Why drug testing works

Drug	$Y_1, Y_2, \ldots, Y_n$
Placebo	$Z_1, Z_2, \ldots, Z_n$

Drug $Y_1, Y_2, \ldots, Y_n$

Placebo $Z_1, Z_2, \ldots, Z_n$

$$\Delta := \left| \frac{Y_1 + Y_2 + \cdots + Y_n}{n} - \frac{Z_1 + Z_2 + \cdots + Z_n}{n} \right|$$

Drug	$Y_1, Y_2, \ldots, Y_n$
Placebo	$Z_1, Z_2, \ldots, Z_n$

$$\Delta := \left| \frac{Y_1 + Y_2 + \cdots + Y_n}{n} - \frac{Z_1 + Z_2 + \cdots + Z_n}{n} \right|$$

Is the observed discrepancy Δ larger than we would expect from random variation if the drug were truly safe?

$$\begin{array}{rl} \text{Drug} & Y_1, Y_2, \dots, Y_n \\ \text{Placebo} & Z_1, Z_2, \dots, Z_n \end{array}$$

$$\Delta := \left| \frac{Y_1 + Y_2 + \cdots + Y_n}{n} - \frac{Z_1 + Z_2 + \cdots + Z_n}{n} \right|$$

Is the observed discrepancy Δ larger than we would expect from random variation if the drug were truly safe?

The basis of a statistical test:
Assess the likelihood of the observed discrepancy *if the drug were indeed truly harmless* by comparison with a *gedanken* experiment consisting of a double sample of placebo patients.

A *gedkanken* or thought experiment: a double placebo sample

$$\left.\begin{array}{l} X_1, X_2, \dots, X_n \\ X'_1, X'_2, \dots, X'_n \end{array}\right. \sim \mathrm{Bernoulli}(p)$$

A *gedkanken* or thought experiment: a double placebo sample

$$\left.\begin{array}{l} X_1, X_2, \ldots, X_n \\ X'_1, X'_2, \ldots, X'_n \end{array}\right\} \sim \text{Bernoulli}(p) \quad \Rightarrow \quad \left.\begin{array}{l} S_n = X_1 + X_2 + \cdots + X_n \\ S'_n = X'_1 + X'_2 + \cdots + X'_n \end{array}\right\} \sim \text{Binomial}(n, p)$$

A *gedkanken* or thought experiment: a double placebo sample

$$\begin{array}{l} X_1, X_2, \ldots, X_n \\ X'_1, X'_2, \ldots, X'_n \end{array} \sim \text{Bernoulli}(p) \quad \Rightarrow \quad \begin{array}{l} S_n = X_1 + X_2 + \cdots + X_n \\ S'_n = X'_1 + X'_2 + \cdots + X'_n \end{array} \sim \text{Binomial}(n, p)$$

We know n but p is the *unknown* rate at which (cardiopulmonary or other) complications occur in the general population

A *gedkanken* or thought experiment: a double placebo sample

$$\left.\begin{matrix} X_1, X_2, \dots, X_n \\ X'_1, X'_2, \dots, X'_n \end{matrix}\right. \sim \text{Bernoulli}(p) \quad \Rightarrow \quad \left.\begin{matrix} S_n = X_1 + X_2 + \cdots + X_n \\ S'_n = X'_1 + X'_2 + \cdots + X'_n \end{matrix}\right. \sim \text{Binomial}(n, p)$$

We know n but p is the *unknown* rate at which (cardiopulmonary or other) complications occur in the general population

Key idea: how likely is it that the difference in the relative frequencies of "success" in the *gedanken* double placebo sample exceeds the observed discrepancy Δ?

A *gedkanken* or thought experiment: a double placebo sample

$$\left.\begin{matrix} X_1, X_2, \dots, X_n \\ X'_1, X'_2, \dots, X'_n \end{matrix}\right\} \sim \mathrm{Bernoulli}(p) \quad \Longrightarrow \quad \left.\begin{matrix} S_n = X_1 + X_2 + \cdots + X_n \\ S'_n = X'_1 + X'_2 + \cdots + X'_n \end{matrix}\right\} \sim \mathrm{Binomial}(n, p)$$

We know n but p is the *unknown* rate at which (cardiopulmonary or other) complications occur in the general population

Key idea: how likely is it that the difference in the relative frequencies of "success" in the *gedanken* double placebo sample exceeds the observed discrepancy Δ?

We want to evaluate

$$\mathbf{P}\left\{ \left| \frac{S_n}{n} - \frac{S'_n}{n} \right| > \Delta \right\}$$

A *gedkanken* or thought experiment: a double placebo sample

$$\left.\begin{matrix} X_1, X_2, \dots, X_n \\ X'_1, X'_2, \dots, X'_n \end{matrix}\right\} \sim \text{Bernoulli}(p)$$

$$\left.\begin{matrix} S_n = X_1 + X_2 + \cdots + X_n \\ S'_n = X'_1 + X'_2 + \cdots + X'_n \end{matrix}\right\} \sim \text{Binomial}(n, p)$$

We know n but p is the *unknown* rate at which (cardiopulmonary or other) complications occur in the general population

Key idea: how likely is it that the difference in the relative frequencies of "success" in the *gedanken* double placebo sample exceeds the observed discrepancy Δ?

We want to evaluate

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S'_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{ \left| \frac{S_n}{n} - \frac{S_n}{n} \right| > \Delta \right\}$$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$|S_n - S'_n| = \big|(S_n - np) - (S'_n - np)\big|$$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S_n' &= X_1' + X_2' + \cdots + X_n' \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S_n'| > n\Delta\}$$

$$|S_n - S_n'| = \big|(S_n - np) - (S_n' - np)\big|$$

Triangle inequality: $|a - b| \leq |a| + |b|$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$\begin{aligned} |S_n - S'_n| &= \big|(S_n - np) - (S'_n - np)\big| \\ &\leq |S_n - np| + |S'_n - np| \end{aligned}$$

Triangle inequality: $|a - b| \leq |a| + |b|$

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > {\color{red}\Delta}\right\} = \mathbf{P}\{|S_n - S'_n| > n{\color{red}\Delta}\}$$

$$\begin{aligned} |S_n - S'_n| &= \big|(S_n - np) - (S'_n - np)\big| \\ &\le |S_n - np| + |S'_n - np| \end{aligned}$$

Triangle inequality: $|a - b| \le |a| + |b|$

If $|a - b| > c$ then $|a| > c/2$ or $|b| > c/2$

$$S_n = X_1 + X_2 + \cdots + X_n$$
$$S'_n = X'_1 + X'_2 + \cdots + X'_n$$
$$\sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$|S_n - S'_n| = \big|(S_n - np) - (S'_n - np)\big|$$
$$\leq |S_n - np| + |S'_n - np|$$

Triangle inequality: $|a - b| \leq |a| + |b|$

If $|a - b| > c$ then $|a| > c/2$ or $|b| > c/2$

If the event $|S_n - S'_n| > n\Delta$ occurs then either $|S_n - np| > n\Delta/2$ or $|S'_n - np| > n\Delta/2$ or both occur

$$\left.\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ S'_n &= X'_1 + X'_2 + \cdots + X'_n \end{aligned}\right\} \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{S_n}{n}\right| > \Delta\right\} = \mathbf{P}\{|S_n - S'_n| > n\Delta\}$$

$$\begin{aligned} |S_n - S'_n| &= \big|(S_n - np) - (S'_n - np)\big| \\ &\leq |S_n - np| + |S'_n - np| \end{aligned}$$

Triangle inequality: $|a - b| \leq |a| + |b|$

If $|a - b| > c$ then $|a| > c/2$ or $|b| > c/2$

If the event $|S_n - S'_n| > n\Delta$ occurs then either $|S_n - np| > n\Delta/2$ or $|S'_n - np| > n\Delta/2$ or both occur

$$\{|S_n - S'_n| > n\Delta\} \subseteq \{|S_n - np| > n\Delta/2\} \cup \{|S'_n - np| > n\Delta/2\}$$