Binary Responses: an Introduction
Coffee taste test

- A fast food chain is considering a change in the blend of coffee beans they use to make their coffee.
- What to know: do the customers prefer the new blend?
- Experiment: select n customers, have them taste two cups of coffee, and say which one they prefer
- Want to know: is there a difference between the two blends?
Statistical Model

• $Y_i$: customer $i$ prefers the new blend

• $Y_i = Bernoulli(\pi)$
  - $P(Y_i = 1|\pi) = \pi$, $P(Y_i = 0|\pi) = 1 - \pi$
  - $P(y_i|\pi) = \pi^{y_i} (1 - \pi)^{1-y_i}$

• All the $Y$s are mutually independent

• $L = P(y_1, y_2, \ldots, y_{100}|\pi) = \prod_{i=1}^{n} P(y_i|\pi)$

• The MLE* is $\hat{\pi}$ s.t. $\prod_{i=1}^{n} P(y_i|\hat{\pi})$ is maximized
  - Maximum likelihood estimator
The MLE of \( \pi \)

\[
\frac{\partial}{\partial \pi} \log L = \frac{\partial}{\partial \pi} \log \prod_{i=1}^{n} P(y_i|\pi)
\]

\[
= \frac{\partial}{\partial \pi} \log \prod_{i=1}^{n} \pi^{y_i} (1 - \pi)^{1-y_i}
\]

\[
= \frac{\partial}{\partial \pi} \sum_{i=1}^{n} (\log \pi^{y_i} + \log(1 - \pi)^{1-y_i})
\]

\[
= \frac{\partial}{\partial \pi} \sum_{i=1}^{n} (y_i \log \pi + (1 - y_i) \log(1 - \pi))
\]
\[
\frac{\partial}{\partial \pi} \sum_{i=1}^{n} \left( y_i \log \pi + (1 - y_i) \log(1 - \pi) \right) = 0
\]

\[
\sum_{i=1}^{n} \frac{y_i}{\pi} = \sum_{i=1}^{n} \frac{1 - y_i}{1 - \pi}
\]

\[
\frac{1}{\pi} \sum_{i=1}^{n} y_i = \frac{1}{1 - \pi} \sum_{i=1}^{n} (1 - y_i)
\]

\[
\frac{\#(y_i = 1)}{N - \#(y_i = 1)} = \frac{\pi}{1 - \pi}
\]

\[
\frac{\#(y_i = 1)/N}{1 - \#(y_i = 1)/N} = \frac{\pi}{1 - \pi}
\]

\[
\pi = \frac{\#(y_i = 1)}{N} = \bar{y}
\]
MLE

• The MLE is then $\hat{\pi} = \bar{y}$

• $\bar{Y} = \frac{\sum_i Y_i}{n}$ so $\text{var}(\bar{Y}) = \frac{1}{N^2} \sum_i \text{var}(Y_i) = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}$
  • (Why)

• For large $n$, $\bar{Y} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$ due to the CLT
Null Hypothesis

• For large $n$, $\bar{Y} \sim N \left( \pi, \frac{\pi(1-\pi)}{n} \right)$
• Null Hypothesis: $\pi = 0.5$
(in R)

• Simulate the experiment