Simple Linear Regression

Yanrong Yang

Research School of Finance, Actuarial Studies and Statistics
The Australian National University

May 15th, 2017

Content

Common Population Mean Model

The common population mean model (CPM) is as follows.

$$y_i = \mu + e_i, \quad i = 1, \ldots, n, \tag{1}$$

where $e_1, \ldots, e_n \sim i.i.d. N(0, \sigma^2)$. In this model, y_i, $i = 1, \ldots, n$ are observed.

Remark

CPM model is equivalent to the model that there is a random sample $y_1, \ldots, y_n$ which is from the population $N(\mu, \sigma^2)$ with μ being unknown.

Statistical Inference for CPM

As σ^2 is known, we have the following statistical inference for μ.

1. Point estimator: $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.
2. Interval estimator: $\left(\bar{y} \pm z_{\tau/2}\sigma/\sqrt{n}\right)$.
3. Hypothesis test: $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$. p-value is $2P\left(Z > \left|\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\right|\right)$.

As σ^2 is unknown, we have the following statistical inference for μ.

1. Point estimator: $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.
2. Interval estimator: $\left(\bar{y} \pm t_{\tau/2}(n-1)s_y/\sqrt{n}\right)$.
3. Hypothesis test: $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$. p-value is $2P\left(T_{n-1} > \left|\frac{\bar{y}-\mu_0}{s_y/\sqrt{n}}\right|\right)$.

Simple Linear Regression

Simple linear regression model is defined as

$$y_i = \mu_i + e_i, \quad \mu_i = \alpha + \beta x_i, \quad i = 1, \ldots, n, \tag{2}$$

where

1. error components: $e_1, \ldots, e_n \sim i.i.d. N(0, \sigma^2)$;
2. independent variables or covariate variables: $x_1, \ldots, x_n$ are observed constants;
3. dependent variables: $y_1, \ldots, y_n$ are observed;
4. intercept parameter: α;
5. slope parameter: β.

Least Square Estimation

The goal is to estimate α and β in SLR model. LSE procedure

1. $(a, b) = \arg\min_{\alpha,\beta} SSE$ with $SSE = \sum_{i=1}^{n}(y_i - a - bx_i)^2$.
2. $0 = \frac{\partial SSE}{\partial a} = -\sum_{i=1}^{n} 2(y_i - a - bx_i)$ and
   $0 = \frac{\partial SSE}{\partial b} = -\sum_{i=1}^{n} 2(y_i - a - bx_i)x_i$.
3. From the two equations we can get $a = \bar{y} - b\bar{x}$ and
   $a = \frac{\sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2}{n\bar{x}}$.
4. Equating the two expressions about a, we get $b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$. And then $a = \bar{y} - b\bar{x}$.

Properties of Least Square Estimators

Theorem (Theorem 1 and 2)

$a = \bar{y} - b\bar{x}$ and $b = S_{xy}/S_{xx}$ are unbiased estimators of α and β respectively.

Proof.

1. $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x}) y_i - \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x}) y_i$. $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \bar{x}) x_i$.
2. $\mathbb{E}b = \frac{\mathbb{E}S_{xy}}{S_{xx}} = \frac{(x_i - \bar{x})\mathbb{E}y_i}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\alpha + \beta x_i)}{S_{xx}} = \frac{\beta S_{xx}}{S_{xx}} = \beta$.
3. $\mathbb{E}a = \mathbb{E}(\bar{y} - b\bar{x}) = \mathbb{E}\bar{y} - \bar{x}\mathbb{E}b = (\alpha + \beta\bar{x}) - \bar{x}\beta = \alpha$.
4. $\mathbb{E}(\bar{y}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}y_i = \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) = \alpha + \beta\bar{x}$.

$\square$

Properties continuing

Theorem (Theorem 3, 5 and 6)

$Var(b) = \frac{\sigma^2}{S_{xx}}$, $Var(a) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{nS_{xx}}$ and $Cov(a,b) = \frac{-\bar{x}\sigma^2}{S_{xx}}$.

Proof.

1. $V(S_{xy}) = V\left(\sum_{i=1}^n (x_i - \bar{x})y_i\right) = \sum_{i=1}^n (x_i - \bar{x})^2 V(y_i) = S_{xx}\sigma^2$.
2. $V(b) = \frac{S_{xx}\sigma^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}$.
3. $V(a) = V(\bar{y} - b\bar{x}) = V(\bar{y}) + \bar{x}^2 V(b) - 2\bar{x}C(\bar{y}, b)$.
4. $V(\bar{y}) = \frac{1}{n^2}\sum_{i=1}^n V(y_i) = \frac{\sigma^2}{n}$.
5. $C(\bar{y}, b) = C\left(\frac{1}{n}\sum_{i=1}^n y_i, \frac{1}{S_{xx}}\sum_{j=1}^n (x_j - \bar{x})y_j\right) =$ $\frac{1}{nS_{xx}}\sum_{i=1}^n \sum_{j=1}^n (x_j - \bar{x})C(y_i, y_j) = \frac{1}{nS_{xx}}\sum_{i=1}^n (x_i - \bar{x})\sigma^2 = 0$.
6. $C(a,b) = C(\bar{y} - b\bar{x}, b) = C(\bar{y}, b) - \bar{x}C(b,b) = -\frac{\bar{x}\sigma^2}{S_{xx}}$.

□

Properties continuing

Theorem (Theorem 7)

Let $\lambda = u + v\alpha + w\beta$, where u, v and w are finite constants. Then

1. $\hat{\lambda} = u + va + wb$ *is an unbiased estimator of* λ.
2. $V(\hat{\lambda}) = \frac{\sigma^2}{S_{xx}} \left(v^2 \frac{1}{n} \sum_{i=1}^{n} x_i^2 + w^2 - 2vw\bar{x}\right)$.

Proof.

1. $\mathbb{E}(\hat{\lambda}) = u + v\mathbb{E}(a) + w\mathbb{E}(b) = u + v\alpha + w\beta = \lambda$.
2. $V(\hat{\lambda}) = v^2 V(a) + w^2 V(b) + 2vwC(a, b)$.

□

Statistical Inference for SLR Model

Under the assumption that $e_1, \ldots, e_n \sim i.i.d. N(0, \sigma^2)$, we have

$$\frac{a-\alpha}{\sqrt{V(a)}} \sim N(0,1), \quad \frac{b-\beta}{\sqrt{V(b)}} \sim N(0,1), \quad \frac{\hat{\lambda}-\lambda}{\sqrt{V(\hat{\lambda})}} \sim N(0,1).$$

Inference on β:

1. $1-\tau$ CI for β is $\left(b \pm z_{\tau/2}\sqrt{V(b)}\right)$.
2. p-value for testing $H_0 : \beta = \beta_0$ vs $H_1 : \beta \neq \beta_0$ is $2P\left(Z > \left|\frac{b-\beta_0}{\sqrt{V(b)}}\right|\right)$.

Inference on $\mu = \alpha + x\beta$:

1. $1-\tau$ CI for μ is $\left(\hat{\mu} \pm z_{\tau/2}\sqrt{V(\hat{\mu})}\right)$.
2. p-value for testing $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ is $2P\left(Z > \left|\frac{\hat{\mu}-\mu_0}{\sqrt{V(\hat{\mu})}}\right|\right)$.

Prediction

The goal is to estimate a new independent single value

$$y = \alpha + \beta x + e,$$

where $e \sim N(0, \sigma^2)$ is an error term that is independent of $e_1, \dots, e_n$.
A reasonable estimator for y is $\hat{y} = a + bx$.

Remark

For the parameter $\mu = \alpha + \beta x$, we provided the estimator $\hat{\mu} = a + bx$.

Prediction Inference (I)

Statistical inference (or prediction inference) on $\hat{y}$:

1. $\hat{y} - y = a + bx - \alpha - \beta x - e$ is normal distributed.
2. $\frac{\hat{y}-y}{\sqrt{V(\hat{y}-y)}} \sim N(0,1)$.
3. an exact $1-\tau$ prediction interval (PI) for y is
$$\left(\hat{y} \pm z_{\tau/2}\sqrt{V(\hat{y}-y)}\right) = \left(a + bx \pm z_{\tau/2}\sigma\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right).$$

Remark

Recall that the exact $1-\tau$ CI for $\mu = \alpha + \beta x$ is
$$\left(\hat{\mu} \pm z_{\tau/2}\sqrt{V(\hat{\mu})}\right) = \left(a + bx \pm z_{\tau/2}\sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right).$$

Techniques to calculate $V(\hat{y} - y)$

1. $V(\hat{y} - y) = V(a + bx - \alpha - \beta x - e) = V(a + bx - e) = V(a + bx) + V(e) - 2C(a + bx, e) = V(\hat{\mu}) + V(e)$.
2. $V(\hat{\mu}) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right)$ and $V(e) = \sigma^2$.

Prediction Inference (II)

The case of σ^2 is unknown:

1. An unbiased and consistent point estimator for σ^2 is $s^2 = \frac{SSE}{n-2} = \frac{1}{n-2}\sum_{i=1}^n \hat{e}_i^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - a - bx_i)^2$.
2. $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$.
3. s^2 is independent of both a and b.
4. $\frac{a-\alpha}{\sqrt{\hat{V}a}} \sim t(n-2)$, $\frac{b-\beta}{\sqrt{\hat{V}b}} \sim t(n-2)$, $\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{V}\hat{\lambda}}} \sim t(n-2)$.
5. an exact $1-\tau$ prediction interval (PI) for y is
$\left(\hat{y} \pm t_{\tau/2}(n-2)\sqrt{V(\hat{y}-y)}\right) =$
$\left(a + bx \pm t_{\tau/2}(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$.

Remark

$\hat{V}a$, $\hat{V}b$ and $\hat{V}\hat{\lambda}$ are Va, Vb and $V\hat{\lambda}$ with σ^2 replaced by s^2.

Techniques for $\mathsf{E}s^2 = \sigma^2$

1. $\mathbb{E}\hat{e}_i = \mathbb{E}y_i - \mathbb{E}a - x_i\mathbb{E}b = (\alpha + \beta x_i) - \alpha - \beta x_i = 0.$
2. $\mathbb{E}\hat{e}_i^2 = V(\hat{e}_i) =$
$V(y_i) + V(a) + x_i^2 V(b) - 2C(y_i, a) - 2x_i C(y_i, b) + 2x_i C(a, b).$

Prediction Inference (III)

Under the assumptions of

1. the sample from a general distribution (instead of normal distribution) and
2. the sample size is large,

we have statistical inference for SLR from CLT

1. if σ is known, the result is the same as Prediction Inference (I);
2. if σ is unknown, the result is the same as Prediction Inference (II) with $t_{\tau/2}(n-2)$ and T_{n-2} replaced by $z_{\tau/2}$ and Z respectively.

SLR and Correlation Analysis

Simple linear regression is

1. to explore the relation between a r.v. y and a non-random variable x; and
2. the relation is reflected by $b = \frac{S_{xy}}{S_{xx}}$ which is an estimator of β.

Correlation analysis is

1. to study the relation between two r.v.'s y and x; and
2. the relation is reflected by $r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$ which is an estimator of the correlation $\rho = \frac{C(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$.

The relation between them is

1. $r = b\sqrt{\frac{S_{xx}}{S_{yy}}}$;
2. the coefficient of determination $r^2 = \frac{S_{yy} - SSE}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$.

Summary

1. Understanding simple linear regression models.
2. How to make statistical inference on unknown parameters in SLR model.
3. Prediction with SLR model.
4. Compare between statistical inferences on predictions and estimations.

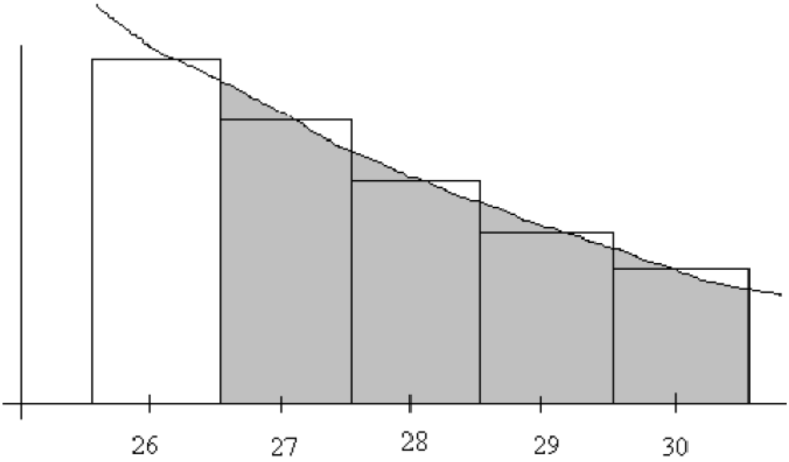
Appendix 1: Continuity Correction

Example: A die is rolled $n = 120$ times. Find the probability that at least 27 sixes come up.

Analysis:

1. $Y \sim Bin(120, 1/6)$.
2. $Bin(n, p) \dot{\sim} N(np, np(1-p))$.
3. $P(Y \geq 27) \approx P(U \geq 27)$.
4. $P(Y \geq 27) = \sum_{y=27}^{120} \begin{pmatrix} 120 \\ y \end{pmatrix} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{120-y} = 0.0597$.
5. $P(U \geq 27) = P\left(Z \geq \frac{27-20}{\sqrt{16.667}}\right) = 0.0436$.
6. $P(U \geq 27 - 0.5) = P\left(Z \geq \frac{27-0.5-20}{\sqrt{16.667}}\right) = P(Z \geq 1.59) = 0.0559$.

Graphs Comparison



Appendix 2: Buffon's needle problem

Problem: A kitchen floor has a pattern of parallel lines that are 10 cm apart. You have a needle in your hand that is also 10 cm long. If you randomly throw the needle onto the floor, what is the probability p that it will cross a line?

Analysis: Monte Carlo method

1. Throw the needle on the floor $n = 1000$ times and find that the needle crosses a line 651 times.
2. An estimator for p is $\hat{p} = \frac{651}{1000} = 0.651$.
3. A 95% CI for p is $\left(0.651 \pm 1.96\sqrt{0.651(1-0.651)/1000}\right) = (0.621, 0.681)$.

Analytical Method of finding p

Analysis:

1. X: perpendicular distance from centre of needle to nearest line in units of 5 cm. Y: acute angle between lines and needle in radians. A: needle crosses a line.
2. $X \sim U(0,1)$, $Y \sim U(0,\pi/2)$, $X \perp Y$.
3. $f(x) = 1, 0 < x < 1$, $f(y) = 2/\pi, 0 < y < \pi/2$,
   $f(x,y) = f(x)f(y) = \frac{2}{\pi}, 0 < x < 1, 0 < y < \pi/2$.
4. $A = \{(x,y) : x < \sin(y)\}$.
5. $p = P(A) = \int\int_A f(x,y)dxdy = \frac{2}{\pi}\int_{y=0}^{\pi/2}\int_{x=0}^{\sin(y)} dxdy = \frac{2}{\pi}$.

Graph