UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO - UFRPE
Programa de Pós-Graduação em Biometria e Estatística Aplicada
Inferência Estatística
Prof. Frank Sinatra

Valores P

Augusto César F. de Miranda Oliveira

RECIFE
2019

1 Introdução

O valor P é derivado da perspectiva de um teste de hipótese, no qual a estatística de um teste é calculada a partir dos resultados de um determinado conjunto de dados sob a hipótese nula ser verdadeira. A distribuição da estatística de teste é utilizada para para obter a menor probabilidade de observar esse resultado (hipótese nula ser verdadeira) ou um resultado mais extremo.

Assim, o valor P é uma medida de evidência contra a hipótese nula. O valor P é baseado na análise de variáveis aleatórias, por isso o mesmo é uma variável aleatória cuja distribuição, para estatísticas de testes contínuas, é conhecida por ser uniforme no intervalo [0,1] sob a hipótese nula. Por esse fato, um ponto de corte para o valor P é utilizado. Geralmente 0.05 é utilizado para controlar as chances de que, para qualquer experimento, o valor P possa ser 0.05 ou menor, sendo a hipótese nula é verdadeira.

Esse conceito, na teoria de Neyman-Pearson de testes de hipóteses, é conhecido como taxa de erro do tipo I, que é uma taxa de erro que determina a região de rejeição e busca controlar a frequência geral de rejeições errôneas da hipótese nula. Outras medidas estatísticas, como intervalos de confiança, podem subsidiar este entendimento, porém não será objeto de estudo deste trabalho.

2 Valores P

P-valor ou valor p é outro meio de relatar os resultados de um teste de hipótese, expondo, assim, um valor associado a um tipo de estatística de teste.

Definição 2.1. Um *valor p* $p(\mathbf{X})$ é uma estatística de teste que satisfaz $0 \leq p(\mathbf{x}) \leq 1$ para cada ponto amostral $\mathbf{x}$. Pequenos valores de $p(\mathbf{X})$ fornecem evidências de que H_1 é verdadeira. Um valor p é *válido* se, para cada $\theta \in \Theta_0$ e cada $0 \leq \alpha \leq 1$,

$$P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha. \tag{1}$$

Onde,

- P_θ declara a probabilidade de ocorrência de eventos extremos (rejeitar a hipótese nula sabendo que é verdadeira);
- $p(\mathbf{X})$ é uma estatística de teste sob a população em estudo e
- α denota o nível de significância estipulado.

Se $p(\mathbf{X})$ for um valor válido, é fácil criar um teste de nível α baseado em p($\mathbf{X}$). O teste que rejeita H_0, se e somente se, $p(\mathbf{X}) \leq \alpha$ é um teste de nível α por causa de (1). Um a vantagem de relatar o resultado de teste por intermédio de um valor p é que cada leitor pode escolher o α que considerar apropriado, e então, pode comparar o $p(\mathbf{x})$ relatado a α e saber se esses dados levam à aceitação ou rejeição de H_0. Deste modo, um valor p relata os resultados de um teste em uma escala mais contínua, em vez de apenas "aceitar H_0" ou "rejeitar H_0". O meio mais comum para definir um valor p válido é dado pelo Teorema 2.2.

Teorema 2.2. *Seja $W(X)$ uma estatística de teste, de modo que grande valores de W dão evidências de que H_1 é verdadeira. Para cada ponto amostral $\boldsymbol{x}$, defina*

$$p(\boldsymbol{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\boldsymbol{X}) \geq W(\boldsymbol{x})) \tag{2}$$

Então, $p(\boldsymbol{x})$ é um valor p válido.

Demonstração. Fixe $\theta \in \Theta_0$. Seja $F_\theta(w)$ denotando a FDA de $-W(\mathbf{X})$. Definimos:

$$\begin{aligned} p_\theta(\mathbf{x}) &= P_\theta(W(\mathbf{X}) \geq W(\mathbf{x})) \\ &= P_\theta(-W(\mathbf{X}) \leq -W(\mathbf{x})) \\ &= F_\theta(-W(\mathbf{x})). \end{aligned}$$

Então, a variável aleatória $p_\theta(\mathbf{X})$ é igual a F_θ(-W($\mathbf{X}$)). Deste modo, de acordo com a **Transformação Integral de Probabilidade** definida por Magalhães (2015, p. 150) como, Seja X uma variável com função de distribuição F. A transformação de X tal que $Y = F(X)$ é denominada *transformação integral de probabilidade.* $\square$

Assim, por exemplo, se temos $X \sim Exp(\lambda)$, $F(x) = 1 - e^{-\lambda x}$ e obtemos a transformação integral $Y = 1 - e^{-\lambda X}$. Em simulações, o uso da transformação depende da possibilidade de inverter F. A inversa tem domínio em [0,1], mas se F tiver saltos ou for em escada, ela não existirá. A função inversa também, por abuso de notação, pode ser representada por F^{-1}.

Logo, uma distribuição de $p_\theta(\mathbf{X})$ é estocasticamente maior ou igual a uma distribuição uniforme (0,1). Isto é, para cada $0 \leq \alpha \leq 1$, $P_\theta(p_\theta(\mathbf{X}) \leq \alpha) \leq \alpha$. Uma vez que $p(\mathbf{x}) = \sup_{\theta' \in \Theta_0} p_{\theta'}(\mathbf{x}) \geq p_\theta(\mathbf{x})$ para cada $\mathbf{x}$.

$$P_\theta(p(\mathbf{X} \leq \alpha) \leq P_\theta(p_\theta(\mathbf{X}) \leq \alpha) \leq \alpha).$$

Isto é verdadeiro para cada $\theta \in \Theta_0$ e para cada $0 \leq \alpha \leq 1; p(\mathbf{X})$ é um valor p válido. $\blacksquare$

O cálculo do supremo em (2) pode ser difícil, pois "[...] alguns conjuntos limitados de números racionais não possuem supremo (ou ínfimo). Este fato está ligado à inexistência de raízes quadradas racionais de certos números inteiros, mas é uma dificuldade que vai muito além dessa falta" (LIMA, 2019, p. 78). Os dois exemplos seguintes ilustram situações comuns nas quais isto não é muito difícil. No primeiro, nenhum supremo é necessário; no segundo, é fácil determinar o valor de θ em que o supremo ocorre.

Exemplo 2.3 (Valor p bilateral da normal)**.** Sejam $X_1,...,X_n$ uma amostra aleatória de uma população n(μ, σ^2). Considere testar $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Onde,

- μ_0 denota a média amostral do experimento.
- $H_0 : \mu = \mu_0$ declara que não houve alteração no experimento.
- $H_1 : \mu \neq \mu_0$ declara que houve alteração no experimento.

De acordo acordo com o exercício 8.38 (CASELLA; BERGER, 2002, p. 408), o Teste de Razão de Verossimilhança (TRV) rejeita H_0 para grandes valores de $W(\mathbf{X}) = |\bar{x} - \mu_0|/(S/\sqrt{n})$, em que $W(\mathbf{X})$ é uma estatística de teste para média. Sabendo-se que, σ é desconhecida, por definição temos.

$$W(\mathbf{X}) = \frac{(\bar{x} - \mu_0)}{(S/\sqrt{n})}, \text{ em que } W(\mathbf{X}) \sim T_{n-1} \tag{3}$$

Calculando (2) para encontrar um valor p válido, temos:

$$\begin{aligned} p(\mathbf{x}) &= \sup_{\mu \in \Theta_0} P_\mu(W(\mathbf{X}) \geq W(\mathbf{x})) \\ &= P_\mu(W(\mathbf{X}) \geq W(\mathbf{x})) \\ &= 2P_\mu(T_{n-1} \geq W(\mathbf{x})) \\ &= 2[1 - P_\mu(T_{n-1} < W(\mathbf{x}))] \\ &= 2\left\{1 - F^{-1}_{T_{n-1}}[W(\mathbf{x})]\right\} \end{aligned} \tag{4}$$

Em que $F^{-1}_{T_{n-1}}$, é o inverso da Função de Distribuição Acumulada (FDA) t-Student com $n-1$ graus de liberdade. Portanto, ao calcular (4), a probabilidade é a mesma para todos os valores de μ, desconsiderando, assim, o cálculo do supremo.

A Figura 1 ilustra o p valor para um teste de hipótese bilateral da normal, onde a ocorrência de valores mais extremos sob a hipótese nula ser verdadeira pode ocorrer nas duas caudas da distribuição.

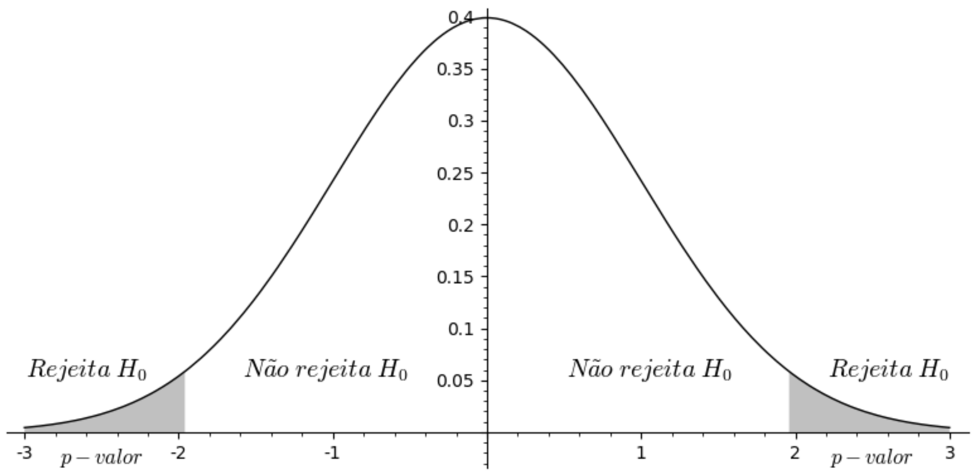

Figura 1: Valor p bilateral da normal

Exemplo 2.4 (Valor p unilateral da normal)**.** Mais uma vez, tenha em conta o modelo normal do Exemplo 2.3, mas considere testar $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. Onde,

- μ_0 é a proporção máxima aceitável de itens com defeito.
- $H_0 : \mu \leq \mu_0$ declara que a proporção de itens com defeito é aceitável.
- $H_1 : \mu > \mu_0$ declara que a proporção de itens com defeito é inaceitavelmente alta.

De acordo com o exercício 8.37 (CASELLA; BERGER, 2002, p. 407), o TRV rejeita H_0 para grandes valores de $W(\mathbf{X}) = (\overline{X} - \mu_0)/(S/\sqrt{n})$. O seguinte argumento mostra que, para esta estatística, o supremo em (2) sempre ocorre em um parâmetro (μ_0, σ), e o valor de σ utilizado não faz diferença. Considere qualquer $\mu \leq \mu_0$ e qualquer σ:

$$
\begin{aligned}
P_{\mu,\sigma}(W(\mathbf{X}) \geq W(\mathbf{x})) &= P_{\mu,\sigma}\left(\frac{\overline{x} - \mu_0}{S/\sqrt{n}} \geq W(\mathbf{x})\right) \\
&= P_{\mu,\sigma}\left(\frac{\overline{x} - \mu_0 + \mu - \mu}{S/\sqrt{n}} \geq W(\mathbf{x})\right) \\
&= P_{\mu,\sigma}\left(\frac{\overline{x} - \mu}{S/\sqrt{n}} + \frac{\mu - \mu_0}{S/\sqrt{n}} \geq W(\mathbf{x})\right) \\
&= P_{\mu,\sigma}\left(\frac{\overline{x} - \mu}{S/\sqrt{n}} \geq W(\mathbf{x}) - \left(\frac{\mu - \mu_0}{S/\sqrt{n}}\right)\right) \\
&= P_{\mu,\sigma}\left(\frac{\overline{x} - \mu}{S/\sqrt{n}} \geq W(\mathbf{x}) + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&= P_{\mu,\sigma}\left(T_{n-1} \geq W(\mathbf{x}) + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&\leq P(T_{n-1} \geq W(\mathbf{x}))
\end{aligned}
$$

Aqui, mais uma vez, T_{n-1} tem uma distribuição t-Student com $n-1$ graus de liberdade. A desigualdade na última linha é verdadeira porque $\mu_0 \geq \mu$ e $(\mu_0 - \mu)/(S/\sqrt{n})$ é uma variável aleatória não negativa. Aqui, o subscrito em P é eliminado, porque esta probabilidade não depende de (μ, σ). Além disso,

$$P(T_{n-1} \geq W(\mathbf{x})) = P_{\mu_0,\sigma}\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \geq W(\mathbf{x})\right) = P_{\mu_0,\sigma}(W(\mathbf{X}) \geq W(\mathbf{x})),$$

e esta probabilidade é uma daquelas consideradas no cálculo do supremo em (2) porque $(\mu_0,\sigma) \in \Theta_0$. Portanto, o valor p de (2) para este teste unilateral t é $p(\mathbf{x}) = P(T_{n-1} \geq W(\mathbf{x})) = P(T_{n-1} \geq (\bar{x} - \mu_0/(s/\sqrt{n}))$. A Figura 2 ilustra o p valor para um teste de hipótese unilateral da normal, onde a ocorrência de valores mais extremos sob a hipótese nula ser verdadeira pode ocorrer apenas na cauda direita da distribuição.

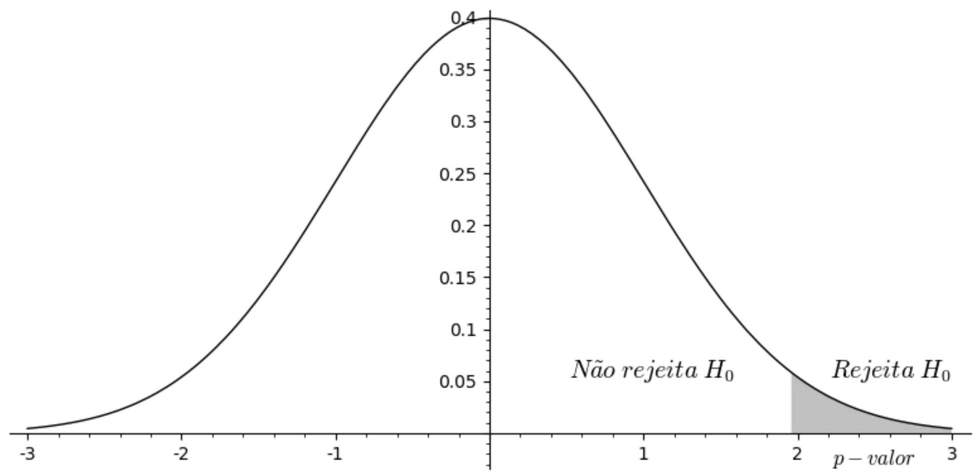

Figura 2: Valor p unilateral da normal

Outro método para definir um valor p válido, uma alternativa à utilização de (2), envolve condicionar sobre uma estatística suficiente. Suponha que $S(\mathbf{X})$ seja uma estatística suficiente para o modelo $\{f(\mathbf{x}|\theta) : \theta \in \Theta_0\}$. (Para evitar testes com poder baixo, é importante que S seja suficiente somente para o modelo nulo, não para todo modelo $\{f(\mathbf{x}|\theta) : \theta \in \Theta\}$.) Se a hipótese nula for verdadeira, a distribuição condicional de $\mathbf{X}$ dado $S = s$ não depende de θ. Mais uma vez, seja $W(\mathbf{X})$ denotando uma estatística de teste para a qual grandes valores dão evidência de que H_1 é verdadeira. Então, para cada ponto amostral $\mathbf{x}$ definimos:

$$p(\mathbf{x}) = P(W(\mathbf{X}) \geq W(\mathbf{x})|S = S(\mathbf{X})) \tag{5}$$

Argumentando como no Teorema 2.2, mas considerando somente a distribuição única, que é a distribuição condicional de $\mathbf{X}$ considerando que $S = s$, vemos que, para qualquer, $0 \leq \alpha \leq 1$,

$$P(p(\mathbf{X}) \leq \alpha|S = s) \leq \alpha$$

Então, para qualquer $\theta \in \Theta_0$, temos, incondicionalmente,

$$P_\theta(p(\mathbf{X}) \leq \alpha) = \sum_s P(p(X) \leq \alpha|S = s)P_\theta(S = s) \leq \sum_n \alpha P_\theta(S = s) \leq \alpha$$

Deste modo, $p(\mathbf{X})$ definido por (5) é um valor p válido. Somas podem ser substituídas pelas integrais para S contínuo, mas este método geralmente é utilizado para S discreto, como no exemplo seguinte.

Exemplo 2.5 (Teste Exato de Fisher)**.** Sejam S_1 e S_2 observações independentes, com $S_1 \sim Binomial(n_1,p_1)$ e $S_2 \sim Binomial(n_2,p_2)$. Considere testar $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$. De acordo com H_0, supondo que p denote o valor comum de $p_1 = p_2$, a fp conjunta de (S_1,S_2) é

$$\begin{aligned} f(s_1,s_2|p) &= \binom{n_1}{n_2} p^{s_1}(1-p)^{n_1-n_2} \binom{n_2}{p_2} p^{s_2}(1-p)^{n_2-p_2} \\ &= \binom{n_1}{n_2}\binom{n_2}{s_2} p^{s_1+s_2}(1-p)^{n_1+n_2-(s_1+s_2)} \end{aligned}$$

Portanto, $S = S_1 + S_2$ é uma estatística suficiente, de acordo com H_0. Considerando o valor de $S = s$, é razoável utilizar S_1 como uma estatística de teste e rejeitar H_0 em favor de H_1 para grandes valores de S_1, porque grandes valores de S_1 correspondem a pequenos valores de $S_2 = s - S_1$. A distribuição condicional S_1 dado $S = s$ é hipergeométrica(n_1+n_2,n_1,s), vejamos:

Sabendo-se que, por definição, a distribuição condicional pode ser reescrita em notação de função de probabilidade da seguinte forma:

$$P_{x|y}(x|y) = \frac{P_{x,y}(x,y)}{P_y(y)}$$

E, se X,Y são independentes, então $P_{x,y}(x,y) = P_x(x)\cdot P_y(y)$. Sendo assim, temos:

$$\begin{aligned} P(S_1 = s_1|S_1+S_2 = s) &= \frac{P(S_1 = s_1, S_1+S_2 = s)}{P(S_1+S_2 = s)} \\ &= \frac{P(S_1 = s_1, S_2 = s-S_1)}{P(S_1+S_2 = s)} \\ &= \frac{P(S_1 = s_1, S_2 = s-s_1)}{P(S_1+S_2 = s)} \\ &= \frac{P(S_1 = s_1)P(S_2 = s-s_1)}{P(S_1+S_2 = s)} \end{aligned} \tag{6}$$

Onde,

- $S_1 \sim Binomial(n_1,s_1) = \binom{n_1}{s_1} p^{s_1}(1-p)^{n_1-s_1}$
- $S_2 \sim Binomial(n_2,s-s_1) = \binom{n_2}{s-s_1} p^{s-s_1}(1-p)^{n_2-(s-s_1)}$
- $S_1+S_2 \sim Binomial(n1+n2,s) = \binom{n_1+n_2}{s} p^{s}(1-p)^{n_1+n_2-s}$

Retomando a equação (6), temos:

$$\begin{aligned} &= \frac{\binom{n_1}{s_1} p^{s_1}(1-p)^{n_1-s_1}\cdot\binom{n_2}{s-s_1} p^{s-s_1}(1-p)^{n_2-s+s_1}}{\binom{n_1+n_2}{s} p^{s}(1-p)^{n_1+n_2-s}} \\ &= \frac{\binom{n_1}{s_1}\binom{n_2}{s-s_1} p^{s_1+s-s_1}(1-p)^{n_1-s_1+n_2-s_1+s_1}}{\binom{n_1+n_2}{s} p^{s}(1-p)^{n_1+n_2-s}} \\ &= \frac{\binom{n_1}{s_1}\binom{n_2}{s-s_1} p^{s}(1-p)^{n_1+n_2-s}}{\binom{n_1+n_2}{s} p^{s}(1-p)^{n_1+n_2-s}} \\ &= \frac{\binom{n_1}{s_1}\binom{n_2}{s-s_1}}{\binom{n_1+n_2}{s}} \\ &= hipergeométrica(n1+n2,s) \end{aligned}$$

Assim, o valor p condicional em (5) é

$$p(s_1,s_2) = \sum_{j=s_1}^{min(n_1,s)} f(j|s),$$

a soma das probabilidades hipergeométricas. O teste definido por este valor p é chamado de Teste Exato de Fisher.

Referências

CASELLA, George; BERGER, Roger L. **Statistical inference**. Pacific Grove: Duxbury, 2002.

LIMA, Elon Lages. Curso de Análise vol. 1–15ª edição. **Rio de Janeiro: IMPA–Projeto Euclides**, 2019.

MAGALHÃES, Marcos Nascimento. **Probabilidade e variáveis aleatórias**. São Paulo: Edusp, 2015.