EDGE Excellence Demands Great Education

CHARTERED FINANCIAL ANALYST - LEVEL II

A PROFESSIONAL'S FIRST CHOICE IN EXAM TRAINING & PREPARATION

SAMPLE FORMULA BOOK

Business Village, Block B, Office 926, UAE
Tel: +97143688966 , Email: info@edge-education.com
Web: www.edge-education.com

STUDY SESSION 3

QUANTITATIVE METHODS

I FINTECH

Term	Description
Activation Function	A formula, typically non-linear, that is applied at nodes of neural networks to the sum of the weighted values received by the node.
Alternative Data	Non-traditional data types generated by the use of electronic devices, social media, satellite and sensor networks, and company exhaust.
Artificial Intelligence	Computer systems that have thinking and decision-making ability comparable (or superior) to that of humans
Big Data	The vast amount of data being generated by industry, governments, individuals, and electronic devices that arises from both traditional and non-traditional data sources.
Bitcoin	A cryptocurrency using blockchain technology that was created in 2009
Blockchain	A type of digital ledger in which information is recorded sequentially and then linked together and secured using cryptographic methods.
Causal Inference	Inference that focuses on establishing that a change in one variable causes a change in another variable.
Classification	Data analytics that focuses on sorting observations into distinct categories.
Clustering	Data analytics that focuses on sorting observations into groups (clusters) such that observations in a cluster are more similar to each other than they are to observations in other clusters.
Composite Variable	A variable that combines two or more variables that are statistically and strongly related to each other.
Cryptocurrency	An electronic medium of exchange that lacks physical form.
Cryptography	An algorithmic process to encrypt data, making the data unusable if received by unauthorized parties.
Data Science	An interdisciplinary field that brings computer science, statistics, and other disciplines together to analyze and produce insights from Big Data
Data Visualization	Data visualization refers to how the data will be formatted, displayed, and summarized in graphical form
Dimension Reduction	Data analytics that focuses on reducing the number of independent variables while retaining the variation across observations (to preserve the information contained in that variation).
Deep Learning (Deep Learning Nets)	Machine learning using neural networks with many hidden layers.
Distributed Ledger	A type of database that may be shared among entities in a network.

Term	Description
Distributed Ledger Technology	Technology based on a distributed ledger
Ensemble Learning	A class of techniques that incorporate the output of a collection of models (such as random forests) to produce classifications that have better signal-to-noise ratios than the individual classifiers.
Features	The name used in machine learning contexts for what are known as independent variables in regression analysis.
Feature Engineering	The curation of a dataset of features (independent variables) for machine learning processing.
FinTech	Technological innovation in the design and delivery of financial services and products in the financial industry.
High Frequency Trading	A form of algorithmic trading that makes use of vast quantities of data to execute trades on ultra-high-speed networks in fractions of a second.
Hyperparameters	Parameters (constants) that the human researcher determines before the learning process begins.
Initial coin offering	An unregulated process whereby companies raise capital by selling crypto tokens to investors in exchange for fiat money or another agreed-upon cryptocurrency
Internet of Things	A network arrangement of structures and devices whereby the objects on the network are able to interact and share information.
Machine Learning	Diverse approaches by which computers are programmed to improve performance in specified tasks with experience.
Natural Language Processing	Computer programs developed to analyze and interpret human language.
Neural Networks	Computer programs based on how our own brains learn and process information.
Overfitting	An undesirable result from fitting a model so closely to a dataset that it does not perform well on new data.
Permissioned networks	Networks that are fully open only to select participants on a DLT network.
Permissionless networks	Networks that are fully open to any user on a DLT network.
Regularization	Any method that tamps down statistical variability in high-dimensional estimation or prediction problems.
Reinforcement Learning	Machine learning approach in which a computer learns from interacting with itself (or data generated by the same algorithm).
Smart Contracts	A computer program that is designed to self-execute on the basis of pre-specified terms and conditions agreed to by parties to a contract.

Term	Description
Supervised Learning	A machine learning approach that makes use of labeled training data.
Target Variable (Tag variable)	The name used in machine learning contexts for what is known as the dependent variable in regression analysis
Text Analytics	The use of computer programs to analyze and derive meaning from typically large, unstructured text- or voice-based datasets.
Training Sample	A sample of correctly labeled target data used to train or fit the algorithm.
Tokenization	The process of representing ownership rights to physical assets on a blockchain or distributed ledger
Unsupervised Learning	A machine learning approach that does not make use of labeled training data.
Validation Sample	A data subset that is used to evaluate how well the model works out of sample.

II CORRELATION AND REGRESSION

Sample covariance

$$\text{Sample covariance} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n - 1}$$

OR

$$\text{Sample covariance} = \text{Cov}(X, Y) = E[(X_i - E(X))(Y_i - E(Y))]$$

Where: n = Sample size
$\overline{X}$ = mean observation of Variable X
$\overline{Y}$ = mean observation of Variable Y
X_i = *i*th observation of Variable X
Y_i = *i*th observation of Variable Y

Sample correlation coefficient

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

Population Variance

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

Sample Variance

$$\text{Sample variance} = s_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$

Population Standard Deviation

$$\sigma_X = \sqrt{\sigma_X^2}$$

Sample standard deviation

$$\text{Sample Standard Deviation} = s_X = \sqrt{s_X^2} = \sqrt{\text{Sample Variance}}$$

Hypothesis Testing – SIX STEPS

(1) State the hypothesis i.e. Null (H_0) Alternative (H_a)

The null always has the 'equal to' sign. The alternative hypothesis is the researcher's 'claim'. We always perform hypothesis test on the Null and determine whether the ***Null*** is rejected or not.

The hypothesis can be one tailed and two-tailed.

Testing Slope coefficient significance

Two tailed test	**One tailed test**		
$H_0: b_1 = 0$	$H_0: b_1 \geq 0$		$H_0: b_1 \leq 0$
$H_a: b_1 \neq 0$	$H_a: b_1 < 0$	OR	$H_a: b_1 > 0$

'Global' or 'Joint' test for all slopes

Two tailed test

$H_0: b_1 = b_2 = b_i = \cdots = 0$

$H_a: atleast\ one\ slope \neq 0$

(2) Select the appropriate test statistic

Select from 'z, t, F, DW'

z and t are used for hypothesis test on mean and each slope coefficient in regression

F is used to test all slope coefficients 'jointly'

Durban Watson (DW) is used to test serial correlation in errors in simple or multiple regression

(3) Specify the level of significance

$\alpha = 1 - \text{confidence level}$

If the hypothesis is two tailed, divide the alpha by half

If the hypothesis is one tailed, do not divide alpha

If the hypothesis test is for F statistic, do not divide alpha (even if it is two tailed test)

(4) State the decision rule regarding the hypothesis

RULE 1 : Reject H_0 if : test statistic calc. > upper critical value OR
test statistic calc. < lower critical value

RULE 2: Construct confidence interval around the sample value.

Reject H_0 if : If the hypothesized value falls *outside* the acceptance range.

$$\text{CI} = \text{Sample statistic} \pm (\text{Reliability factor})(\text{Std. Error of sample})$$

RULE 3: Compare the p-value with the significance level.

Reject H_0 if: p-value < α (one-tailed) OR p-value < $\frac{\alpha}{2}$ (two-tailed)

(5) Collect the sample and calculate the sample statistic

$$\text{Test Statistic}_{calc} = \frac{\text{Sample statistic} - \text{Hypothesized value } H_0}{\text{Std. error of sample statistic}}$$ ← This is used for 'z' or 't' test on regression slope coefficients i.e. intercept and slope

$$t_{correlation} = \frac{r\sqrt{n-2} - \text{hypothesis value}}{\sqrt{1-r^2}}$$ ← For a test on correlation between two variables

$$F_{calc} = t^2_{calc}$$ ← For a simple regression to test single coefficient on one independent variable.

$$F_{calc} = \frac{MSR}{MSE} = \frac{RSS/k}{SSE/n-k-1}$$ ← For multiple regression to test all slope coefficients simultaneously on all independent variables

Critical values will come from respective z, t, or F table. look up the critical values, you will need:

Confidence Level	$Z_{\alpha/2\%}$ Two tailed	$Z_{\alpha\%}$ One-tailed
68%	± 1	±0.465
90%	± 1.645	±1.28
95%	± 1.96	±1.645
99%	± 2.58	±2.33

1. Use significance level (α *or* $\alpha/2$) to select values from z table.
2. Use the significance level (α *or* $\alpha/2$) and degrees of freedom (n – 1) for the t table when doing hypothesis test for mean.
3. Use the significance level (α *or* $\alpha/2$) and degrees of freedom (n – 2) for the t table when doing hypothesis test for correlation.
4. Use the significance level (α *or* $\alpha/2$) and degrees of freedom or (n – k – 1) for the t table when doing hypothesis test for regression coefficients.
5. Use the significance level (α) and two sets of degrees of freedom k and $n - k - 1$ for the F test when performing joint hypothesis test of regression coefficients.

(6) Make a decision regarding the hypothesis
Either reject the null (i.e. accept alternative) Or fail to reject the null (i.e. accept the null)

(7) Make a decision based on the results of the test
Interpret the decision in an economic context. i.e. was the hypothesis claim true or false?

Statistical Errors

Type I error (α) – The probability of rejecting a TRUE null.
Type II (β) – The probability of not rejecting a FALSE null.

	True Condition	
Decision	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error β = P(Type II error)
Reject H_0	Type I error α = P(Type I error)	Correct decision Power of test = 1- β

Correlation test statistic

$$t_{correlation} = \frac{r\sqrt{n-2} - \text{hypothesized value}}{\sqrt{1-r^2}}$$

Note: degrees of freedom = n – k – 1

Where: n = Number of observations r = Sample correlation
k = number of independent variables. Hypothesized value = value of null usually is 0

Test Statistic Calculated

$$z_{calc} \text{ or } t_{calc} = \frac{\text{Observed value} - \text{Hypothised value}}{\text{Standard error}}$$

Linear Regression with One Independent Variable

$$\text{Regression model equation} = Y_i = b_0 + b_1 X_i + \varepsilon_i$$

When performing the regression:
b_1 and b_0 = regression coefficients. b_1 = slope coefficient b_0 = intercept term
ε = error term that represents the variation in dependent variable not explained by independent variable.