

(一社) 日本官能評価学会ワークショップ

ベイズ統計学と官能評価

Q and A

(日本女子大学)

岡本 安晴

A: 識別確率の年齢変化の事例で、例えば中間にピークがあるようなデータで経験年数と老化の両方の要素が考えられるような事例の場合、どのようなモデルを設定すればよいのでしょうか。最大値がある式として年齢の一次式に変えて二次式を置くような考えでよいのでしょうか。

B: 3点識別法でベイズ分析を行う場合、得られた事後確率の信頼範囲がに含まれなければ有意といえるのでしょうか？実際には危険率も求めるようにしたほうが良いのではないのでしょうか？それともd'を求めて判断したほうが良いのでしょうか？今までの官能評価で行われていた統計解析の問題点をお教えいただければと思います。同等性を言いたい場合の考え方、多重比較検定でのベイズ分析の方法と利点。：業務上の応用ではしばしば識別テストで同等性が課題になります（特に、本当は違っているのに（識別されているのに）、同じ（識別されていない）と誤って判断するケース）。ベイズ統計を識別テストに応用することで、得られた Θ やPdに関する知見からサンプル間の同等性の判断より正確に行うためには、どのような点に留意すべきとお考えでしょうか？

C: ②-1 第1回資料p15から、 β 関数を識別テストの共役事前分布として公式をあてはめれば、stanによるシミュレーションを省略して Θ の事後分布を推定できると考えて良いのでしょうか？

D: ③-2 できないと思いますが、Pdの事後分布はstanをつかわずに公式で導出することは（ Θ の時のように）難しいのでしょうか？

E: 様々な課題を解くにあたっては確率モデルを式として表す必要があると思うので、官能評価で用いる様々な統計検定での確率モデル式を紹介してほしいです。また、そのような確率モデルを学ぶのに良い書籍があれば紹介してほしいです。

F:

CATA法でのチェック数の解析など、様々な官能評価手法に対する具体的な統計モデリング例を例示いただけるとありがたいです。官能評価以外の場面での活用事例など、具体的なスクリプトをもとにご説明していただきたいです。例えば、

①梅雨時&雨天→売り上げは通常、夏&雨天→売上アップ、などのように2つ以上の条件を加味した上での売上予測はどのようになるのか？

②雨傘に加えて日傘も加えた上で傘の売上予測をしたい場合は、そこに傘の種類も加わってくるかと思います。日傘&夏&雷雨→売上は変わらない、雨傘&夏&雷雨→売上アップ、というような、分布の異なるような複雑な系存在下での売上予測上記のような複雑な例のとき、どのように考えていけばよいのか実例をもとにご講義いただけると助かります。

1. 統計学の考え方
2. 官能評価で用いる確率モデル
3. Stanを使わずに数学的に求めること
4. 中間にピークがあるモデル
5. CATAの確率モデル
6. 雨傘・日傘モデル：ダミー変数

1. 統計学の考え方

P値の問題

R. L. Wasserstein & N. A. Lazar (2016) “The ASA Statement on p-values: Context, Process, and Purpose.” *The American Statistician*, 70:2, 129-133

While p-value can be a useful statistical measure, it is commonly misused and misinterpreted. (p.131)

差の有無 いずれかに割り切ることが問題

ベイズ分析 事後分布
差の大きさについての確かさ・不確かさ (uncertainty)

ROPE (Region Of Practical Equivalence)

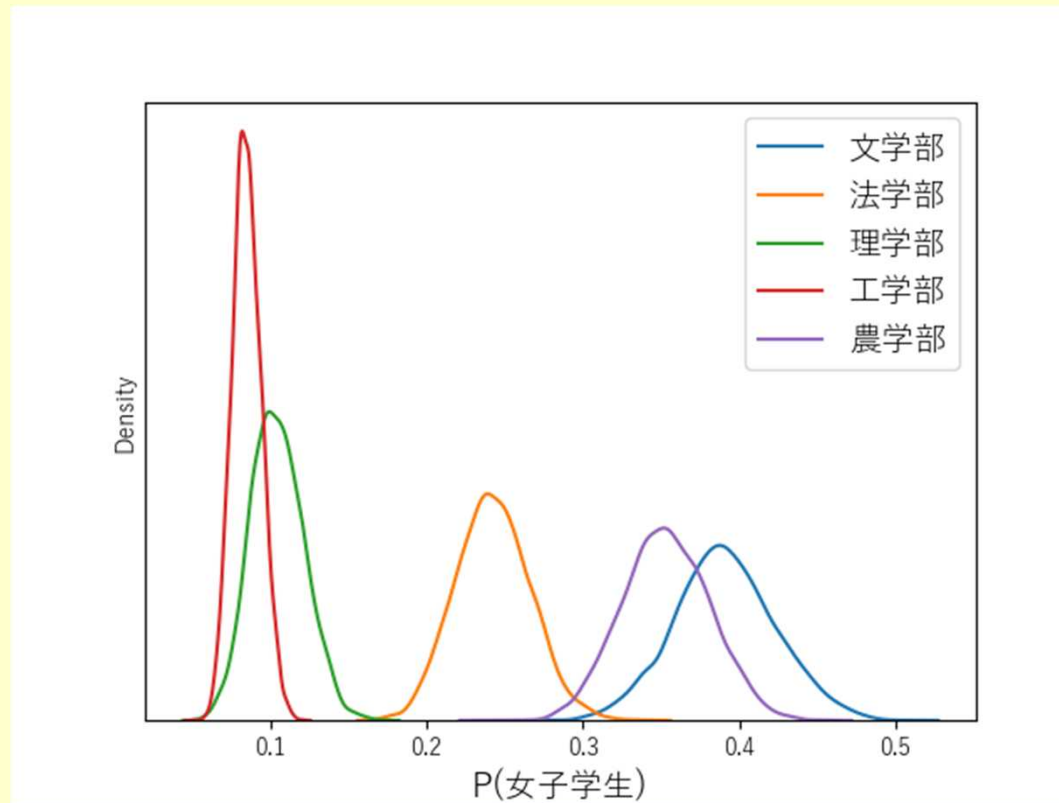
O. Martin (2016) “Bayesian Analysis with Python”

多重比較

表B.1 平成31年度男女別合格者（K大学ホームページより）

合計が200名以上の学部、内数で「認」と「他」を除く）

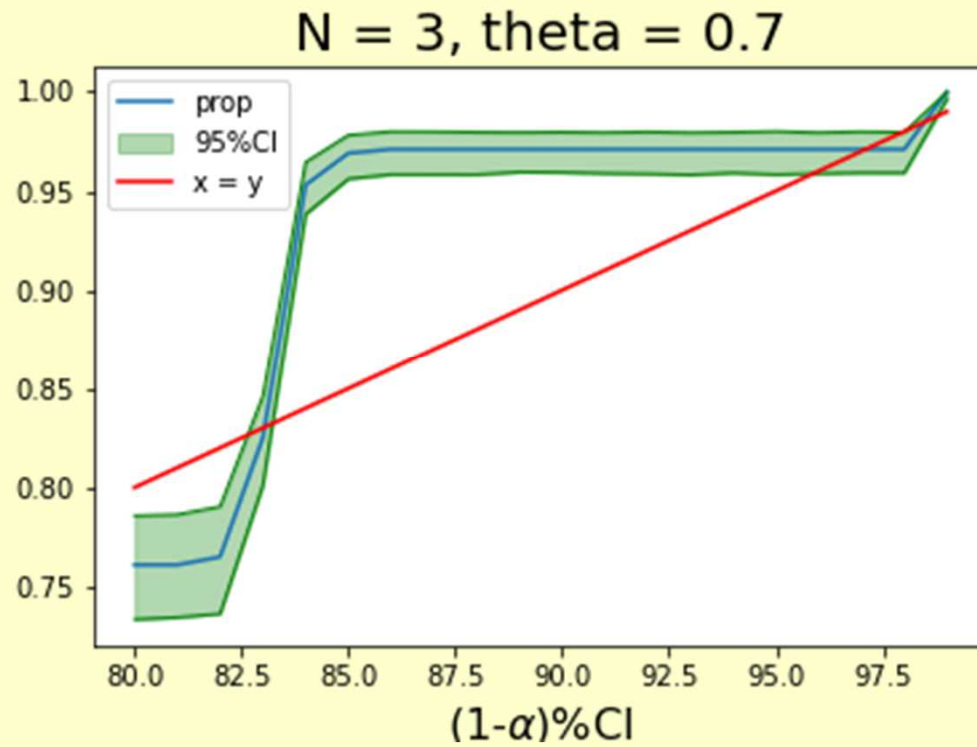
	文学部	法学部	理学部	工学部	農学部
男性	130	236	281	858	189
女性	83	75	32	79	103



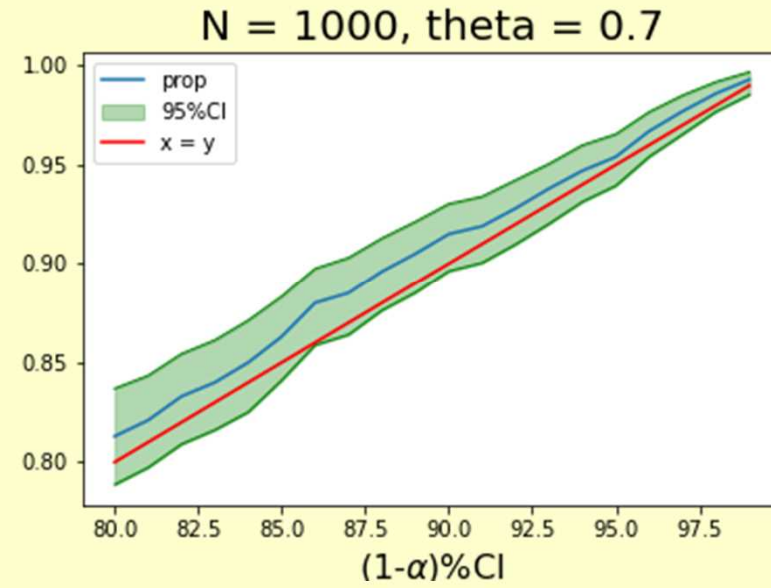
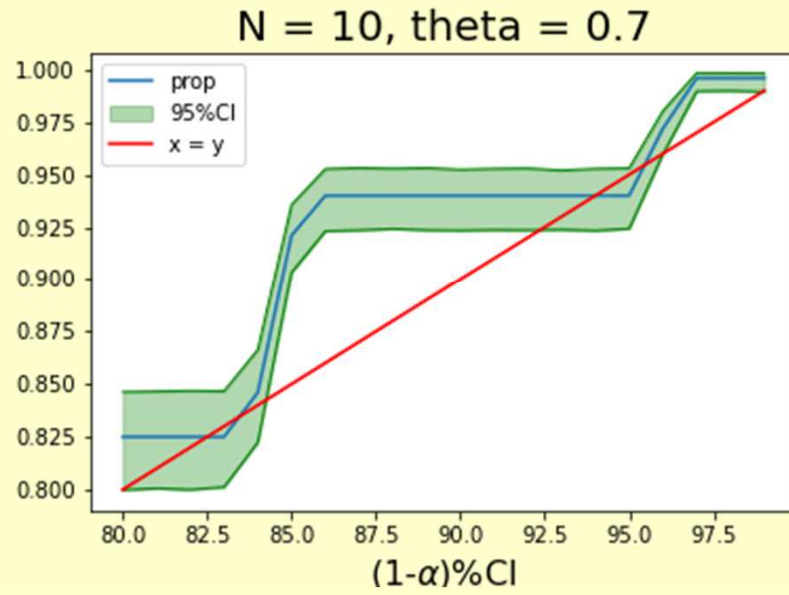
事後分布とFact

2 項分布 $\text{Bin}(N, \theta)$

$(1-\alpha)\text{CI}$ (credible Interval)

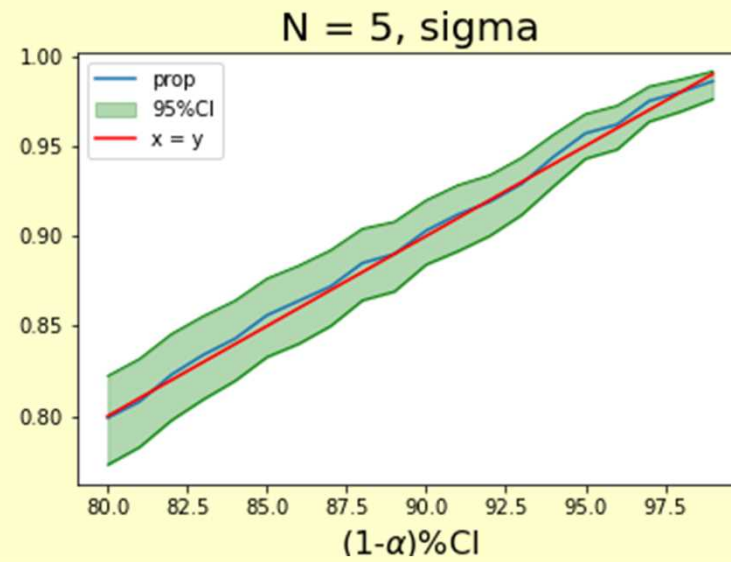
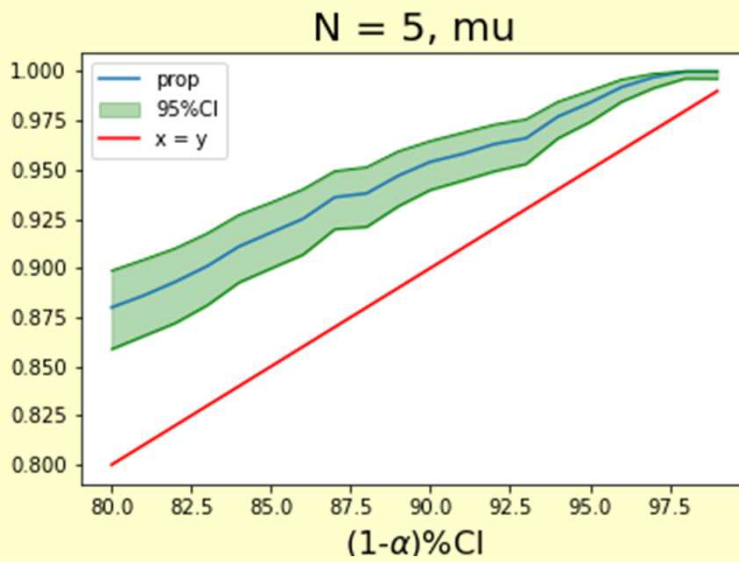


事後分布とFact

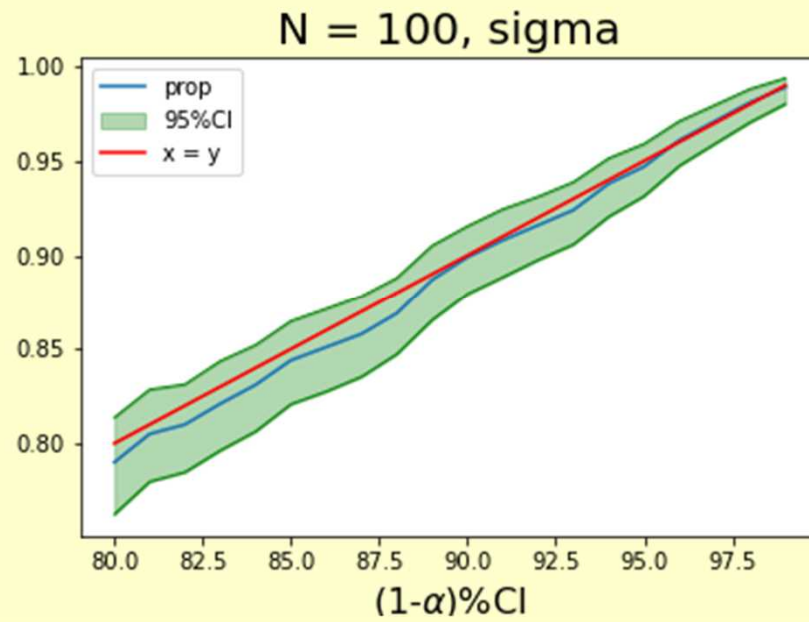
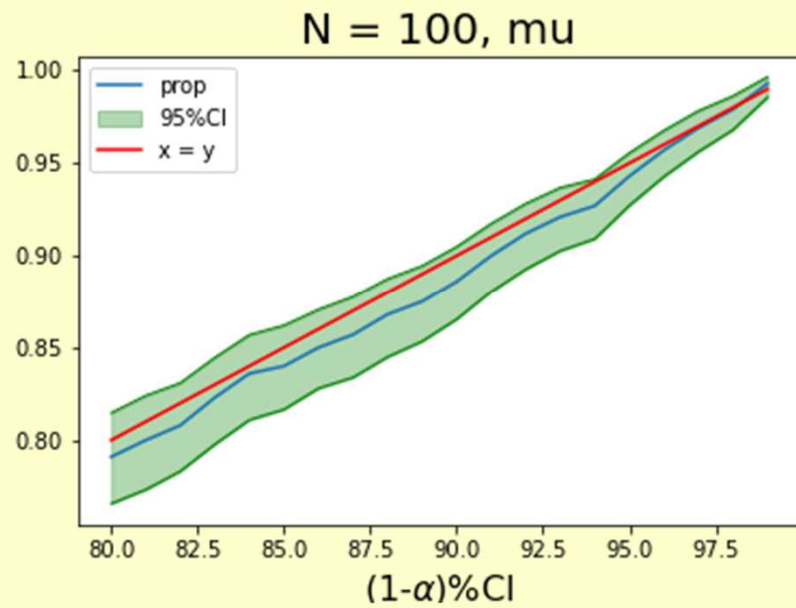


事後分布と Fact

正規分布 $N(0, 1)$



事後分布と Fact



2. 官能評価で用いる確率モデル

心理物理学／精神物理学、計量心理学

F. A. A. Kingdom & N. Prins (2016) Psychophysics.

J. K. Kruschke (2015) Doing Bayesian Data Analysis, second ed.

岡本安晴 (2014) 心理学データ分析と測定

3. Stanを使わずに数学的に求めること

$$P(P_c|k) \propto P_0(P_c) \binom{N}{k} P_c^k (1 - P_c)^{N-k}$$

$$\propto \text{Beta}(P_c; 1 + k, 1 + N - k) \quad P_0(P_c) = \text{const.}$$

$$\propto \frac{\text{Beta}(P_c; 1 + k, 1 + N - k)}{\int_{1/3}^1 \text{Beta}(P_c; 1 + k, 1 + N - k) dP_c} \quad \frac{1}{3} \leq P_c \leq 1$$

事後分泌を表す数式に基づいて統計量、平均値、中央値、
区間推定などを数学的計算で求める。

P_c に回帰モデルが設定されているときは、回帰モデルのパラメータの
事後分布関数を扱うことになるので、複雑な式になると予想される。

$$P_c = P_d + (1 - P_d) \frac{1}{3}$$

$f(P_d)$: P_d の確率密度関数

$g(P_c)$: P_c の確率密度関数

$$0 \leq P_d \leq 1$$

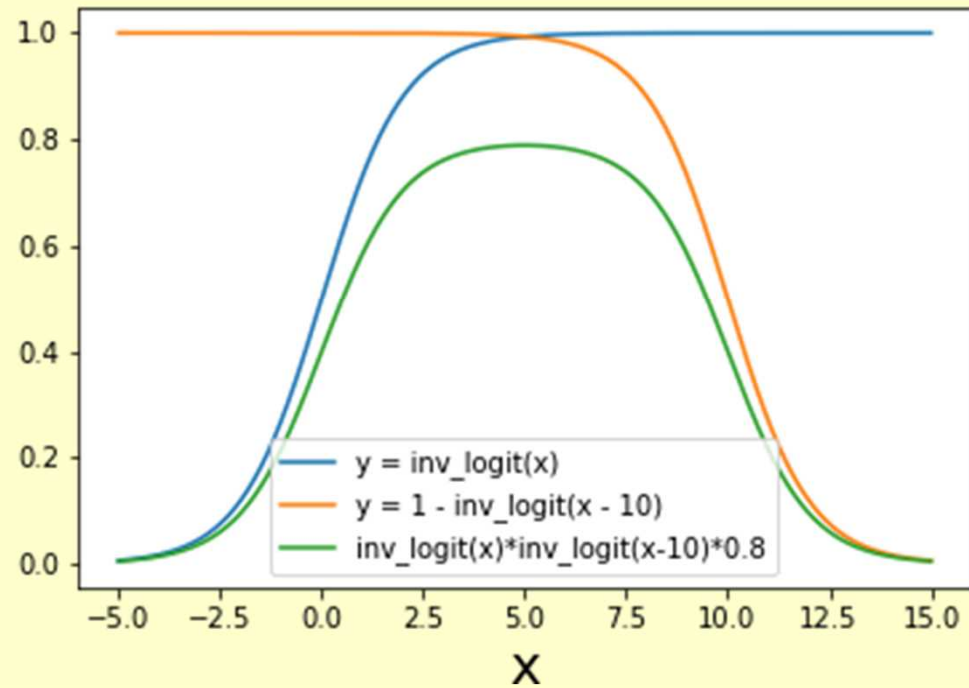
$$\frac{1}{3} \leq P_c \leq 1$$

$$f(s) = \frac{2}{3} g\left(\frac{2}{3}s + \frac{1}{3}\right)$$

4. 中間にピークがあるモデル

$$y = \text{inv_logit}(x) = \frac{1}{1 + \exp(-x)}$$

積モデル



```
import matplotlib.pyplot as plt
import numpy as np

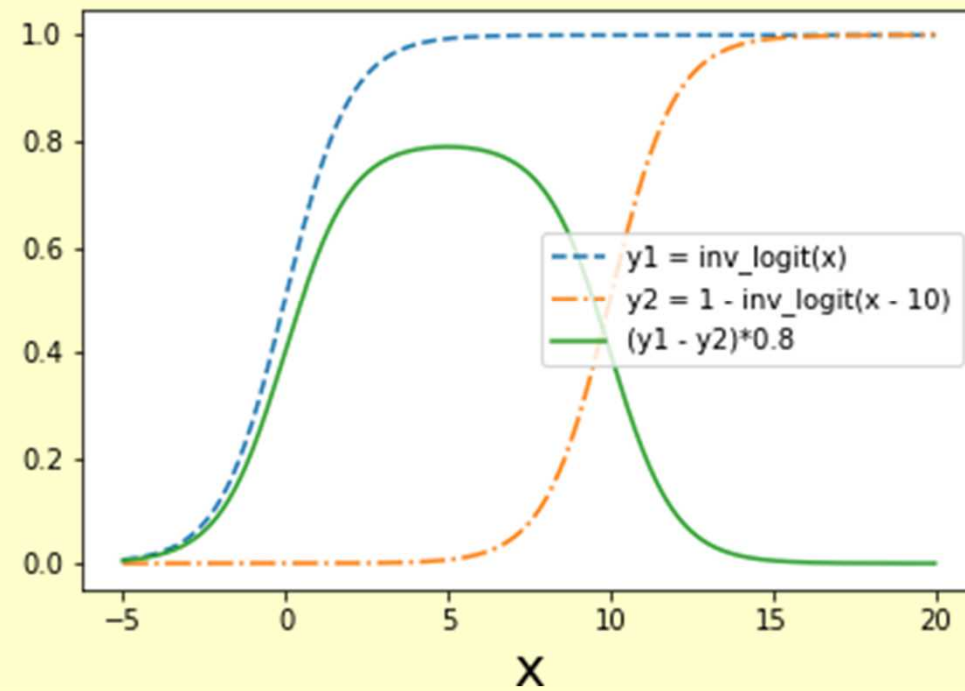
def inv_logit(x):
    return 1.0 / (1.0 + np.exp(-x))

x_coord = np.linspace(-5, 15, 1001)
y_coord = inv_logit(x_coord)
y1_coord = 1.0 - inv_logit(x_coord - 10)
yy_coord = y_coord * y1_coord * 0.8
plt.plot(x_coord, y_coord, label = 'y = inv_logit(x)')
plt.plot(x_coord, y1_coord, label = 'y = 1 - inv_logit(x - 10)')
plt.plot(x_coord, yy_coord, label = 'inv_logit(x)*inv_logit(x-10)*0.8')
plt.xlabel('x', fontsize = 20)
plt.legend()
plt.savefig('FigPeak.png')
plt.show()
```

4. 中間にピークがあるモデル

$$y = \text{inv_logit}(x) = \frac{1}{1 + \exp(-x)}$$

差モデル



```
import matplotlib.pyplot as plt
import numpy as np

def inv_logit(x):
    return 1.0 / (1.0 + np.exp(-x))

x_coord = np.linspace(-5, 20, 1001)
y_coord = inv_logit(x_coord)
y1_coord = inv_logit(x_coord - 10)
yy_coord = (y_coord - y1_coord) * 0.8
plt.plot(x_coord, y_coord, linestyle = 'dashed', label = 'y1 = inv_logit(x)')
plt.plot(x_coord, y1_coord, linestyle = 'dashdot', label = 'y2 = 1 - inv_logit(x - 10)')
plt.plot(x_coord, yy_coord, label = '(y1 - y2)*0.8')
plt.xlabel('x', fontsize = 20)
plt.legend()
plt.savefig('FigPeak_Diff.png')
plt.show()
```


5. CATAの確率モデル

CATA (Check-All-That-Apply)のデータ

3相データ： パネリスト × 評価用語 × 資料

SD法 (Semantic Differential)

確率モデル

展開法の確率モデル

$$\text{パネリスト} : N_p \text{人} \quad R_{j,i}^{(s)} = \begin{cases} 1 & \checkmark \\ 0 & \checkmark \text{なし} \end{cases}$$

$$\text{資料} : N_t \text{個} \quad T_j = (x_1^{(j)}, \dots, x_q^{(j)})$$

$$\text{評価用語} : N_a \text{個} \quad A_i = (\mu_1^{(i)}, \dots, \mu_q^{(i)})$$

$$P(R_{j,i}^{(s)}) = \phi(\|T_j, A_i\|; 0, 1)^{R_{j,i}^{(s)}} \left(1 - \phi(\|T_j, A_i\|; 0, \sigma_i^2)\right)^{1-R_{j,i}^{(s)}}$$

$$\prod_{s=1}^{N_p} \prod_{j=1}^{N_t} \prod_{i=1}^{N_a} P(R_{j,i}^{(s)})$$

6. 雨傘・日傘モデル：ダミー変数

ダミー変数： 該当 $\rightarrow 1$, 非該当 $\rightarrow 0$

$$\text{売上} = a_0 + a_1 \times \text{夏} + a_2 \times \text{梅雨} + a_3 \times \text{雷雨} + e$$

$$\text{日傘売上} = a_{01} + a_{11} \times \text{夏} + a_{21} \times \text{梅雨} + a_{31} \times \text{雷雨} + e$$

$$\text{雨傘売上} = a_{02} + a_{12} \times \text{夏} + a_{22} \times \text{梅雨} + a_{32} \times \text{雷雨} + e$$

Pythonによるベイズ分析スクリプト例

<http://y-okamoto-psy1949.la.cocacn.jp/Python/IntroPyStat/>