統計学II

早稲田大学政治経済学術院
西郷 浩

本日の目標

- 片側検定と両側検定
- 母平均の検定
  - 正規母集団からの標本の場合
  - 正規分布に従わない母集団からの標本の場合
- 母比率（成功確率）の検定

利用データについて

- JGSS2010年
  - 抽出実験に用いたデータは、東京大学社会科学研究所付属社会調査・データアーカイブ研究センターSSJデータアーカイブのリモート集計システムを利用し、同データアーカイブが所蔵する[JGSS2010]の個票をデータを集計したものである。

片側検定（復習）

- 標本抽出
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n.$
- 前提
  - 母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0\colon\ \mu = \mu_0 \\ H_1\colon\ \mu = \mu_1 \end{cases}$ ただし、$\mu_0 < \mu_1$ であるとする。
- 有意水準0.05の棄却域
  - $D = \left\{ \bar{x}\colon\ \bar{x} > \mu_0 + 1.64\sqrt{\sigma^2/n} \right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$ として、$D = \{z\colon z > 1.64\}$
- 仮説設定の拡張

片側検定（仮説の拡張）(1)

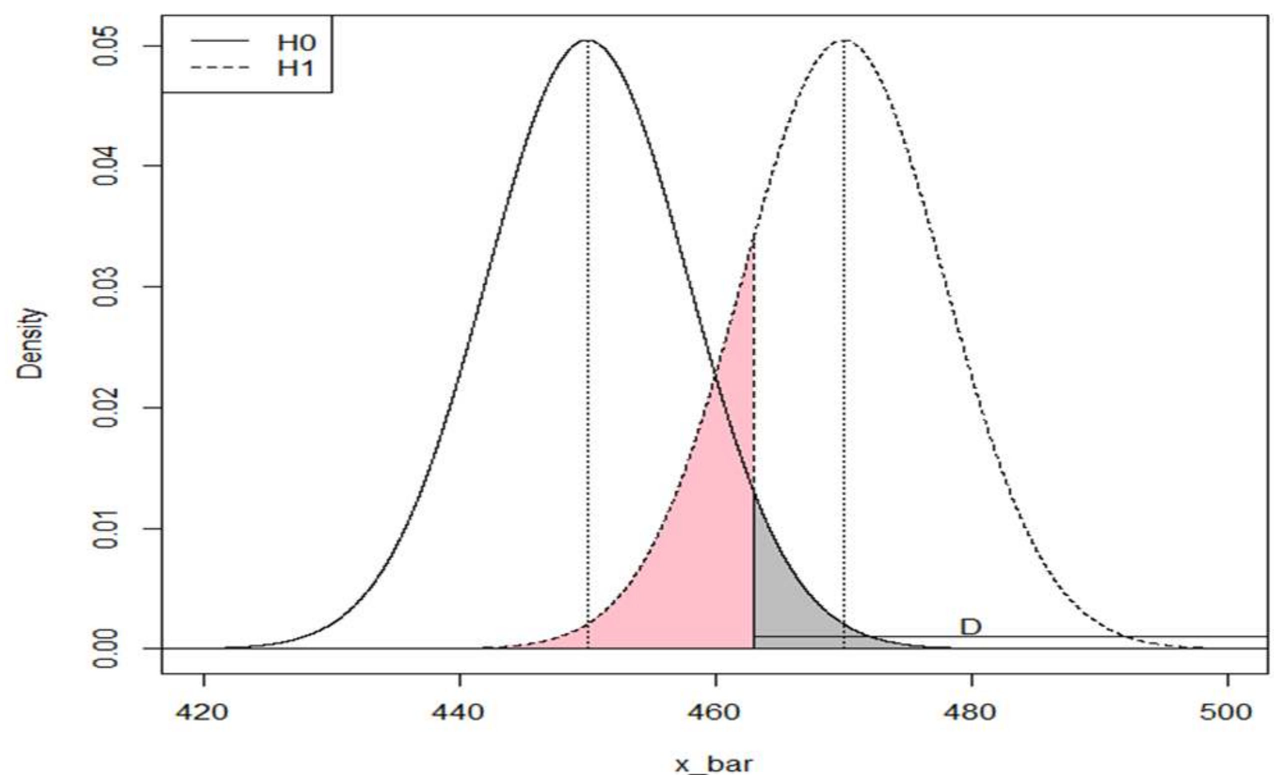
- 標本抽出
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n.$
- 前提
  - 母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$
- 有意水準0.05の棄却域
  - $D = \left\{\bar{x}\colon\ \bar{x} > \mu_0 + 1.64\sqrt{\sigma^2/n}\right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$ として、$D = \{z\colon z > 1.64\}$

片側検定（仮説の拡張）(2)

- 例 : 果樹の例の変形
  - $X_i \sim_{iid} N(\mu, 50^2)\ i = 1, 2, \dots, n.$
  - $\begin{cases} H_0\colon \mu = 450 \\ H_1\colon \mu = 470 \end{cases} \rightarrow \begin{cases} H_0\colon \mu = 450 \\ H_1\colon \mu = 465 \end{cases}$
  - 棄却域の決め方
    - 第1種の過誤（H_0が正しいにもかかわらず、それを棄却する）の確率αを0.05以下にしつつ、
    - 第2種の過誤（H_0が正しくないにもかかわらず、それを棄却しない）の確率βを最小にする。

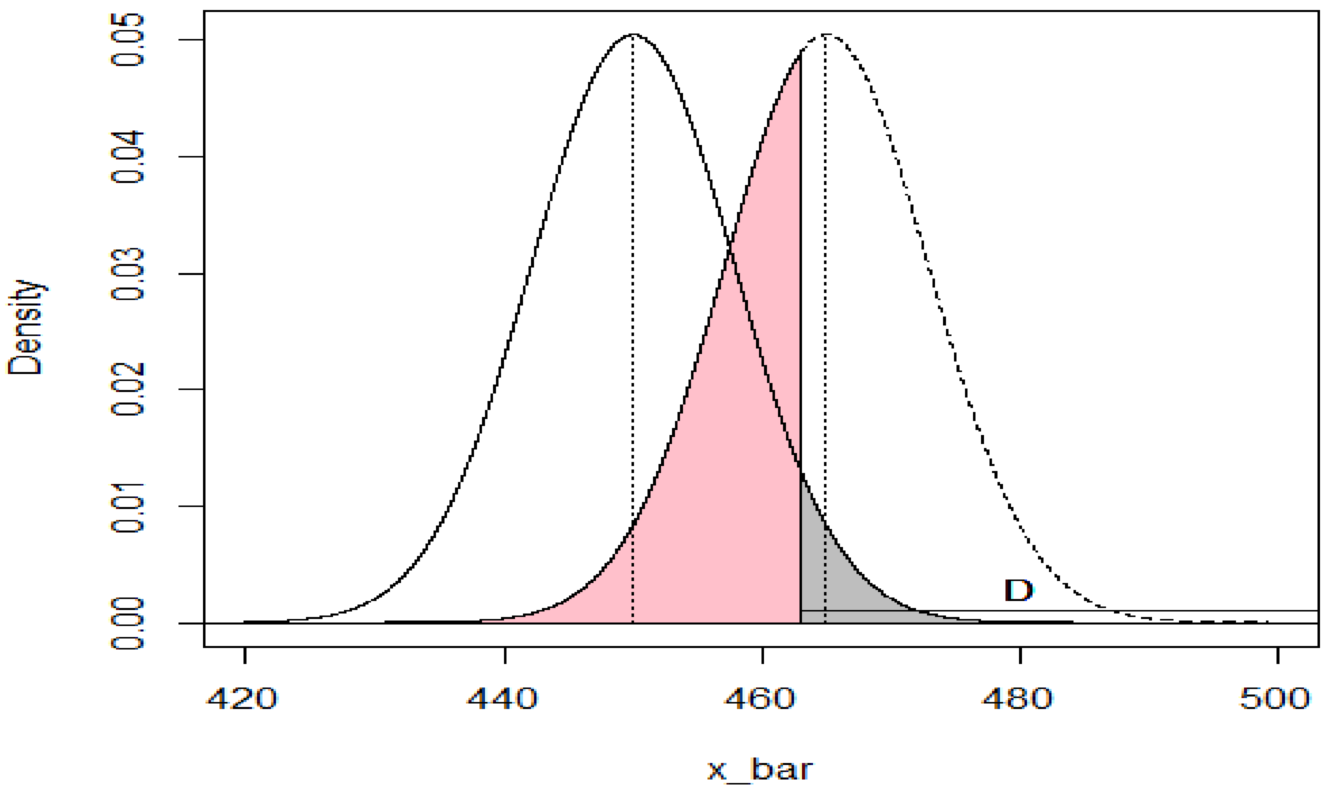
片側検定（仮説の拡張）(3)

図1：棄却域の決定（H_0: $\mu = 450$ vs H_1: $\mu = 470$）



片側検定（仮説の拡張）(4)

図2：棄却域の決定（$H_0: \mu = 450$ vs $H_1: \mu = 465$）



片側検定（仮説の拡張）(5)

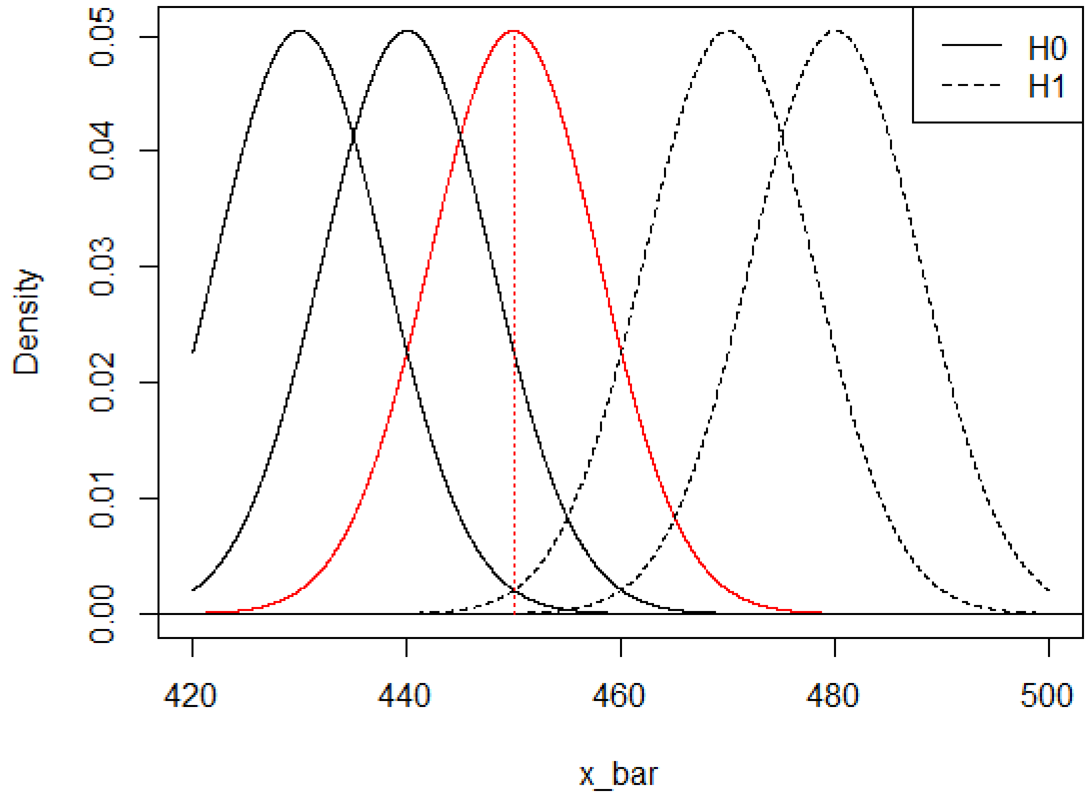
- 結論
  - 棄却域の決定に際して
    - 「対立仮説の下での標本平均$\bar{X}$ の標本分布」が、「帰無仮説の下での標本平均$\bar{X}$ の標本分布」よりも右側にあることが重要である。
      - 右側のどこに位置していても、導かれる棄却域は同じ領域になる。
  - 有意水準0.05の $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$ の棄却域：
    - $D = \left\{\bar{x}\colon \bar{x} > \mu_0 + 1.64\sqrt{\sigma^2/n}\right\}$

片側検定（仮説の拡張）(6)

- 標本抽出
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \ldots, n.$
- 前提
  - 母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0\colon \mu \leq \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$
- 有意水準0.05の棄却域
  - $D = \left\{\bar{x}\colon \bar{x} > \mu_0 + 1.64\sqrt{\sigma^2/n}\right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$として、$D = \{z\colon z > 1.64\}$

片側検定（仮説の拡張）(7)

図3：棄却域の決定（$H_0: \mu \leq 450$ vs $H_1: \mu > 450$）



片側検定（仮説の拡張）(8)

- 帰無仮説 $H_0: \mu \leq 450$
  - その中で、対立仮説 $H_1: \mu > 450$ ともっとも区別しにくい仮説：
    - $H_0': \mu = 450$
  - $H_0': \mu = 450$ を帰無仮説として検定する。
  - 検定方法：
    - H_0' が棄却される。→ H_0 も棄却される。
    - H_0' が棄却されない。→ H_0 も棄却されない。

右側検定（まとめ）

- 標本抽出と前提
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \ldots, n$.母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu = \mu_1 \end{cases} or \begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases} or \begin{cases} H_0\colon \mu \le \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$
    - ただし、$\mu_0 < \mu_1$ であるとする。
- 有意水準0.05の棄却域
  - $D = \left\{\bar{x}\colon \bar{x} > \mu_0 + 1.64\sqrt{\sigma^2/n}\right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$ として、$D = \{z\colon z > 1.64\}$

左側検定（まとめ）

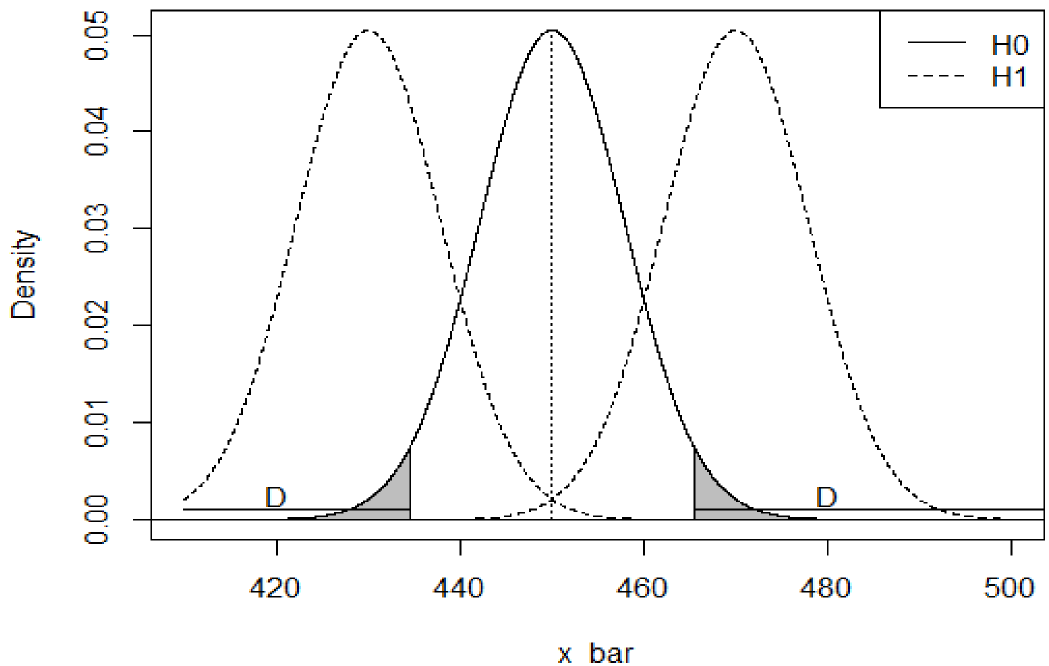
- 標本抽出と前提
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n$.母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu = \mu_1 \end{cases}$ or $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ or $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$
    - ただし、$\mu_0 > \mu_1$ であるとする。
- 有意水準0.05の棄却域
  - $D = \left\{ \bar{x}:\ \bar{x} < \mu_0 - 1.64\sqrt{\sigma^2/n} \right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$ として、$D = \{z: z < -1.64\}$

両側検定(1)

- 標本抽出と前提
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n.$母分散 σ^2 が既知である。
- 帰無仮説と対立仮説
  - $\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu \neq \mu_0 \end{cases}$
- 有意水準0.05の棄却域
  - $D = \left\{\bar{x}\colon \bar{x} < \mu_0 - 1.96\sqrt{\sigma^2/n}\ or\ \bar{x} > \mu_0 + 1.96\sqrt{\sigma^2/n}\right\}$
    - 次のようにも表現できる。$Z = \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}$ として、$D = \{z\colon |z| > 1.96\}$
    - 対立仮説のもとでは、標本平均 $\bar{X}$ （または、それを標準化した Z ）が、帰無仮説で想定される母平均μ_0 （または、Zの期待値0）よりも、小さくなりすぎる可能性と、大きくなりすぎる可能性とがある。
      - そのような状況では、両側にH_0 の棄却域を設ける方が安全である。

両側検定(2)

図4：両側検定の棄却域



両側検定(3)

- X大学Y学部
  - 入学時に英語能力試験を実施している。
  - 過去の実績(得点の分布)
    - 平均点500点、標準偏差100点の正規分布。
  - 入学者900人に対して実施した試験の実際の平均点が495点だった。
  - 入学者900人の平均的な英語の能力が例年と異なるといえるか。

両側検定(4)

- 解答
  - 仮説の設定
    - データ発生の仕組み：$X_i \sim N(\mu, 100^2)$
    - 帰無仮説：H_0: $\mu = 500$
    - 対立仮説：H_1: $\mu \neq 500$
    - 有意水準を0.05とする。
  - 棄却域
    - $|Z| = \frac{|\bar{X}-500|}{\sqrt{100^2/900}} > 1.96$

両側検定(5)

– 結論

- $|Z|_{obs} = \frac{|495-500|}{\sqrt{100^2/900}} = 1.5$ なので棄却域に入らない。
  → H_0は棄却できない。
- つまり:
  – 今年度の入学者の平均点で測った能力が例年と異なるという証拠は得られなかった。

母平均の検定(1)

- 母分散 σ^2 が未知のとき
  - ケース1：正規母集団からの標本の場合：$X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n$
    - t 分布の利用
  - ケース2：非正規母集団からの標本の場合：$X_i \sim_{iid} (\mu, \sigma^2)\ i = 1, 2, \dots, n$
    - 標本の大きさ n が大きければ、正規近似（中心極限定理）を利用する。

母平均の検定(2)

- ケース1：正規母集団からの標本の場合
  - $X_i \sim_{iid} N(\mu, \sigma^2)\ i = 1, 2, \dots, n$
  - $\begin{cases} H_0\colon\ \mu = \mu_0 \\ H_1\colon\ \mu \neq \mu_0 \end{cases}$
  - 一般に、真の μ を使えば、以下が成り立つ。
    - $T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t(n-1)$ 教科書104-106ページ。
  - H_0 が正しいとき、以下が成り立つ。
    - $T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \sim t(n-1)$
  - 帰無仮説と対立仮説の組み合わせから、両側検定になる。
    - $|T| > t_{0.025}(n-1)$ なら、H_0 を棄却する。

母平均の検定(3)

- 例：NB10（名目重量が10g の重り）
  - 5回の測定結果：9.999591g, 9.999600g, 9.999594g, 9.999601g, 9.999598g
  - 仮説の設定：$\begin{cases} H_0: \mu = 10 \\ H_1: \mu \neq 10 \end{cases}$
  - 検定統計量：$|T| = \frac{|\bar{X} - \mu_0|}{\sqrt{S^2/n}}$
  - 有意水準0.05の棄却域（意思決定規則）
    - $|T| > t_{0.025}(4) = 2.776$ であれば、H_0 を棄却する。
  - 検定統計量の値
    - $|T|_{obs} = \frac{|9.999597 - 10|}{\sqrt{1.77 \times 10^{-11}/5}} = 214.197 > 2.776$
  - 検定の結果
    - H_0 は棄却される。

母平均の検定(4)

- ケース2：非正規母集団からの標本の場合
  - $X_i \sim_{iid} (\mu, \sigma^2)\ i = 1, 2, \dots, n$
  - 標本平均が近似的に正規分布にしたがう。
    - $T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \to_d \ N(0, 1)$
    - 標本の大きさn がある程度大きい（30以上）のときは、T が正規分布に従うとみなして、臨界値を決める。
  - 両側検定
    - 仮説の設定：$\begin{cases} H_0\colon\ \mu = \mu_0 \\ H_1\colon\ \mu \neq \mu_0 \end{cases}$
    - 検定統計量：$|T| = \frac{|\bar{X} - \mu_0|}{\sqrt{S^2/n}}$
    - 有意水準0.05の棄却域
      - $|T| > 1.96$ であれば、H_0 を棄却する。

母平均の検定(5)

– 例：JGSS2010 における就労年数
  - 大きさ $n = 30$ の無作為標本
    – 31, 0, 11, 23, 1, 45, 8, 7, 25, 5, 36, 9, 20, 17, 6, 10, 3, 4, 38, 25, 12, 30, 16, 35, 1, 35, 5, 7, 5, 41.（教科書と異なるサンプルであることに注意）
  - 仮説の設定：$\begin{cases} H_0\colon\ \mu = 10 \\ H_1\colon\ \mu \neq 10 \end{cases}$
  - 検定統計量：$|T| = \frac{|\bar{X}-10|}{\sqrt{S^2/n}}$
  - 有意水準 0.05 の棄却域
    – $|T| > 1.96$ であれば、H_0 を棄却する。
  - 検定統計量の値
    – $|T|_{obs} = \frac{|17.0-10|}{\sqrt{190.4/30}} = 2.77 > 1.96$
  - 検定の結果
    – H_0 は棄却される。

成功確率（母比率）の検定(1)

- 成功確率（母比率）の検定
  - 標本抽出：$X_i \sim Bernoulli(p)\ i = 1,2,\ldots,n.$
    - 有限母集団における母比率の検定にも利用する。
      - 例：有権者の政党支持率
  - 真の p を使えば、以下の性質が成り立つ。
    - $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$（近似的に）ただし、$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$（標本比率）
  - 右側検定
    - $\begin{cases} H_0: p \le p_0 \\ H_1: p > p_0 \end{cases}$
  - 検定統計量と棄却域（有意水準0.05）
    - $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$　もし、$Z > 1.64$ なら、H_0 を棄却する。

成功確率（母比率）の検定(2)

- 例：JGSS2010 における、労働組合に入っていない人の割合
  - 母集団から50人を無作為抽出して母比率について検定する。
  - 仮説の設定（6割よりも多くの人が労働組合に入っていないかどうか）
    - $\begin{cases} H_0: p \leq 0.6 \\ H_1: p > 0.6 \end{cases}$
  - 検定統計量と棄却域
    - $Z = \frac{\hat{p} - 0.6}{\sqrt{\frac{0.6(1-0.6)}{50}}}$ を検定統計量とする。これが1.64を超えるときに帰無仮説を棄却する。
  - 標本抽出の結果、39人が「労働組合に入っていない」と回答した。
  - 検定統計量の値
    - $Z = \frac{\frac{39}{50} - 0.6}{\sqrt{\frac{0.6(1-0.6)}{50}}} = 2.60 > 1.64$
    - H_0 は棄却される（労働組合に入っていない人の割合は60%より高い）