ADVANCED STATISTICS

HW-2

Priya Iddalgi
Rahul Rajput

> Linear Invariance of Regression Coefficients

1) Transformed x_i : $\tilde{x_i} = \frac{x_i - \bar{x}}{s_x}$ (standardized

Original Linear Equation : $y_i = \beta_0 + \beta_1 x_i + \varepsilon$

Linear eqn on the standardized parameter will be of the form : $y_i = \alpha_0 + \alpha_1 \tilde{x_i} + \varepsilon$

The least sq. estimates can be found as :

$$\alpha_0 = \bar{y_i} - \alpha_1 \bar{\tilde{x}}$$

$$\alpha_1 = \frac{\sum (y_i - \bar{y_i})(\tilde{x_i} - \bar{\tilde{x}})}{\sum (\tilde{x_i} - \bar{\tilde{x}})^2}$$

where, $\bar{y_i}$ & $\bar{\tilde{x}}$ are sample means of y_i and $\tilde{x_i}$

Note that α_0 & α_1 describe changes in y_i with respect to $\tilde{x_i}$ (and not x_i). They can be converted back to their original scale as :

$$\beta_0 = \alpha_0 + \alpha_1 \frac{\bar{x}}{s_x} \quad , \quad \beta_1 = \frac{\alpha_1}{s_x}$$

→ α_0 and α_1 reflect standard deviation of y for a one standard deviation change in x.

2) Relationship b/w r (yi and xi) and α_1 :

We know that,

$$r = \frac{\Sigma (\tilde{y}_i - \tilde{x}_i)}{n-1} \longrightarrow (1)$$

where, $\tilde{y}_i = \frac{y_i - \bar{y}}{s_y}$

and $\tilde{x}_i = \frac{x_i - \bar{x}}{s_x}$

Substituting in (1), we have,

$$r = \Sigma \left[\left(\frac{y_i - \bar{y}}{s_y} \right) \right] * \left[\left(\frac{x_i - \bar{x}}{s_x} \right) \right] / n-1$$

Also, $\alpha_1 = \frac{\Sigma (\tilde{y}_i - \tilde{x}_i)}{\Sigma (\tilde{x}_i)^2} \longrightarrow (2)$

Dividing by n-1

we have $\boxed{\alpha_1 = r \frac{s_y}{s_x}}$

3) Sampling variance of α_0 :

$$var(\alpha_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{n-1} \Sigma (\tilde{x}_i)^2 \right]$$

$$var(\alpha_1) = \frac{\sigma^2}{(n-1)} \Sigma (\tilde{x}_i)^2$$

Sampling covariance b/w slope and intercept :

$$cov(\alpha_0, \alpha_1) = \frac{\sigma^2 \bar{x}}{(n-1)} \Sigma (\tilde{x}_i)^2$$

σ^2 = common variance of errors ε_i

4)

Yes, we can use standardized slope & intercept-estimates (α_0, α_1) for regression of x_i and y_i

Original: $y_i = \beta_0 + \beta_1 x_i + \varepsilon$ → ①

Standardized: $\tilde{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 \tilde{x}_i + \varepsilon$ → ②

We know that, $\tilde{x}_i = \frac{x_i - \bar{x}}{s_x}$] × α_1

Multiplying by α_1,

$$\tilde{x}_i \hat{\alpha}_1 = \hat{\alpha}_1 \left(\frac{x_i - \bar{x}}{s_x} \right)$$

Substituting in ②,

$$y_i = \hat{\alpha}_0 + \hat{\alpha}_1 \left(\frac{x_i - \bar{x}}{s_x} \right) + \varepsilon$$

↳ $$y_i = \hat{\alpha}_0 + \frac{\hat{\alpha}_1 x_i}{s_x} - \frac{\alpha_1 \bar{x}}{s_x} + \varepsilon$$

Thus, $\boxed{\beta_1 = \frac{\hat{\alpha}_1}{s_x}}$ (coefficient term with x_i)

$$y_i = \hat{\alpha}_0 - \frac{\hat{\alpha}_1}{s_x} \bar{x} + \beta_1 x_i + \varepsilon$$

then $\boxed{\beta_0 = \hat{\alpha}_0 s_x - \hat{\alpha}_1 \bar{x}}$

thus, we can use $\hat{\alpha}_0$ and $\hat{\alpha}_1$ by changing back their scale to predict x_i from x_i.

5. If each variable gets multiplied by 100, there would be no change to $\hat{\alpha}_0$ and $\hat{\alpha}_1$ as they are "standardized" estimates.

Consider,

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad \longrightarrow (1)$$

$$\beta_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \quad \longrightarrow (2)$$

Now $x \rightarrow x_i' = 100 x_i$ & $\bar{x}' = 100 \bar{x}$

Substituting,

$$\beta_0' = \bar{y} - \beta_1' \bar{x}'/s_x$$

$$\beta_0' = \bar{y} - \beta_1' (100 \bar{x}')/100 s_x$$

$$\beta_1' = \frac{\Sigma (100 x_i - 100 \bar{x})(y_i - \bar{y}) \times 100 s_x}{\Sigma (100 x_i - 100 x_i)}$$

$$\therefore \beta_0' = \bar{y} - \beta_1 \bar{x}'/s_x$$

$$\beta_1' = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y}) s_x}{\Sigma (x_i - \bar{x})^2}$$

$$\therefore \alpha_0' = \alpha_0 \quad \& \quad \alpha_1' = \alpha_1.$$

Thus, scaling will not change α_0 or α_1.

Practical implications:

i) Improved Interpretability: Standardizing all predictors will make their scale same & effect of each predictor can be compared.

ii) Robustness to Outliers & Improved Stability: Scaling can make regression estimates

sensitive to outliers. Standardizing can help remove this sensitivity. Standardizing also makes the predictors more stable.

6.)

Examples where x_i & ϵ_i may not be independent:

i) Autocorrelation / Time series data:

In many data collected over time, the errors ϵ_i are correlated with previous errors. Sequences are often correlated to themselves, for eg., advertising affecting sales for each month.

ii) Measurement Errors:

If independent variables (x_i) are measured with error, these errors can be correlated with other errors, violating independence assumptions.

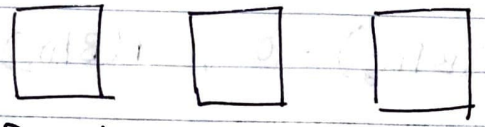
iii) Spatial / Cluster Data:

If data is collected from different location, ϵ_i maybe correlated with x_i. When data are collected from groups of individuals (clusters), the ϵ_i within the same group might be correlated with x_i.

In all these cases, we may get biased and inefficient estimates.

Probability Review

1)



→ At the start, we assume that there is equal probability of the car being behind each door.

⇒ $P(D_1) = P(D_2) = P(D_3) = 1/3$ [Prob of car being behind a given door]

→ This is Unconditional probability.

→ We choose Door 1, ⇒ $P(D_1) = 1/3$

but, $P(\sim D_1) = 2/3$

→ Host, has prior knowledge, opens door 3, which has a goat.

Event 2 = 'B', Host opening door 3 to reveal a goat.

→ We want to calculate probability of car being behind Door 1, given the Host opened Door 3 to reveal a goat, i.e., Event 'B'.

⇒ $P(D_1|B) = ?$ Using Bayes Theorem:

$$P(D_1|B) = \frac{P(B|D_1)P(D_1)}{P(B|D_1) \times P(D_1) + P(B|\sim D_1)P(\sim D_1)} =$$

$$\frac{P(B|D_1)P(D_1)}{P(B|D_1)P(D_1) + P(B|D_2)P(D_2) + P(B|D_3)P(D_3)}$$

$P(D_1) = P(D_2) = P(D_3) = 1/3$

$P(B|D_1) = 1/2$, $P(B|D_3) = 0$, $P(B|D_2) = 1$

→ $P(B|D_1) = 1/2$, because if money is behind Door 1, then the host could have only opened Door 2 or Door 3. Since Door 3 does not have the car, it is either behind Door 1 or Door 2.

→ $P(B|D_3) = 0$, because we now know it has a goat.

→ $P(B|D_2) = 1$, because Prob of opening Door 3 given car is behind Door 2. Host has only 2 options Door 2 & Door 3, Since we have chosen Door 1 & host cannot open that. Door 3 is opened to reveal goats, which only leaves Door 2 as the definite option

$$\Rightarrow P(D_1|B) = \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3} = \frac{1/6}{3/6} = 1/3$$

⇒ Probability of winning if we stay = $1/3$ OR,

Probability of winning if we switch = $2/3$

MLE Estimator of Poisson Distribution.

$x_1, x_2, x_3, \ldots, x_n \sim \text{Poisson}(\lambda)$, where

$$P(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \{0, 1, 2 \ldots\}$$

1) Mean, Expected value $= E(x) = \sum_{i=1}^{n} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$

$$= e^{-\lambda} \sum_{i=1}^{\infty} \frac{x \cdot \lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{i=1}^{\infty} \frac{x}{x!} \times \frac{\lambda}{\lambda} \times \lambda^x$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

Let $z = x - 1$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^z}{z!}$$ [using Taylor series Expansion]

$$= \lambda e^{-\lambda} \times e^{\lambda}$$

$$= \underline{\lambda}.$$

$\text{Var}(x) = E(x^2) - (E(x))^2$, $(E(x))^2 = \lambda^2$

$$E(x^2) = \sum_{i=1}^{\infty} x^2 \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \cdot \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \left[\sum_{x=1}^{\infty} \frac{(x-1)}{(x-1)!} \cdot \lambda^{x-1} + \sum_{x=1}^{\infty} \frac{1}{(x-1)!} \lambda^{x-1} \right]$$

$$= \lambda e^{-\lambda} \left[\lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right]$$

Substituting $(x-2)$ & $(x-1)$ with j & k & applying Taylor Series expansion,

$$= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda})$$
$$= \lambda^2 + \lambda$$

$$\Rightarrow \text{Var}(x) = (\lambda^2 + \lambda) - \lambda^2 = \underline{\underline{\lambda}}.$$

2) MLE Estimator.

$$P(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \{0, 1, 2, 3 \ldots\}$$

→ MLE is a way to assess the distribution of the data from which the random sample is generated. We try to estimate the value of the parameter of the original data distribution, such that the probability of getting the same values as the observed values is the highest.

Likelihood function

$$f(x_1, x_2 \ldots x_n | \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^x}{x!} \cdot x \qquad \left(p(x|\lambda) \cdot x \right)$$

$$\Rightarrow \ln f(x_1, x_2, \ldots x_n | \lambda) = \ln \left[\frac{e^{-n\lambda} \times \lambda^{\sum_{i=1}^{n} x_i}}{(x_1! \times \ldots \times x_n!)} \right]$$

$$\Rightarrow \ln f(x_1, x_2, \ldots x_n | \lambda) = \ln(e^{-n\lambda}) + \left[\ln \left(\lambda^{\sum_{i=1}^{n} x_i} \right) - \ln(x_1! \times \ldots \times x_n!) \right]$$

$$\Rightarrow \ln f(x_1, x_2, \ldots x_n | \lambda) = -n\lambda + \sum_{i=1}^{n} x_i \cdot \ln\lambda - \ln(x_1! \times \ldots \times x_n!)$$

Now, to find maxima, $\frac{d}{d\lambda} \ln f(x_1, x_2, \ldots x_n | \lambda) = 0$.

$$\Rightarrow -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\Rightarrow \boxed{\lambda = \bar{x}}$$

$\therefore$ MLE for $\lambda = \bar{x}$ OR $\hat{\lambda} = \bar{x}$

3) Now, $E(\hat{\lambda}) = E(\bar{x}) = E(y) = \lambda$,

Since $E(\hat{\lambda}) = \lambda$, Unbiased.

MLE and Inadmissible Estimators.

$X \rightarrow$ binary random variable, 0 or 1.
prob $\rightarrow p$ for 0, $p-1$ for 1.

1) $$f(x_i|p) = p^{x_i} \cdot (1-p)^{1-x_i}$$

$$\text{Likelihood} = L(p) = \prod_{i=1}^{n} f(x_i|p) = p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{x_n}$$

$$\text{or, } L(p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

Taking log,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Differentiating & setting derivative to 0 to locate maxima.

$$\frac{d}{dp} \ln L(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i - np + p\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - np = 0$$

$$\Rightarrow p = \frac{\sum_{i=1}^{n} x_i}{n}$$

MLE of p $\Rightarrow \hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$.

2) $E[\hat{p}] = E[\bar{x}] = E\left(\frac{x_1 + x_2 + \cdots x_n}{n}\right)$

$= \frac{1}{n}\left(E(x_1) + E(x_2) + \cdots + E(x_n)\right)$

$= \frac{(p + p + \cdots p)}{n} n = \frac{np}{n} = p$

Since $E(\hat{p}) = p$, the MLE estimator, $\hat{p}$, is unbiased.

3) Expected Square Error.

$MSE = E(\hat{p} - p)^2 = E(\hat{p}^2) + E(p^2) - 2pE(\hat{p}) = Var(\hat{p}) + E(\hat{p}) - p]^2$

$= E\left[\left(\frac{X}{n} - p\right)^2\right] = E\left[\frac{X^2}{n^2} - \frac{2pX}{n} + p^2\right]$

$= \frac{1}{n^2} E[X^2] - \frac{2p}{n} E[X] + p^2$

for Bernoulli Distr, $E(X) = np$ & $E[X^2] = Var(X) + E[X]^2$

$= np(1-p) + np^2$

$= \frac{1}{n^2}\left(np(1-p) + np^2\right) - \frac{2p}{n} \cdot np + p^2$

$= \boxed{\frac{p(1-p)}{n}}$

4) from the prev part, we get the expected MSE for a Bernoulli distribution as,

$$\frac{p(1-p)}{n}$$

=) if $p = 0$ or 1, the MSE $= 0$.

i.e, in a distribution where there can only be success or failure, the MSE $= 0$.

Now if $p \in \left[\frac{1}{4}, \frac{3}{4}\right]$, $n = 3$

=) if $p = 1/4$, $MSE = \frac{1/4 \times 3/4}{3} = \frac{1}{16}$

if $p = 3/4$, $MSE = \frac{3/4 \times 1/4}{3} = \frac{1}{16}$

for $p \in [1/4, 3/4]$, since $R(\hat{p}, p)$ is the same for both values of p & $E(\hat{p}) = p$, there is no minima for MSE among all values of $\hat{p}$.

∴ we can claim that the MLE estimator is inadmissible.