Metodologias Experimentais em Informática

Introdução ao método experimental:

Porque precisamos de experimentar:

- Investigadores: para validar hipóteses, modelos e confirmar teorias.
- Engenheiros: para afinar sistemas, para comparar e escolher entre as diferentes opções nos projetos e para verificar se os requisitos ou as especificações são atendidas.

Planeamento de experiências:

O objetivo é projetar e realizar experiências válidas que permitam boas decisões técnicas.

Ideia base: introduzir alterações controlados nas variáveis de entrada, a fim de estudar o seu efeito sobre as variáveis observadas e obter o máximo de informações sobre as relações causa e efeito.

Etapas:

1. Declaração do problema.
2. Identificar as variáveis (dependentes são as variáveis de saída e as independentes são as variáveis de entrada que podem ser alteradas).
3. Gerar a hipótese (descrever as relações entre as variáveis de entrada e as variáveis de saída).
4. Definir o cenário experimental.
5. Desenvolver ferramentas e procedimentos para a experiência.
6. Executar a experiência e recolher dados.
7. Realizar a análise dos dados.
8. Tirar conclusões (muitas vezes volta-se ao inicio do processo).

Medições:

Métricas de desempenho:

- Speedup: $\beta = T_1/T_2$, o sistema 2 é β vezes mais rápido/lento que o sistema 1.
- Mudança relativa: $\Delta = (T_1 - T_2)/T_2 = \beta - 1$, o sistema 2 é Δ% mais rápido/lento que o sistema 1.

Métricas de velocidade:

- Throughput: número de eventos completados num intervalo de tempo.
- Speedup: $\beta = R_2/R_1$, o sistema 2 é β vezes mais rápido/lento que o sistema 1.
- Mudança relativa: $\Delta = (R_2 - R_1)/R_1 = \beta - 1$, o sistema 2 é Δ% mais rápido/lento que o sistema 1.

Propriedades para caraterizar as métricas:

- Linearidade: se o valor da métrica for modificado, o desempenho da máquina deve mudar na mesma proporção.
- Confiabilidade: uma métrica de desempenho é confiável, se um sistema A supera sempre o sistema B quando os valores correspondentes da métrica para ambos os sistemas indicam que o sistema A deva superar o sistema B.
- Repetibilidade: uma métrica de desempenho é repetível se o mesmo valor da métrica é obtido a cada vez que a mesma experiência é realizada.
- Consistência: uma métrica é consistente se tiver o mesmo valor num sistema diferente.

Incertezas:
Se repetirmos uma medição iremos obter resultados ligeiramente diferentes. Dois tipos de incertezas:
- Incertezas aleatórias: as variações nas medições ocorrem sem um padrão previsível.
- Incertezas sistémicas: as variações que causam que o valor medido seja menor ou maior que o valor exato.

Modelo de erros gaussiano:
O valor medido para X pode ser X + E ou X – E com a mesma probabilidade.

Introdução á interferência:

População: é um conjunto de entidades com alguma caraterística em comum.
Amostra: representa um subconjunto de uma população.

Distribuição normal:
É uma distribuição de probabilidade contínua com dois parâmetros μ e σ.
Permite calcular a probabilidade num dado intervalo.
A distribuição normal pode ser transformada para a distribuição normal standard pela fórmula $Z = (x - \mu)/\sigma$.
Exemplo: para $\mu = 6$ e $\sigma = 4/5$, $P(4 \leq X \leq 5) \Leftrightarrow P(-2.45 \leq Z \leq -1.25) = 9.94\%$
A probabilidade associada à distribuição normal padrão está tabulada.

Calcular intervalos de confiança:

Cálculo de intervalos de confiança:

$x \pm z_{1-\alpha/2} \times \frac{s}{\sqrt{n}}$, para n (número de amostras) ≥ 30

$h = t_{n-1,1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}$, para $n < 30$

Intervalo de confiança para a diferença entre médias:
É usado para estimar a diferença entre as médias de duas populações.
Amostras independentes: se a seleção da primeira amostra não alterar a seleção da segunda amostra.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2}{n_1} + \frac{(n_2-1)s_2^2}{n_2}}$$

$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \times S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, se ambos os grupos tem $n \geq 30$

$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,1-\alpha/2} \times S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, se pelo menos um grupo tem $n < 30$

Amostras dependentes: se para cada observação numa amostra existe uma outra observação correspondente na outra.

$\bar{x}_D \pm z_{1-\alpha/2} \times \frac{S_D}{\sqrt{n}}$, para $n \geq 30$

$\bar{x}_D \pm t_{n-1,1-\alpha/2} \times \frac{S_D}{\sqrt{n}}$, para $n < 30$

Proporções:

$\mu_{\bar{p}} = p$, média $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$, desvio padrão

Intervalo de confiança para a proporção:

$\bar{p} \pm z_{1-\alpha/2} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ se $\min(n\bar{p}, n(1 - \bar{p})) > 5$

Intervalo de confiança para a diferença entre proporções:

$(\bar{p}_1 - \bar{p}_2) \pm z_{1-\alpha/2} \times \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$ se $\min(n_1\bar{p}_1, n_1(1 - \bar{p}_1), n_2\bar{p}_2, n_2(1 - \bar{p}_2)) > 5$

Teste de hipóteses:

Hipótese:

- É uma proposta de explicação para um determinado fenómeno.
- É uma suposição sobre a eficiência de um dado componente/sistema.

Uma hipótese requer avaliação para poder ser considerada verdadeira.

Caraterização de uma hipótese:

- Tópico: área de interesse onde a dificuldade a ser resolvida está incluída.

Exemplo: qualidade do código (ausência de bugs) produzido pelos programadores.

- Problema: é o objeto de estudo, pressupõe perguntas claras que formulam o problema a ser resolvido. Exemplo: se o método utilizado para o desenvolvimento do software está relacionado com o número de bugs.
- Hipótese: resposta provisória à pergunta, ou seja se a hipótese for confirmada, a resposta é considerada correta (para um determinado grau de certeza). Exemplo: o desenvolvimento do software utilizando o método A produz menos bugs que o método B.

Etapas dos testes de hipóteses:

1. Estado da hipótese a ser testada:

- Hipótese nula (H_0): é uma afirmação sobre um parâmetro da população (por exemplo a média da população) que se presume ser verdadeiro.
- Hipótese alternativa (H_1): é uma afirmação que contradiz a hipótese nula.

2. Selecionar o critério de decisão:

Definir um critério (5%), significa indicar o nível de significância para o teste.

3. Calcular os testes estatísticos:

Escolher uma amostra aleatória da população e medir a média da amostra.
O teste estatístico é uma fórmula para determinar se a hipótese nula é verdadeira.

$$Z_c = \frac{M - \mu}{\sigma / \sqrt{n}}$$

M – média da amostra, μ – média da população

4. Tomar uma decisão:

O valor da estatística de teste (Z_c) é a chave para tomar uma decisão sobre a hipótese nula. Se $P < 5\%$ rejeita-se a hipótese nula, se $P > 5\%$ mantêm-se a hipótese nula.

Medir o tamanho de um efeito:

$$\textbf{Cohen's d} = \frac{M - \mu}{\sigma}$$

M – média da amostra, μ – média da população,
σ - desvio padrão da população.

Efeitos: pequeno: $|d| < 0.2$, médio: $0.2 < |d| < 0.8$, grande: $|d| > 0.8$

Se o valor de d é negativo, o efeito deslocou abaixo da média da população.

Teste T:

Segue a distribuição T-Student. É utilizado quando o número de amostras é < 30.

- Uma amostra: utilizado para comparar uma média de uma amostra com a média de uma população conhecida. (graus de liberdade: df = n – 1)

$$t_c = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$\bar{x}$ – média da amostra, μ – média da população,
s – desvio padrão da amostra, n – número de elementos da amostra

- Duas amostras independentes: utilizado para comparar duas amostras.
(graus de liberdade: $df = n_1 + n_2 - 2$)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad S_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

- Duas amostras dependentes: utilizado para comparar duas amostras.
(graus de liberdade: $df = n - 1$)

$$t = \frac{\bar{x}_D - \mu}{\frac{s_D}{\sqrt{n}}}$$

$\overline{X}_D$ – média das diferenças entre as duas amostras,
S_D – desvio padrão das diferenças

Testes não paramétricos:

Os testes de hipóteses não paramétricos são usados quando:
- Os dados têm um ranking e nenhuma interpretação numérica é possível.
- Estamos interessados em um parâmetro da população para o qual a distribuição é desconhecida (média, mediana).

Análise da variância:

Utilizada quando temos mais de duas amostras para comparar. Por exemplo: qual é o impacto do tamanho dos elementos no tempo de ordenamento.
O objetivo é testar se a diferença entre as médias das várias amostras é significativa. ANOVA pode fazer isso.

ANOVA com um fator:

Testa as hipóteses seguintes:
H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ (todas as médias dos grupos são iguais)
H_1: nem todas as médias dos grupos são iguais.
Não indica quais é que diferem, para isso é preciso acompanhar com comparações múltiplas (testes post-hoc, quando a hipótese H_0 é rejeitada).
Tabela ANOVA:
É um resumo de todos os elementos necessários para o cálculo do P value.

ANOVA etapas:

- Formulação de hipóteses (H_0 e H_1).
- Calcular o teste estatístico (preenchimento da tabela ANOVA).
- Obter o P value (está na tabela ANOVA).
- Tomar uma decisão (se o P value < α rejeita-se a hipótese H_0).

Análise de variância com 2 fatores:

Utilizada para testar a igualdade de duas ou mais médias populacionais quando duas variáveis independentes são utilizados (fator A e fator B).
ANOVA com 2 fatores testa 3 hipóteses simultaneamente:
- Nenhuma diferença nas médias devido ao fator A - H_0^A $\mu_1 = \mu_2 = \ldots = \mu_a$
- Nenhuma diferença nas médias devido ao fator B - H_0^B $\mu_1 = \mu_2 = \ldots = \mu_b$
- Nenhuma interação dos fatores A e B - H_0^{AB}: $\forall(i,j)$: γ_{ij}

Regressão linear:

Um modelo de regressão linear representa a relação entre uma variável y e uma variável x. Exemplo: o tempo de resposta de um servidor web com o aumento do número de pedidos.

Forma de um modelo de regressão linear:

$y = a + bx$, onde x é a variável de entrada, y é a resposta de saída prevista e a e b são os parâmetros de regressão que queremos estimar a partir de um conjunto de medições.

Correlação:

O coeficiente de determinação r^2 dá a fração de variação total explicada pelo modelo de regressão. Se existir uma relação perfeita entre a entrada e a saída então $r^2 = 1$.

Transformações:

Transformar os dados não lineares em uma forma linear:

- Regressão linear padrão: $y = a + bx$, $\hat{y} = a + bx$
- Modelo exponencial: $\ln(y) = a + bx$, $\hat{y} = ab^x$
- Modelo quadrático: $\sqrt{y} = a + bx$, $\hat{y} = (a + bx)^2$
- Modelo recíproco: $1/y = a + bx$, $\hat{y} = 1/(a + bx)$
- Modelo logarítmico: $y = a + b.\ln(x)$, $\hat{y} = a + b.\ln(x)$

Experiências com pessoas:

Áreas de aplicação:

- Testes de usabilidade (como é que as pessoas usam os sistemas).
- Avaliação do software ou produto no mercado.

Técnicas experimentais:

- Observação e análise de dados: ver como os utilizadores se comportam.
- Experiências controladas: ver como os utilizadores realizam tarefas predefinidas.
- Entrevistas e questionários: tentar entender as preferências e necessidades dos utilizadores.

Experiências controladas com utilizadores (etapas num estudo típico):

1. Definir os objetivos do sistema ou o módulo sob avaliação.
2. Criar um conjunto de tarefas a fim de atender a esses objetivos.
3. Escolher os testadores.
4. Observá-los a executar as tarefas.

Definir as tarefas para os testes de usabilidade:

- Fazer uma lista de tarefas e avaliá-las pela importância numa escala de 1 a 6.
- Definir os objetivos.
- Ser específico e claro sobre o que queremos que o utilizador faça.
- Criar uma sequência razoável.
- Evitar o uso de palavras que aparecem na interface.

Aspetos experimentais:

- Medir o tempo para executar as diferentes tarefas.
- Gravar em vídeo para análise posterior.
- Recolher informações sobre a experiência do testador.