



Departamento de Estatística de Stanford

Curso: Introduction to Statistics

Professor Palestrante: Guenther Walther

Carga Horária Total : 32 horas

*curso on-line sem créditos autorizado pela Stanford University e ministrado através da plataforma da Coursera.

Introdução a estatísticas descritivas para a exploração de dados

Este curso "Introduction to Statistics" de Stanford ensina conceitos de raciocínio estatístico que são essenciais para aprender com os dados e comunicar percepções. Ao final do curso, o(a) senhor(a) será capaz de realizar análises exploratórias de dados, compreender os princípios fundamentais de amostragem e selecionar testes apropriados de significância para vários contextos. Os tópicos incluem Estatística Descritiva, Amostragem e Experimentos Controlados Aleatórios, Probabilidade, Distribuições de Amostragem e o Teorema do Limite Central, Regressão, Testes Comuns de Significância, Reamostragem, Comparações Múltiplas.

Atividades programáticas do Curso

Semana 1: Introdução e estatísticas descritivas para a exploração de dados.

Gráfico de pizza, gráfico de barras e histogramas

Gráfico de Box-and-Whisker e gráfico de dispersão

Fornecer contexto é fundamental para as análises estatísticas

Armadilhas na visualização de informações

Média e mediana, moda

Percentis, resumo de cinco números e desvio padrão

[EXTRA] Insight do setor: Introdução a Andrew Radin

Semana 2: Produção de dados e amostragem e definição de probabilidade, regras essenciais de probabilidade.

Amostragem aleatória simples e amostragem aleatória estratificada

Viés e erro casual

Observação vs. experimento, confusão e efeito placebo

A lógica dos experimentos controlados e aleatórios

[EXTRA] Informações sobre o setor: Registro de uma patente para twoXAR

Produção de dados e amostragem

Probabilidade e regras de probabilidade.

Aplicação de probabilidade.

Semana 3: Regra empírica e a aproximação normal dos dados, distribuição binomial, conceitos básicos de variáveis aleatórias. Distribuições de amostragem e o Teorema do Limite Central(a Lei dos Grandes Números e o Teorema do Limite Central). Os diferentes tipos de histogramas presentes na análise estatística.

A interpretação da probabilidade

Complemento, resultados igualmente prováveis, adição e multiplicação

Exemplo de quatro regras: Como lidar com "Pelo menos um"

Resolução de problemas por enumeração total

Regra de Bayes

Análise Bayesiana

Modelo de resposta aleatória da Warner

[EXTRA] Informações sobre o setor: Descoberta de medicamentos na twoXAR

Distribuições de amostragem

Teorema do Limite Central(a Lei dos Grandes Números e o Teorema do Limite Central).

Histogramas para análise estatística.

Semana 4: Inferência. Regressão e como fazer diagnósticos de regressão.

A curva normal

A regra empírica

Padronização de dados e a curva normal padrão

Aproximação normal

Cálculo de percentis com a aproximação normal

A configuração binomial e o coeficiente binomial

A fórmula binomial

Variáveis aleatórias e histogramas de probabilidade

Aproximação normal ao binomial; amostragem sem substituição

[EXTRA] Informações sobre o setor: Oportunidades em Ciências da Vida

A aproximação normal para dados e a distribuição binomial
Parâmetro e estatística
Valor esperado e erro padrão
EV e SE da soma, porcentagens e ao simular
A lei da raiz quadrada
A distribuição de amostragem
Três histogramas
A lei dos grandes números
O Teorema do Limite Central
Quando se aplica o Teorema do Limite Central?
Distribuições de amostragem e o Teorema do Limite Central
A previsão é uma tarefa fundamental da estatística
O coeficiente de correlação
Medidas de correlação Associação linear
Linha de regressão e o método dos mínimos quadrados
Regressão à média, a falácia da regressão
Previsão de y a partir de x e x a partir de y
Aproximação normal Dado
Gráficos de resíduos, heterocedasticidade e transformações
Pontos discrepantes e influentes
[EXTRA] Insights do setor: Desafios ao uso da ciência de dados na medicina
Regressão para analisar a relação entre uma variável dependente (ou resposta) e uma ou mais variáveis independentes (ou preditoras). Diagnósticos de regressão.

Semana 5: Construção e interpretação de intervalos de confiança em situações padrão.

Testes de significância.

Interpretação de um intervalo de confiança
Usando o Teorema do Limite Central para encontrar um intervalo de confiança
Estimativa do erro padrão com o princípio de bootstrap
Tópicos sobre intervalos de confiança
A ideia por trás do teste de hipóteses
Configuração de uma estatística de teste
valores de p como medidas de evidência•
O teste t
Significância estatística vs. importância
Testes de Significância

O teste z para duas amostras

Pares combinados

[EXTRA] Informações sobre o setor: Contratação de talentos em ciência de dados

Semana 6: Reamostragem e os dois principais métodos usados na inferência estatística com uso intensivo de computador (O método Monte Carlo e o método Bootstrap). Análise de dados categóricos (Teste Qui-Quadrado de Adequação, Teste Qui-Quadrado de Homogeneidade e Teste Qui-Quadrado de Independência)

Uso de simulações de computador no lugar de cálculos

Uso da lei dos grandes números para aproximar quantidades de interesse

Princípio do plug-in

O bootstrap paramétrico e os intervalos de confiança bootstrap

Bootstrapping em regressão

Reamostragem

Relações entre duas variáveis categóricas

O teste qui-quadrado (de adequação, homogeneidade, independência)

Análise de dados categóricos

Semana 7: Análise de variância unidirecional (ANOVA) e como os testes F funcionam em exemplos de ANOVA unidirecional.

Comparação de várias médias

A ideia da análise de variância

Usando a distribuição F para avaliar a ANOVA

Outros tópicos sobre ANOVA

[EXTRA] Informações sobre o setor: Iniciando sua carreira em ciência de dados

Análise de variância unidirecional

Semana 8: Comparações múltiplas. Questões muito importantes que surgiram na era do big data: espionagem de dados e a falácia dos múltiplos testes. Os motivos por trás dos

desafios da reprodutibilidade e aplicabilidade dos dados e como evitar esses problemas no próprio trabalho.

Snooping de dados e a falácia dos testes múltiplos, reprodutibilidade e replicabilidade

Correção de Bonferroni, taxa de descoberta falsa e divisão de dados

Comparações múltiplas

Teste de Avaliação Final