Institute of Statistical Research and Training
University of Dhaka

Syllabus

PhD Program in Applied Statistics
Session : 2018–2019

www.isrt.ac.bd/academics/graduate

PhD Program in Applied Statistics

The Doctor of Philosophy (PhD) program in Applied Statistics is a FOUR academic year program. According to the Dhaka University regulations for PhD admission in Applied Statistics, candidates who do not have either four years B.Sc./B.S. honours and one year M.Sc./M.S. degree from the University of Dhaka or MPhil equivalent degree (approved by the Dhaka University MPhil equivalent committee) are required to take course work as part of his/her PhD degree. **A candidate graduated in Applied Statistics/Statistics is eligible for admission into the program**. The syllabus is designed for these candidates as follows. The candidates are required to take a total of 9 credit hours of theoretical courses from two groups of courses: A and B. The group A consists of courses related to basic topics and the group B consists of courses related to advanced topics in Applied Statistics. Courses from both groups are of 3 credit hours each. Candidates are required to take THREE courses to make a total of 9 credit hours for theoretical courses, preferably maximum TWO courses from each group. The choice of courses will depend on the availability of teaching faculties of the institute. In addition, there will be a 3 credit hours oral comprehensive course, altogether a total of 12 credit hours.

Breakdown of the credit hours

Courses	Credit Hour
Theoretical Courses	9
Oral	3
Total	12

The marks allocation for courses will be as follows:

Theoretical		
Attendance	:	05
In-course exam	:	25
Final exam	:	70

There will be two in-course examinations for each of the courses. Candidate should be present at least 60% of the classes to attend the final exam. The Dhaka University grading policy will be followed for finalizing the result. Candidate should get minimum GPA 2.5 for theoretical courses and oral course, separately. However, no 'F' grade in any theoretical courses will be acceptable. The candidate must have to sit for the examination for the course(s) for which the candidates got 'F' grade.

Courses in Group A

Course ID	Course Title	Credit Hour
AST 701	Applied Bayesian Statistics	3
AST 702	Advanced Classical Inference	3
AST 703	Advanced Multivariate Techniques	3
AST 704	Generalized Linear Models	3

Courses in Group B

Course ID	Course Title	Credit Hour
AST 710	Advanced Survival Analysis	3
AST 711	Environmental and Spatial Statistics	3
AST 712	Advanced Time Series Analysis	3
AST 713	Actuarial Techniques	3
AST 714	Advanced Operations Research	3
AST 715	Advanced Econometric Methods	3
AST 716	Advanced Population Studies	3
AST 717	Queueing Theory and Stochastic Processes	3
AST 718	Advanced Epidemiology	3
AST 719	Analysis of Longitudinal Data	3
AST 720	Adaptive Sampling	3
AST 721	Optimum Experimental Designs	3
AST 722	Statistical Signal Processing	3
AST 723	Meta Analysis	3
AST 724	Clinical Trials	3
AST 725	Statistical Machine Learning	3

Viva

Course ID	Course Title	Credit Hour
AST 740	Oral	3

Detailed Syllabus

AST 701: Applied Bayesian Statistics | **Credit 4**

Bayesian thinking: background, benefits and implementations; Bayes theorem, components of Bayes theorem - likelihood, prior and posterior; informative and non-informative priors; proper and improper priors; discrete priors; conjugate priors; semi-conjugate priors; exponential families and conjugate priors; credible interval; Bayesian hypothesis testing; building a predictive model.

Bayesian inference and prediction: single parameter models - binomial model, Poisson model, normal with known variance, normal with known mean; multi-parameter models - concepts of nuisance parameters, normal model with a non-informative, conjugate, and semi-conjugate priors, multinomial model with Dirichlet prior, multivariate normal model; posterior inference for arbitrary functions; methods of prior specification; method of evaluating Bayes estimator.

Summarizing posterior distributions: introduction; approximate methods: numerical integration method, Bayesian central limit theorem; simulation method: direct sampling and rejection sampling, importance sampling; Markov Chain Monte Carlo (MCMC) methods - Gibbs sampler, general properties of the Gibbs sampler, Metropolis algorithm, Metropolis-Hastings (MH) sampling, relationship between Gibbs and MH sampling, MCMC diagnostics - assessing convergence, acceptance rates of the MH algorithm, autocorrelation; evaluating fitted model - sampling from predictive distributions, posterior predictive model checking.

Linear model: introduction, classical and Bayesian inference and prediction in the linear models, hierarchical linear models - Bayesian inference and prediction, empirical Bayes estimation; generalized linear model - Bayesian inference and prediction (logit model, probit model, count data model).

Text Books

1. Hoff PD (2009). A First Course in Bayesian Statistical Methods. Springer.
2. Gelman A, Carlin JB and Stern HS, Dunson DB, Vehtari A, and Rubin DB (2013). Bayesian Data Analysis, 3^{rd}*edition*. Chapman and Hall.
3. Gill J (2007). Bayesian Methods: A Social and Behavioral Sciences Approach, 2^{nd}*edition*. Chapman and Hall.

AST 702: Advanced Classical Inference | **Credit 3**

Statistical inference: Parametric, nonparametric and semiparametric inference.

Approximate and computationally intensive methods for statistical inference: The general problem of inference; likelihood functions; maximum likelihood estimation; optimization techniques- Newton type methods; EM algorithm- simple form, properties, uses in analysing missing data, fitting mixture models and latent variable model; restricted maximum likelihood (REML) method of estimation; Multi-stage maximization; Efficient maximization via profile likelihood; confidence interval and testing hypothesis in these complex cases; Bayesian method of inference: prior and posterior distribution, different types of prior, credible intervals and testing hypothesis; analytical approximations- asymptotic theory, Laplace approximation; numerical integral methods- Newton-Cotes type methods; Monte carlo methods; simulation methods- Markov chain Monte Carlo.

Resampling techniques: Bootstrap- confidence intervals, test, parametric bootstrap, advantages and disadvantages of parametric bootstrap; jackknife- confidence interval, test and permutation test.

Text Books

1. Casella G and Berger RL (2003). Statistical inference, 2^{nd} *edition*. Duxbery.
2. Millar RB (2011). Maximum Likelihood Estimation and Inference: with Examples in R, SAS and ADMB. Wiley.
3. Hogg RV, McKean J and Craig AT (2010). Introduction to Mathematical Statistics, 7^{th} *edition*. Pearson.

AST 703: Advanced Multivariate Techniques | **Credit 3**

Principal Components: Population principal components, summarizing sample variations by principal components, graphing the principal components, large sample inference.

Factor Analysis: The orthogonal factor models, methods of estimation (maximum likelihood estimates and principal factor analysis), selection of loadings and factor (factor rotation, varimax rotation, quartimax rotation, oblimin rotations), factor scores, structural equations models.

Canonical Correlation Analysis: Canonical variates and canonical correlations, sample canonical variates and sample canonical correlations, large sample inference.

Discrimination and Classification: Separation and classification two populations, classification of two multivariate normal populations, evaluating classification functions, Fisher's discriminant function, classification with several populations, Fisher's method for discriminating several populations.

Clustering: Similarity measures, hierarchical clustering methods, nonhierarchical clustering methods, multidimensional scaling.

Text Books

1. Johnson RA, Wichern W (2002). Applied Multivariate Statistical Analysis. Pearson.
2. Srivastava KS (2002). Methods of Multivariate Statistics. Wiley.

AST 704: Generalized Linear Models | **Credit 3**

Generalized linear models: Exponential family of distributions; estimation: method of maximum likelihood, method of least squares, estimation of generalized linear models; inference: sampling distribution for scores, sampling distribution for maximum likelihood estimators, confidence intervals for model parameters, adequacy of a model, sampling distribution for log-likelihood statistic, log-likelihood ratio statistic (deviance), assessing goodness of fit, hypothesis testing; multiple regression: maximum likelihood estimation, log-likelihood ratio statistic.

Models for binary responses: Probability distributions, generalized linear models, dose response models, general logistic regression, maximum likelihood estimation and log-likelihood ratio statistic, other criteria for goodness of fit, least square methods; multinomial distributions; nominal logistic regression models; ordinal logistic regression models.

Models for count data, Poisson regression and log-linear models: probability distributions, maximum likelihood estimation, hypothesis testing and goodness of fit.

Text Books

1. Dobson A (2008). An Introduction to Generalized Linear Models, 3^{rd} *edition.* Chapman & Hall.
2. McCullagh, P. and Nelder, J A. (1989). Generalized Linear Models, 2^{nd} *edition.* CRC Press.

AST 710: Advanced Survival Analysis

Credit 3

Estimating the Survival and Hazard Functions: Introduction and notation, the Nelson-Aalen and Kaplan-Meier estimators, counting process and martingals, properties of Nelson-Aalen estimator.

Semiparametric Multiplicative Hazards Regression Model: Introduction, estimation of parameters, inclusion of strata, handling ties, sample size determinations, counting process form of a Cox model, time-dependent covariates, different types of residuals for Cox models, checking proportionality assumption.

Multiple Modes of Failure: Basic characteristics of model specification, likelihood function formulation, nonparametric methods, parametric methods, semiparametric methods for multiplicative hazards model.

Analysis of Correlated Lifetime Data: Introduction, regression models for correlated lifetime data, representation and estimation of bivariate survivor function.

Text Books

1. Therneau TM and Grambsch PM (2000). Modeling Survival Data: Extending the Cox Model, Springer.
2. Kalbfleisch JD and Prentice RL (2002). The Statistical Analysis of Failure Time Data, 2^{nd} *edition.* Wiley.
3. Hougaard P (2000). Analysis of Multivariate Survival Data. Springer.

AST 711: Environmental and Spatial Statistics | **Credit 3**

Review of non-spatial statistics and stochastic process, overview of different types of spatial data; random field and spatial process - geostatistical/point reference process, areal/lattice process and point process; spatial data concern.

Geostatistical data: real data examples, measure of spatial dependence- variogram and covariance, stationarity and isotropic, variograms and covariance functions, fitting the variograms functions; Kriging, linear geostatistical model - formulation, simulation, estimation and prediction, generalized linear geostatistical model - formulation, simulations, estimation and prediction. Areal data: neighborhoods, testing for spatial association, autoregressive models (CAR, SAR), estimation/inference; grids and image analysis, disease mapping. Point pattern data: locations of events versus counts of events, types of spatial patterns, CSR and tests - quadrat and nearest neighbor methods, K-functions and L-functions, point process models- estimation and inference, health event clustering.

Special topics in spatial modeling: Hierarchical models, Bayesian methods for spatial statistics, Bayesian disease mapping, Spatio-temporal modeling, more on stationarity. Use of R and GIS software to give emphasis on analysis of real data from the environmental, geological and agricultural sciences.

Text Books

1. Cressie N (1993). Statistics for Spatial Data, *Revised edition.* Wiley.
2. Banerjee S, Carlin BP, and Gelfand AE (2014). Hierarchical Modelling and Analysis for Spatial Data.2^{nd}*edition* Chapman and Hall.
3. Cressie N and Wikle CK (2011). Statistics for Spatio-Temporal Data. Wiley.
4. Illian J, Penttine A, Stoyan H and Stoyan D (2008). Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.

AST 712: Advanced Time Series Analysis

Credit 3

Introduction: Forecasting time series, estimation of transfer functions, stochastic and deterministic dynamic mathematical models, stationary and nonstationary stochastic models for forecasting and control, basic ideas in model building. Time series and stochastic processes, stationary stochastic processes.

Seasonal models: parsimonious models for seasonal time series, fitting versus forecasting, seasonal models involving adaptive sines and cosines, general multiplicative seasonal model, some aspects of more general seasonal ARIMA models, structural component models and deterministic seasonal components.

Nonlinear and long memory models: Autoregressive conditional heteroscedastic (ARCH) models, generalized ARCH (GARCH) models, model building and parameter estimation, nonlinear time series models, long memory time series processes.

Multivariate time series analysis: Stationary multivariate time series, vector autoregressive-moving average (ARMA) models and representations, relation of vector ARMA to transfer function and ARMAX model forms, forecasting for vector autoregressive-moving average processes.

Text Books

1. Box GEP, Jenkins GM and Reinsel GC (2008). Time Series Analysis: Forecasting and Control, 4^{th} *edition*. Wiley.
2. Brockwell PJ and Davis RA (2013). Introduction to Time Series and Forecasting, 2^{nd} *edition*. Springer.

AST 713: Actuarial Techniques | **Credit 3**

Theory of interest in continuous time. Forces of interest and discount (constant and varying). Present and accumulated value calculations using non-level interest rates. Continuous annuities, valuation of continuous streams of payment, including the case in which interest conversion period differs from the payment period, continuous varying annuities. Bonds and related securities.

Definition and application of standard mortality probability symbols and force of mortality; relationship between survival distribution and life table functions; Continuous life annuities. Multiple decrement models. Net premiums, fully continuous premiums. Net premium reserves. Valuation theory for pension plans. The expanse factor and dividends. Introduction to risk theory– Purpose of the theory of risk; main problems in risk theory; individual risk models for a short term; applications of risk theory.

Principles of actuarial modeling. Familiarity with actuarial models– survival models, credibility models, risk theory models, ruin theory models, etc and their applications.

Text Book

1. Kellison SG (1991). Theory of Interest. Irwin.

AST 714: Advanced Operations Research

Credit 3

Special Types of Liner Programming Problems: Transportation problem, Transshipment problem, Assignment problem, Multidimensional problems.

Network Analysis: Terminology of networks, shortest path problem, minimum spanning tree problem, maximum flow problem, minimum cost flow pattern, network simplex method, project planning and control with PERT-CPM.

Dynamic Programming: Characteristics of dynamic programming problems, deterministic dynamic programming, probabilistic dynamic programming.

Non-linear Programming: Sample application, Graphical illustration of non-linear programming problems, types of non-liners programming problems, one-variable unconstrained optimization, multivariate unconstrained optimization, Karush-Kuhn Tucker (KKT) conditions for constrained optimization, quadratic programming, separable programming, convex programming, non-convex programming.

Inventory Models: The ABC inventory system, a generalized inventory model, deterministic models, probabilistic models, just-in-time manufacturing system.

Text Book

1. Hillier FS, Lieberman GJ, Nag B and Basu P (2001). Introduction to Operations Research, 9^{th} *edition*. McGraw-Hill.

AST 715: Advanced Econometric Methods

Credit 3

Three-stage least squares estimation: The three-stage least squares estimator (3SLS), comparison between GMM 3SLS and traditional 3SLS.

Linear unobserved effects panel data models: Strict exogeneity assumptions on the explanatory variables, some examples of unobserved effects panel data models, estimating unobserved effects models by pooled OLS, random effects (RE) methods, estimation and inference under the basic random effects, a general FGLS analysis, fixed effects (FE) methods, consistency of the fixed effects estimator, asymptotic inference with fixed effects. The Hausman test comparing the RE and FE estimators.

Nonlinear models: Discrete response models, the linear probability model for binary response, probit and logit, maximum likelihood estimation of binary response models, specification issues in binary response models, neglected heterogeneity, continuous endogenous explanatory variables, a binary endogenous explanatory variable, heteroskedasticity and nonnormality in the latent variable model, estimation under weaker assumptions, binary response models for panel data, pooled probit and logit.

Multinomial response models: Multinomial Logit, probabilistic choice models.

Ordered response models: Ordered logit and ordered probit, applying ordered probit to interval-coded data, corner solution outcomes and censored regression models– derivations of expected values, inconsistency of OLS, estimation and inference with censored tobit, pooled tobit, applying censored regression to panel data.

Text Books

1. Wooldridge JM (2010). Introductory Econometrics: A Modern Approach, 5^{th} *edition*. Cengage Learning.
2. Greene WH (2011). Econometric Analysis, 7^{th} *edition*. Prentice Hall.

AST 716: Advanced Population Studies

Credit 3

Fertility Determinants and models. Davis-Blake intermediate variables. Proximate determinants. Bongaarts model. Birth interval dynamics. Fecundity. Application of these models in the Bangladesh setting.

Estimation of population parameters from incomplete data. Estimation of mortality from census. Survivorship ratio. Estimate of infant and child mortality by indirect techniques such as Brass, Sullivan, Trussell and Feeney. Estimation of adult mortality from information on widowhood and orphanhood. Estimation of fertility by indirect techniques such as Brass, Hill, Coale-Trussell, relational Gompertz and reduced Gompertz model.

Population and Development: Inter-relation between population and development as envisaged value. Various population theories such as demographic transition theory. Emerging theories of population. Micro-economic theory of population. Recent contribution of Eastisliu, Becker, Caldwell etc.

Morbidity: Morbidity differentials and trends in Bangladesh. Health expectancy and burden of disease.

Manpower planning: Manpower in Bangladesh. Factors effecting manpower. Working life tables. Statistical analysis of manpower planning. Labor migration and its impact on the economy.

Population program of Bangladesh: Population policy. Population control and family planning. Evaluation of family planning programs. Use effectiveness and cost effectiveness. Multiple decrement life tables for measuring use-effectiveness. FP target setting models and impact assessment models.

Text Books

1. Chiang CL (1984). The Life Table and Its Applications. Krueger.
2. Bongaarts J and Potter RG (1983). Fertility, Biology and Behavior: An Analysis of the Proximate Determinants of Fertility. Academic Press, New York.

AST 717: Queueing Theory and Stochastic Processes

Credit 3

Queueing theory: Classical M/M/1 queue, global and local balance, performance measures, Poisson arrivals see time averages (PASTA) property, M/M/1/S queueing systems, blocking probability, performance measures, multi-server systems M/M/m, performance measures, waiting time distribution of M/M/m, performance measures of M/M/m/m with finite customer population, Erlang loss systems, a more general queueing models: M/G/1, M/G/m, G/M/1 queueing systems and analysis.

Queueing networks: Open queueing networks, analysis of tandem queues, applications of tandem queues in data networks, Jackson queueing networks, performance measures for open networks, closed queueing networks: Jackson closed queueing networks, steady-state probability distribution, application of closed queueing networks.

Reliability theory: Structure functions, minimal path and minimal cut sets, reliability of systems of independent components, bounds on the reliability function, system life as a function of component lives, expected system lifetime, systems with repair.

Brownian motion and stationary processes: Brownian motion, hitting times, maximum variable, variations on Brownian motion, Brownian motion with drift, geometric Brownian motion, pricing stock options, white noise, Gaussian processes, stationary and weakly stationary processes, harmonic analysis of weakly stationary processes.

Text Books

1. Chee-Hock N and Boon-Hee S (2008). Queueing Modelling Fundamentals with Applications in Communication Networks. John Wiley & Sons, Chichester.
2. Ross SM (2010). Introduction to Probability Models, 10^{th} *edition*. Academic press, New York.

AST 718: EPIDEMIOLOGY | **Credit 3**

Introduction: Disease processes, statistical approaches to epidemiological data, study design, binary outcome data, causality.

Measures of disease occurrence: Prevalence and incidence, disease rates, hazard function; review of simple random samples, probability, conditional probabilities, and independence of two events.

Measures of disease-exposure association: Relative risk, odds ratio, relative hazard, excess risk, attributable risk.

Study designs: Population-based studies, cohort studies, case-control studies, case-cohort studies; Assessing significance of 2×2 tables obtained from chohort designs, case-control designs.

Estimation and inference for measures of assocaition: Odds ratio, sampling distribution and confidence interval for odds ratio, relative risk, excess risk, attributable risk.

Confounding and interaction: Causal inference, counterfactuals, confounding variables, control of confounding variables by stratification, causal graphs, controllong confounding in causal graphs; Cochran-Mantel-Haenszel test, summary estimates and confidence intervals for odds ratio and relative risk after adjusting for confounding factors.

Interaction: Multiplicative and additive interaction, interaction and counterfactuals, test of consistency of association across strata, overall test of association, a test for trend in risk.

Text Books

1. Jewell NP (2003). Statistics for Epidemiology. Chapman and Hall.
2. Kleinbaum DG, Kupper LL and Morgenstern H. (1982). Epidemiologic Research: Principles and Quantitative Methods. Wiley.

AST 719: Analysis of Longitudinal Data

Credit 3

Longitudinal data: Concepts, examples, objectives of analysis, problems related to one sample and multiple samples, sources of correlation in longitudinal data, exploring longitudinal data.

Linear model for longitudinal data: Introduction, notation and distributional assumptions, simple descriptive methods of analysis, modelling the mean, modelling the covariance, estimation and statistical inference.

ANOVA for longitudinal data: Fundamental model, one sample model, sphericity condition; multiple samples models.

Linear mixed effects models: Introduction, random effects covariance structure, prediction of random effects, residual analysis and diagnostics.

Extension of GLM for longitudinal data: Review of univariate generalized linear models, quasi-likelihood, marginal models, random effects models, transition models, comparison between these approaches; the GEE methods: methodology, hypothesis tests using wald statistics, assessing model adequacy; GEE1 and GEE2.

Introduction to the concept of conditional models, joint models, their applications to bivariate binary and count data. Estimation, inference and test of independence.

Generalized Linear Mixed Models (GLMM): Introduction, estimation procedures–Laplace transformation, penalized quazi-likelihood (PQL), marginal quazi likelihood (MQL).

Numerical integration: Gaussian quadrature, adaptive gaussian quadrature, Monte Carlo integration; markov chain Monte Carlo sampling; comparison between these methods.

Statistical analysis with missing data: Missing data, missing data pattern, missing data mechanism, imputation procedures, mean imputation, hot deck imputation. estimation of sampling variance in the presence of non-response, likelihood based estimation and tests for both complete and incomplete cases, regression models with missing covariate values, applications for longitudinal data.

Text Books

1. Verbeke G and Molenberghs G (2000). Linear Mixed Model for Longitudinal Data. Springer.
2. Molenberghs G and Verbeke G (2005). Models for Discrete Longitudinal Data. New York: Springer-Verlag.
3. Islam MA and Chowdhury RI (2017). Analysis of Repeated Measures Data. Springer.
4. Diggle PJ, Heagerty P, Liang K-Y, and Zeger SL (2002). Analysis of Longitudinal Data, 2^{nd} *edition*. Oxford.

AST 720: Adaptive Sampling

Credit 3

Design and model unbiased estimators; fixed and stochastic population sampling theory.

Adaptive sampling deigns; Detectability in adaptive sampling; constant and unequal detectabilities for adaptive design.

Adaptive cluster sampling; initial random sample with and without replacement; initial unequal probability sampling; expected sample size and cost; comparative efficiencies of adaptive and conventional sampling.

Systematic and strip adaptive cluster sampling; stratified adaptive cluster sampling; adaptive allocation in stratified sampling; sample sizes based on observations in each strata and from previous strata; comparison of systematic and stratified adaptive sampling with conventional sampling procedures; adaptive cluster sampling based on order statistics.

Multivariate aspects of adaptive sampling; multivariate conditions for adding neighbourhoods; design-unbiased estimation for multivariate approach.

Text Books

1. Thompson SK and Seber GAF (1996). Adaptive Sampling. Wiley.
2. Thompson SK (1992). Sampling. Wiley.

AST 721: Optimum Experimental Designs

Credit 3

Optimum design theory: Continuous and exact designs, the general equivalence theorem, algorithms for continuous designs and general equivalence theorem, function optimization and continuous design.

Criteria of optimality: A-, D-, and E-optimality; D_A-optimality, D_S-optimality, c-optimality, linear optimality; compund design criteria.

D-optimum designs: Properties of D-optimum designs, sequential construction of optimum designs, polynomial regression in variable, second-order model with several variables.

Algorithms for constructing of exact D-optimum designs: The exact design problem, basic formulae for exchange algorithm, sequential algorithms, non-sequential algorithms, the KL and BLKL exchange algorithms.

Experiments with both qualitative and quantitative factors, blocking response surface designs, mixture experiments, non-linear models, Bayesian optimum designs, model checking and designs for discriminating between models, compund design criteria, generalized linear models.

Text Book

1. Atkinson AC, Donev AN, and Tobias RD (2008). Optimum Experimental Designs with SAS. Oxford.

AST 722: Statistical Signal Processing

Credit 3

Introduction to signals: Signals and their classification; real world analog signals: audio, video, biomedical (EEG, ECG, MRI, PET, CT, US), SAR, microarray, etc; digital representation of analog signals; role of transformation in signal processing. Orthogonal representation of signals. Review of exponential Fourier series and its properties.

Signal estimation theory: Estimation of signal parameters using ML, EM algorithm, minimum variance unbiased estimators (Rao-Blackwell theorem, CRLB, BLUE), Bayesian estimators (MAP, MMSE, MAE), linear Bayesian estimators.

Signal detection theory: Detection of DC signals in Gaussian noise: detection criteria (Bayes risk, Probability of error, Neyman-Pearson), LRT; detection of known signals in Gaussian noise: matched filter and its performance, minimum distance receiver; detection of random signals in Gaussian noise: the estimator correlator.

Applications: Scalar quantization, image compression, pattern recognition, histogram equalization, segmentation, application of signal estimation and detection theory to signal communication, signal recovery from various types of linear and nonlinear degradations, copyright protection, enhancement, etc.

Text Books

1. Soliman SS and Mandyam DS (1998). Continuous and Discrete Signals and Systems, 2^{nd} *edition.* Prentice-Hall.
2. Kay SM (1993). Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall.
3. Kay SM (1998). Fundamentals of Statistical Signal Processing: Detection Theory. Prentice Hall.
4. Gonzalez RC and Woods RE (2008). Digital Image Processing, 3^{rd} *edition.* Pearson Education, Inc.

AST 723: Meta analysis | **Credit 3**

Introduction to systematic review and meta analysis: Motivation, strengths and weakness of meta-analysis, problem formulation (why study meta analysis), systematic review process.

Types of results to summarize; overview of effect size; effect size calculation for both continuous and discrete data.

Combining effect size from multiple studies; fixed effect and random effects models and their estimation; heterogeneity between studies and its estimation techniques; test of homogeneity in meta analysis; prediction intervals; subgroup analysis, Meta regression: random effect meta regression, baseline risk regression.

Publication bias in meta analysis; Power analysis for meta analysis; Effect size rather than P-values; Meta analysis based on direction and P-values, reporting the results of meta analysis.

Introduction to Bayesian approach to meta analysis; Meta analysis for multivariate/longitudinal data; network meta analysis.

Text Books

1. Borenstein M, Hedges LV, Higgins JPT and Rothstein HR (2009). Introduction to Meta-Analysis, John Wiley & Sons, UK.
2. Hartung J and Knapp G and Sinha BK (2011). Statistical Meta-Analysis with Applications. John Wiley & Sons, UK.

AST 724: Clinical Trials | **Credit 3**

Statistical approaches for clinical trials: Introduction, comparison between Bayesian and frequentist approaches and adaptivity in clinical trials. Phases of clinical trials, pharmacokinetics (PK) and pharmacodynamics (PD) of a drug, dose-concentration-effect relationship and compartmental models in pharmacokinetic studies.

Phase I studies: Determining the starting dose from preclinical studies. Rule-based designs: 3+3 design, Storer's up-and-down designs, pharmacologically-guided dose escalation and design using isotonic regression. Model-based designs: continual reassessment method and its variations, escalation with overdose control and PK guided designs.

Phase II studies: Gehan and Simon's two-stage designs. Seamless phase I/II clinical trials: TriCRM, EffTox and penalised D-optimum designs for optimum dose selection.

Phase III studies: Randomised controlled clinical trial, group sequential design and multi-arm multi-stage trials in connection with confirmatory studies.

Text Books

1. Berry SM, Carlin BP, Lee JJ, and Muller P (2010). Bayesian Adaptive Methods for Clinical Trials. CRC press.
2. Rosenbaum SE (2012). Basic Pharmacokinetics and Pharmacodynamics: An Integrated Textbook and Computer Simulations. John Wiley & Sons.

AST 725: Statistical Machine Learning | **Credit 3**

Statistical learning: Statistical learning and regression, curse of dimensionality and parametric models, assessing model accuracy and bias-variance trade-off, classification problems and K-nearest neighbors.

Linear regression: Model selection and qualitative predictors, interactions and nonlinearity.

Classification: Introduction to classification, logistic regression and maximum likelihood, multivariate logistic regression and confounding, case-control sampling and multiclass logistic regressionl, linear discriminant analysis and Bayes theorem, univariate linear discriminant analysis, multivariate linear discriminant analysis and ROC curves, quadratic discriminant analysis and naive bayes.

Resampling methods: Estimating prediction error and validation set approach, k-fold cross-validation, cross-validation- the right and wrong ways, the bootstrap, more on the bootstrap.

Linear model selection and regularization: Linear model selection and best subset selection, forward stepwise selection, backward stepwise selection, estimating test error using mallow's Cp, AIC, BIC, adjusted R-squared, estimating test error using cross-validation, shrinkage methods and ridge regression, the Lasso, the elastic net, tuning parameter selection for ridge regression and lasso, dimension reduction, principal components regression and partial least squares.

Moving beyond linearity: Polynomial regression and step functions, piecewise polynomials and splines, smoothing splines, local regression and generalized additive models.

Tree-based methods: Decision trees, pruning a decision tree, classification trees and comparison with linear models, bootstrap aggregation (Bagging) and random forests, boosting and variable importance.

Support vector machines: Maximal margin classifier, support vector classifier, kernels and support vector machines, example and comparison with logistic regression.

Text Books

1. James G, Witten D, Hastie T and Tibshirani R (2013). An Introduction to Statistical Learning: with Applications in R, $1^{st} edition$. Springer.
2. Hastie T, Tibshirani R and Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction, $2^{nd} edition$. Springer.

AST 740: Oral | **Credit 3**

Each student must be examined orally by a committee of selected members at the end of the academic year.
