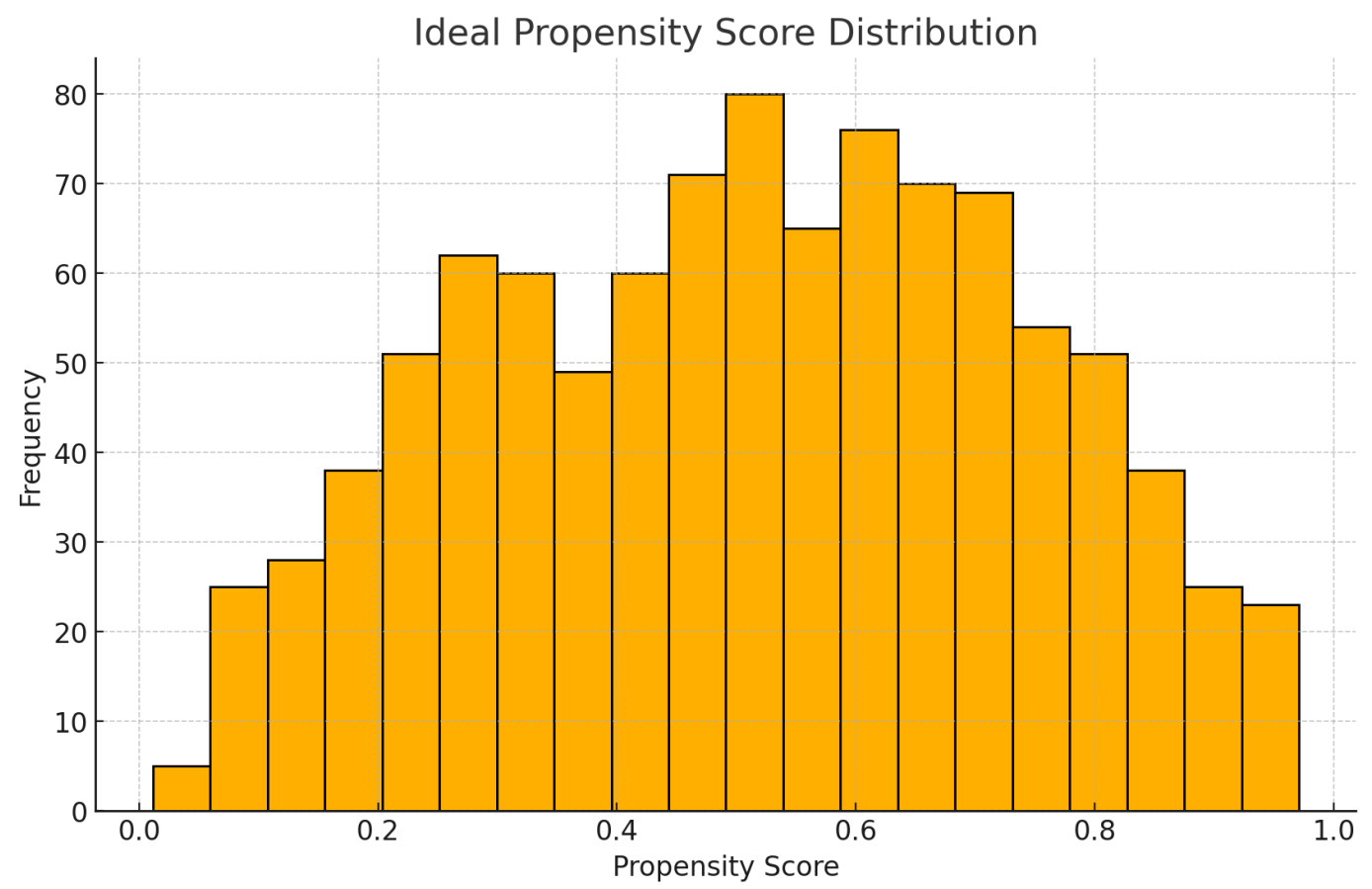


傾向スコアに関する質問を回答

Q. 今回の設問ではあまりマッチしていないとの事でしたが、マッチした場合はどのような分布になるのでしょうか？

A. 本来期待されるヒストグラムの例を提示します。



追記。

そもそも交絡因子の値に問題あり。妥当な交絡因子の値をSMD（以下）を使い、評価。以下に教科書で使ったデータに対するSMD値を掲載。理想は SMD値<0.1 が望ましい。

```
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
smoke      drink      age      ps
2.0000000  0.3924414 -0.2008718  2.5832924
```

SMD: Standardized Mean Difference 標準化平均差

アルゴリズムとコード。

1. 連続変数の SMD

1. グループごとの平均と分散を求める

- $\bar{X}_t, s_t^2$: 処置群 (treated) の平均と分散
- $\bar{X}_c, s_c^2$: 対照群 (control) の平均と分散

2. プールド標準偏差を計算

$$s_p = \sqrt{\frac{s_t^2 + s_c^2}{2}}$$

3. SMD を計算

$$\text{SMD} = \frac{\bar{X}_t - \bar{X}_c}{s_p}$$

2. 二値変数の SMD

1. グループごとの発生率を求める

- p_t : 処置群における事象の割合
- p_c : 対照群における事象の割合

2. プールド分散を計算

$$v_p = \frac{p_t(1 - p_t) + p_c(1 - p_c)}{2}$$

3. SMD を計算

$$\text{SMD} = \frac{p_t - p_c}{\sqrt{v_p}}$$

#— 0. データ準備 —

```
df <- data.frame(  
  smoke      = c(1,0,0,0,1,0,1,1,0,0,0,1,0,0,0,1,0,0,0,0),  
  drink      = c(1,1,1,1,1,0,0,0,0,1,0,0,0,0,0,0,0,0,1,1),  
  age        = c(51,63,52,43,56,66,63,60,84,64,70,82,47,50,77,52,60,71,69,78),  
  treatment  = c(1,0,0,1,1,0,1,1,0,0,0,1,0,0,0,1,0,0,1,1)  
)
```

#— 1. 傾向スコアモデルの構築 —

```
ps_model <- glm(
  treatment ~ smoke + drink + age,
  family = binomial(link = "logit"),
  data = df
)
df$ps <- predict(ps_model, type = "response")
```

#— 2. 調整前のバランス診断 (SMD 計算) —

```
compute_smd <- function(x, grp) {
  tr <- x[grp == 1]; ct <- x[grp == 0]
  if (all(unique(x) %in% c(0,1))) {
    # 二値変数
    p_t <- mean(tr); p_c <- mean(ct)
    v_p <- (p_t*(1-p_t) + p_c*(1-p_c)) / 2
    return((p_t - p_c) / sqrt(v_p))
  } else {
    # 連続変数
    m_t <- mean(tr); v_t <- var(tr)
    m_c <- mean(ct); v_c <- var(ct)
    s_p <- sqrt((v_t + v_c)/2)
    return((m_t - m_c) / s_p)
  }
}

vars <- c("smoke", "drink", "age", "ps")
smd_before <- sapply(vars, function(v) compute_smd(df[[v]], df$treatment))
print(smd_before)
```

Warning message:

"glm.fit: fitted probabilities numerically 0 or 1 occurred"

smoke	drink	age	ps
2.0000000	0.3924414	-0.2008718	2.5832924