ケーススタディ (2)

「音声認識の性能に男女で差があるか？」

- 男性の被験者10名が10数字を発声
  - 数字音声認識率：70%
- 女性の被験者3名が10数字を発声
  - 数字音声認識率：75%

ケーススタディ (2) (つづき-2)

- 帰無仮説 H_0：
  「　　　　　　　　　　　　　　　　　　　」
  i.e. 母集団が共通
- 認識率の確率分布は？
  - 各発声データが「正しく認識されたか否か」
    ⇒
  - n が大きいときは　　　分布で近似

ケーススタディ (2) (つづき-3)

- 認識率の確率分布は？
  - n 個の音声を、比率 π で正しく認識
    ⇒ 正答数 X は二項分布 B(n, π) に従う

$$E(X) = n \cdot \pi, \quad V(X) = n \cdot \pi(1-\pi)$$

  - 認識率 P (=X/n) に変換

$$E(P) = \pi, \quad V(P) = \frac{\pi(1-\pi)}{n}$$

ケーススタディ (2) (つづき-4)

- 認識率 P の確率分布を正規分布で近似
  - 男性の音声認識率 P_1 $\quad N\left(\pi, \quad \frac{\pi(1-\pi)}{n_1}\right)$
  - 女性の音声認識率 P_2 $\quad N\left(\pi, \quad \frac{\pi(1-\pi)}{n_2}\right)$
  - 認識率の差 $P_2 - P_1$ $\quad N\left(0, \quad \pi(1-\pi)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$

補足 (1)

- 独立性
  - 確率変数 X, Y が独立であるとは

$$h(x,y)=f(x)g(y)$$

- 共分散 (covariance)

$$\begin{aligned} Cov(X,Y) &= E((X-E(X))(Y-E(Y))) \\ &= E(X\cdot Y)-E(X)\cdot E(Y) \end{aligned}$$

補足 (2)

- 確率変数 X-Y の平均

$$E(X-Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-y)\cdot h(x,y)dxdy$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\cdot h(x,y)dxdy-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\cdot h(x,y)dxdy$$

↓　X, Y は独立と仮定する

$$=\int_{-\infty}^{\infty}g(y)\,dy\int_{-\infty}^{\infty}x\cdot f(x)\,dx-\int_{-\infty}^{\infty}f(x)\,dx\int_{-\infty}^{\infty}y\cdot g(y)\,dy$$

$$=E(X)-E(Y)$$

補足 (3)

- 確率変数 X-Y の分散

$$V(X-Y)$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((x-\overline{x})-(y-\overline{y}))^2\cdot h(x,y)\,dxdy$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((x-\overline{x})^2-2(x-\overline{x})(y-\overline{y})+(y-\overline{y})^2)\cdot h(x,y)\,dxdy$$

↓　X, Y は独立と仮定する

$$=\int_{-\infty}^{\infty}(x-\overline{x})^2\cdot f(x)\,dx+\int_{-\infty}^{\infty}(y-\overline{y})^2\cdot g(y)\,dy$$
$$=V(X)+V(Y)$$

ケーススタディ (2) (つづき-5)

「音声認識の性能に男女で差があるか？」

- 男性の被験者10名が10数字を発声
  - 数字音声認識率：70％

    $n_1=100$，$p_1=0.7$
- 女性の被験者3名が10数字を発声
  - 数字音声認識率：75％

    $n_2=30$，$p_2=0.75$

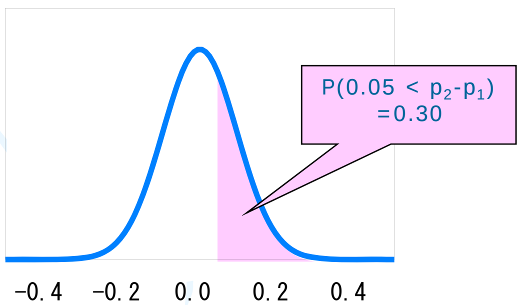
$p_2-p_1=0.05$

$\pi=(100\times0.7+30\times0.75)/(100+30)=0.71$

ケーススタディ (2) (つづき-6)

- 男女の音声認識率の差 $P_2 - P_1$ の分布

$$N\left(0,\quad 0.71\cdot(1-0.71)\left(\tfrac{1}{100}+\tfrac{1}{30}\right)\right)= N(0,\quad 0.095)$$



仮説 H_0 のもとで
p_2-p_1 が 0.05
以上になる確率は
結構大きい

ケーススタディ (2) (つづき-7)

- 仮説 H_0 のもとで認識率の差 p_2-p_1 が 0.05 以上になる確率は 30％

⇒ 帰無仮説 H_0:「男女で認識率に差はない」を、

-