thing)"] = 1 - df_c06_83["P(Bad thing)"]
df_c06_83["Proportion"] = np.round(df_c06_83["Number of respondents"] /  df_c06_83["Number of respondents"].sum(), 6)

display(df_c06_83)

print("The probability that one randomly selected would say the U.K. leaving EU is a bad thing is ",
      np.round((df_c06_83["P(Bad thing)"] * df_c06_83["Proportion"]).sum(), 4))
```

	Country	Number of respondents	P(Bad thing)	P(Good thing)	Proportion
0	France	630	0.62	0.38	0.370588
1	Germany	590	0.74	0.26	0.347059
2	Italy	480	0.57	0.43	0.282353

```
The probability that one randomly selected would say the U.K. leaving EU is a bad thing is  0.6475
```

The proportion of people from different countries is rounded to the six decimal space; The answer is rounded to the four decimal space.

The probability that one randomly selected would say the U.K. leaving EU is a bad thing is $0.62 \times 0.370588 + 0.74 \times 0.347059 + 0.57 \times 0.282353 = 0.22976456 + 0.25682366 + 0.16094121 = 0.64752869 \simeq 0.6475$.

6.91

10

$P(BAC \text{ is greater than } 0.09) = 12 \% = P(BAC \text{ is greater than } 0.09 | \text{Crash with fatality}) + P(BAC \text{ is greater than } 0.09 | \text{Crash without fatality}) \\= 0.084 + P(BAC \text{ is greater than } 0.09 | \text{Crash without fatality}) \\ \Rightarrow P(BAC \text{ is greater than } 0.09 | \text{Crash without fatality}) = 0.12 - 0.084 = 0.036.$

The probability of a crash with at least one fatality if a driver drives while legally intoxicated()$BAC \gt 0.09$ is $$\frac{P(BAC \text{ is greater than } 0.09 | \text{Crash with fatality})}{P(BAC \text{ is greater than } 0.09 | \text{Crash with fatality}) + P(BAC \text{ is greater than } 0.09 | \text{Crash without fatality})} = \frac{0.084}{0.12} = 0.7$$

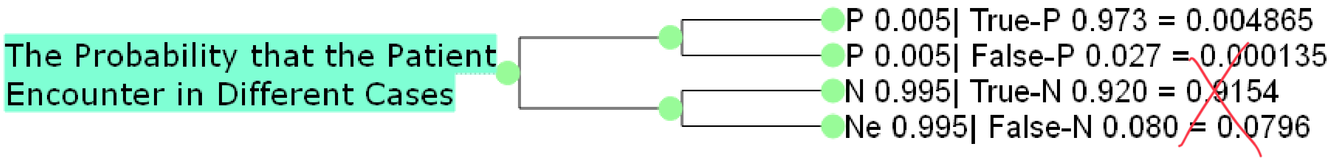
6.99 -10

In [18]:

```
from ete3 import Tree, faces, TreeStyle, TextFace, NodeStyle
t = Tree("((P 0.005| True-P 0.973 = 0.004865, P 0.005| False-P 0.027 = 0.000135), (N 0.995| True-N 0.920 = 0.9154, Ne 0.995| False-N 0.080 = 0.0796));")

Name1 = TextFace("The Probability that the Patient")
Name2 = TextFace("Encounter in Different Cases")
Name1.background.color = "aquamarine"
Name2.background.color = "aquamarine"
t.add_face(Name1, column=0, position = "branch-top")
t.add_face(Name2, column=0, position = "branch-bottom")

for n in t.traverse():
    nstyle = NodeStyle()
    nstyle["fgcolor"] = "palegreen"
    nstyle["size"] = 15
    n.set_style(nstyle)

ts.show_leaf_name = True
ts.show_scale = False

print("P = Positive ; N = Negative")
t.render("%%inline", tree_style =ts)
```

```
P = Positive ; N = Negative
```

Out[18]:



The probability that the patient actually has HIV is $0.005 \times 0.973 + 0.995 \times 0.08 = 0.004865 + 0.0796 = 0.084465. $