Data File
n rows : cases
Columns : variables
- Age
- Gender
- height
- weight
- income
- Diagnosis
- Disease Severity
- Fam Hist of HD
- Family Doctor - YN
- 5 year survival - YN

Vars can be
Categorical
- unordered : Nominal scale
- ORDERED : Ordinal scale

Continuous / Quantitative
- Blood pressure
- height

Explanatory vs Response
x vars in reg. | y in regression
Ind. vars. | dependent var

Distributions

Bernoulli

Binomial

Multinomial

Poisson

Poisson process

Events happening randomly in space or time

Independent increments

In small region

- Prob of 2 or more $\to 0$
- Prob is roughly proportional to size of interval

\# of events in ~~on~~ a region

$$p(y) = \frac{e^{-\lambda} \lambda^{y}}{y!}, \quad y = 0, 1, \ldots$$

Model

Reality

Model: $y_1, \ldots, y_n \overset{iid}{\sim} \text{Bernoulli}(\pi)$

Task: Estimate π

Model Ind for $i = 1, \dots, n$

$P(New) = \pi \quad P(Old) = 1 - \pi$

$$P(y_i) = \pi^{y_i}(1-\pi)^{1-y_i}, \text{ for } y = 0, 1$$

Estimate π by max likelihood

Have $y_1, \dots, y_n$

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

$$= \frac{\partial}{\partial \pi} \log\left(\pi^{\sum_{i=1}^{n} y_i} (1-\pi)^{n - \sum_{i=1}^{n} y_i} \right)$$

$$= \frac{\partial}{\partial \pi}\left(\sum_{i=1}^{n} y_i \log \pi + (n - \Sigma y) \log(1-\pi) \right)$$

$$= \frac{\Sigma y_i}{\pi} + \frac{n - \Sigma y_i}{1-\pi}(-1) = \frac{n\bar{y}}{\pi} - \frac{n(1-\bar{y})}{1-\pi} \overset{set}{=} 0$$

$$\Rightarrow \frac{n\bar{y}}{\pi} = \frac{n(1-\bar{y})}{1-\pi} \Rightarrow \cancel{n}\pi - \cancel{n\pi\bar{y}} = \cancel{n}\bar{y} - \cancel{n\pi\bar{y}}$$

$$\Rightarrow \hat{\pi} = \bar{y} = P = \frac{\sum_{i=1}^{n} y_i}{n}$$

2nd derivative test

Likelihood is written $\ell(\pi)$

$$\frac{d^2}{d\pi^2} \log \ell(\pi)$$

$$= \frac{\partial}{\partial \pi}\left(\frac{n\bar{y}}{\pi} - \frac{n(1-\bar{y})}{1-\pi} \right)$$

$$= \frac{\partial}{\partial \pi}\left(n\bar{y}\pi^{-1} - n(1-\bar{y})(1-\pi)^{-1} \right)$$

$$= n\bar{y}(-1)\pi^{-2} - n(1-\bar{y})(-1)(1-\pi)^{-2}(-1))$$

$$= \frac{-n\bar{y}}{\pi^2} - \frac{n(1-\bar{y})}{(1-\pi)^2}$$

$$= -n\left(\frac{\bar{y}}{\pi^2} + \frac{1-\bar{y}}{(1-\pi)^2} \right) < 0$$

concave down

MAX

Hypothesis test to find out if new coffee is preferred.

$$H_0 : \pi = \tfrac{1}{2} \qquad \alpha = 0.05$$
$$H_1 : \pi \neq \tfrac{1}{2}$$

CLT says $\dfrac{\bar{y} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \to \infty} N(0,1)$

Here, $\mu = \pi$, $\sigma^2 = \pi(1-\pi)$, so

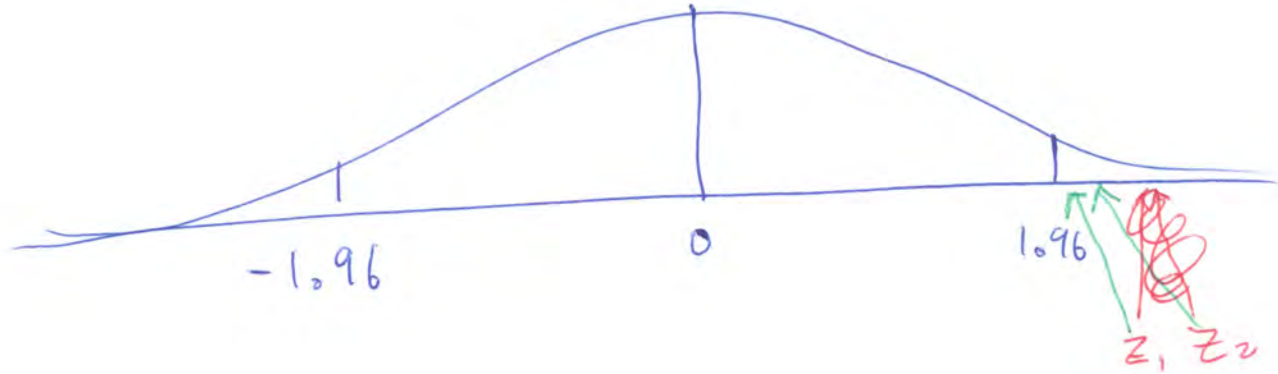
$$\frac{\sqrt{n}(p - \pi)}{\sqrt{\pi(1-\pi)}} \longrightarrow N(0,1)$$

Spse H_0 is true, & $\pi = 0.5 = \pi_0$

$$Z_1 = \frac{\sqrt{n}(p - \pi_0)}{\sqrt{\pi_0(1-\pi_0)}} = \frac{\sqrt{100}(0.6-0.5)}{\sqrt{0.5(1-0.5)}} = 2$$

$$Z_2 = \frac{\sqrt{n}(p - \pi_0)}{\sqrt{p(1-p)}} = \frac{\sqrt{100}(.6-.5)}{\sqrt{.6(1-.6)}} = 2.04$$

Critical value $= z_{\alpha/2} = 1.96$



Conclusions

A) In symbols: $\pi \neq \frac{1}{2}$. Specifically $\pi > \frac{1}{2}$

B) In words Plain Language

More consumers prefer the new coffee blend.

What if $Z = 1.84$ >

Please don't accept H_0 just bec it's not rejected. Say

"There's not enough evidence to conclude that preference for the 2 coffees is different."

"These results are consistent with no difference in preference for the two coffees"

confidence intervals

$$1-\alpha \approx P_r\left(-z_{\alpha/2} < Z_2 < z_{\alpha/2}\right)$$

$$= P_r\left(-z_{\alpha/2} < \frac{\sqrt{n}\,(p-\pi_0)}{\sqrt{p(1-p)}} < z_{\alpha/2}\right)$$

$$= P_r\left(\frac{-z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} < p - \pi < \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

$$= P_r\left(-p - \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} < -\pi < -p + \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

$$= P_r\left(p + \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} > \pi > p - \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

$$= P_r\left(p - \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} < \pi_0 < p + \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

could say $p \pm \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}$ 95% Margin of error

For coffee taste test 95% CI is

$$\left(.6 - \frac{1.96\sqrt{.6 * .4}}{\sqrt{100}},\ .6 + \frac{1.96\sqrt{.6 * .4}}{10}\right)$$

$$= (0.504, 0.696)$$

Connection between tests & CIs

$H_0: \pi = \pi_0$ will be rejected at level α if & only if the $(1-\alpha)100\%$ confidence interval does not contain π_0.

To show, just put π_0 in place of π in derivation of the confidence interval