Business Analytics

Lecture 9

Revision

Dr Yufei Huang

Lecture 1 Understanding the Data

Measures of central tendency

mean	$m = \bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$
median	1. Sort the data according to its order 2. Odd number of points: middle point 3. Even number of points: mean of the two middle values
mode	Most frequent value There can be more than one modes

Measures of dispersion/spread

range	largest value - smallest value
variance	$s^2 = \frac{1}{N-1}\left[(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_N - m)^2\right] = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2$
standard deviation	$sd = s = \sqrt{\frac{1}{N-1}\left[(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_N - m)^2\right]}$

UCL SCHOOL OF MANAGEMENT

Measures of risk

historical volatility	standard deviation
Return-to-risk ratio	Return / sd Return is a profit of an investment
coefficient of variation	sd / mean CV is a dimensionless measure of variation

Lecture 2 Probability

Definitions

- An experiment is any process or procedure for which more than one outcome is possible.
- Sample Space is the set of all possible outcomes of an experiment.
- An event is a subset of a sample space.

    Example: Flip a coin twice

    Sample space S = {HH, HT, TH, TT}

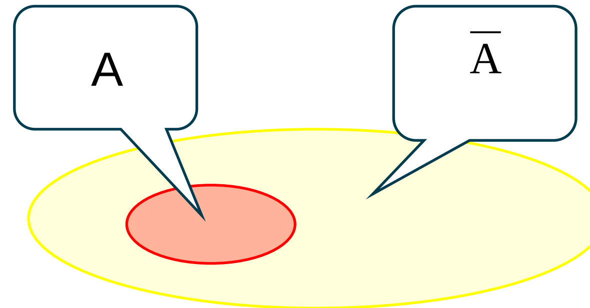
    An event E = {HH, HT}, i.e., first flip results in a Head

- Probability measure is a function P from the set of all of the events of a sample space Ω to [0,1], which satisfies for all disjoint events A_i, i.e., A_1, $A_2, \ldots, A_n$:

1. $P(A_i) \geq P(\Phi) = 0$
2. $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$.
3. $P(\Omega) = 1$

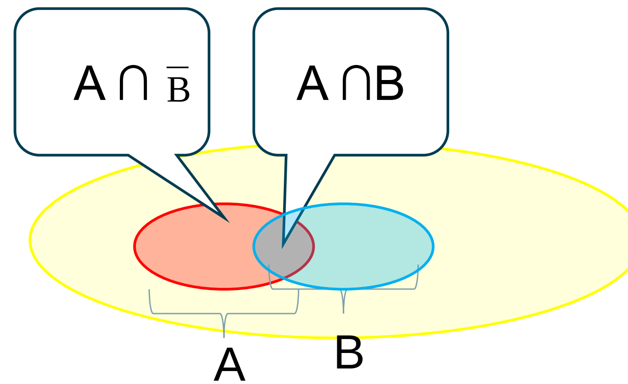
Probability rules

- Theorem. $P(A)+P(\bar{A})=1$.
- Theorem. $P(A \cap B)+P(A \cap \bar{B})=P(A)$.
- Theorem. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



- Definition. Conditional probability is the probability of the occurrence of an event A, given information about the occurrence of an event B.

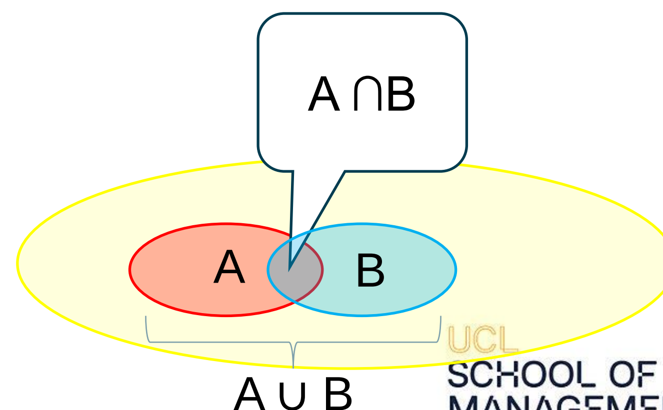
  If $P(B) \neq 0$ then $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



- Bayes' Theorem.

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})}$$

- Definition. Two events, A and B are called independent if $P(A|B)=P(A)$.



Lecture 3 Random Variables

Probability Distribution

- Definition. A probability distribution for a discrete random variable is a mutually exclusive listing of all possible numerical outcomes for that variable such that a particular probability of occurrence is associated with each outcome.

- If x is a possible outcome of a discrete random variable X, we can denote the probability of x, by P(x) or P(X=x).

- The probability distribution of a discrete random variable X must satisfy the following two conditions.

  1. $P(x) \geq 0$ for all possible values x of a random variable X
  2. $\sum_{\text{all } x} P(x) = 1$

- We use $F(x)=P(X\leq x)$ to denote the probability of all outcomes from X that are smaller or equal to x. F(x) is called the cumulative distribution function.

Expectation, variance and standard deviation of a random variable

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i).$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{N} (X_i - \mu)^2 P(X_i).$$

$$\sigma^2 = E(X^2) - [E(X)]^2$$

$$\sigma = SD(X) = \sqrt{V(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{\sum_{i=1}^{N} (X_i - \mu)^2 P(X_i)}$$

$$= \sqrt{E(X^2) - (E(X))^2}$$

UCL
SCHOOL OF
MANAGEMENT

Binomial Distribution

- A Bernoulli variable with parameter p is a random variable which has only two possible outcomes
- A binomial random variable is a random variable that counts the number of successes in many independent, identical Bernoulli trials.

Theorem. Let X~B(n,p) be a binomial random variable. Then, for x=0,1,2,…n:

$$1.\ P(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x},$$

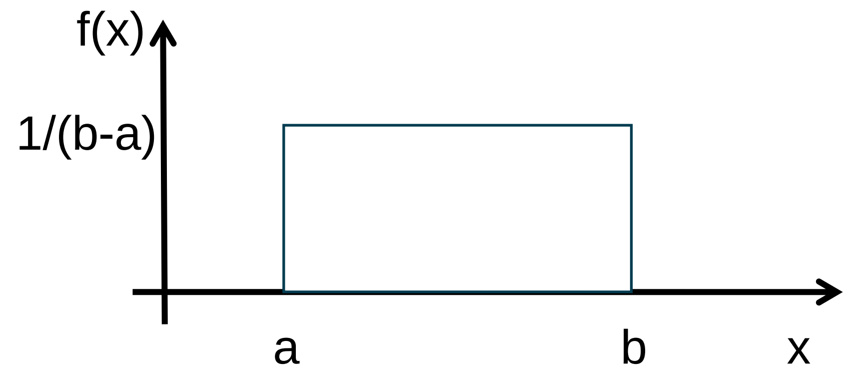
$$2.\ E(X) = np,$$

$$3.\ V(X) = np(1-p),$$

Uniform Distribution

- The probability density function of the uniform distribution:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{Otherwise} \end{cases}$$



- Notation. If X has a uniform distribution, then we write: X~U(a,b).
- Theorem. If X~U(a,b) and a ≤ c ≤ d ≤ b then:

1. P(c ≤ x ≤ d) = (d-c) / (b-a)
2. E(X) = (a+b) / 2
3. V(X) = $(b-a)^2$ / 12

Lecture 4 Normal Distribution

Key Point 1:

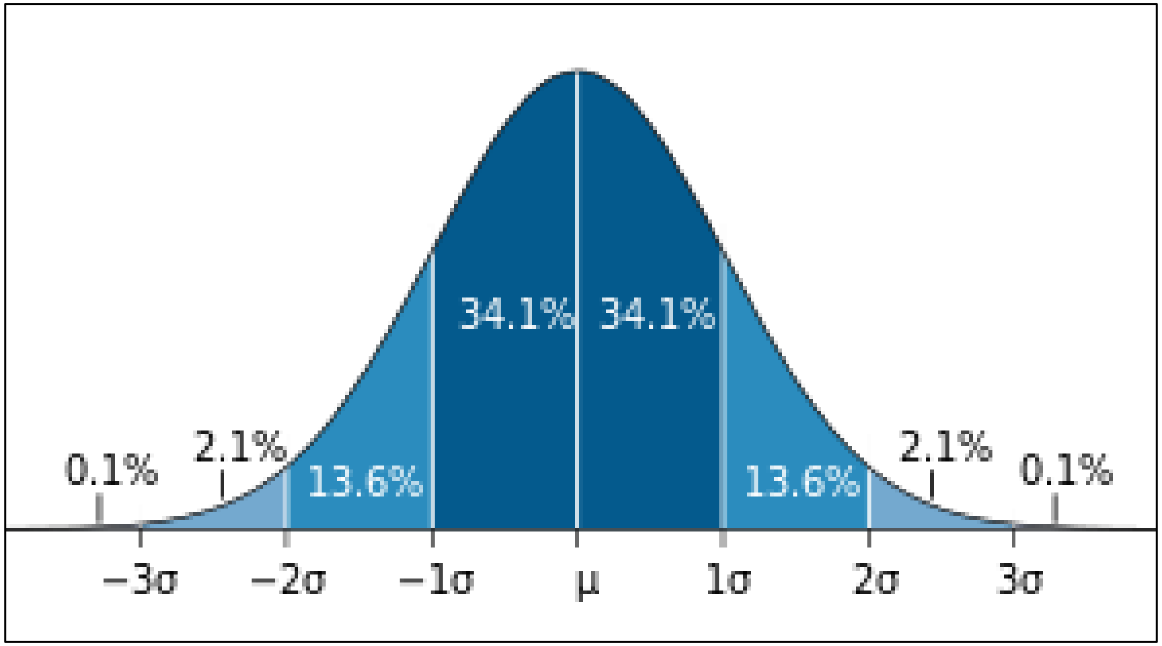
- The function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

is the probability density function of a normal distribution with **mean μ** and **standard deviation σ**.

Key Point 2:

- The total area under the density curve = 1
- 68.26% of the area under the curve is between **μ-σ, μ+σ**,
- 95.44% of the area under the curve is between **μ-2σ, μ+2σ**,
- 99.72% of the area under the curve is between **μ-3σ, μ+3σ**.



Key Point 3:

Theorem. Let X_1, X_2,…, X_n be normally distributed random variables, which are independent. Then $Q=a_1X_1+a_2X_2+\ldots+a_nX_n+b$ is also normally distributed with

$$E(Q)=a_1E(X_1)+a_2E(X_2)+\ldots+a_nE(X_n)+b$$

and

$$V(Q)=a_1^2V(X_1)+a_2^2V(X_2)+\ldots+a_n^2V(X_n).$$

Key Point 4:

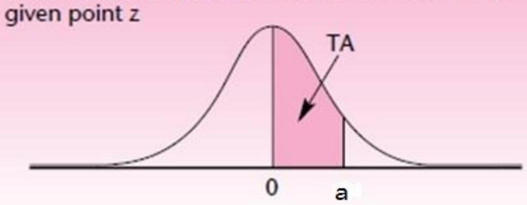
- Definition. The standard normal random variable Z is the normal random variable with **mean 0** and **standard deviation 1**. That is:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Notation**. Z~N(0,1²)**
- How to use the table of P(0<Z<a) in different cases, assume a>0

1. P(0<Z<a) is the value written in the table.
2. P(-a<Z<0)=P(0<Z<a)
3. P(Z<0)=P(-∞<Z<0)= P(0<Z< ∞)=P(Z>0)=0.5
4. P(Z<a)=P(-∞<Z<0)+P(0<Z<a)=0.5+P(0<Z<a).
5. P(Z>a)=1-P(Z<a)

The area given in the standard normal probability table is the area under the curve between 0 and a given point z



The inverse approach for Point 4.

- Find z such that the P(0<Z<z) is 0.3315.

- Find z such that the P(0<Z<z) is 0.2224.

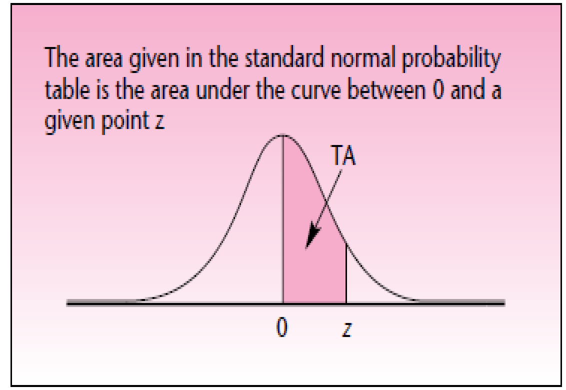


TABLE 4–1 Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389

Key Point 5.

The transformation of a general normally distributed random variable to the standard normal distribution, and back:

$X = \mu + \sigma Z \rightarrow Z = (X - \mu)/\sigma$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

Lecture 5 Sampling

The Central Limit Theorem

- Assume that the mean of a certain property of a population is μ, and that its standard deviation is σ.

- Then the distribution of the sample mean $\overline{X}$ tends to a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, as the sample size n becomes large.

- That is: for large n (n≥30):

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

Confidence Intervals

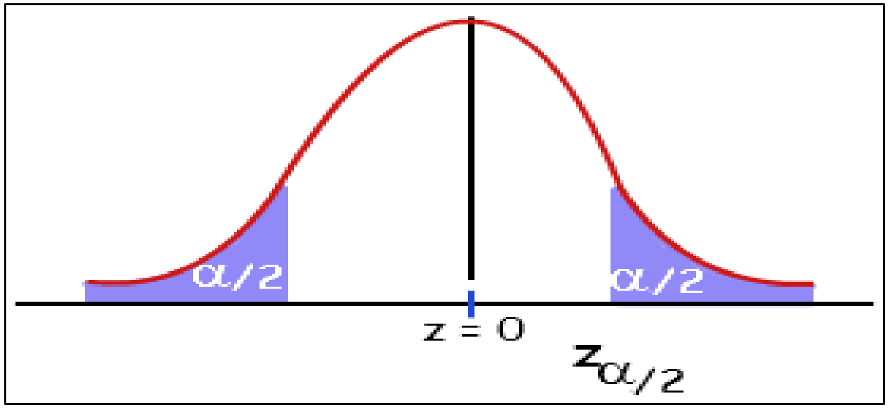
- Definition. A confidence interval is a range of numbers, which is believed to include an unknown population parameter with a certain probability.

- A 95% confidence interval for μ when the standard deviation of the population is known:

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

(1-α)·100% Confidence interval

Notation. We denote by $z_{\alpha/2}$ the z value that cuts off a right-tail area of α/2 under the standard normal curve.



- A (1-α)·100% confidence interval for μ when σ is known and sampling is done from a normal population, or with a large sample, is

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

UCL
SCHOOL OF
MANAGEMENT

Confidence intervals when the standard deviation of the population is unknown

- A $(1\text{-}\alpha)\cdot 100\%$ confidence interval for μ when σ is not known (assuming a normally distributed population) is

$$\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right],$$

where $t_{\alpha/2}$ is the value of the t distribution with n-1 degrees of freedom that cuts off a tail area of $\alpha/2$ to its right.

Sampling Distribution of Sample Proportion

p population proportion
$\hat{p}$ sample proportion
n sample size

Expected value of $\hat{p}$ is $\mathrm{E}[\hat{p}] = p$

Standard deviation of $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$

Sample Proportion

- Suppose np ≥ 5 and n(1 – p) ≥ 5.

- The sampling distribution of p approaches a Normal distribution with mean p and standard deviation $\sqrt{p(1-p)/n}$ as the sample size becomes large.

- 95% confidence interval for population proportion

$$\hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

- If p in unknown, then 95% confidence interval is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Lecture 6 Hypothesis Testing

UCL
SCHOOL OF
MANAGEMENT

Four Steps of hypothesis testing

- Step 1. Formulate the hypothesis
  - Null Hypothesis (H_0)
  - Alternate Hypothesis (H_1)

- Step 2. Set the criteria for a decision

- Step 3. Acquire an objective test statistic (e.g. from evidence)

- Step 4. Does the objective test statistic represent an overwhelming evidence against the Null Hypothesis? (i.e., p-value $< \alpha$? or equivalently, compare the test statistic with the critical z-value or t-value)
  - If yes, reject the null (H_0) and accept the alternative (H_1) as the truth.
  - If no, **cannot reject** the null (H_0) .

UCL SCHOOL OF MANAGEMENT

P-value & Significance

- P-value: probability of observing the statistic (<u>or more extreme</u>) given that the null hypothesis is true.
  - Must compute
- Significance level α: a threshold probability (e.g. 0.05, or 0.1) that determines whether or not the evidence is overwhelming.
  - Typically given
- If the p-value is less than the significance level α, then you can reject the null hypothesis
- How do we find the p-value? Use the sample distribution (normal)!

T-Statistics

- Instead of comparing p-value with significance level α, we can also compare the test statistics with z-value (or t-value, if use t-distribution)
- We compare the test-statistic with the z-value, z_α, which corresponds to the required significance level α. Criteria depend on the type of test.
- Then, we decide to accept H_0 or not.

One-tailed and two tailed tests

Right tailed	Left tailed	Two-tailed
H_0: μ≤M, H_1: μ>M statistic=$(\bar{X}$ -M)/(s/√n) P-value=P(Z>statistic) 1. If P<α, accept H_1 2. If statistic>Z_α, accept H_1	H_0: μ≥M, H_1: μ<M statistic=$(\bar{X}$-M)/(s/√n) P-value= P(Z<statistic) 1.If P<α, accept H_1 2.If statistic<Z_α, accept H_1	H_0: μ=M, H_1: μ≠M statistic=$(\bar{X}$-M)/(s/√n) If the statistic is positive: P-value=2*P(Z>statistic) If the statistic is negative: P-value=2*P(Z<statistic) 1.If P<α, accept H_1 2.If |statistic|>$Z_{\alpha/2}$, accept H_1 3.If M is outside the (1-α) level confidence interval for μ, accept H_1

Lecture 7 Simple Linear Regression

The Simple Linear Regression Model

The population simple linear regression model:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\text{Nonrandom or Systematic Component}} + \underbrace{\varepsilon}_{\text{Random Component}}$$

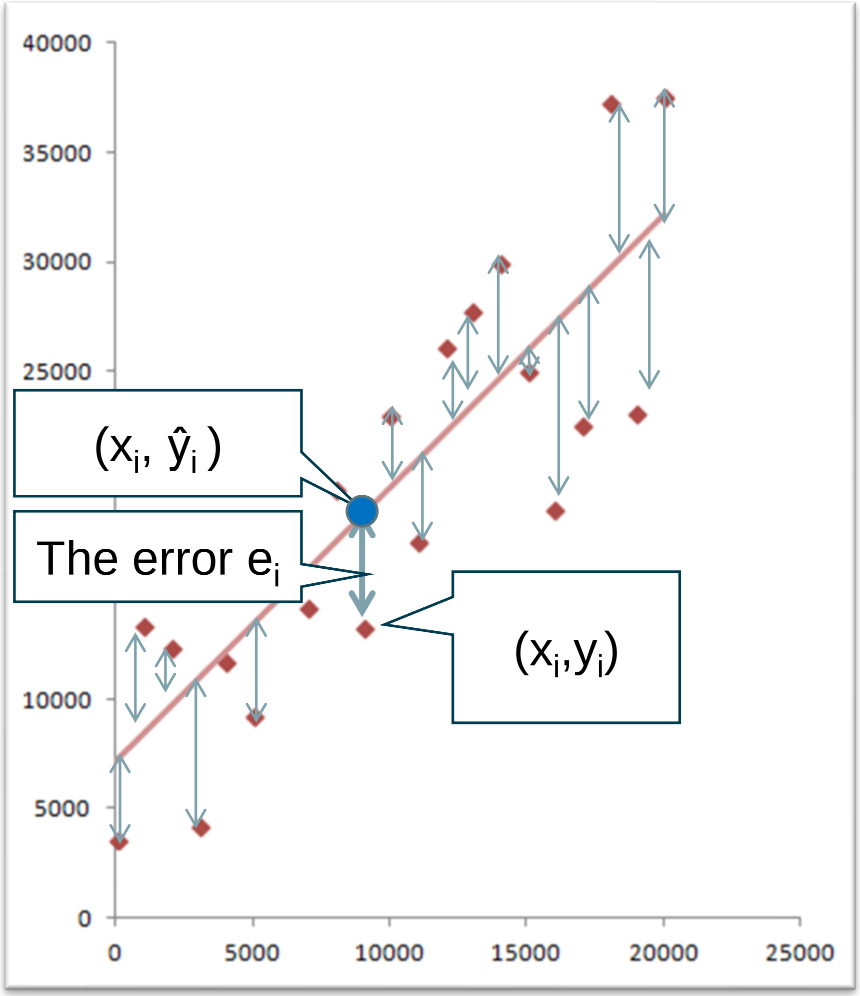
where

- Y is the dependent variable, the variable we wish to explain or predict
- X is the independent variable, also called the predictor variable
- ε is the error term, the only random component in the model, and thus, the only source of randomness in Y

- β_0 is the intercept of the systematic component of the regression relationship
- β_1 is the slope of the systematic component

The method of least squares

- To find coefficients b_0, b_1,
- we denote each data point by (x_i,y_i).
- The line gives us an approximated value: $\hat{y}_i = b_0 + b_1 x_i$.
- The approximation error of each point is $e_i = |y_i - \hat{y}_i|$.
- The sum of squares for errors in regression is:



$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

UCL SCHOOL OF MANAGEMENT

To find b_0, b_1, which minimise SSE

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

Theorem. The following b_0 and b_1 minimise SSE:

(Least Squares Estimator)

$$\boxed{\begin{aligned} b_1 &= \frac{SS_{xy}}{SS_x}, \\ b_0 &= \bar{y} - b_1 \bar{x}, \end{aligned}}$$

where $\bar{x} = \text{mean}(X), \bar{y} = \text{mean}(Y)$

$$SS_x = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right).$$

Hypothesis test for a linear relationship

Hypotheses:

H_0: $b_1=0$

H_1: $b_1 \neq 0$.

The test statistic for the existence of a linear relationship between X and Y can be calculated in Excel.

UCL SCHOOL OF MANAGEMENT

Summary

- Regression is useful in testing the relationship between two variables and in forecasting. Excel can generate the regression results.

- How to interpret them:

1. Write the equation of the estimated line
   - Sales = b_0 + b_1 *(student population) $+ \; \varepsilon$

2. Is the coefficient, b_1, significant? Check,
   - p-value < 0.05?
   - t-stats > Z-value from normal distribution (or t-value from t-distribution)
   - does the 95% interval for the coefficient contain 0?

3. What is the point forecast for the mean and the 95% prediction interval?

$$\boxed{\hat{y} \pm 1.96 \text{ standard error of the estimate}}$$

4. How good is the fit? Look at the R^2.

UCL SCHOOL OF MANAGEMENT

Lecture 8 Multidimensional and Non-Linear Regression

Nonlinear Regression

- A nonlinear relationship may be a better model than a linear relationship.
- A widely used regression for nonlinear relationship is multiplicative regression

$$\text{The multiplicative model:}$$
$$Y = \beta_0 X^{\beta_1} \varepsilon$$

$$\text{The logarithmic transformation:}$$
$$\ln Y = \ln \beta_0 + \beta_1 \ln X + \ln \varepsilon$$

Interpreting Multiplicative Models

- $y = b_0 + b_1 LN(x_1) + \varepsilon$ Model (1)
  - If x_1 increases by 1%, then y increases by approximately $0.01\ b_1$ units.

- $LN(y) = b_0 + b_1x_1 + \varepsilon$ Model (2)
  - If x_1 increases by 1 unit, then y increases by approximately $100\ b_1$%.

- $LN(y) = b_0 + b_1 LN(x_1) + \varepsilon$ Model (3)
  - If x_1 increases by 1%, then y increases by approximately b_1%.

- Interpretation of the coefficient b_1 is of managerial use. For example, if y is **sales** or demand and x_1 is **price** then in Model 3, the coefficient b_1 measures the elasticity of sales with respect to price. That is, in Model 3, a 1% change in price leads to approximately b_1% change in sales.

Multidimensional Regression

- Where $X_1,\ldots,X_p$ are p independent variables and $b_0,\ldots,b_p$ are the coefficients obtained by the Least Squares Method.

$$Y = b_0 + b_1 X_1 + \ldots + b_p X_p + \varepsilon$$

- Interpretation of b_i: The magnitude of b_i represents an estimate of the change in Y corresponding to a one unit change in X_i when all other independent variables are held constant.

Summary

- **Nonlinear Models (e.g. multiplicative models)**
  - Linear regression can be applied to capture nonlinear relationships using the multiplicative models.
  - "Ln" denotes the percentage change

- **Multidimensional Regression**
  - You can add more independent variables to improve the understanding of different factors that influence the outcome.
  - Significance for overall model: "significance of F"
  - Goodness of fit: R^2 , more variable usually leads to higher R^2
  - But watch out for multicollinearity