UNIVERSIDADE DE PERNAMBUCO

ESCOLA POLITÉCNICA DO RECIFE

TRADUÇÃO DO CAPÍTULO 8 DO LIVRO

Spiegel, M, Schiller, J & Srinivisan, A. SHAUM’S EASY OUTLINES: Probability and Statistics. Crash Course. McGraw-Hill, New York, 2001

RECIFE – 2010

CAPITULO 8

TESTE DE HIPÓTESES E SIGNIFICÂNCIA

**Sumário**

* Decisões Estatísticas
* Hipóteses Estatísticas
* Testes de Hipóteses e Significância
* Erros do Tipo I e Tipo II
* Nível de Significância
* Testes Envolvendo Distribuição Normal
* Testes Unicaudal e bicaudal
* Valor de P
* Testes Especiais
* Relação entre a Teoria da Estimação e o Teste de Hipóteses

**DECISÕES ESTATÍSTICAS**

 Frequentemente precisamos tomar decisões sobre populações, com base em informações sobre amostras das mesmas. Tais decisões chamam-se *decisões estatísticas.* Por exemplo, com base em resultados amostrais, podemos querer decidir se determinada vacina é eficiente na cura de determinada doença, se um processo educacional é melhor do que outro, se uma moeda é viciada ou não, etc.

**HIPÓTESES ESTATÍSTICAS**

 Na tomada de decisões, se faz necessário formular hipóteses ou suposições sobre as populações em jogo. Estas hipóteses, que podem ou não serem verdadeiras, chamam-se *hipóteses estatísticas* e, em geral, consistem de afirmações sobre as distribuições de probabilidade das populações. Por exemplo, se queremos decidir se determinada moeda é viciada, formulamos a hipótese de que a moeda é verdadeira, isto é p = 0,5, onde *p* é a probabilidade de ‘caras”. Analogamente, se desejamos decidir se determinado processo é melhor do que outro formulamos a hipótese de que *não há diferença* entre os dois processos, isto é, quaisquer diferenças observadas são devidas a meras flutuações de amostragem da mesma população. Tais hipóteses são chamadas de *hipóteses nulas* e se denotam pro H0.

Qualquer hipótese que difere de uma dada hipótese nula é chamada de *hipótese alternativa*. Por exemplo, se a hipótese nula é p = 0.5, a hipótese alternativa possível é p = 0.7, p ≠ 0.5 ou p > 0.5. A hipótese alternativa é denotada por H1.

**TESTES DE HIPOTESES E SIGNIFICÂNCIA**

Se na suposição de que uma determinada hipótese é verdadeira nós achamos que os resultados observados em uma amostra aleatória diferem daquelas esperadas sob a hipótese com base no puro acaso usando a teoria de amostragem, podemos dizer que as diferenças observadas são *significativas* e que estaria inclinado a rejeitar a hipótese (ou pelo menos não aceitá-la com base nos dados obtidos). Por exemplo, se em 20 jogadas de uma moeda obtemos 16 caras, estaríamos inclinados a aceitar a hipótese de que a moeda é viciada, embora possamos estar errados.

**VOCÊ PRECISA SABER!**



Procedimentos que nos permitem decidir entre aceitar ou rejeitar uma hipótese ou determinar se as amostras observadas diferem significativamente dos resultados esperados são chamados de *testes de hipóteses, testes de significância ou regras de decisão.*

**ERROS DO TIPO I E DO TIPO II**

Se rejeitarmos uma hipótese, quando deveríamos aceita-lá, dizemos que um erro do tipo I foi cometido. Se, por outro lado, aceitamos uma hipótese quando ela deveria ser rejeitada, dizemos que foi cometido um erro do tipo II. Em ambos os casos, ocorre uma decisão errada ou erro de julgamento.

Para que qualquer teste de hipótese ou regras de decisão seja considerado bom, eles devem ser concebidos de modo a minimizar os erros de decisão. Esta não é uma questão simples, já que, para uma determinada dimensão da amostra, uma tentativa de diminuir um tipo de erro, em geral, é acompanhada por um aumento no outro tipo de erro. Na prática, um tipo de erro pode ser mais grave do que o outro, e assim um objetivo deverá ser alcançado em favor de uma limitação do erro mais grave. A única maneira de reduzir os dois tipos de erros é aumentar o tamanho da amostra, que pode ou não pode ser possível.

**NÍVEL DE SIGNIFICÂNCIA**

Ao testar uma hipótese dada, a probabilidade máxima com a qual estaríamos dispostos a arriscar um erro do tipo I é denominada nível de significância do teste. Essa probabilidade é frequentemente especificada antes de qualquer amostra ser colhida para que os resultados obtidos não influenciem na decisão.

**PROBABILIDADE E ESTATÍSTICA**

Na prática, um nível de significância de 0,05 ou 0,01 é habitual, embora outros valores também sejam utilizados. Se, por exemplo, um nível de 0,05 ou 5% de significância é escolhido na concepção de um teste de uma hipótese, então há cerca de 5 chances em 100 de que iríamos rejeitar a hipótese quando ela deve ser aceita, ou seja, sempre que a hipótese nula é verdadeira, somos cerca de 95% confiantes de que iríamos tomar a decisão certa. Nesses casos, dizemos que a hipótese foi rejeitada com um nível de significância de 0,05, o que significa que pode estar errado com probabilidade 0,05.

**NOTA!**

Escolher o nível de significância, antes de começar os testes em muitos dos casos pode ajudá-lo na escolha entre aceitar ou rejeitar uma hipótese nula.

**TESTE ENVOLVENDO A DISTRIBUIÇÃO NORMAL**

Para ilustrar as idéias apresentadas acima, suponha que sob a hipótese dada, a distribuição de amostragem de uma estatística S é uma distribuição normal com médias µS e desvio padrão σS. A distribuição desse padrão variável Z = (S - µS) / σS é a distribuição normal padrão (média 0, variância 1) mostrado na Figura 8.1, e os valores extremos de Z levaria à rejeição da hipótese.

 Região Crítica 0,95 Região Crítica

 ***0,025 0,025***

 z =-1,96 z=1,96

 Figura 8.1: Mostra os valores extremos de Z.

Como indicado na figura, podemos estar 95% confiantes de que, se a hipótese for verdadeira, o escore z de uma amostra estatística S real estará entre - 1,96 a 1,96 (pois a área sob a curva normal entre esses dois valores é 0,95).

No entanto, se na escolha de uma única amostra aleatória, verificamos que o escore z esta fora da faixa de - 1,96 a +1,96, iríamos concluir que um evento como esse poderia acontecer com a probabilidade de apenas 0,05 (área diferenciada na figura) se a hipótese apresentada for verdadeira. Nós, então, dizemos que este escore z difere *significativamente* do que seria esperado sob a hipótese, e que estaríamos inclinados a rejeitar a hipótese.

A área total diferenciada na Figura 8.1, 0,05 é o nível de significância do teste. Representa a probabilidade de erro na rejeição de uma hipótese, ou seja, a probabilidade de cometer um erro do tipo I. Portanto, dizemos que a hipótese é rejeitada a um nível de significância de 0,05 ou que o escore Z da estatística de amostra dado é significativo a um nível de significância de 0,05.

O conjunto de escores z fora do intervalo de - 1,96 a +1,96 constitui o que se chama de *região crítica ou região de rejeição da hipótese ou ainda região de significância.* O conjunto de escores z dentro do intervalo de - 1,96 a +1,96 poderia ser chamado de *região de aceitação da hipótese ou região de não-significância.*

Com base nas observações acima, podemos formular a seguinte regra de decisão:

1. Rejeitar a hipótese ao nível de significância de 0,05, se o escore Z da estatística de S está fora da faixa - 1,96 a +1,96 (ou seja, se z > 1,96 ou z < - 1,96). Isto equivale a dizer que a estatística da amostra observada é insignificante ao nível de 0,05.
2. Caso contrário, aceitar a hipótese (ou, se preferir, não tomar nenhuma decisão).

Perceba que outros níveis de significância poderiam ter sido utilizados. Por exemplo, se um nível de 0,01 fosse utilizado nos iríamos substituir 1,96 por 2,58 (ver Tabela 8.1).

**TESTES UNICAUDAL E BICAUDAL**

No teste acima nós demonstramos interesse em valores extremos da estatística S, ou seja, em seus escores z em ambos os lados da média, isto é, em ambas as caudas da distribuição. Por esta razão, tais testes são denominados testes bilaterais ou testes de duas faces.

Muitas vezes, porém, podemos estar interessados apenas em valores extremos de um lado da média, ou seja, em uma cauda da distribuição como, por exemplo, quando estamos testando a hipótese de que um processo é melhor que outro (que é diferente de testar se um processo é melhor ou pior que o outro). Esses testes são chamados de testes unicaudal ou testes unilaterais. Nesses casos, a região é uma região crítica para um dos lados da distribuição, com área igual ao nível de significância.

Tabela 8.1: Mostra os valores de significância para testes unicaudal e bica**udal.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Níveis de Significância** | **0,10** | **0,05** | **0,01** | **0,005** | **0,002** |
| Valores críticos de z para testes bicaudal |  -1,28ou 1,28 |  -1,645ou 1,645 |  -2,33ou 2,33 |  -2,58ou 2,58 |  -2,88ou 2,88 |
| Valores críticos de z para testes bicaudal |  -1,645e 1,645 |  -1,96e 1,96 |  -2,58e 2,58 |  -2,81e 2,81 |  -3,08e 3,08 |

**VALOR DE P**

Na maioria dos testes, vamos considerar a hipótese nula Ho será uma afirmação de que um parâmetro da população tem um valor específico, e a hipótese alternativa H1 será uma das seguintes afirmações:

1. O parâmetro é maior do que o valor declarado (teste bicaudal da direita).
2. O parâmetro é menor que o valor declarado (teste bicaudal da esquerda).
3. O parâmetro pode ser superior ou inferior ao valor declarado (teste bicaudal).

Nos casos (i) e (ii), H1 possui uma única direção no que diz respeito ao parâmetro e, no caso (iii), H1 é bi-direcional. Após o ensaio do teste da estatística de S computadorizada ter sido realizado, o valor de P no teste é a probabilidade de que um valor de S na(s) direção(s) H1 é tão extremo tanto quanto ocorreria se H0 fosse verdadeira.

Por exemplo, suponha que o desvio-padrão σ de uma população normal é 3, e H0 afirma que a média de µ é igual a 12. Uma amostra aleatória de tamanho 36, oriundos da população produz uma média da amostra x = 12,95. A estatística de teste é:



o qual se H0 for verdadeira, é a variável normal padrão. O valor do teste de Z é o seguinte:



O valor de P para o teste, então depende da hipótese alternativa H1 como segue:

(i) Para H1: µ > 12 [caso (i) acima], o valor P é a probabilidade de que uma amostra aleatória de tamanho 36 renderia uma amostra média de 12,95 ou mais, se a média de verdade fosse 12, ou seja, P ( Z > 19) = 0,029. Em outras palavras, as chances são de aproximadamente 3 em cada 100 que X ≥ 12,95 se µ = 12.

(ii) Para H1: µ < 12 [caso (ii) acima], o valor P é a probabilidade que uma amostra aleatória de tamanho 36 renderia uma amostra média de 12.95 ou menor se a verdadeira média for 12, i. e. P ( Z ≤ 19) = 0.971. Em outras palavras, as chances são de 97 em 100 de que se 

(iii) Para H1: [caso (iii) acima], o valor P é a probabilidade de que uma amostra aleatória de 0.95 ou mais unidades acima de 12, i. e. ou , se a verdadeira média for 12. Aqui o valor P é P () + P() = 0.057, o que diz que as chances são de 6 em 10 para que se 

Pequenos valores de P providenciam evidências para rejeitar a hipótese nula em favor de hipóteses alternativas, e grandes valores de P providenciam evidências para não rejeitar a hipótese nula em favor das hipóteses alternativas. No caso (i) do exemplo acima, o pequeno valor P 0.029 é um indicador razoavelmente forte de que a média da população é maior que 12, enquanto que no caso (ii), o grande valor P 0.971 sugere fortemente que H0: não deve ser rejeitado em favor de H0: . No caso (iii), o valor P 0.057 providencia evidências para rejeitar H0 em favor de H0: , mas não com a mesma evidência com que forneceu para rejeitar H0 em favor de H0: .

 Deve-se ter em mente que o valor P e o nível de significância não fornecem critério para rejeitar ou não a hipótese nula por si mesma, mas para rejeitar ou não a hipótese nula em favor de hipóteses alternativas. Como o exemplo anterior ilustra, resultados idênticos de testes e níveis de significância diferentes podem levar a diferentes conclusões sobre a mesma hipótese nula em relação a diferentes hipóteses alternativas.

 Quando o teste estatístico S é a variável aleatória normal, a tabela no apêndice B é suficiente para computar o valor P, mas quando S é uma das t, F ou bi-quadrado variáveis aleatórias, todas estas têm diferentes distribuições dependendo de seus graus de liberdade, qualquer software de computador ou tabelas mais extensas que aquelas nos Apêndices C, D e E serão necessárias para computar o valor P.

**Exemplo 8.1.** A vida média de uma amostra de 100 lâmpadas fluorescentes produzidas por uma companhia é computada como 1570 horas com um desvio padrão de 120 horas. Se  é a vida média de todas as lâmpadas produzidas pela companhia, teste a hipótese  horas contra a hipótese alternativa  horas. Use um nível de significância de 0.05 e encontre o valor P do teste.

Nós precisamos decidir entre duas hipóteses:

H0:  horas H0:  horas

Um teste bilateral deve ser utilizado aqui desde que  inclua ambos os valores maiores e menores que 1600.

Para um teste bilateral com nível de significância 0.05, nós temos a seguinte regra de decisão:

1. Rejeitar H0 se o escore z da média da amostra estiver fora da faixa -1.96 a 1.96
2. Aceitar H0 (ou reter uma decisão) caso contrário.

A estatística sob consideração é a média da amostra . A distribuição da amostra de X tem uma média  e desvio-padrão , onde  e  são a média e o desvio-padrão da população de todas as lâmpadas produzidas pela companhia.

Sob a hipótese H0, nós temos  e , usando o desvio-padrão da amostra como uma estimativa de . Como  está fora do intervalo de -1.96 a 1.96, nós rejeitamos H0 no nível de significância 0.05.

O valor P do teste bilateral é , com a probabilidade de que uma vida média abaixo de 1570 horas ou acima de 1630 horas poderia ocorrer caso H0 fosse verdadeira.

## TESTES ESPECIAIS

 Para grandes amostras, muitas estatísticas S têm distribuições quase normais com média  e desvio-padrão . Em tais casos, nós podemos usar os resultados acima para formular regras de decisão ou testes de hipóteses e significância. Os seguintes casos especiais são apenas algumas das estatísticas de interesse prático. Em todos os casos, os resultados são válidos para as populações infinitas ou para amostras com reposição. Para amostras sem reposição em populações finitas, os resultados precisam ser modificados. Nós só consideraremos os casos de grandes amostras ().

### 1. Médias

Aqui , a média da amostra; , a média da população; , onde  é o desvio-padrão da população e n é o tamanho da amostra. A variável normalizada é dada por:

  (1)

quando necessário, o desvio-padrão observado na amostra, s (ou ), é usado para estimar .

 Para testar a hipótese nula H0 de que a média da população é , nós iremos usar a estatística (1). Então, se a hipótese alternativa é , usando um teste bilateral, nós iremos aceitar H0 (ou pelo menos não rejeita-la) no nível 0.05 se uma amostra particular de tamanho n tiver média 

  (2)

e iremos rejeitar caso contrário. Para outros níveis de significância nós iremos alterar (2) apropriadamente. Para testar H0 contra a hipótese alternativa de que a média da população é maior que a, nós usaremos um teste unilateral e aceitaremos H0 (ou pelo menos não rejeita-la) no nível 0.05 se:

 **** (3)

(ver tabela 8.1) e rejeita-la caso contrário. Para testar H0 contra a hipótese alternativa de que a média da população é menor que a, nós aceitaremos H0 no nível 0.05 se:

 **** (4)

### 2. Proporções

Aqui S = P, a proporção de “sucessos” em uma amostra; , onde P é a proporção da população de sucessos e n é o tamanho da amostra; , onde q = 1 – p. A variável normalizada é dada por:

  (5)

No caso P=X/n, onde X é o atual número de sucessos em uma amostra, (5) torna-se:

  (6)

Observações semelhantes às feitas acima de testes uni e bilaterais para médias que podem ser feitas.

### 3. Diferenças de médias

Sejam  e  as médias amostrais obtidas em amostras grandes de tamanhos n1 e n2 extraídas das respectivas populações de médias  e  e desvios-padrões  e . Considere a hipótese nula de que não há diferença entre as médias das populações, i.e., . Da nossa discussão sobre as distribuições de amostragem de diferenças e somas (Capítulo 6), na colocação  vemos que a distribuição da amostragem de diferenças nas médias é aproximadamente normal com média e desvio-padrão dados por:

  (7)

Onde nós podemos, se necessário, usar os desvios-padrões observados da amostra s1 e s2 (ou  e ) como estimativas de  e .

Usando a variável normalizada dada por:

  (8)

De modo similar ao descrito na Parte 1 acima, nós podemos testar a hipótese nula contra uma hipótese alternativa (ou a significância de uma diferença observada) em um nível de significância apropriado.

### 4. Diferenças de proporções

Sejam P1 e P2 as proporções amostrais obtidas em amostras grandes de tamanhos n1 e n2 extraídas das respectivas populações que têm proporções p1 e p2. Considere a hipótese nula de que não há diferença entre as proporções das populações, i.e., p1 = p2, e, assim, as amostras são realmente extraídas da mesma população.

 Da nossa discussão sobre diferenças de proporções no Capítulo 6, na colocação p1=p2=p, nós vemos que a distribuição amostral de diferenças em proporções é aproximadamente normal com média e desvio-padrão dados por:

  (9)

Onde  é usado como uma estimativa da proporção p da população.

Usando a variável normalizada:

  (10)

Nós podemos observar diferenças em um apropriado nível de significância e, assim, testar a hipótese nula.

 Testes envolvendo outras estatísticas podem ser concebidos de modo similar.

## RELAÇÃO ENTRE A TEORIA DA ESTIMAÇÃO E O TESTE DE HIPÓTESES

 A partir das observações acima, não se pode deixar de notar que existe uma relação entre a teoria que envolve estimativa de intervalos de confiança e a teoria dos testes de hipóteses. Por exemplo, nós notamos que o resultado (2) para aceitação de H0 no nível de 0.05 é equivalente ao resultado (1) no Capítulo 7, levando para o intervalo de confiança de 95 %.

  (11)

Então, pelo menos no caso de testes bilaterais, nós podemos atualmente empregar os intervalos de confiança do Capítulo 7 para testar a hipótese. Um resultado similar para testes unilaterais irá requerer intervalos de confiança unilaterais.

**Exemplo 8.2.** Considere o exemplo 8.1. Um intervalo de confiança de 95 % para o exemplo 8.1. é o seguinte:



O que é:



Isto leva para um intervalo de (1546.48, 1593.52). Note que este não contém a média alegada de 1600, então leva-nos a rejeitar H0.