Decision Science Course Review

Descriptive Statistics, Probability Distributions and Portfolio Analysis

random variable: measurements may take range of values; discrete: finite possible values; continuous: infinite possible values
probability distributions. *Skewness*: degree of asymmetry; *Kurtosis*: degree of flatness; *Bimodality*: distribution has two peaks
cumulative probability distribution: Gives the probability of being less than a certain value $P(X_0<X)$
Arithmetic: *Mean* (μ) = average; *Median*: middle value; *Mode*: Most frequent value
Variance(σ^2): avg squared distance from μ; *Standard Dev.*(σ) = √variance; *Covariance*: avg deviations x + y from their means
Correlation: linear relationship between variables; avg of std. variables $\frac{(x_i-\mu_x)}{\sigma_x}$ and $\frac{(y_i-\mu_y)}{\sigma_y}$ $r = corr(x,y) = \frac{\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)}{n\sigma_x\sigma_y} = \frac{\text{cov}(x,y)}{\sigma_x\sigma_y}$
-1<r<1 (0: no relationship; 1:perfect positive; -1:perfect negative
Normal Distribution: probabilities of std normal distribution (μ=0, σ=1) can be found in tables; with $\mu \neq 0$ and/or $\mu \neq 1$ calculate Z (# std dev. from mean): $Z = \frac{X-\mu}{\sigma}$ and look up probability of Z in the table.
Portfolios: mixture assets; *Dominating investment*: higher return(μ); lower risk(σ); Efficient frontier: not dominated by any inv.
Aim of portfolio: combine variables (assets) with different means (returns) and std devs (risks) for new better random variable
Expected Return: $E(w_1X + w_2Y) = w_1E(X) + w_2E(Y)$ *Risk*: $\text{Var}(w_1X + w_2Y) = w_1^2\sigma_x^2 + w_2^2\sigma_y^2 + 2w_1w_2\sigma_x\sigma_y\text{Corr}(X,Y)$
Expected Return (3 stocks): $E(w_1X + w_2Y + w_3Z) = w_1E(X) + w_2E(Y) + w_3E(Z)$
Risk (3 stocks): $Var(w_1X + w_2Y + w_3Z) = w_1^2\sigma_x^2 + w_2^2\sigma_y^2 + w_3^2\sigma_z^2 + 2w_1w_2\sigma_x\sigma_yCorr(X,Y) + 2w_1w_3\sigma_x\sigma_zCorr(X,Z) + 2w_2w_3\sigma_y\sigma_zCorr(Y,Z)$

Statistical Sampling

Population (μ,σ,p): entire set of items under consideration Sample: subset or part of population -> Statistics ($\bar{x}, s, \hat{p}$
Inference: Infer population characteristics from sample stats; estimation, confidence intervals, hypothesis testing (sampling error)
Sample Data Assumptions: sample is properly representative of population; data collected is reliable and accurate
Sample Mean: mean of the sampling distribution (mean of all sample means) is μ, the population mean.
Std Dev of sample mean (standard error): $\frac{\sigma}{\sqrt{n}}$ σ: std dev of population, s: std dev of sample, n: sample size
Central Limit Theorem: distribution of sample means approach normal distribution as sample size (n) increases.
If σ is unknown, estimate it from the sample: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(xi-\bar{x})^2}$ If σ estimated, use t-dist.; If n>30, use normal dist.
Since $\bar{X}$ is estimated, use n-1 degrees of freedom
Confidence Intervals: use probability to assess sample mean's likely relation to population mean (sampling distribution is normal)
95% Confidence Interval: $95\%CI = \pm 1.96\frac{\sigma}{\sqrt{n}}$ There is a 95% probability that this interval contains the population mean.
Hypothesis Testing: assume hypothesis is true; calculate plausible range for mean; reject hypothesis if estimate is not in range
T-Statistic Approach: how many standard errors is $\bar{X}$ from null hypothesis value of μ: $t-stat = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
P-Value: area in both tails (2 x prob. value in table); compare to significance level (<0.05)
Control Charts: at each time interval, sample size n taken and sample mean $\bar{X}$ plotted; if $\bar{X}$ outside upper/lower warning limits (2σ) process may be out of control; If $\bar{X}$ outside action limits (3σ) process is out of control.

Simple Regression

Model Building: 1.Scatterplots, correlation analysis, 2.Model estimation, 3.Evaluation (validity: significance, anything left out? usefulness: fit)
Simple linear regression: relationship between 2 variables, model: y=a+bx+e b:coefficient, a:constant, x:independent vble, y: dependent vble
Residual: Vertical distance from line, line minimizes sum of squared residuals
Standard error (std dev of residuals, depends on scale)
$R^2 = 1 - \frac{s^2}{s_y^2}$. Proportion of variation in dependent variable explained. Adjusted R^2 penalizes for few observations.
Test: coefficients: |t-stat|>2, p-value <0.05; Residuals: random, no autocorrelation; CI: use std dev of residuals (std error)

Multiple Regression

Multiple explanatory variables: $y = a + b_1x_1 + b_2x_2 + \dots + b_kx_k + e$
Use p-values/t-stats to decide whether variables should be kept in the model
Multicollinearity: correlation between two or more explanatory variables -> check correlations between explanatory variables
Durbin-Watson (time series): Autocorrelation for residuals placed in same order as observations $DW \approx 2(1-r)$, r=corr(e_i,e_{i-1}): autocorrelation.
DW = 2 -> no, DW > 2 -> negative, DW < 2 -> positive
Trending variables have high correlation, fit even if unrelated, use changes

Regression Checklist:
- Get feel for data using scatterplots & correlations
- Fit model to data
- Check coefficients different from 0: t-stats, p-values
- Check fit of model: Adjusted R^2, std. dev. of residuals
- Check residuals: autocorrelation, normality
- Check for multicollinearity (correlations)
- Judgmentally assess the model
- Use for description and prediction

Risk Analysis

Risk analysis: method for assessing impact of risk on decision situations
Objectives: Sensitivity testing of assumptions, better perception of risks and interactions, anticipation and contingency planning, overall reduction of risk exposure, insight, knowledge and confidence for better decision making and improved risk management
Scenario Analysis: optimistic, pessimistic and mostly likely scenarios.
Sensitivity Analysis: explore robustness of results to variations in model parameters, understand and challenge assumptions, identify variables to which results are sensitive, range over which results might vary; What-if Questions, Tornado diagram, one-way and two-way sensitivity analysis, spiderplots (% change from base vs. target variable)
Monte Carlo Simulation: define distributions for uncertain variables, pick random numbers according to distributions, calculate target variable, repeat many times -> distribution of target variable

- Replacing uncertain inputs with average values does not result in expected value of output unless function is linear -> Monte Carlo useful for expected value

Decision Analysis

Methodology: Problem Structuring, Deterministic Analysis, Probabilistic Analysis, Evaluation, Communication, Answers/Insights
Decision trees: decision square, chance circle; determine all possible outcomes, payoffs, probabilities; calculate values at nodes
Expected value criterion is standard approach but risk attitude: Risk aversion, seeking, neutral -> Alternative criteria: maximin, maximax, utility functions (monetary values replaced by relative preference; useful for qualitative outcomes)
Marginal Probability P(A) Joint Probability P(AnB) Conditional Probability P(A|B): A if B occurred $P(A|B) = \frac{P(A\cap B)}{(B)}$
Bayes Theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ *Value of perfect info* (value of observing an uncertain event beforehand) sets upper bound on value of new info. Should also consider risk profiles and sensitivity analysis