TESTE DE HIPÓTESES

March 13, 2003

1 Introdução

Existem duas grandes áreas na inferência estatística: estimação de parâmetros e teste de hipóteses. Estudaremos a segunda área neste capítulo. Nosso objetivo será apresentar os conceitos e desenvolver métodos gerais para testes de hipóteses e aplicá-los a alguns problemas comuns.

Em uma pesquisa experimental o objetivo pode ser simplesmente o de estimar parâmetros; por exemplo, podemos estar interessados em estimar a produtividade média de uma nova linhagem de milho híbrido. Porém, o objetivo final pode ser a utilização desta estimativa para outras inferências. Podemos, por exemplo, comparar esta produtividade com uma linhagem padrão e talvez recomendar que a linhagem padrão seja substituída pela nova. Esta é uma situação comum na pesquisa. Em outras situações podemos estar interessados em determinar se uma nova liga para filamentos de lâmpadas é melhor que o utilizado atualmente; se um novo germicida é mais eficaz que um padrão; se o novo método empregado pela concorrência para preservar comidas é mais efetivo em termos de manutenção de vitaminas; e assim por diante.

Observe que em alguns exemplos anteriores não foram colocados de forma clara o que determina se um novo método é melhor do que o anterior. Por exemplo, no caso das lâmpadas, poderíamos levar em consideração vários objetivos: tempo médio de vida, custo médio de produção, proporção de lâmpadas que duram pelo menos 1000 horas, consumo de energia, etc. Neste capítulo e durante o curso não discutiremos com detalhes a escolha das características a serem testadas, sendo sempre considerado que sabemos de antemão a característica de interesse e que esta característica é unidimensional. Isto deve-se somente aos objetivos do curso e não por acreditarmos que este é o ponto mais importante; geralmente a outra discussão é mais importante.

Dado ao exposto no parágrafo anterior considere que no caso das lâmpadas estamos interessados no tempo médio de vida das lâmpadas com filamentos padrão e o novo. Fica claro que no modelo estatístico estamos tratando de duas populações de lâmpadas. Se sabemos de experiências passadas que o tempo médio de vida da população padrão é de 1400 horas a questão é saber se o tempo médio da nova população é maior ou menor do que 1400 horas. Tradicionalmente, para responder a esta questão nós estamos procurando evidências em favor da hipótese de que o novo filamento é melhor. Para testar a hipótese, uma certa quantidade de lâmpadas é fabricada com o novo filamento e medidos os seus tempos de vida. Suponha que o tempo médio da amostra seja de 1550 horas. A indicação é de que o novo processo é melhor, mas suponha que a estimativa do desvio padrão da média, $\hat{\sigma}/\sqrt{n} = 150$, (n sendo igual ao tamanho da amostra). Então o IC 95% para a média da segunda população (assumindo normalidade caso n seja pequeno e utilizando o TCL caso n seja suficientemente grande) é aproximadamente 1300 a 1800 horas. A média amostral poderia ter vindo facilmente de uma população com média 1400 horas, e, portanto, não temos grande evidência para rejeitar a hipótese de que o novo tipo de liga não é melhor do que o padrão. Se, por outro lado, $\hat{\sigma}/\sqrt{n} = 25$, poderíamos com bastante confiança dizer que a nova liga é superior. Novamente lembramos que é superior levando-se em consideração somente o tempo médio de vida. O exemplo mostra que teste de hipóteses está estreitamente relacionado com o problema de estimação. No entanto, apenas por uma questão didática, será interessante, pelo menos no início, desenvolver a teoria de teste de hipóteses independentemente da de estimação. No próximo exemplo consideramos o tratamento de um problema do ponto de vista puramente de teste de hipóteses.

Exemplo 1.1: Suponha que você trabalhe em uma empresa que necessita um parafuso especial importado que satisfaça a certas exigência de resistência à tração (R). Suponha que um dado bem conhecido é que os parafusos americanos tem resistência à tração $R \sim N(145, 144)$ e os parafusos japoneses têm $R \sim N(155, 400)$. Um lote de origem desconhecida será leiloado. O edital afirma que, pouco antes do leilão será divulgada a resistência média $\bar{r}$ de uma amostra de 25 peças do lote. A fim de fazer um lance, você precisa saber de que procedência é o lote. Que regra de decisão deve-se usar para decidir a procedência do lote?

Regra de decisão: Iniciamos vamos considerar uma regra bastante intutitiva: "Se $\bar{r} \leq C_0$ decido que a procedência é americana" e "se $\bar{r} > C_0$ decido que a procedência é japonesa".

Suponha que no dia do leilão fomos informados que $\bar{r} = 148$. Para podermos decidir é necessário que o valor C_0 seja fixado. De qualquer forma sabemos que $\bar{R}$ é uma variável aleatória e qualquer que seja a escolha de C_0 ela não vai me garantir uma decisão correta em todos os casos.

Para facilitar a análise considere que seja escolhido $C_0 = 150$. Depois ficará claro como este valor deve ser escolhido. Com base nesse valor de C_0 e no valor de $\bar{r}$ encontrado decidimos que a procedência é americana.

Para melhor entendermos esta regra de decisão devemos estudar os tipos de erros possíveis:

erro A: Dizer que os parafusos são americanos quando, na realidade, são japoneses;
erro B: Dizer que os parafusos são japoneses quando, na realidade, são americanos.

Portanto, podemos dizer que:

$$\begin{aligned}
\mathbf{P}[\text{erro A}|\text{parafusos japoneses}] &= \mathbf{P}[\bar{R} \leq 150|\text{parafusos japoneses}] \\
&= \mathbf{P}[\bar{R} \leq 150|\bar{R} \sim N(155; 16)] \\
&= \mathbf{P}[\frac{\bar{R} - 155}{4} \leq \frac{150 - 155}{4}] \\
&= \mathbf{P}[Z \leq -1,25] = 0,1056 = \alpha
\end{aligned}$$

e

$$\begin{aligned}
\mathbf{P}[\text{erro B}|\text{parafusos americanos}] &= \mathbf{P}[\bar{R} > 150|\text{parafusos americanos}] \\
&= \mathbf{P}[\bar{R} > 150|\bar{R} \sim N(145; 144/25)] \\
&= \mathbf{P}[\frac{\bar{R} - 145}{12/5} > \frac{150 - 145}{12/5}] \\
&= \mathbf{P}[Z > 2,08] = 0,0187 = \beta
\end{aligned}$$

Com a regra de decisão adotada temos que $\mathbf{P}$[erro A|parafusos japoneses] > $\mathbf{P}$[erro B|parafusos americanos]. Ou seja, de certa forma estamos privilegiando a decisão de dizer que os parafusos são americanos.

Note que mudando o valor de C_0 mudamos os valores de α e β. Se $C_0 < 150$, aumentamos α e diminuímos β; Se $C_0 > 150$, diminuímos α e aumentamos β. Daí, por exemplo, podemos fazer com que $\alpha = \beta$ com $C_0 = 148,75$ e com isto, $\alpha = \beta = 0,0594$. Isto é, podemos escolher C_0 para conseguir diferentes valores de α e β. Outra forma é fixar α e encontrar C_0, tal que $\mathbf{P}$[erro A|parafusos japoneses] $= \alpha$. Por exemplo, se $\alpha = 0,05$, fazemos

$$\begin{aligned} 0,05 &= \mathbf{P}[\text{erro A}|\text{parafusos japoneses}] \\ &= \mathbf{P}[\bar{R} \leq c_0 | \bar{R} \sim N(155;16)] \\ &= \mathbf{P}[Z \leq -1,64] \end{aligned}$$

$$-1,64 = \frac{c_0 - 155}{4} \Rightarrow C_0 = 148,4\ ,$$

que dá a seguinte *Regra de decisão:* "Se $\bar{r} \leq 148,4$ decido que a procedência é americana" e "se $\bar{r} > 148,4$ decido que a procedência é japonesa". Daí,

$$\begin{aligned} \mathbf{P}[\text{erro A}|\text{parafusos japoneses}] &= \mathbf{P}[\bar{R} > 148,4 | \bar{R} \sim N(145;5,76)] \\ &= \mathbf{P}[Z > 1,425] = 0,0793. \end{aligned}$$

Por outro lado poderíamos fixar β. Já mencionamos que com esta regra de decisão ao diminuirmos a probabilidade de um tipo de erro aumentamos a probabilidade do outro tipo de erro. Logo, uma pergunta que devemos fazer é: existe uma regra de decisão, não necessariamente do tipo: $\bar{R} > C_0$, que leve a um melhor balanço entre as probabilidades de erros?. Por exemplo, fixado α ou β existe uma regra que minimize a probabilidade de outro tipo de erro? Será que devo dar a mesma importância a estes 2 tipos de erros? Algumas destas perguntas serão respondidas neste capítulo.

2 Conceituação de Teste de Hipóteses

Cientistas, engenheiros de controle de qualidade, pesquisadores de mercado, técnicos governamentais, entre outros profissionais, frequentemente fazem hipóteses sobre seus campos de interesse, hipóteses estas que necessitam serem verificadas ou substanciadas. Para este fim, eles coletam dados e deixam os dados confirmarem ou rejeitarem esta hipótese. Este processo é chamado de "teste de hipóteses" e é uma das áreas principais da inferência estatística. Para facilitar a discussão incialmente serão introduzidas algumas linguagens, notações e definições. Assim como em estimação assumiremos que podemos obter uma amostra aleatória de uma densidade $f(\cdot;\theta)$.

Definição 2.1 Hipótese Estatística: Uma hipótese estatística é uma afirmação ou conjectura sobre a distribuição de uma ou mais variáveis aleatórias. Se a hipótese especifica completamente a distribuição, então ela é chamada de *hipótese simples*; caso contrário ela é chamada de *hipótese composta.*

Definição 2.2 Teste de Uma Hipótese Estatística : Um teste de uma hipótese estatística H é uma regra ou procedimento, para decidir se rejeitamos ou não a hipótese H. Vamos sempre considerar que a regra de decisão será baseada nos dados amostrais

Exemplo 3.1: Julgamento Um juri deve decidir com base em evidências se o acusado é culpado ou inocente.

Neste caso é evidente que existem 2 hipóteses sendo consideradas. Na primeira hipótese o réu é inocente e na segunda o réu é o culpado. Partindo da máxima da justiça que o a réu é considerado "inocente" a menos que haja forte evidência de culpa vemos que inicialmente a primeira hipótese é aceita antes do julgamento. Além disso nós ficamos com ela a menos que haja forte evidência em contrário. Por outro lado, o julgamento é realizado exatamente para procurar evidências de que a outra hipótese é verdadeira. O que está por trás deste comportamento é que é preferível cometer o erro de não condenar um culpado (não rejeita a primeira hipótese quando ela é falsa) do que cometer o erro de condenar um inocente (rejeitar primeira hipótese quando ela é verdadeira).

Exemplo 2.2: Estamos testando um novo remédio (NEW) a ser utilizado no tratamento de AIDS. O medicamento já estabelecido é o AZT e sua utilização prolonga o tempo de vida do paciente, em média por k_0 meses. O laboratório que produz NEW diz que seu medicamente é melhor e um teste foi feito com 12 pacientes. Dizemos que:

H_A: NEW não é melhor que AZT, i.e., $\mu_{NEW} \leq \mu_{AZT}$;
H_B: NEW é melhor que AZT, i.e., $\mu_{NEW} > \mu_{AZT}$.

Note que um erro mais grave é dizer que o NEW é melhor que AZT quando na realidade não é (e todos os pacientes mudam para um medicamento menos eficaz), do que dizer que NEW não é melhor que AZT quando na realidade ele é (e todos os pacientes continuam com um medicamento que é comprovado). Veja que, só desejamos mudar para um medicamento novo quando há fortes evidências que este seja melhor.

Na maioria dos problemas de teste de hipóteses, como os discutidos até aqui, duas hipóteses são discutidas o que dá origem a dois tipos de erros. A discussão dos dois exemplos anteriores mostrou que os dois tipos de erros normalmente levam a custos diferentes. Na tomada de decisão as hipóteses são denominadas de hipóteses nula e alternativa. Antes da realização do experimento temos uma hipótese que é considerada como correta, e ela deve ser mantida a menos que os dados mostrem evidências suficientes (suficientemente forte) para rejeitá-la. Por exemplo, em um julgamento o réu é considerado inocente a menos que haja evidências que o júri ou juiz acreditem serem suficientes para que a hipótese de inocência possa ser refutada, enquanto no caso das drogas a droga NEW não é considerada melhor a menos que haja prova em contrário. Esta hipótese conservadora é normalmente chamada de hipótese nula e denotada por H_0. Esta denominação originou-se em experimentos comparativos em que um novo produto ou nova técnica é comparado com um padrão para verificar se a sua superioridade pode ser comprovada através de evidências experimentais. Neste contexto, a hipótese nula é a afirmativa de que a diferença entre o produto padrão e o novo é nulo ou zero. A mudança para um novo produto ou técnica geralmente requer uma grande despesa inicial, e o tomador de decisão não deve fazer isto a menos que o novo produto mostre ser claramente melhor do que o antigo. Então a perda potencial por rejeitar incorretamente a hipótese nula é maior do que aceitar incorretamente a hipótese nula. Desta forma, procuramos manter a probabilidade de se rejeitar a hipótese nula, quando ela é verdadeira, sob controle.

Outra forma de distinguir as hipóteses é verificar que todo o experimento (julgamento no *Exemplo 2.1* e teste no *Exemplo 2.2*) é realizado com o objetivo de verificar se existem evidências em favor da outa hipótese, a chamada hipótese alternativa. Ou seja, desejamos provar que a hipótese alternativa é verdadeira. A hipótese nula seria a negação da hipótese que queremos provar. A hipótese alternativa é normalmente denotada por H_1 ou A.

Definição 2.3: Se um teste é baseado em uma certa estatística, $T = t(X_1, \ldots, X_n)$, então T é chamada de *estatística do teste.* A decisão pode ou não depender apenas do valor assumido pela estatística do teste.

Definição 2.4: Teste Não Aleatorizado e Região Crítica : Seja um teste δ, de uma hipótese estatística H_0, definida como: Rejeite H_0 se e somente se $(x_1, \ldots, x_n) \in C_\delta$, onde C_δ é um subconjunto de R^n; então δ é chamado de *Teste Não Aleatorizado* e C_δ é chamada de *Região Crítica* do teste δ.

Definição 2.5: Teste Aleatorizado : Um teste δ, de uma hipótese estatística H_0, é definido como sendo um *Teste Aleatorizado* se δ é definido por uma função

$$\psi_\delta(x_1, \ldots, x_n) = \text{P}[H_0 \text{ ser rejeitado}|(x_1, \ldots, x_n) \text{ é observado}]$$

A função $\psi_\delta(x_1, \ldots, x_n)$ é chamada *Função Crítica do Teste.*

A aplicação de um teste não aleatorizado é bastante simples. Se o resultado amostral cair na região crítica rejeita-se a hipótese nula, caso contrário não se rejeita a hipótese nula. No caso do teste aleatorizado, observada uma amostra $(x_1, \cdots, x_n)$ e calculado $\psi_\delta(x_1, \cdots, x_n)$ é necessário realizar um ensaio de Bernoulli cuja probabilidade de sucesso é igual a $\psi_\delta((x_1, \ldots, x_n)$. Se o resultado for sucesso rejeita-se a hipótese nula. Este procedimento é discutível dado que o resultado do ensaio e, portanto, da decisão, é independente da informação dada pelo experimento e de quaisquer informações sobre o(s) processo(s). Por esta razão o teste aleatorizado não é normalmente adotado. Quando adotado, geralmente divide o espaço amostral em três regiões; uma onde a hipótese nula é rejeitada, outra onde não é rejeitada e uma terceira onde a aleatorização é realizada. Nas duas primeira regiões existe confiança para rejeitar ou não a hipótese nula, enquanto na terceira, que geralmente fica entre ou na fronteira das duas primeiras, fica difícil decidir por uma ou outra. Uma alternativa é chamar a terceira região de não conclusiva e realizar testes adicionais. Assim como no teste não aleatorizado a região crítica definia o teste, a função crítica define o teste aletorizado. Observe também que um teste não aleatorizado também pode ser definido por uma função crítica que toma apenas valores 0 ou 1. Neste texto, a menos que colocado explicitamente, sempre trataremos de testes não aleatorizados.

O procedimento descrito pelas duas definições anteriores, que será utilizado inicialmente na construção de certos testes, considera teste de hipóteses do ponto de vista da teoria da decisão. O desenvolvimento da sua teoria deve-se muito ao seu mais ilustre defensor, o Professor Neyman. Uma forma alternativa de procedimento, que será utilizado com mais frequência nas análises realizadas durante o curso, são os *testes de significância.* Nesta abordagem não se trabalha com regras de decisões, mas procura se dar uma medida de "evidência" da hipótese alternativa ser verdadeira, e é a abordagem de Fisher.

Definição 2.6 Tipos de Erros e "Tamanhos" de Erros : A rejeição de H_0 quando ela é verdadeira é chamada de erro do tipo I, e a aceitação de H_0 quando ela não é verdadeira é chamada de erro do tipo II. O tamanho do erro do tipo I (II) é definida como sendo a probabilidade de se cometer um erro do tipo I (II) quando H_0 (não) é verdadeira.

Definição 2.7 Função Poder : Seja δ um teste da hipótese H_0 . A função poder do teste δ, denotada por $\beta(\theta, \delta)$ é definida pela probabilidade de que H_0 seja rejeitada quando a amostra foi obtida de uma distribuição parametrizada por θ.

Dada uma função crítica $\psi_\delta(X_1, \ldots, X_n)$ temos que $\beta(\theta, \delta) = E_\theta[\psi_\delta]$

Definição 2.8 Tamanho de um teste : Seja δ um teste para a hipótese $H_0 : \theta \in \Theta_0$, onde $\Theta_0 \subset \Theta$, isto é, Θ_0 é um subconjunto do espaço paramétrico. O *Tamanho do Teste* é definido como:

$$t(\delta) = \sup_{\theta \in \Theta_0} \beta(\theta, \delta)$$

Muitas vezes $t(\delta)$ é chamado o *Nível de Significância* do Teste δ.

O tamanho do teste para um teste não aleatorizado também é referido como sendo o tamanho da região crítica.

O papel da função poder em teste de hipóteses é análogo ao do erro quadrático médio em estimação. Ela é normalmente utilizada na comparação entre dois testes competitivos. Uma função poder ideal é igual a zero quando θ pertence à hipótese nula, e igual a 1 quando θ pertence à hipótese alternativa. Este teste dificilmente existe.

MGB comentam na pag. 407 a utilização dos termos nível de significância e tamanho do teste. Segundo MGB os dois termos são intercambiáveis para muitos autores, mas que eles preferem utilizar o termo nível de significância para ser utilizado em testes de significância. A abordagem de teste de significância terá prioridade neste curso. Os termos nível de significância e tamanho do teste como intercambiáveis, enquanto, ao se adotar a abordagem de teste de significância, utilizaremos os termos probabilidade de significância, valor descritivo do teste, valor-p e p-value.

Em estimação chegamos que, sob certas considerações, podemos nos restringir a funções de estatísticas suficientes. O mesmo ocorre em teste de hipóteses quando a função poder é utilizada como base de comparação, como mostra o teorema a seguir.

Teorema 2.1: Se $X_1, \ldots, X_n$ é uma amostra aleatória de $f(\cdot;\theta)$, onde $\theta \in \Theta$, e $S = (s_1(X_1, \ldots, X_n), \ldots, S_r = s_r(X_1, \ldots, X_n))$ é um conjunto de estatísticas suficientes, então para qualquer teste δ com função crítica ψ_δ , existe um teste, digamos, δ', e uma correspondente função crítica, digamos $\psi_{\delta'}$, dependendo somente do conjunto de estatísticas suficientes, que satisfazem $\beta(\theta, \delta) = \beta(\theta, \delta')$ para todo $\theta \in \Theta$.

Assim como em estimação o problema de teste de hipóteses tem dois aspectos: primeiro são necessários métodos para encontrar testes e depois uma forma de compará-los. Embora estejamos interessados nestes dois problemas a discussão não seguirá esta ordem. Na *Seção 3* será discutido o teste entre duas hipóteses simples. Serão utilizadas duas abordagens: na primeira a função poder será utlizada como critério de bondade, enquanto no segundo será utilizada a função perda. Depois é apresentado e provado o importante lema de Neyman-Pearson. Todos estes testes, que são ótimos segundo algum critério, podem ser expressas na forma da razão de verossimilhança simples, que é definida na *subseção 3.1.*

Os testes para hipóteses compostas são discutidos na *Seção 4*. A *subseção 4.1* apresenta o princípio da razão de verossimilhança generalizada. Esta técnica tem um papel central em teste de hipóteses, assim como os estimadores de máxima verossimilhança em estimação. Sua importância vem do fato de que a maioria dos testes derivados a partir deste princípio são bons. O conceito de teste uniformemente mais poderoso e alguns exemplos são apresentados na *subseção 4.2.* Como muitas vezes estes testes não existem, assim como em estimação, restringimos a classe de testes de tal forma que possamos encontrar o "melhor" dentro desta classe. Uma forma de restringir esta classe é utlizar a propriedade de invariância e de não viciado. Apenas o segundo seá discutido superficialmente no curso.

A *Seção 5* discutirá vários exemplos de testes de hipóteses em distribuições normais. A *Seção 6* discutirá alguns testes classificados sob a designação genérica de testes chi-quadrados. Entre eles estão incluídos a distribuição assintótica do teste da razão de verossimihança generalizada, o teste de bondade de ajuste, o teste de igualdade entre duas ou mais distribuições e testes de independência em tabelas de contingência. Finalmente, na *Seção 7* será discutida de forma mais formal a relação entre testes de hipóteses e intervalos de confiança. Na discussão, é normalmente é introduzido os intervalos de confiança "ótimos"; porém eles não são objeto de estudo no curso.

Exemplo 2.3: Seja $X_1, \ldots, X_{25}$ i.i.d. $N(\theta; 25)$. Queremos testar:
$H_0 : \theta \leq 17$ versus $H_1 : \theta > 17$. Um teste possível é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0 & \text{se } \bar{X} > 17 + 5/\sqrt{25} \\ \text{Não rejeita } H_0 & se\bar{X} \leq 17 + 5/\sqrt{25} \end{cases}$$

Portanto,

$$\begin{aligned} \beta(\theta, \delta) &= \mathbf{P}_\theta(\bar{X} > 18) \\ &= \mathbf{P}(\frac{\bar{X} - \theta}{5/\sqrt{25}} > \frac{18 - \theta}{5/\sqrt{25}}) \\ &= 1 - \Phi(18 - \theta) \end{aligned}$$

θ	$\beta(\theta, \delta)$
15	0.0013
16	0.0228
17	0.1587
18	0.5000
19	0.8413
20	0.9772
21	0.9987

Exemplo 2.4: $X_1, \ldots, X_n$ i.i.d. $b(1, \theta)$, $n = 10$. Queremos testar $H_0 : \theta \leq 1/2$ versus $H_1 : \theta > 1/2$.

Vamos utilizar o teste: $\delta(\mathbf{X}) = 1$ isto é, rejeito H_0, se, e somente se, $\sum_{i=1}^{10} X_i \geq 6$. Temos: H_0 : hipótese composta; H_1 : hipótese composta.

Região crítica: $\Gamma_\delta = \{(x_1, \ldots, x_{10}); \sum_{i=1}^{10} x_i \geq 6, x_i = 0 \text{ ou } 1\}$.

Erro tipo I: Rejeitar H_0 dado que H_0 é verdadeira;
Erro tipo II: Não rejeitar H_0 quando H_0 é falsa.

$$\begin{aligned} \mathbf{P}(\text{erro tipo I}|\theta = \theta_0 \leq 1/2) &= \mathbf{P}(\sum X_i \geq 6|\theta = \theta_0 \leq 1/2) \\ \mathbf{P}(\text{erro tipo II}|\theta = \theta_1 \leq 1/2) &= \mathbf{P}(\sum X_i < 6|\theta = \theta_1 > 1/2) \end{aligned}$$

Função poder:

$$\beta(\theta, \delta) = \mathbf{P}_\theta(\text{rejeitar } H_0)$$

$$\begin{aligned} &= \mathbf{P}_\theta(\sum_{i=1}^{10} X_i \geq 6) \\ &= \mathbf{P}_\theta(Y=6) + \ldots + \mathbf{P}_\theta(Y=10), Y \sim b(10,\theta) \end{aligned}$$

Neste caso, $\Theta = [0,1]$ e $\Theta_0 = [0, 1/2]$. O tamanho do teste é:

$$t(\delta) = \sup_{\theta \leq 1/2} \beta(\theta, \delta)$$

θ	$\beta(\theta,\delta)$
0.1	0.000
0.2	0.006
0.3	0.047
0.4	0.166
0.5	0.377
0.6	0.633
0.7	0.850
0.8	0.964
0.9	0.998

Daí, $t(\delta) = .377$.

Exercício 2.1: Considere o seguinte teste para aceitação de lote: para cada lote recebido são escolhidos aleatoriamente 20 peças e rejeitado o lote quando existe pelo menos uma peça defeituosa na amostra. Segundo o fabricante a proporção máxima de peças defeituosas é de 0,07. Especifique e/ou calcule: hipóteses nula e alternativa, estatística de teste, região crítica, nível de significância, poder e tamanho do teste.

Exercício 2.2: Suponha que aproximadamente 50% das crianças que nascem são do sexo masculino. Um novo tratamento é proposto para aumentar a probabilidade de nascer uma criança do sexo masculino. Um experimento é realizado com 20 mulheres que desejam filhos do sexo masculino.
a. Discuta quais as hipóteses nula e alternativa adequadas.
b. Dados os seguintes testes definidos pelas regiões críticas:

$$RC1 = \{x; x \geq 13\} \quad \text{e} \quad RC2 = \{x; x \geq 14 \text{ ou } x \leq 6\},$$

onde X é o número de crianças do sexo masculino (discarte a possibilidade de gêmeos) discuta qual o melhor teste.
c. Discuta o resultado caso no experimento tenham nascido 12 crianças do sexo masculino.

3 Hipótese Simples VS Hipótese Alternativa Simples

Embora este caso dificilmente ocorra na prática ele será bastante útil na introdução dos conceitos e princípios utilizados em teste de hipóteses. Assumiremos que a amostra é proveniente de uma entre duas distribuições completamente especificadas. Nosso objetivo é descobrir qual. Mais precisamente assumiremos que uma amostra aleatória $X_1, \ldots, X_n$ veio de uma densidade $f_0(x)$ ou $f_1(x)$ e desejamos testar a hipótese $H_0 : X_i$ distribuído como $f_0(x)$, abreviado $X_i \sim f_0(\cdot)$ versus $H_1 : X_i \sim f_1(\cdot)$.

3.1 Teste da Razão de Verossimilhança

Assumimos que temos uma amostra aleatória que pode ter vindo de uma de duas distribuições de probabilidade completamente especificadas pelas nossas hipóteses. Nosso objetivo é determinar de qual distribuição provém a amostra. Isto é, queremos testar:

$$H_0 : X_i \sim f_0(\cdot), \forall i = 1, \ldots, n \quad \text{versus} \quad H_1 : X_i \sim f_1(\cdot), \forall i = 1, \ldots, n.$$

Se nós tivéssemos somente uma observação x^* e $f_0(\cdot)$ e $f_1(\cdot)$ fossem da forma:

Um critério muito razoável seria dizer que :

$$\begin{aligned} X_i &\sim f_0(\cdot), \quad \text{se } f_0(x^*) \geq f_1(x^*) \\ X_i &\sim f_1(\cdot), \quad \text{se } f_0(x^*) < f_1(x^*) \end{aligned}$$

Definição 3.1: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma densidade (ou função de probabilidade) que pode ser $f_0(\cdot)$ ou $f_1(\cdot)$. Um teste δ para testar $H_0 : X_i \sim f_0(\cdot), \forall i = 1, \ldots, n$ versus $H_1 : X_i \sim f_1(\cdot), \forall i = 1, \ldots, n$ é dito ser um *teste da razão de verossimilhança simples* se:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \lambda(\mathbf{X}) \leq k \\ \text{Não rejeita } H_0, & \text{se } \lambda(\mathbf{X}) > k \end{cases}$$

onde k é uma constante não negativa e

$$\lambda(x_1, \ldots, x_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} = \frac{L_0(x_1, \ldots, x_n)}{L_1(x_1, \ldots, x_n)}.$$

Para cada k temos um teste diferente. Para um k fixo, o teste nos diz que devemos rejeitar H_0 se a razão das verossimilhanças é pequena, isto é, rejeitamos H_0 se é muito mais provável ($L_1 > L_0$) que a amostra tenha vindo de $f_1(\cdot)$ do que de $f_0(\cdot)$.

3.2 Testes mais poderosos

Seja $X_1, \ldots, X_n$ uma amostra aleatória proveniente de $f_0(\cdot)$ ou de $f_1(\cdot)$. Vamos escrever $f_0(x) = f(x, \theta_0)$ e $f_1(x) = f(x, \theta_1)$. Portanto queremos testar:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1$$

Para qualquer teste δ para as hipóteses acima, temos associado uma função poder $\beta(\theta, \delta)$. Um bom teste deve ter:

$$\beta(\theta_0, \delta) = \mathbf{P}[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] \quad \text{pequeno}$$

$$\beta(\theta_1, \delta) = \mathbf{P}[\text{rejeitar } H_0 | H_0 \text{ falsa}] \quad \text{grande.}$$

Daí é razoável usar os valores de $\beta(\theta_0,\delta) = \mathbf{P}$[erro tipo I] e $\beta(\theta_1,\delta) = 1 - \mathbf{P}$[erro tipo II] como critérios para se definir um bom teste. Nosso critério seria fazer com que as duas probabilidades de erro fossem pequenas. A partir desse parágrafo, apenas por facilidade, ao colocar $\mathbf{P}$[erro tipo I] ou $\mathbf{P}$[erro tipo II] estaremos considerando probabilidade condicional.

Por exemplo, $\beta(\theta_0,\delta) + 1 - \beta(\theta_1,\delta)$ seja pequeno, ou então queremos um teste δ fixando $\beta(\theta_0,\delta) = \alpha$ e $1 - \beta(\theta_1,\delta)$ mínimo.

Definição 3.2: Um teste δ^* para

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1$$

é dito ser o *teste mais poderoso de tamanho* α se, e somente se:
(i) $\beta(\theta_0,\delta^*) = \alpha$;
(ii) $\beta(\theta_1,\delta^*) \geq \beta(\theta_1,\delta)$ para todo teste δ com $\beta(\theta_0,\delta) \leq \alpha$.

Teorema 3.1 Lema de Neyman Pearson: Seja $X_1,\ldots,X_n$ uma amostra aleatória de uma densidade (ou função de probabilidade) $f(\cdot,\theta)$ com $\theta = \theta_0$ ou θ_1, e seja $0 < \alpha < 1$ fixo. Seja k^* uma constante positiva e C^* um subconjunto de valores possíveis da amostra, tal que:
(i) $\mathbf{P}_{\theta_0}[(X_1,\ldots,X_n) \in C^*] = \alpha$;
(ii) $\lambda(\mathbf{x}) = \frac{L(\theta_0;x_1,\ldots,x_n)}{L(\theta_1;x_1,\ldots,x_n)} = \frac{L_0}{L_1} \leq k^*$ se $(x_1,\ldots,x_n) \in C^*$ e $\lambda(\mathbf{x}) > k^*$ se $(x_1,\ldots,x_n) \notin C^*$.

Então o teste δ^*

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \mathbf{X} \in C^* \\ \text{Não rejeita } H_0, & \text{se } \mathbf{X} \notin C^* \end{cases}$$

ou equivalentemente,

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \lambda(\mathbf{X}) \leq k^* \\ \text{Não rejeita } H_0, & \text{se } \lambda(\mathbf{X}) > k^* \end{cases}$$

é o teste mais poderoso de tamanho α para testar:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1$$

Nota: A versão acima do Lema de Neyman-Pearson é para o teste não aleatorizado. Para o caso do teste aleatorizado fazemos uma aleatorização quando $\lambda(\mathbf{X}) = k^*$, com probabilidade $\phi(\mathbf{X})$, de se rejeitar a hipótese nula; onde a probabilidade $\phi(\mathbf{x})$ não é a mesma para todos os pontos do espaço amostral onde $\lambda(\mathbf{x}) = k^*$. Desta forma, o teste aleatorizado mais poderoso é dado por:

$$\psi_{\delta^*}(\mathbf{X}) = \begin{cases} 1, & \text{se } \lambda(\mathbf{X}) < k^* \\ \phi(\mathbf{X}), & \text{se } \lambda(\mathbf{X}) = k^* \\ 0, & \text{se } \lambda(\mathbf{X}) > k^* \end{cases}$$

e o nível de significância dada por:

$$\alpha = E[\psi_\delta(\mathbf{X})/H_0] = P[\lambda(\mathbf{X}) < k^*] + \sum_{\forall \mathbf{x} t.q. \lambda(\mathbf{x}) = k^*} \phi(\mathbf{x}) P[\mathbf{X} = \mathbf{x}]$$

Note que no caso contínuo temos $P[\lambda(\mathbf{X}) = k^*] = 0$. Porisso no caso contínuo não tem sentido falar em teste aleatorizado.

Exemplo 3.1 Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição exponencial com taxa de falha que sabemos deve ser ou 50 ou 100. Isto é, queremos testar:

$$H_0 : \theta = 50 \quad \text{versus} \quad H_1 : \theta = 100$$

Sabemos que:

$$\lambda(\mathbf{x}) = \frac{L_0}{L_1} = \frac{(50)^n \exp(-50 \sum_{i=1}^n x_i)}{(100)^n \exp(-100 \sum_{i=1}^n x_i)} = (1/2)^n \exp(50 \sum_{i=1}^n x_i)$$

e de acordo com o Lema de Neyman-Pearson o teste mais poderoso é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } (1/2)^n \exp(50 \sum_{i=1}^n X_i) \leq k^* \\ \text{Não rejeita } H_0, & \text{se } (1/2)^n \exp(50 \sum_{i=1}^n X_i) > k^* \end{cases}$$

i.e.,

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \sum_{i=1}^n X_i \leq \frac{\log(2^n k^*)}{50} = k' \\ \text{Não rejeita } H_0, & \text{se } \sum_{i=1}^n X_i > \frac{\log(2^n k^*)}{50} = k' \end{cases}$$

e k' é escolhido de modo que:

$$\mathbf{P}_{\theta_0}[\sum_{i=1}^n X_i \leq k'] = \int_0^{k'} \frac{1}{\Gamma(n)} \theta_0^n x^{n-1} e^{-\theta_0 x} dx = \alpha.$$

Exemplo 3.2: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição Bernoulli com probabilidade de "Sucesso", que sabemos deve ser ou θ_0 ou θ_1, onde $\theta_0 < \theta_1$. Queremos testar:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1.$$

Neste caso,

$$\lambda(\mathbf{x}) = \frac{L_0}{L_1} = \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\theta_1^{\sum x_i}(1-\theta_1)^{n-\sum x_i}}$$

e de acordo com o Lema de Neyman-Pearson o teste mais poderoso é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\theta_1^{\sum x_i}(1-\theta_1)^{n-\sum x_i}} \leq k^* \\ \text{Não rejeita } H_0, & \text{se } \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\theta_1^{\sum x_i}(1-\theta_1)^{n-\sum x_i}} > k^* \end{cases}$$

mas,

$$\begin{aligned} \lambda(\mathbf{x}) \leq k^* \quad & \Leftrightarrow \quad \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\theta_1^{\sum x_i}(1-\theta_1)^{n-\sum x_i}} \leq k^* \\ & \Leftrightarrow \quad \sum x_i \underbrace{\log\left\{\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right\}}_{\leq 0} \leq \log\left\{k^*(\frac{1-\theta_1}{1-\theta_0})^n\right\} \\ & \Leftrightarrow \quad \sum x_i \geq k' \end{aligned}$$

Assim, o teste mais poderoso é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \sum_{i=1}^n X_i \geq k' \\ \text{Não rejeita } H_0, & \text{se } \sum_{i=1}^n X_i < k' \end{cases}$$

e k' é escolhido de modo que:

$$\mathbf{P}_{\theta_0}[\sum_{i=1}^n X_i \geq k'] \leq \alpha.$$

Isto é:

$$\begin{aligned} k' &= \min\{k; \mathbf{P}_{\theta_0}[\text{rejeitar } H_0] \leq \alpha\} \\ &= \min\{k; \mathbf{P}_{\theta_0}[\sum_{i=1}^n X_i \geq k] \leq \alpha\} \end{aligned}$$

que pode ser facilmente calculado através da distribuição binomial.

O tamanho do teste é dado por:

$$\alpha' = \mathbf{P}_{\theta_0}[\sum_{i=1}^n X_i \geq k']$$

Observe que como $\sum X_i$ assume valores inteiros entre 0 e n, α' pode assumir apenas $(n+1)$ valores diferentes. Desta forma, não existe um teste não aleatorizado mais poderoso de tamanho α para todo α. No entanto, existe um teste mais poderoso não aleatorizado de tamanho α dado por:

$$\psi_{\delta^*}(\mathbf{X}) = \begin{cases} 0, & \text{se } \sum X_i < k' - 1 \\ (\alpha - \alpha')/P[\sum X_i = k' - 1], & \text{se } \sum X_i = k' - 1 \\ 1, & \text{se } \sum X_i \geq k' \end{cases}$$

onde α' e k' são definidos como anteriormente.

Exemplo 3.3: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição normal com variância conhecida e igual a 1 e com apenas 2 possíveis valores para a média, μ_0 e μ_1 com $\mu_0 < \mu_1$. Queremos testar:

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu = \mu_1.$$

Vamos utilizar este exemplo para introduzir a abordagem de testes de significância em teste de hipóteses.

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{L_0}{L_1} = \frac{(2\pi)^{-n/2}\exp\{-\sum(x_i - \mu_0)^2\}}{(2\pi)^{-n/2}\exp\{-\sum(x_i - \mu_1)^2\}} \\ &= \exp\{2n\bar{x}(\mu_0 - \mu_1) + n(\mu_1^2 - \mu_0^2), \} \end{aligned}$$

e de acordo com o Lemma de Neyman-Pearson o teste ótimo é dado pela seguinte regra de decisão: rejeito H_0 caso $\exp\{2n\bar{x}(\mu_0 - \mu_1) + n(\mu_1^2 - \mu_0^2)\} \leq k$, que é equivalente a rejeitar H_0 quando $\bar{x} \geq k'$ dado que $\mu_0 < \mu_1$; ou seja, o teste mais poderoso é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0, & \text{se } \bar{X} \geq k' \\ \text{Não rejeita } H_0, & \text{se } \bar{X} > k' \end{cases}$$

e k' é escolhido de modo que:

$$\mathbf{P}[\bar{X} \geq k' | X_i \sim N(\mu_0; 1)] = \alpha.$$

cuja solução é $k' = \mu_0 + z_{1-\alpha}/\sqrt{n}$; isto é, **rejeito a hipótese nula** se

$$\bar{x} \geq \mu_0 + z_{1-\alpha}/\sqrt{n}, \quad \text{ou equivalentemente, se} \quad \mu_0 \leq \bar{x} - z_{1-\alpha}/\sqrt{n}.$$

Observe que esta regra de decisão independe do valor de μ_1, desde que μ_1 seja maior do que μ_0.

Considere agora que temos $n = 16$, $\mu_0 = 3,0$, $\alpha = 0,05$ e $\bar{x} = 3,40$. Neste caso $k' = 3,411$ e como $3,40 < 3,411$ não rejeitamos a hipótese nula ao nível de significância 0,05. Isto não quer dizer que estamos falando que existem evidências de que a hipótese nula seja verdadeira. Na verdade qualquer valor de μ_0 tal que $\mu_1 > \mu_0 > 2,989$ não é rejeitado ao nível de significância 0,05; daí ser necessário ter um certo cuidado na interpretação e apresentação dos resultados. Não dizemos que a hipótese nula seja verdadeira, mas que ela não é rejeitada, ou então que não existe evidência estatística ao nível de significância de 0,05 para rejeitar a hipótese nula .

Suponha agora que a média amostral tenha sido um pouco maior, por exemplo, em vez de 3,40 fosse igual a 3,42. Nesse caso rejeitaríamos a hipótese nula. Qualquer um, estatístico ou não, deveria estranhar que com resultados tão próximos, que podem inclusive ser causados por arredondamentos levem a decisões diferentes. Considere agora que a média amostral seja igual a 3,80; ou seja, também neste caso a hipótese nula é rejeitada. Mas declarar simplesmente que a hipótese nula é rejeitada ao nível de significância 0,05 tanto nesse caso como quando a média amostral é 3,42 é jogar fora parte da informação; afinal, o resultado 3,80 dá uma evidência maior de que a hipótese alternativa é verdadeira.

Estas críticas levantadas desaparecem na utilização da abordagem de testes de significância, que é a defendida por Fisher. Nesta abordagem procura-se dar uma medida de evidência de que a hipótese nula não seja verdadeira. Verifique que com o resultado da média amostral igual a 3,45 rejeitamos a hipótese nula de que $\mu_0 = 3,0$, ao nível de significância α se $z_{1-\alpha} \leq \sqrt{16}(\bar{x} - 3,0)$; isto é para $\alpha > 0,0359$. Da mesma forma com o resultado 3,80 rejeitaríamos a hipótese nula para todo $\alpha > 0,0007$. Os valores encontrados dão, em cada caso, o menor valor do nível de significâncica que permite a rejeição da hipótese nula com o valor experimental observado. Este valor é chamado de probabilidade de *significância, p-value, valor p, nível descritivo do teste.*

Definição 3.3 Valor-p: Dada uma estatística de teste T, que ordena os pontos do espaço amostral em termos de evidência em favor da hipótese alternativa, o *valor-p* é uma estatística definida como o menor nível de significância em que a hipótese nula é rejeitada com base nos dados experimentais.

Chamamos a atenção para o fato do valor-p ser uma estatística; logo, não pode depender de parâmetros desconhecidos. Isto é verdade porque condicionamos em relação à hipótese nula. Vemos também que se o nível de significância é menor do que o valor-p não rejeitamos a hipótese nula; caso contrário ela é rejeitada. Mais tarde daremos outra definição de valor-p.

Do exposto, com o valor observado 3,45 só rejeitaríamos a hipótse nula se, ao definirmos o nível de significância, estivermos dispostos a correr o risco de rejeitar a hipótese nula incorretamente com probabilidade, no mínimo igual a 0,0359. Já, no segundo caso, a hipótese nula seria rejeitado, mesmo que se queira correr um risco pequeno, mas maior do que 0,0007. Desta forma, os valores de probabilidades de significância encontrados, 0,0359 e 0,0007, podem ser utilizados para medir a evidência de que a hipótese alternativa seja verdadeira. Quanto menor o valor-p maior a evidência.

Para que esta abordagem possa ser utilizada é necessário que possamos ordenar os valores do espaço amostral em termos de evidência. No presente caso quanto maior a distância da média amostral de μ_0 maior a evidência de que a hipótese nula não seja verdadeira.

Nem sempre existe um teste não aleatorizado mais poderoso. Por exemplo, existem casos em que não existe um teste que tenha o tamanho igual ao tamanho especificado. Quando trabalhamos com distribuções contínuas sempre existe um teste do tamanho especificado. Observe que o teste uniformemente mais poderoso de nível α é necessariamente um teste da razão de verossimilhança simples.

Exemplo 3.4: Considere uma amostra aleatória de uma das funções densidades abaixo e que estamos interessados em testar

$$H_0 : X \sim f_0 \quad \text{versus} \quad H_1 : X \sim f_1$$

x	0	1	5	7	10
$f_0(x)$	0,30	0,20	0,10	0,30	0,10
$f_1(x)$	0,20	0,20	0,30	0,05	0,25
$f_0(x)/f_1(x)$	1,50	1,00	0,33	6,00	4,00

Na discussão a seguir vamos sempre considerar o teste mais poderoso dado pelo Lema de Neyman-Pearson. Vamos colocar em ordem decrescente os valores amostrais que dão maior evidência de que a hipótese alternativa é verdadeira. De acordo com o Lema de Neyman-Pearson o ponto que dá mais evidência em favor da hipótese alternativa é o valor 5 pois tem a menor relação de verossimilhanças. A seguir vem o ponto 1, 0, 10 e finalmente 7.

x	5	1	0	10	7
$f_0(x)/f_1(x)$	0,33	1,00	1,50	4,00	6,00
$f_0(x)$	0,10	0,20	0,30	0,10	0,30
Prob. Acumulada	0,10	0,30	0,60	0,70	1,00

Os valores da última linha dão os valores exatos dos possíveis testes não aleatorizados mais poderosos. Desta forma, caso estejamos dispostos a cometer erro do tipo I com probabilidade no máximo igual a 0,20 a região crítica seria dado por $RC = \{5\}$, que tem tamanho igual a 0,10. Um teste de tamanho igual a 0,20 teria que ser aleatorizado e é dado por:

$$\psi_\delta(\mathbf{X}) = \begin{cases} 1, & \text{se } x = 5 \\ 1/2, & \text{se } x = 1 \\ 0, & \text{se } x \neq 1 \ e \ 5 \end{cases}$$

Suponha agora que foi observado o valor 0. Neste caso, tanto no teste aleatorizado como no teste não aleatorizado iríamos rejeitar a hipótese nula caso o máximo risco que estamos dispostos a cometer o erro de tipo I seja igual a 0,20. Porém qual seria o valor-p neste caso? Observe que segundo o teste mais poderoso, dado pelo Lema de Neyman-Pearson, o conjunto de valores que dão tanta ou mais evidências em favor da hipótese alternativa, do que o valor observado é dado pelo conjunto:

$$A = \{5, 1, 0\}$$

Uma das definições de valor-p é exatamente a probabilidade deste conjunto sob a hipótese nula. Logo

$$\text{valor-p} = P[A/H_0] = 0,10 + 0,20 + 0,30 = 0,60$$

Observe que só podemos calcular o valor-p porque, dada a estatística do teste, podemos ordenar os pontos do espaço amostral em termos de evidência em favor da hipótese alternativa, o que não ocorre, por exemplo, no *Exercício 4.4*.

Exercício 3.1: No *Exemplo 3.1* calcule o valor-p quando n = 20 e $\bar{x} = 0,025$ utilizando a distribuição exata e pela aproximação normal.

Exercício 3.2: No *Exemplo 3.4* encontre o valor-p quando temos uma amostra de tamanho 2 e os valores observados são 0 e 1.

Exercício 3.3: Suponha que no Departamento de Manutenção existiam 6 ordens de serviço no início do dia. Estes serviços devem ser realizados o mais rápido possível porque outros serviços podem aparecer. O supervisor sorteou 2 ordens de serviços e entregou a João e os restantes a José. Ao final do dia o supervisor verificou que, entre os 6 serviços, exatamente os 2 de João foram os que mais demoraram para serem terminados. Apenas baseados nestes dados o supervisor resolveu punir João. Argumente em termos estatísticos se o supervisor está correto ou não ao tomar esta decisão.

3.3 Função Perda

Analogamente à sub-seção anterior queremos decidir entre $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$. Podemos tomar uma das duas decisões, d_j, $j = 0$ ou 1, que significa escolher a hipótese H_j . Vamos assumir que a função perda seja conhecida.

Definição 3.4 Função Perda : Ao testar versus $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ defina $l(d_i; \theta_j) =$ perda ao se tomar a decisão i (decidir que a hipótese H_i é verdadeira), quando o valor verdadeiro é θ_j para $i = 0, 1$ e $j = 0, 1$. Adotaremos a convenção de que $l(d_i; \theta_i) = 0$ para $i = 0, 1$ e $l(d_i; \theta_j) > 0$ para $i \neq j$

Observe que definir uma função decisão é definir uma região onde se adota a decisão d_0 , tomando-se a decisão d_1 na região complementar. Desta forma, se consideramos apenas os testes não aleatorizados, que são testes definidos pelas regiões críticas, também teremos definidas as funções decisões. Podemos, portanto, utilizar os mesmos conceitos e notações.

Na comparação entre dois ou mais testes é claro que preferimos o de menor perda. Infelizmente, dificilmente existe um teste que tenha a menor perda para as duas possíveis decisões e para θ_0 e θ_1 . Analogamente ao caso de estimação definimos a função risco que é a perda média.

Definição 3.5 Função Risco : Para uma amostra aleatória $X_1, \ldots, X_n$ de $f(\cdot; \theta_0)$ ou de $f(\cdot; \theta_1)$, seja δ um teste de $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ tendo a região crítica C_δ . Para uma dada função perda $l(\cdot; \cdot)$, a função risco de δ, denotada por $R_\delta(\theta)$, é definida como a perda média, isto é,

$$
\begin{aligned}
R_\delta(\theta) \quad = \quad & \underbrace{\int \ldots \int}_{C_\delta} l(d_1; \theta) \left[\prod_{i=1}^{n} f(x_i; \theta) dx_i \right] \\
& + \underbrace{\int \ldots \int}_{\overline{C_\delta}} l(d_0; \theta) \left[\prod_{i=1}^{n} f(x_i; \theta) dx_i \right]
\end{aligned}
$$

Observe que

$$
\begin{aligned}
R_T(\theta) &= l(d_1;\theta).P[(X_1,\ldots,X_n) \in C_T] + l(d_0;\theta).P[(X_1,\ldots,X_n) \in \overline{C_T}] \\
&= l(d_1;\theta).\beta(\theta,\delta) + l(d_0;\theta).[1-\beta(\theta,\delta)],
\end{aligned}
$$

como θ assume apenas dois valores temos

$$
R_\delta(\theta_0) = l(d_1;\theta_0).\beta(\theta_0,\delta) \quad \text{e} \quad R_\delta(\theta_1) = l(d_0;\theta_1).(1-\beta(\theta_1, delta))
$$

Infelizmente dificilmente existe um teste que minimize a função risco para ambas as hipóteses. Um critério, que não é tão bom quanto este é o critério minimax definido por:

Definição 3.6 Teste Minimax: Um teste δ_m de $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ é definido como sendo minimax se e somente se

$$
\max[R_{\delta_m}(\theta_0), R_{\delta_m}(\theta_1)] \leq \max[R_\delta(\theta_0), R_\delta(\theta_1)]
$$

para qualquer outro teste δ.

Neste caso o seguinte teorema auxilia a encontrar o teste minimax

Teorema 3.3: Considere uma amostra aleatória $X_1,\ldots,X_n$ de $f(\cdot;\theta_0)$ ou de $f(\cdot;\theta_1)$ para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Se δ_m tem uma região crítica dada por $C_m = (x_1,\ldots,x_n) : \lambda \leq k_m$ onde k_m é uma constante positiva tal que $R_{\delta_m}(\theta_0) = R_{\delta_m}(\theta_1)$, então T_m é minimax.

Se $f_0(\cdot)$ e $f_1(\cdot)$ são funções densidades discretas então pode não existir k_m que satisfaça a igualdade do teorema se considerarmos apenas os testes não aleatorizados.

Observe que tanto o teste mais poderoso de Neyman-Pearson quanto o teste minimax são testes razão de verossimilhança simples.

Na abordagem Bayesiana, que não será vista agora, além das funções perda e risco temos uma distribuição à priori para θ. Neste caso também pode ser visto que a região crítica também é dada por um teste da razão de verossimilhança simples.

4 Hipóteses Compostas

Suponha que temos uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição de probabilidade com densidade (ou função de probabilidade) $f(\cdot, \theta)$, onde $\theta \in \Theta$. Com base nestes dados, queremos testar:

$$H_0 : \theta \in \Theta_0 \qquad \text{versus} \qquad H_1 : \theta \in \Theta_1,$$

onde, $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$. (Em geral, mas não sempre $\Theta_0 \cup \Theta_1 = \Theta$).

4.1 Teste Geral da Razão de Verossimilhança

Baseados em uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição de probabilidade com densidade (ou função de probabilidade) $f(\cdot, \theta)$, onde $\theta \in \Theta$, queremos testar as hipóteses:

$$H_0 : \theta \in \Theta_0 \qquad \text{versus} \qquad H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0.$$

Definição 4.1 razão de verossimilhança generalizada: Seja $L(\theta, x_1, \ldots, x_n)$, a função de verossimilhança para a amostra aleatória $X_1, \ldots, X_n$ tendo distribuição conjunta $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n, \theta)$, onde $\theta \in \Theta$. A *razão de verossimilhança generalizada*, denotada por λ, é dada por:

$$\lambda = \lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta, x_1, \ldots, x_n)}{\sup_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n)}.$$

Note que λ é uma função de $x_1, \ldots, x_n$. Quando as observações são trocadas com as respectivas v.a.'s $X_1, \ldots, X_n$, então temos

$$\Lambda = \lambda(X_1, \ldots, X_n)$$

que é uma v.a. e mais ainda é uma estatística.

Obs.:
(1) $0 \leq \Lambda \leq 1$;
(2) θ pode ser um vetor;
(3) O denominador de Λ é a função de verossimilhança avaliada no EMV (irrestrito) de θ;
(4) O numerador de Λ é a função de verossimilhança avaliada no EMV (restrito, ou sob H_0) de θ;
(5) No caso de termos uma amostra aleatória, as v.a's são i.i.d. e neste caso, $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$.

A estatística Λ é usada para formular o teste da razão de verossimilhança:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \Lambda \leq k \\ \text{Não rejeito } H_0, & \text{se } \Lambda > k \end{cases}$$

onde k é uma constante fixa $0 < k < 1$ e é geralmente especificada fixando-se o tamanho do teste α.

Geralmente, o teste da RV é um bom teste. A aplicação deste teste pode esbarrar em duas dificuldades. A primeira pode ser encontrar $\sup_{\theta \in \Theta_0} L(\theta, x_1, \ldots, x_n)$ e a segunda encontrar a distribuição de Λ, que é necessária para se calcular a função poder do teste. No segundo caso muitas vezes é mais fácil encontrar a distribuição de $g(\Lambda)$, onde $g(\cdot)$ é uma função monótona.

Exemplo 4.1: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição exponencial com parâmetro θ. Isto é, $f(x,\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$ e $\Theta = (0, \infty)$. Queremos testar $H_0 : \theta \leq 1000$ versus $H_1 : \theta > 1000$. Sabemos que o EMV de θ é: $1/\bar{X}$. Daí,

$$\begin{aligned} \sup_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n) &= \sup_{\theta > 0} \theta^n \exp\{-\theta \sum_{i=1}^n x_i\} \\ &= (\frac{1}{\bar{x}})^n e^{-n} \end{aligned}$$

e

$$\begin{aligned} \sup_{\theta \in \Theta_0} L(\theta, x_1, \ldots, x_n) &= \sup_{0 < \theta \leq 1000} \theta^n \exp\{-\theta \sum_{i=1}^n x_i\} \\ &= \begin{cases} (\bar{x})^{-n} e^{-n}, & \text{se } \frac{1}{\bar{x}} \leq 1000 \\ 1000^n e^{-1000 n \bar{x}}, & \text{se } \frac{1}{\bar{x}} > 1000 \end{cases} \end{aligned}$$

Daí,

$$\lambda(x_1, \ldots, x_n) = \begin{cases} 1, & \text{se } \frac{1}{\bar{x}} \leq 1000 \\ \frac{1000^n \exp\{-1000 n \bar{x}\}}{(1/\bar{x})^n e^{-n}}, & \text{se } \frac{1}{\bar{x}} > 1000 \end{cases}$$

Se $0 < k < 1$ então o teste da RV é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \frac{1}{\bar{X}} > 1000 \text{ e } (1000\bar{X})^n \exp(-1000n\bar{X} + n) \leq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Isto é,

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } 1000\bar{X} < 1 \text{ e } (1000\bar{X})^n \exp\{-n(1000\bar{X} - 1)\} \leq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Escreva, $Y = 1000\bar{X}$ e note que $Y^n e^{-n(Y-1)}$ tem um máximo em $Y = 1$ e é crescente em $(0, 1)$. Assim, $Y^n e^{-n(Y-1)} \leq 1$ e $Y^n e^{-n(Y-1)} \leq k$ se, e somente se, $Y \leq k'$ onde k' é uma constante $0 < k' < 1$.

Daí,

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } 1000\bar{X} \leq k' \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Se queremos o teste de tamanho α, k' deve ser a solução de:

$$\begin{aligned}
\alpha &= \mathbf{P}[1000\bar{X} \leq k'] \\
&= \mathbf{P}_{1000}[1000\sum_{i=1}^{n} X_i \leq nk'] \\
&= \int_0^{nk'} \frac{1}{\Gamma(N)} u^{n-1} e^{-u} du.
\end{aligned}$$

(Note que: $\mathbf{P}_\theta[1000\bar{X} \leq k']$ é uma função crescente de θ.

4.2 Testes Uniformementes Mais Poderosos (Testes UMP)

Definição 4.2 Teste uniformemente mais poderoso: Um teste δ^* para $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, é dito ser um *teste uniformemente mais poderoso* de tamanho α se, e somente se:
(i) $\sup_{\theta \in \Theta_0} \beta(\theta, \delta^*) = t(\delta^*) = \alpha$;
(ii) $\beta(\theta, \delta^*) \geq \beta(\theta, \delta)$ para todo $\theta \in \Theta_1$ e qualquer teste δ tal que $t(\delta) \leq \alpha$.

Exemplo 4.2: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição exponencial com parâmetro $\theta \in (0, \infty)$. Encontre o teste UMP para $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Para $\theta_1 > \theta_0$ fixo, vimos no *Exemplo 3.1*, que pelo Lema de Neyman-Pearson, o teste mais poderoso para $\bar{H}_0 : \theta = \theta_0$ versus $\bar{H}_1 : \theta = \theta_1$ é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } \bar{H}_0, & \text{se } \sum X_i \leq k \\ \text{Não rejeito } \bar{H}_0, & \text{se } \sum X_i > k \end{cases}$$

onde k é solução de

$$\alpha = \int_0^k \frac{1}{\Gamma(n)} \theta_0^n u^{n-1} e^{\theta_0 u} du.$$

Como o teste acima não depende do valor de θ_1 mas somente do fato que $\theta_1 > \theta_0$, temos que o teste δ dado acima é UMP para testar:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \quad H_1 : \theta > \theta_0.$$

O teste é também UMP para testar:

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \quad H_1 : \theta > \theta_0$$

pois a função poder é crescente em θ. Observe que para testar:

$$H_0' : \theta = \theta_0' \leq \theta_0 \qquad \text{versus} \quad H_1' : \theta = \theta_1 > \theta_0$$

o teste é da mesma forma, independente do valor de θ_0'. No caso da hipótese composta o tamanho do teste é dado por:

$$\sup_{\theta in \Theta_0} \beta(\theta, \delta),$$

e como a função poder é crescente em θ temos que o tamanho do teste é dado por $\alpha = \beta(\theta_0, \delta)$, que produz o mesmo teste dado anteriormente.

Teorema 4.1: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $f(x, \theta)$, $\theta \in \Theta$, onde Θ é um intervalo. Assuma que f pertença à família exponencial:

$$f(x, \theta) = a(\theta) b(x) \exp[c(\theta) d(x)]$$

e coloque $t(x_1, \ldots, x_n) = \sum_{i=1}^n d(x_i)$, isto é $t(X_1, \ldots, X_n)$ é a estatística suficiente, completa e minimal para θ.

(i) Se $c(\theta)$ é uma função monótona crescente de θ e se existe k^* tal que $\mathbf{P}_{\theta_0}[t(X_1, \ldots, X_n) \geq k^*] = \alpha$ então o teste:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } t(X_1, \ldots, X_n) \geq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

é um teste UMP para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ ou $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

(ii) Se $c(\theta)$ é uma função monótona decrescente de θ e se existe k^* tal que $\mathbf{P}_{\theta_0}[t(X_1, \ldots, X_n) \leq k^*] = \alpha$ então o teste:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } t(X_1, \ldots, X_n) \leq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

é um teste UMP para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ ou $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Exemplo 4.3: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição exponencial com parâmetro $\theta \in (0, \infty)$. Encontre o teste UMP para $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

$$f(x, \theta) = \underbrace{\theta}_{a(\theta)} \underbrace{I_{(0,\infty)}(x)}_{b(x)} \exp(\underbrace{-theta}_{c(\theta)} \underbrace{x}_{d(x)})$$

assim,

$$t(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$$

e $c(\theta) = -\theta$ é decrescente em θ. Portanto,

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \sum_{i=1}^n X_i \leq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde k^* é a solução de

$$\alpha = \mathbf{P}_{\theta_0}[\sum_{i=1}^n Xi \leq k^*] = \int_0^{k^*} \frac{1}{\Gamma(n)} \theta_0^n u^{n-1} e^{-\theta_0 u} du$$

é um teste UMP para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

Definição 4.3: Uma família de densidades $\{f(x, \theta), \theta \in \Theta\}$, onde Θ é um intervalo, é dita ter **razão de verossimilhança monótona** se existe uma estatística, digamos, $T = t(X_1, \ldots, X_n)$ tal que

$$\frac{L(\theta', x_1, \ldots, x_n)}{L(\theta", x_1, \ldots, x_n)}$$

é uma função **crescente** de $t(x_1, \ldots, x_n)$ se $\theta' < \theta"$, ou; é uma função **decrescente** de $t(x_1, \ldots, x_n)$ se $\theta' < \theta"$.

Exemplo 4.4: Se $f(x,\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, temos

$$\begin{aligned}
\frac{L(\theta', x_1, \ldots, x_n)}{L(\theta", x_1, \ldots, x_n)} &= \frac{\theta'^n \exp(-\theta' \sum x_i)}{\theta"^n \exp(-\theta" \sum x_i)} \\
&= (\frac{\theta'}{\theta"})^n \exp[-(\theta' - \theta') \sum x_i]
\end{aligned}$$

que é função crescente de $\sum_{i=1}^n X_i$ se $\theta' < \theta"$.

Exemplo 4.5: Se $f(x,\theta) = (1/\theta) I_{(0,\theta)}(x)$, para $\theta > 0$, temos

$$\begin{aligned}
\frac{L(\theta', x_1, \ldots, x_n)}{L(\theta", x_1, \ldots, x_n)} &= \frac{(1/\theta')^n \prod_{i=1} n I_{(0,\theta')}(x_i)}{(1/\theta")^n \prod_{i=1} n I_{(0,\theta")}(x_i)} \\
&= \frac{(1/\theta')^n I_{(0,\theta')}(x_{(n)})}{(1/\theta")^n I_{(0,\theta")}(x_{(n)})} \\
&= \begin{cases} (\theta"/\theta')^n, & \text{se } 0 < x_{(n)} < \theta' \\ 0, & \text{se } \theta' < x_{(n)} < \theta" \end{cases}
\end{aligned}$$

que é função não decrescente de $x_{(n)}$ para $\theta' < \theta"$.

Teorema 4.2: Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x,\theta)$, onde $\theta \in \Theta$ e Θ é um intervalo. Assuma que a família de densidades $\{f(x,\theta); \theta \in \Theta\}$ tem razão de verossimilhança monótona em $T = t(X_1, \ldots, X_n)$.
(i) Se a razão de verossimilhança é monótona não-decrescente em $T = t(X_1, \ldots, X_n)$ e se k^* é tal que $\mathbf{P}_{\theta_0}(T \leq k^*) = \alpha$, então

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } T \leq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

é teste UMP de tamanho α para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.
(ii) Se a razão de verossimilhança é monótona não-crescente em $T = t(X_1, \ldots, X_n)$ e se k^* é tal que $\mathbf{P}_{\theta_0}(T \geq k^*) = \alpha$, então

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } T \geq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

é teste UMP de tamanho α para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

Exemplo 4.6: Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x,\theta) = (1/\theta) I_{(0,\theta)}(x)$, para $\theta > 0$ e queremos testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Já vimos que a família de densidades tem RVM não crescente em $X_{(n)} = \max\{X_1, \ldots, X_n\}$. De acordo com (ii), o teste UMP de tamanho α é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } X_{(n)} \geq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde k^* é tal que:

$$\begin{aligned}
\mathbf{P}_{\theta_0}[X_{(n)} \geq k^*] &= \int_{k^*}^{\theta_0} n(\frac{y}{\theta_0})^{n-1} \frac{1}{\theta_0} dy \\
&= \frac{1}{\theta_0^n}[\theta_0^n - (k^*)^n] \\
&= 1 - \frac{(k^*)^n}{\theta_0^n}
\end{aligned}$$

portanto, $k^* = \theta_0(1-\alpha)^{1/n}$.

Obs.:
(i) Nos teoremas acima temos $H_0 : \theta \leq \theta_0$. Se tivéssemos $H_0 : \theta \geq \theta_0$, as regiões críticas seriam dadas da mesma forma exceto que reverteríamos as desigualdades.
(ii) Os teoremas consideram apenas testes unilaterais.

Portanto temos que os testes UMP existem para hipóteses unilaterais se a densidade amostrada tiver razão de verossimilhança monótona em alguma estatística. Existem muitos problemas onde não é possível se obter testes UMP. Neste caso devemos restringir a classe de testes a fim de encontrar um teste ótimo nesta classe. Uma destas classes são os testes não viciados, que é definida a seguir.

Exercício 4.1: No *Exemplo 4.2* considere que $\theta_0 = 0,01$, $n = 100$, e que foi observado um tempo amostral médio igual a 107. Discuta os resultados em termos de valor-p e IC do tempo médio.

Exercício 4.2: Segundo um acordo entre comprador e vendedor a proporção de peças defeituosas não pode ultrapassar 0,05. O comprador recebe um lote e examina aleatoriamente 50 peças encontrando 2 peças defeituosas. Discuta qual deve ser a decisão do comprador se ele decide só devolver os lotes quando ele tem evidências de que ele não satisfaz as especificações.

Exercício 4.3: Suponha que em um dado instante a percentagem de eleitores da cidade de Campinas favoráveis ao projeto de renda mínima seja igual a 40%. Depois de realizar uma campanha de esclarecimento a prefeitura resolve realizar uma pesquisa de opinião pública para verificar se a campanha surtiu efeito. Neste caso qual deve ser o tamanho mínimo da amostra para que:
a. A probabilidade de se chegar a conclusão de que a campanha teve efeito, quando na verdade isto não ocorreu seja no máximo igual a 0,10; e,
b. A percentagem de se chegar a conclusão que a campanha teve efeito, caso a proporção de eleitores favoráveis tenha crescido para 0,50 seja no mínimo igual a 0,60.

Exercício 4.4: Suponha que certo pesquisador deseja testar uma hipótese nula H_0 contra a alternativa H_1, ambas compostas. Para tanto ele realiza um experimento, cujos resultados são variáveis aleatórias que podem ser representadas por $X_1, \ldots, X_k$. A região crítica adotada é dada por:

$$RC = \{(X_1, \ldots, X_k); T(X_1, \ldots, X_k) \in A\}$$

onde T é uma estatística de teste conhecida e A um conjunto, ou região, definido e conhecido.
a. Mostre como se calcula o nível de significância do teste.
b. Encontrado os valores $X_1 = x_1, \ldots, X_k = x_k$, qual seria a decisão?
c. Discuta porque, dado os valores amostrais, dependendo da forma que é dado a RC você não poderia discutir os resultados em termos de valor-p.

Exercício 4.5: Discuta os resultados em termos de valor-p quando temos os seguintes valores amostrais:
Exemplo 4.1: $\bar{x} = 0.0015$ e $n = 20$.
Exemplo 4.2: $\theta_0 = 1000$, $\bar{x} = 0.0015$ e $n = 20$.
Exemplo 4.6: $\theta_0 = 1$, $x_n = 0.96$ e $n = 10$.

Exercício 4.6: Considere uma amostra aleatória de tamanho 25 de uma Poisson com média θ.
a. Encontre o teste UMP para $H_0 : \theta = 1$ versus $H_1 : \theta > 1$. Esboce o gráfico da função poder

utilizando a aproximação normal e calcule o valor exato para alguns valores. Comente se a aproximação é boa.
b. Considere agora que a hipótese alternativa é bilateral, isto é, $H_1 : \theta \neq 1$. Mostre como deve ser a forma da região crítica do teste da razão de verossimilhança generalizada. Coloque a região crítica em termos da média amostral.

Exercício 4.7: Considere uma amostra aleatória de tamanho n_1 de uma exponencial com taxa de falha θ_1. Considere outra amostra aleatória de tamanho n_2 de uma exponencial com taxa de falha θ_2.
a. Encontre o teste da razão de verossimilhança generalizada para $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$.
b. Mostre que a região crítica não depende do valor da taxa de falha verdadeira.

4.3 Testes não viciados

Definição: Um teste δ para testar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ é dito ser um *teste não viciado* se, e somente se

$$\sup_{\theta \in \Theta_0} \beta(\theta, \delta) \leq \inf_{\theta \in \Theta_1} \beta(\theta, \delta).$$

Consequentemente, em um teste não viciado a probabilidade de rejeitar H_0 quando ela é falsa é pelo menos tão grande quanto a probabilidade de rejeitar H_0 quando esta é verdadeira. Dentre os testes não viciados sempre existe um teste UMP.

Na *Seção 5* será dado um exemplo de um teste que não é UMP, mas o mais poderoso dentre os testes não viciados.

4.4 Métodos para Encontrar Testes

Vários métodos já foram apresentados para realizar testes de hipóteses. Nesta seção eles serão rapidamente revistos e na última parte da subseção será apresentada a relação entre testes de hipóteses e intervalos de confiança. A discussão servirá também para colocar alguns cuidados necessários na utilização de testes de hipóteses e intervalos de confiança na análise estatística.

4.4.1 Teste da Razão de Verossimilhança

Já vimos que sempre podemos encontrar um teste através da razão de verossimilhnça. Se queremos testar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ baseados em uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição de probabilidade com densidade (ou função de probabilidade) $f(\cdot, \theta)$, onde $\theta \in \Theta$.

A *razão de verossimilhança generalizada*, denotada por λ, é dada por:

$$\lambda = \lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta, x_1, \ldots, x_n)}{\sup_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n)}.$$

e

$$\Lambda = \lambda(X_1, \ldots, X_n)$$

é uma v.a. e mais ainda é uma estatística.

A estatística Λ é usada para formular o teste da razão de verossimilhança:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \Lambda \leq k \\ \text{Não rejeito } H_0, & \text{se } \Lambda > k \end{cases}$$

onde k é uma constante fixa $0 < k < 1$ e é geralmente especificada fixando-se o tamanho do teste α.

Para família de densidades com *razão de verossimilhança monótona* o *Teorema 4.2* é bastante útil para encontrar testes UMP.

4.4.2 Relação entre Teste de Hipóteses e Intervalos de Confiança

Podemos utilizar um intervalo de confiança para um parâmetro unidimensional θ para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ e vice-versa.

Suponha que temos uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição $f(\cdot, \theta)$ e queremos testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Se temos $IC_\gamma(\mathbf{X})$ = o intervalo de confiança de nível γ para θ podemos definir a seguinte regra intuitiva: rejeitamos H_0 se θ_0 não pertencer ao IC_γ. O tamanho do teste pode ser facilmente calculado. Sabemos que se $IC_\gamma(\mathbf{X})$ é um IC de nível γ então:

$$\mathbf{P}_\theta(\theta \in IC_\gamma(\mathbf{X})) = \gamma,$$

e consequentemente a probabilidade de rejeitar H_0 quando H_0 é verdadeira é:

$$t(\delta) = \mathbf{P}_{\theta_0}(\theta_0 \notin IC_\gamma(\mathbf{X})) = 1 - \gamma.$$

Assim, um intervalo com 95% de confiança nos dá um teste com tamanho 5%.

Na realidade, temos uma dualidade aqui, pois uma família de teste pode ser usada para construir intervalos de confiança. Suponha que $C(\theta_0)$ é a região crítica de um teste de tamanho α para $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, definido para cada θ_0:

$$\mathbf{P}_{\theta_0}[\mathbf{X} \in C(\theta_0)] = \alpha.$$

Então, defina para cada ponto amostral $\mathbf{x}$,

$$IC(\mathbf{x}) = \{\theta | \mathbf{x} \notin C(\theta)\}$$

Há uma completa equivalência entre os eventos:

$$\{\theta \in IC(\mathbf{X})\} \qquad \text{e} \qquad \{\mathbf{X} \notin C(\theta)\},$$

o que significa que as probabilidades são as mesmas

$$\mathbf{P}_\theta\{\theta \in IC(\mathbf{X})\} = 1 - \mathbf{P}_\theta\{\mathbf{X} \in C(\theta)\} = 1 - \alpha,$$

Assim, $IC(\mathbf{X})$ é uma região de confiança para θ com nível $1 - \alpha$, é dado pelo conjunto de todos os valores de θ que não podem ser rejeitados em um teste de hipóteses com nível de significância α, com os resultado amostral $\mathbf{X}$ observado.

A região de confiança obtida através de um teste pode ou não ser um intervalo. Se for utilizada uma região crítica unicaudal para uma estatística do teste T, então geralmente obtemos cotas de confiança para θ.

Os IC construídos através de testes de hipóteses podem ter uma largura grande ou não. Se a largura do intervalo for grande isto implica que existem valores de θ distantes de θ_0, que não podem ser rejeitados, por serem considerados consistentes com os valores amostrais, dentro do risco de erro escolhido. Isto mostra que o teste não tem poder, e a falta de poder pode ser a única razão por não podermos rejeitar a hipótese nula, e porisso seria aconselhável novos estudos. Por outro lado, pode ocorrer do IC não cobrir o valor de θ_0, mas a largura do intervalo ser bastante pequena e todos os valores do IC estarem próximos do valor de θ_0. Neste

Embora não seja discutido aqui, os intervalos de confiança construídos através de testes ótimos (teste UMP e melhor entre os não viciados, que será apresentado mais tarde) tem também certas propriedades ótimas.

Exemplo 4.7: Considere a família de distribuições normais com média desconhecida μ e variância 1, se utilizarmos o teste da razão de verossimilhança para $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, obtemos o seguinte teste:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } |\bar{X} - \mu_0| \geq z/\sqrt{n} \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde z é tal que $\Phi(z) - \Phi(-z) = 1 - \alpha$. Podemos definir nosso intervalo de confiança:

$$\begin{aligned} IC(\mathbf{X}) &= \{\mu | -\frac{z}{\sqrt{n}} < \bar{X} - \mu < \frac{z}{\sqrt{n}} \\ &= \{\mu | \bar{X} - \frac{z}{\sqrt{n}} < \mu < \bar{X} + \frac{z}{\sqrt{n}}\} \end{aligned}$$

Assim obtemos um IC para μ de nível $1 - \alpha$.

Exercício 4.8: Quais os IC mencionados na discussão do *Exemplo 3.3*? Qual seria a relação entre coeficiente de confiança e valor-p?

4.5 Métodos para Encontrar Testes

Vários métodos já foram apresentados para realizar testes de hipóteses. Nesta seção eles serão rapidamente revistos e na última parte da subseção será apresentada a relação entre testes de hipóteses e intervalos de confiança. A discussão servirá também para colocar alguns cuidados necessários na utilização de testes de hipóteses e intervalos de confiança na análise estatística.

4.5.1 Teste da Razão de Verossimilhança

Já vimos que sempre podemos encontrar um teste através da razão de verossimilhnça. Se queremos testar $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ baseados em uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição de probabilidade com densidade (ou função de probabilidade) $f(\cdot, \theta)$, onde $\theta \in \Theta$.

A *razão de verossimilhança generalizada*, denotada por λ, é dada por:

$$\lambda = \lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta, x_1, \ldots, x_n)}{\sup_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n)}.$$

e

$$\Lambda = \lambda(X_1, \ldots, X_n)$$

é uma v.a. e mais ainda é uma estatística.

A estatística Λ é usada para formular o teste da razão de verossimilhança:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \Lambda \leq k \\ \text{Não rejeito } H_0, & \text{se } \Lambda > k \end{cases}$$

onde k é uma constante fixa $0 < k < 1$ e é geralmente especificada fixando-se o tamanho do teste α.

Para família de densidades com *razão de verossimilhança monótona* o *Teorema 4.2* é bastante útil para encontrar testes UMP.

4.5.2 Relação entre Teste de Hipóteses e Intervalos de Confiança

Podemos utilizar um intervalo de confiança para um parâmetro unidimensional θ para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ e vice-versa.

Suponha que temos uma amostra aleatória $X_1, \ldots, X_n$ de uma distribuição $f(\cdot, \theta)$ e queremos testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Se temos $IC_\gamma(\mathbf{X})$ = o intervalo de confiança de nível γ para θ podemos definir a seguinte regra intuitiva: rejeitamos H_0 se θ_0 não pertencer ao IC_γ. O tamanho do teste pode ser facilmente calculado. Sabemos que se $IC_\gamma(\mathbf{X})$ é um IC de nível γ então:

$$\mathbf{P}_\theta(\theta \in IC_\gamma(\mathbf{X})) = \gamma,$$

e consequentemente a probabilidade de rejeitar H_0 quando H_0 é verdadeira é:

$$t(\delta) = \mathbf{P}_{\theta_0}(\theta_0 \notin IC_\gamma(\mathbf{X})) = 1 - \gamma.$$

Assim, um intervalo com 95% de confiança nos dá um teste com tamanho 5%.

Na realidade, temos uma dualidade aqui, pois uma família de teste pode ser usada para construir intervalos de confiança. Suponha que $C(\theta_0)$ é a região crítica de um teste de tamanho α para $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, definido para cada θ_0:

$$\mathbf{P}_{\theta_0}[\mathbf{X} \in C(\theta_0)] = \alpha.$$

Então, defina para cada ponto amostral $\mathbf{x}$,

$$IC(\mathbf{x}) = \{\theta | \mathbf{x} \notin C(\theta)\}$$

Há uma completa equivalência entre os eventos:

$$\{\theta \in IC(\mathbf{X})\} \qquad \text{e} \qquad \{\mathbf{X} \notin C(\theta)\},$$

o que significa que as probabilidades são as mesmas

$$\mathbf{P}_\theta\{\theta \in IC(\mathbf{X})\} = 1 - \mathbf{P}_\theta\{\mathbf{X} \in C(\theta)\} = 1 - \alpha,$$

Assim, $IC(\mathbf{X})$ é uma região de confiança para θ com nível $1 - \alpha$, é dado pelo conjunto de todos os valores de θ que não podem ser rejeitados em um teste de hipóteses com nível de significância α, com os resultado amostral $\mathbf{X}$ observado.

A região de confiança obtida através de um teste pode ou não ser um intervalo. Se for utilizada uma região crítica unicaudal para uma estatística do teste T, então geralmente obtemos cotas de confiança para θ.

Os IC construídos através de testes de hipóteses podem ter uma largura grande ou não. Se a largura do intervalo for grande isto implica que existem valores de θ distantes de θ_0, que não podem ser rejeitados, por serem considerados consistentes com os valores amostrais, dentro do risco de erro escolhido. Isto mostra que o teste não tem poder, e a falta de poder pode ser a única razão por não podermos rejeitar a hipótese nula, e porisso seria aconselhável novos estudos. Por outro lado, pode ocorrer do IC não cobrir o valor de θ_0, mas a largura do intervalo ser bastante pequena e todos os valores do IC estarem próximos do valor de θ_0. Neste caso, embora seja detectada estatisticamente uma diferença entre a verdadeira distribuição da população e a dada pela hipótese nula, esta diferença pode não ser importante em termos práticos; por exemplo, existe um ganho de produtividade, mas ele não tem significância prática. Isto mostra que é necessário um certo cuidado ao discutir um resultado experimental, e que muitas vezes um resultado de um teste de hipóteses deve vir necessariamente acompanhado de uma estimativa por IC. Por exemplo, nós não vamos sugerir que se mude um processo de produção simplesmente porque encontramos evidências de que o novo processo é melhor do que o padrão. A melhora pode ser pequena e não compensar os custos das mudanças. Isto não implica que pequenas diferenças não são importantes. Suponha que uma nova droga é testada e mostrou-se estatisticamente melhor do que o padrão, estimando-se a diminuição, com uma confiança de 95%, do tempo médio de recuperaço de 120 dias para $(119,1 \pm 0,5)$ dias. Neste caso a decisão deve-se basear em outros fatores como custo, facilidade de aplicação, contra-indicações, variabilidade do tempo, etc. É importante salientar que a necessidade de apresentar estimativas junto com teste de hipóteses não ocorre simplesmente quando o IC é construído a partir do teste de hipóteses. Outro ponto a considerar é que o valor-p dá a evidência em favor da hipótese alternativa e não estima a diferença entre a distribuição real da população e a distribuição dada pela hipótese nula.

Embora não seja discutido aqui, os intervalos de confiança construídos através de testes ótimos (teste UMP e melhor entre os não viciados, que será apresentado mais tarde) tem também certas propriedades ótimas.

Exemplo 4.7: Considere a família de distribuições normais com média desconhecida μ e variância 1, se utilizarmos o teste da razão de verossimilhança para $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, obtemos o seguinte teste:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } |\bar{X} - \mu_0| \geq z/\sqrt{n} \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde z é tal que $\Phi(z) - \Phi(-z) = 1 - \alpha$. Podemos definir nosso intervalo de confiança:

$$\begin{aligned} IC(\mathbf{X}) &= \{\mu| -\frac{z}{\sqrt{n}} < \bar{X} - \mu < \frac{z}{\sqrt{n}} \\ &= \{\mu|\bar{X} - \frac{z}{\sqrt{n}} < \mu < \bar{X} + \frac{z}{\sqrt{n}}\} \end{aligned}$$

Assim obtemos um IC para μ de nível $1 - \alpha$.

Exercício 4.8: Quais os IC mencionados na discussão do *Exemplo 3.3*? Qual seria a relação entre coeficiente de confiança e valor-p?

5 Teste de Hipóteses para Distribuições Normais

5.1 Testes a respeito da média

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $N(\mu, \sigma^2)$ e estamos interessados em testar hipóteses sobre a média μ. Há uma grande variedade de hipóteses que podem ser feitas a respeito da média. Vamos começar considerando as hipóteses unilaterais.

Caso 1: $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.

Temos dois casos a considerar: σ conhecido e σ desconhecido.

(1) σ conhecido. Neste caso o espaço paramétrico é:

$$\Theta = \{\mu | -\infty < \mu < \infty\}$$

e a densidade pode ser escrita como:

$$\begin{aligned} f(x,\mu) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\frac{\mu}{\sigma})^2\} \exp\{-\frac{1}{2}(\frac{x}{\sigma})^2\} \exp\{\frac{\mu}{\sigma}x\} \end{aligned}$$

que pertence à família exponencial com

$$a(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\frac{\mu}{\sigma})^2\}$$

$$b(x) = \exp\{-\frac{1}{2}(\frac{x}{\sigma})^2\}$$

$$c(\mu) = \frac{\mu}{\sigma} \qquad \text{e} \qquad d(x) = x.$$

Neste caso, $c(\mu)$ é função crescente de μ e a família tem razão de verossimilhança monótona decrescente em $T(\mathbf{X}) = \sum_{i=1}^n X_i$ (ver *Definição 4.3:* se $\mu' < \mu$" então $L(\mu', \mathbf{X})/L(\mu", \mathbf{X})$ é decrescente em $T(\mathbf{X})$) e o teste UMP de tamanho α é dado por:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } T \geq k^* \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde k^* é tal que

$$\begin{aligned} \alpha &= \mathbf{P}_{\mu_0}[\sum_{i=1}^n X_i \geq k^*] \\ &= \mathbf{P}_{\mu_0}\left[\frac{\sum X_i - n\mu_0}{\sqrt{n}\sigma} \geq \frac{k^* - n\mu_0}{\sqrt{n}\sigma}\right] \\ &= 1 - \Phi(\frac{k^* - n\mu_0}{\sqrt{n}\sigma}). \end{aligned}$$

Daí,

$$\frac{k^* - n\mu_0}{\sqrt{n}\sigma} = z_{1-\alpha} \Rightarrow k^* = n\mu_0 + z_{1-\alpha}\sqrt{n}\sigma,$$

e o teste é:

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \sum X_i \geq n\mu_0 + z_{1-\alpha}\sqrt{n}\sigma \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Isto é,

$$\delta^*(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \bar{X} \geq \mu_0 + z_{1-\alpha}\sigma/\sqrt{n} \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

(2) σ **desconhecido.** Neste caso também podemos fazer o teste da razão de verossimilhança , ou podemos encontrar uma estatística que se comporta diferentemente sob as duas hipóteses e basear nosso teste nela. Tal estatística é:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

e sabemos que T tende a ser maior para valores grandes de $\mu > \mu_0$ do que para valores $\mu \leq \mu_0$. Um teste baseado em T poderia ser:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } T \geq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde k é tal que

$$\begin{aligned} \alpha &= \mathbf{P}_{\mu_0}[T \geq k] \\ &= 1 - \mathbf{P}_{\mu_0}[T < k] \end{aligned}$$

onde sob a suposição que H_0 é verdadeira $\mu = \mu_0$ e $T \sim t_{n-1}$ e $k = t_{1-\alpha;n-1}$. E o teste seria:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \bar{X} \geq \mu_0 + t_{1-\alpha;n-1}S/\sqrt{n} \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Exemplo 4.8: Calcule o valor-p no teste anterior.
Segundo a *Definição 3.3* o valor-p é definido como o menor nível de significância em que ainda rejeitamos a hipótese nula com o valor amostral observado. Do teste estatístico temos que rejeitamos a hipótese nula se, e somente se,

$$\begin{aligned} \bar{X} &\geq \mu_0 + t_{1-\alpha;n-1}S/\sqrt{n} \\ &\Leftrightarrow t_{1-\alpha;n-1} \leq \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \\ &\Leftrightarrow \alpha \geq P[t_{1-\alpha;n-1} \geq \frac{\bar{X} - \mu_0}{S/\sqrt{n}}] = \text{valor-p} \end{aligned}$$

Observe que antes de termos de conhecermos o valor amostal o valor-p é uma estatística.

Exemplo 4.9: Foram observados os seguintes tempos de reação dos ratos, em segundos, a um certo estímulo:

2,98 2,65 4,62 2,57 7,21 2,35 13,20 4,07 9,38 4,30 4,47 2,38 4,55 14,99 1,78

fazendo um ramo-e-folha é evidente que não temos uma amostra de uma população normal, mas provavelmente de uma distribuição exponencial (de fato eles são dados simulados pelo MINITAB de uma distribuição exponencial com média 5.2 - infelizmente esqueci o valor da semente). Embora os testes de normalidade não tenham muito poder para tamanhos amostrais igual a 15, o teste de normalidade de Ryan-Joiner produz um valor-p menor do que 0,01 e o de Anderson-Darling um valor igual a 0,001; ou seja os testes confirmam a grande evidência de que a distribuição da população não seja normal. No entanto, vamos utilizar a estatística de teste T e a distribuição t-Student para verificar a robustez do teste e o seu poder. Vamos considerar o teste de:

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0$$

Sabemos que, sendo a distribuição da população uma exponencial o teste mais poderoso rejeita a hipótese quando a soma dos valores é suficientemente grande. Como a distribuição exata da soma de exponencias tem uma distribuição Gama, ou equivalentemente, que exponenciais com média 2 tem distribuição χ^2_2 podemos encontrar o valor exato do valor-p para o teste mais poderoso. Este valor-p apresentado na última coluna da tabela abaixo. O valor-p do teste T poderia ser calculado via simulação. Como exercício calcule os valores-p exatos para o teste T e complete a tabela. Para alguns valores de μ_0 temos:

Tabela: valores-p aproximados e exato do teste T e valor-p exato do teste UMP

μ_0	t	aproximado	exato	valor- teste UMP
4,7	0,70	0,25		0,255
4,2	1,18	0,13		0,131
3,7	1,66	0,060		0,047
3,5	1,85	0,043		0,028

Verifique que até para valores-p próximos de 0,10 os valores-p calculados pela aproximação *t-Student* para o teste T e do teste UMP são muito próximos. Isto indica que provavelmente a aproximação pela distribuição *t-Student* funciona e que o teste tem um poder próximo ao teste UMP. Já para valores nas caudas deve estar acontecendo simultaneamente duas coisas: a aproximação não funciona nos extremos das caudas e o teste perde poder (valor-p menor para o teste UMP). Para separar estas 2 causas seria necessário completar a outra coluna. Aqui valem os mesmos comentários da *Seção 3.1* sobre a aproximação da distribuição exata da quantidade T pela distribuição *t-Student*, o que é confirmado pelo exemplo. Isto mostra que o teste, embora tenha sido desenvolvido para distribuições normais pode ser considerado como um teste mais geral para teste de médias, já que existe uma certa robustez em relação a distribuição da população. Não se esqueça que o teste não é muito robusto a valores aberrantes.

Exercício 4.9: Através da técnica de Monte Carlo complete a tabela do *Exemplo 4.9* e comente os resultados.

Exercício 4.10: Mostre que o teste apresentado é o teste da razão de verossimilhança quando temos uma distribuição normal.

Caso 2: $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

(1) σ **é conhecido**. Sabemos que

$$IC_\gamma(\mu) = [\bar{X} - z_{(1+\gamma)/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{(1-\gamma)/2}\frac{\sigma}{\sqrt{n}}]$$

é um IC de nível γ. Um teste possível seria:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \mu_0 \notin IC_\gamma \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Tal teste tem tamanho $\alpha = 1 - \gamma$.

(2) σ **é desconhecido**. Sabemos que

$$IC_\gamma(\mu) = [\bar{X} - t_{(1+\gamma)/2;n-1}\frac{S}{\sqrt{n}}; \bar{X} + t_{(1-\gamma)/2;n-1}\frac{S}{\sqrt{n}}]$$

é um IC de nível γ. Um teste possível seria:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \mu_0 \notin IC_\gamma \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Tal teste tem tamanho $\alpha = 1 - \gamma$.

Ou então podemos achar o teste da razão de verossimilhança generalizado. Neste caso a verossimilhança é:

$$L(\mu, \sigma, x_1, \ldots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\}$$

e $\Theta = \{(\mu, \sigma); -\infty < \mu < \infty, \sigma > 0\}$ e $\Theta_0 = \{(\mu, \sigma); \mu = \mu_0, \sigma > 0\}$. Daí,

$$\begin{aligned} \sup_{(\mu,\sigma)\in\Theta} L(\mu, \sigma, x_1, \ldots, x_n) &= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\} \\ &= \frac{n^{n/2}}{[2\pi\sum(x_i - \bar{x})^2]^{n/2}} \exp\{-\frac{n}{2}\} \end{aligned}$$

Para maximizar L restrito à Θ_0 coloque $\mu = \mu_0$ e maximize em σ.

$$\begin{aligned} \sup_{(\mu,\sigma)\in\Theta_0} L(\mu, \sigma, x_1, \ldots, x_n) &= \sup_{\sigma>0} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2\} \\ &= \frac{n^{n/2}}{[2\pi\sum(x_i - \mu_0)^2]^{n/2}} \exp\{-\frac{n}{2}\} \end{aligned}$$

A razão de verossimilhança generalizada é:

$$\begin{aligned} \lambda(x_1, \ldots, x_n) &= \left[\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \mu_0)^2}\right]^{n/2} \\ &= \left[\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \bar{x} + \bar{x} - \mu_0)^2}\right]^{n/2} \\ &= \left[\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}\right]^{n/2} \\ &= \left[\frac{1}{1 + n(\bar{x} - \mu_0)^2/\sum(x_i - \bar{x})^2}\right]^{n/2} \end{aligned}$$

Neste caso,

$$\Lambda = \left[\frac{1}{1 + n(\bar{X} - \mu_0)^2 / \sum (x_i - \bar{X})^2} \right]^{n/2}$$

e o teste da RV é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \Lambda \leq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Note que,

$$\begin{aligned} \Lambda \leq k \quad & \Leftrightarrow \quad \left[\frac{1}{1 + n(\bar{X} - \mu_0)^2 / \sum (x_i - \bar{X})^2} \right]^{n/2} \leq k \\ & \Leftrightarrow \quad \frac{1}{1 + n(\bar{X} - \mu_0)^2 / \sum (x_i - \bar{X})^2} \leq k^* \\ & \Leftrightarrow \quad 1 + n(\bar{X} - \mu_0)^2 / \sum (x_i - \bar{X})^2 \geq k' \\ & \Leftrightarrow \quad \frac{(\bar{X} - \mu_0)^2}{\sum (x_i - \bar{X})^2 / (n-1) n} \geq k" \end{aligned}$$

Portanto um teste da RV é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \frac{(\bar{X} - \mu_0)^2}{S^2/n} \geq k" \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

ou

$$\delta(\mathbf{X}) = \begin{cases} \text{Não rejeito } H_0, & \text{se } -c \leq \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}} \leq c \\ \text{Rejeito } H_0, & \text{caso contrário} \end{cases}$$

Sabemos que

$$T = \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}$$

tem uma distribuição t com $n-1$ graus de liberdade quando $\mu = \mu_0$, determinamos c através da tabela da distribuição t por:

$$c = t_{1-\alpha/2; n-1}$$

Observe que este teste é o mesmo que o obtido utilizando-se o método do IC. Pode-se provar que o teste obtido não é o teste UMP, mas o mais poderoso dentre os testes não viciados.

Exercício 4.11: Gere uma amostra aleatória de tamanho 20 de uma distribuição exponencial com média igual a 5,0. Faça uma análise dos dados através de um ramo-e-folha e do gráfico probabilístico normal e aplique qualquer teste de normalidade. Comente os resultados. Construa intervalos de confiança 90%, 95% e 99% para a média populacional utilizando a aproximação pela distribuição t-Student. Utilize a distribuição exata para calcular a confiança em cada um dos intervalos encontrados. Comente os resultados encontrados.

Exercício 4.12: Ache testes para $H_0 : \mu \leq \mu_0$. O que você faria para obter um teste para testar hipóteses da forma $H_0 : \mu_1 \leq \mu \leq \mu_2$ versus $H_1 : \mu < \mu_1$ ou $\mu > \mu_2$?

5.2 Testes de Variância

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $N(\mu, \sigma^2)$ e estamos interessados em testar hipóteses sobre a variância σ^2. Há uma grande variedade de hipóteses que podem ser feitas a respeito da média. Vamos começar considerando as hipóteses unilaterais.

Caso 1:$H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$.

Temos dois casos a considerar: μ conhecido e μ desconhecido.

(1) μ conhecido. Neste caso o espaço paramétrico é o intervalo:

$$\Theta = \{\sigma^2 | \sigma^2 > 0\},$$

a hipótese alternativa é unilateral e podemos encontrar um teste UMP. A densidade pode ser escrita como:

$$f(x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

que pertence à família exponencial com

$$a(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}; \quad b(x) = 1, \quad c(\sigma^2) = -\frac{1}{2\sigma^2}, \quad \text{e} \quad d(x) = (x-\mu)^2$$

Temos que $c(\sigma^2) = -\frac{1}{2\sigma^2}$ é uma função monótona crescente de σ^2 e temos que o teste UMP de tamanho α para $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$ é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \sum_{i=1}^n (X_i - \mu)^2 \geq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde k é tal que

$$\mathbf{P}_{\sigma_0^2}[\sum_{i=1}^n (X_i - \mu)^2 \geq k] = \alpha,$$

isto é, $k = \sigma_0^2 \chi^2_{1-\alpha;n}$.

(1): μ desconhecido. Podemos utilizar a estatística

$$V = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2}$$

a qual tende a ser grande para valores de $\sigma^2 > \sigma_0^2$ e pequena para valores de $\sigma^2 < \sigma_0^2$; assim um teste razoável seria:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } V \geq k \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

O tamanho do teste é:

$$t(\delta) = \mathbf{P}_{\sigma_0^2}[V \geq k]$$

e sabemos que se $\sigma^2 = \sigma_0^2$, $V \sim \chi^2_{n-1}$ e temos

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } V \geq \chi^2_{1-\alpha;n-1} \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Caso 2:$H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$.

Temos dois casos a considerar: μ conhecido e μ desconhecido.

Exercício 4.13: Achar o teste da RV para o caso de μ ser conhecido.

(2): μ **desconhecido**, podemos achar o teste da RV. Também podemos usar o IC. Já vimos que o IC de nível $\gamma = 1 - \alpha$ para σ^2 é dado por:

$$IC_\alpha = \left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}} \right).$$

Assim, um teste de tamanho α para $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$ é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } \sigma_0^2 \in IC_\alpha \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Ao contrário dos testes para médias apresentados o teste da variância depende bastante da suposição de normalidade, ou seja, não é robusto em relação a suposição de normalidade. Outro problema deste teste é o baixo poder para amostras não muito grandes.

5.3 Testes de Várias Médias

Muitas vezes estamos interessados em comparar duas ou mais médias. Por exemplo, queremos verificar a eficácia de um novo medicamento na sobrevida de pessoas com AIDS, comparando-o com o tradicionalmente usado AZT. Para isso retiramos amostras aleatórias das populações recebendo o novo medicamento e o antigo e queremos comparar o tempo médio de sobrevida destas populações. Num outro problema estamos interessados em verificar a eficácia de diferentes pesticidas A, B, C e D no combate à ferrugem do feijão. Para realizar o experimento plantamos diversos exemplares de feijão e tratamos n_1 deles com pesticida A, n_2 deles com pesticida B, n_3 deles com pesticida C e n_4 deles com pesticida D e com base nesses dados queremos comparar a incidência média de ferrugem em cada população. Deseja-se verificar se um novo material utilizado para fabricação de solas de sapatos é tão durável quanto o antigo material, para isso fabrica-se pares de sapatos onde cada pé é revestido de um material e utiliza-se esses sapatos num grupo de 10 crianças.

5.3.1 Igualdade de 2 Médias

Por exemplo, desejamos comparar dois fornecedores de matéria prima para a fabricação de cerâmicas refratárias, onde estamos interessados na resistência média ao calor destas cerâmicas.

Para podermos estudar este problema temos que verificar claramente se os dados vem de populações independentes ou se os dados foram pareados.

Dados pareados: Neste caso temos uma amostra aleatória de uma distribuição normal bivariada. Seja $(X_1, Y_1), \ldots, (X_n, Y_n)$ v.a.'s i.i.d $N_2((\mu_1, \mu_2), \Sigma)$ onde:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

e queremos testar $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Neste caso, observe que se transformamos os dados em $D_i = X_i - Y_i$ temos uma amostra aleatória de uma distribuição $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$ e sabemos testar $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ baseados nesta amostra considerando-se $\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$ como a variância desconhecida.

Amostras independentes: Suponha que temos n_1 observações $X_1, \ldots, X_{n_1}$ de uma distribuição $N(\mu_1, \sigma_1^2)$ e uma amostra independente de n_2 observações $Y_1, \ldots, Y_{n_2}$ de uma distribuição $N(\mu_2, \sigma_2^2)$. Com base nestes dados queremos testar: $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$.

O espaço paramétrico neste caso tem dimensão quatro. O subspaço Θ_0 é tridimensional. A função de verossimilhança é:

$$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2; x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}) =$$
$$= \left(\frac{1}{2\pi\sigma_1^2}\right)^{n_1/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n_1}\left(\frac{x_i - \mu_1}{\sigma_1}\right)^2\right] \left(\frac{1}{2\pi\sigma_2^2}\right)^{n_2/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n_2}\left(\frac{y_i - \mu_2}{\sigma_2}\right)^2\right]$$

e seu máximo para todo o espaço paramétrico é dado por:

$$\sup_{\theta\in\Theta} L(\theta) = \left[\frac{2\pi}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2\right]^{-n_1/2} \left[\frac{2\pi}{n_2}\sum_{i=1}^{n_2}(y_i - \bar{y})^2\right]^{-n_2/2} \exp\{-\frac{1}{2}(n_1 + n_2)\}$$

Se colocamos $\mu_1 = \mu_2 = \mu$ e tentamos encontrar $\sup_{\theta\in\Theta_0} L(\theta)$ veremos que a estimativa de μ é a raiz de uma equação cúbica e muito difícil de ser calculada explicitamente e portanto a razão de verossimilhança terá uma expressão complicada e mais complicada ainda será a sua distribuição. Além disso, a probabilidade de erro tipo I depende dos valores desconhecidos de σ_1 e σ_2 e isso torna impossível calcular exatamente o teste da razão de verossimilhança. Veremos depois que podemos achar testes assintóticos neste caso e assim resolver pelo menos parcialmente este problema.

Entretanto, se supusermos que $\sigma_1 = \sigma_2$, então temos um problema bem mais simples. Neste caso podemos fazer o teste da razão de verossimilhança facilmente (exercício). Também podemos utilizar o método do intervalo de confiança. O IC de nível γ neste caso foi encontrado utilizando-se como pivô:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

a qual tem uma distribuição $t_{n_1+n_2-2}$ e assim o IC de nível $\gamma = 1 - \alpha$ é dado por:

$$IC = \left[\bar{X} - \bar{Y} + t_{\alpha/2;n_1+n_2-2}S_p\sqrt{\frac{n_1 + n_2}{n_1 n_2}}; \bar{X} - \bar{Y} - t_{\alpha/2;n_1+n_2-2}S_p\sqrt{\frac{n_1 + n_2}{n_1 n_2}}\right]$$

e o teste de tamanho α é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } 0 \notin IC \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

Caso estejamos em dúvida se a suposição de variâncias iguais é válida ou não podemos utilizar o pivô modificado com distribuição aproximadamente t-Student (sob H_0). A discussão da *Seção 3.4* continua válida aqui.

O caso de hipóteses alternativas unilaterais é tratado com intervalos de confiança unilaterais. Por exemplo, para testar:

$$H_0 : \mu_1 \geq \mu_2 \ (\mu_1 - \mu_2 \geq 0) \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 \ (\mu_1 - \mu_2 < 0)$$

o IC de nível $\gamma = 1 - \alpha$ é dado por:

$$IC = \left(-\infty; \bar{X} - \bar{Y} + t_{1-\alpha;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

e o teste de tamanho α é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeito } H_0, & \text{se } 0 \notin IC \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

isto é, **Rejeito** H_0 se

$$\bar{X} - \bar{Y} + t_{1-\alpha;n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < 0, \text{ou}$$

$$\frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{1-\alpha;n_1+n_2-2}$$

A interpretação é bastante simples: Se tenho $100\gamma\%$ de confiança de que $\mu_1 - \mu_2 < \bar{x} - \bar{y} + t_{1-\alpha;n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, e este limite é menor do que zero, então tenho $100\gamma\%$ de confiança em H_1.

Para calcular o valor-p basta verificar o caso limite, isto é, o valor de α onde o valor amostral produz um valor no limite do IC, isto é, o valor-p é dado por:

$$P[t_{n_1+n_2-2} < \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{n_1+n_2}{n_1 n_2}}}]$$

Aleatorização: Suponha que n unidades experimentais (ratos, pessoas, lotes de terras, dias, bateladas, etc) estejam disponíveis para a realização de um experimento onde serão comparados 2 tipos de tratamentos (droga padrão e nova; processo A e B; ausência ou não de um reagente, etc). Neste caso, sempre que possível, é interessante que a distribuiçao das unidades experimentais para os tipos de tratamentos seja realizada de forma totalmente aleatória. Este procedimento, chamado de aleatorização, controla possíveis favorecimentos entre os tratamentos e garante a utilização da distribuição t-Student sob a hipótese nula de que não existe diferença entre os tratamentos. A distribuição da estatística de teste sob a hipótese nula (em certos casos no pior caso da hipótese nula) é geralmente chamada de distribuição de referência.

Suponha que os engenheiros desejavam testar se um novo reagente melhora a produtividade de um certo processo. Para tanto eles dispunham de 8 dias e escolheram 4 dias para trabalhar sem o reagente (tratamento A) e os 4 restantes com o reagente (tratamento B). Os valores encontrados de produtividade foram:

dia	1	2	3	4	5	6	7	8
tratamento	A	B	A	B	B	B	A	A
produtiv.	3,1	4,2	1,5	6,3	5,0	3,5	2,6	4,7

Sob a hipótese nula de que o novo reagente não tem nenhum efeito o valor observado 3,1, no primeiro dia, não modificaria, mesmo que o primeiro dia fosse sorteado para o tratamento B, ou seja, sob a hipótese nula os valores observados seriam o mesmo independentemente do sorteio. Existem $8!/(4!4!)$ formas de se escolher 4 unidades para cada tratamento. Como cada combinação tem a mesma probabilidade de ser escolhida a distribuição exata da estatística de teste, sob a hipótese nula, é fácil de ser encontrada. Esta distribuição é chamada de distribuição aleatorizada (randomization distribution). Quando os tamanhos amostrais crecem fica cada vez mais complicado encontrar o valor-p através da distribuição exata caso não tenhamos um programa estatístico como o *StatXact* que calcula este valor. Uma solução é encontrar este valor através de simulações ou então procurar uma distribuição aproximada. Felizmente tal aproximação existe e é dada pela distribuição t-Sudent. Ou seja, a distribuição t-Student aproxima o valor-p exato e a aletorização justifica a utilização da distribuição t-Student. Aqui também vale os mesmos comentários da *Seção 3.4* para a aproximação. A aproximação é melhor para valores amostrais maiores, para valores-p maiores, para conjunto de dados mais simétricos, e depende bastante da proporção e tamanhos relativos dos valores aberrantes.

5.3.2 Igualdade de Várias Médias:

O teste da razão de verossimilhança pode ser estendido para a comparação de k médias de populações normais. Assumimos que temos disponíveis k amostras aleatórias independentes de populações; Seja $X_{j1}, \ldots, X_{jn_j}$ uma amostra aleatória de tamanho n_j de uma população $N(\mu_j, \sigma^2)$. Assuma que as amostras são independentes. A função de verossimilhança é dada por:

$$\begin{aligned} & L(\mu_1, \ldots, \mu_k, \sigma^2; x_{11}, \ldots, x_{1n_1}, \ldots, x_{k1}, \ldots, x_{kn_k}) = \\ = \quad & (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ji}-\mu_j)^2\right], \end{aligned}$$

onde $n = \sum_{j=1}^{k} n_j$.

O espaço paramétrico é $(k+1)$-dimensional, $\Theta = \{(\mu_1, \ldots, \mu_k, \sigma^2); -\infty < \mu_j < \infty; \sigma^2 > 0\}$ e o espaço paramétrico restrito $\Theta_0 = \{(\mu, \ldots, \mu, \sigma^2); -\infty < \mu < \infty; \sigma^2 > 0\}$ é bi- dimensional. Em Θ o EMV dos parâmetros são dados por:

$$\hat{\mu}_j = \bar{X}_j; \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ji}-\bar{X}_j)^2$$

assim,

$$\sup_{\theta\in\Theta} L(\theta) = \left[\frac{2\pi\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2}{n}\right]^{-n/2} e^{-n/2}.$$

Em Θ_0 os EMV de μ e σ^2 são dados [por:

$$\tilde{\mu}_j = \bar{X} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}X_{ji}; \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ji}-\bar{X})^2,$$

e assim,

$$\sup_{\theta\in\Theta_0} L(\theta) = \left[\frac{2\pi\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ji}-\bar{x})^2}{n}\right]^{-n/2} e^{-n/2}.$$

A razão de verossimilhança generalizada é:

$$\lambda = \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \left[\frac{\sum_j \sum_i (x_{ji}-\bar{x})^2}{\sum_j \sum_i (x_{ji}-\bar{x}_j)^2}\right]^{-n/2}$$

$$= \left[\frac{\sum_j \sum_i (x_{ji}-\bar{x}_j+\bar{x}_j-\bar{x})^2}{\sum_j \sum_i (x_{ji}-\bar{x}_j)^2}\right]^{-n/2}$$

$$= \left[\frac{\sum_j \sum_i (x_{ji}-\bar{x}_j)^2+\sum_j n_j(\bar{x}_j-\bar{x})^2}{\sum_j \sum_i (x_{ji}-\bar{x}_j)^2}\right]^{-n/2}$$

$$= \left[1+\frac{k-1}{n-k}\frac{\sum_j n_j(\bar{x}_j-\bar{x})^2/(k-1)}{\sum_j \sum_i (x_{ji}-\bar{x}_j)^2/(n-k)}\right]^{-n/2}$$

Um teste da razão de verossimilhanç rejeita H_o se $\Lambda \leq \lambda_0$, mas $\Lambda \leq \lambda_0$ se, e somente se $R \geq c$, onde

$$R = \frac{\sum_j n_j(\bar{X}_j-\bar{X})^2/(k-1)}{\sum_j \sum_i (X_{ji}-\bar{X}_j)^2/(n-k)}$$

e c é selecionada de forma a que o teste tenha tamanho α. Sob a hipótese nula R tem uma distribuição F com $k-1$ graus de liberdade no numerador e $n-k$ graus de liberdade no denominador.

O problema estudado acima é, em geral, referido como um problema de *análise de variância de um fator*. Por exemplo, queremos saber se há alguma diferença entre os vários tipos de pesticidas disponíveis no mercado.

5.4 Testes de várias variâncias

Note que nos testes de igualdade de médias precisamos supor de que as variâncias das diversas populações são iguais. Para isto precisamos antes de aplicar o teste de igualdade de médias estar certos sobre a igualdade de variâncias.

Exercício 4.14: Algumas enfermeiras estão interessadas em verificar o efeito dos cuidados pré-natais no peso das crianças recém-nascidas. As crianças são divididas em dois grupos de acordo com o número de exames pré-natais das mães, e medidos os pesos em onças:

5 ou menos visitas: 49 108 110 82 93 114 134 114 96 52 101 114 120 116
6 ou mais visitas: 133 108 93 119 119 98 106 87 153 116 129 97 110 131

a. Faça uma análise descritiva comparativa dos dados.
b. Comente os resultados em termos de peso médio e variabilidade utilizando teste de hipóteses e intervalos de confiança.
c. Comente quais foram as suposições realmente necessárias para os testes e IC utilizados no item anterior. Comente também se estas suposições são razoáveis ou não.