Capítulo 4

Testes de Hipóteses

Uma *hipótese* é uma afirmação sobre o estado da natureza. 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.

4.1 Um exemplo

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: "Se $\bar{R} \leq 150$ decido que a procedência é americana" e "se $\bar{R} > 150$ decido que a procedência é japonesa".

Suponha que no dia do leilão fomos informados que $\bar{r} = 148$. Portanto diríamos que o lote é americano.

Podemos estar enganados? É possível que uma amostra de 25 parafusos japoneses apresentem uma média de 148 Kgf? Sim!!!

Para melhor entendermos uma regra de decisão devemos estudar os tipos de erros possíveis:
Erro Tipo I: Dizer que os parafusos são americanos quando, na realidade, são japoneses;
Erro Tipo II: Dizer que os parafusos são japoneses quando, na realidade, são americanos.

Para facilitar a linguagem, defina duas hipóteses:
H_0: os parafusos são japoneses, i.e., $R \sim N(155, 400)$;
H_1: os parafusos são americanos, i.e., $R \sim N(145, 144)$.

Defina a região crítica:

$$\Gamma = \{y \in [0, \infty); y \leq 150\}.$$

Portanto, podemos dizer que:

$$\mathbb{P}[\text{erro tipo I}] = \mathbb{P}[\bar{R} \in \Gamma | H_0 \text{ verdadeira}] = \alpha$$

$$\mathbb{P}[\text{erro tipo II}] = \mathbb{P}[\bar{R} \notin \Gamma | H_1 \text{ verdadeira}] = \beta$$

Quando H_0 é verdadeira temos que $R \sim N(155, 400)$, o que implica $\bar{R} \sim N(155, 16)$;
Quando H_1 é verdadeira temos que $R \sim N(145, 144)$, o que implica $\bar{R} \sim N(155, 144/25)$.

Daí,

$$\begin{aligned}
\mathbb{P}[\text{erro tipo I}] &= \mathbb{P}[\bar{R} \in \Gamma | \bar{R} \sim N(155, 16)] \\
&= \mathbb{P}[\bar{R} \leq 150 | \bar{R} \sim N(155, 16)] \\
&= \mathbb{P}\left[\frac{\bar{R} - 155}{4} \leq \frac{150 - 155}{4} \mid \bar{R} \sim N(155, 16)\right] \\
&= \mathbb{P}[Z \leq -1.25] = 0.1056 = 10.56\% = \alpha
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{P}[\text{erro tipo II}] &= \mathbb{P}[\bar{R} \notin \Gamma | \bar{R} \sim N(145, 144/25)] \\
&= \mathbb{P}[\bar{R} > 150 | \bar{R} \sim N(145, 144/25)] \\
&= \mathbb{P}\left[\frac{\bar{R} - 145}{12/5} > \frac{150 - 145}{12/5} \mid \bar{R} \sim N(145, 144/25)\right] \\
&= \mathbb{P}[Z > 2.08] = 0.0187 = 1.87\% = \beta
\end{aligned}$$

Com a regra de decisão adotada temos que $\mathbb{P}[\text{erro tipo I}] > \mathbb{P}[\text{erro tipo II}]$. Ou seja, de certa forma estamos privilegiando a decisão de dizer que os parafusos são americanos.

Note que mudando a regra de decisão mudamos os valores de α e β. Note que se a **regra de decisão:** "Se $\bar{R} \leq r_c$ decido que a procedência é americana" e "se $\bar{R} > r_c$ decido que a procedência é japonesa".
Se $r_c < 150$, aumentamos α e diminuímos β;
Se $r_c > 150$, diminuímos α e aumentamos β.
Daí, por exemplo, podemos fazer com que $\alpha = \beta$ com $r_c = 148.75$ e com isto, $\alpha = \beta = 5.94\%$. Isto é, escolhendo r_c podemos achar α e β. Por outro lado, podemos fixar α e encontrar r_c, tal que $\mathbb{P}[\text{erro tipo I}] = \alpha$. Por exemplo, se $\alpha = 0.05$, fazemos

$$\begin{aligned} 0.05 &= \mathbb{P}[\text{erro tipo I}] \\ &= \mathbb{P}[\bar{R} \leq r_c | \bar{R} \sim N(155, 16)] \\ &= \mathbb{P}[Z \leq -1.64] \end{aligned}$$

$$-1.64 = \frac{r_c - 155}{4} \Rightarrow r_c = 148.4$$

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,

$$\begin{aligned} \mathbb{P}[\text{erro tipo II}] &= \mathbb{P}[\bar{R} > 148.4 | \bar{R} \sim N(145, 5.76)] \\ &= \mathbb{P}[Z > 1.425] = 7.93\%. \end{aligned}$$

4.2 Introdução

Uma hipótese estatística é uma afirmação sobre o modelo probabilístico que rege o fenômeno que estamos interessados. Esta hipótese pode ser precisa e específica, afirmando que o estado da natureza é descrito completamente através de um modelo de probabilidade; isto é uma hipótese *simples*. Ou ela pode ser mais geral, afirmando que o estado da natureza tem certa propriedade ou satisfaz certa condição que é verdadeira para mais de um modelo de probabilidade; isto é uma hipótese *composta*.

Exemplo: Suponha que estamos interessados em estudar o tempo de vida de certos componentes eletrônicos. Sabemos através de considerações teóricas que o tempo de vidas desses componentes são v.a.'s exponencialmente distribuídas com parâmetro θ. Se queremos testar a afirmação que $\theta = 0.001$ então temos uma hipótese simples pois com isso determinamos completamente o modelo de probabilidade que rege o tempo de vida desses componentes. Por outro lado, se queremos saber somente se a taxa de falha θ é menor que 0.001, temos uma hipótese composta.

Exemplo: No exemplo dos parafusos americanos e japoneses, tanto H_0: os parafusos são japoneses, i.e., $R \sim N(155, 400)$, quanto H_1: os parafusos são americanos, i.e., $R \sim N(145, 144)$, são hipóteses simples.

"Testar" uma hipótese é retirar uma amostra aleatória do fenômeno de interesse e com base nesta amostra aleatória decidir se a hipótese deve ser "rejeitada" ou "não rejeitada". Note que nunca podemos aceitar uma hipótese pois os dados nunca podem provar que uma hipótese esteja certa, o que podemos dizer é que os dados não carregam evidência para se rejeitar a hipótese.

Exemplo: Suponha que estamos interessados em verificar a eficácia de um novo medicamento para hipertensão. Os dados não podem provar que o novo medicamento é melhor que o antigo, entretanto, a amostra pode nos dar evidência a favor do novo medicamento.

Definição: Um *teste de hipóteses* é uma regra que ou procedimento que, com base nos dados amostrais, nos diz se devemos rejeitar ou não rejeitar uma hipótese estatística.

A primeira etapa da construção de um teste de hipóteses é a determinação da hipótese a ser testada. Em geral, temos duas hipóteses H_0 e H_1, isto é assumimos que quando H_0 for falsa H_1 será verdadeira. Dizemos que H_0 é a *hipótese nula* e H_1 é a *hipótese alternativa.*

Exemplo: Quando testamos a eficácia de uma nova pasta de dentes na prevenção de cáries, poderíamos selecionar dois grupos de crianças, um dos quais usaria a pasta de dentes nova e o segundo grupo ("controle") usaria a pasta padrão. A hipótese nula seria H_0: não há diferença entre as pastas de dentes e a hipótese alternativa seria H_1: há diferença entre as pastas de dentes.

Exemplo: Uma expedição arqueológica encontrou 13 amostras de cerâmica presumivelmente todos de um mesmo período. Sabe-se que naquela região viviam somente duas populações A ou B, para as quais se conhece período vivido. Deseja-se testar a hipótese se as cerâmicas pertencem à população A ou B. Assim temos,

$$H_0 : \mu = \mu_A \qquad H_1 : \mu = \mu_B$$

Aqui se assumimos que a idade dos fragmentos é uma v.a. com distribuição normal com varância conhecida ambas hipóteses são simples.

Questão: Como selecionar H_0 e H_1?

Como nosso teste é baseado no resultado aleatório de um experimento, podemos sempre tomar a decisão errada. Claro que existem testes bons e testes ruins; mas mesmo com um teste bom podemos ter a rejeição de H_0 quando esta é verdadeira, ou a não rejeição de H_0 quando esta é falsa. Estes erros podem, então, ser de dois tipos:
Erro tipo I: Rejeitar H_0 quando H_0 é verdadeira;
Erro tipo II: Não rejeitar H_0 quando H_0 é falsa.

Um teste é julgado em termos de suas probabilidades de cada tipo de erro.

Voltando à questão de como selecionar H_0 e H_1.
(a) Quando nosso objetivo é estabelecer uma asserção com suporte substantivo obtido da amostra, a negação da asserção é H_0, a asserção é H_1.
(b) Escolhemos H_0 de modo que o erro tipo I seja o mais grave.

Exemplo: 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 12 meses. O laboratório que produz NEW diz que seu medicamente é melhor e um teste foi feito com 12 pacientes. Dizemos que:
H_0: NEW não é melhor que AZT, i.e., $\mu_{NEW} \leq \mu_{AZT}$;
H_0: 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. O outro erro é 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.

Exemplo: Julgamento Um juri deve decidir com base em evidências se o acusado é culpado ou inocente. A grande máxima da justiça é que o réu é considerado "inocente" a menos que haja evidência forte de culpa.

Julgamento: Precisamos forte evidência de *culpa*. Portanto, H_0: inocente e H_1: culpado. Atitude: O acusado é "inocente" até que se prove o contrário.

Hipótese estatística: Precisamos forte evidência para a *conjectura*. Portanto, H_0: conjectura é falsa e H_1: conjectura é verdadeira. Atitude: Não rejeitamos H_0 a menos que os dados amostrais tenham evidência suficiente.

É preferível não condenar um culpado (não rejeitar H_0 quando H_1 é verdadeira) do que condenar um inocente (rejeitar H_0 quando H_0 é verdadeira).

Hipótese nula simples vs. hipótese alternativa simples

O problema mais simples de analizar é aquele em que ambas as hipóteses nula e alternativa são simples. Nestes casos, é uma questão de escolher entre dois modelos probabilísticos - dois possíveis mmodelos para explicação dos dados. Se $f(\mathbf{x})$ denota a função densidade (ou função de probabilidade) da amostra $\mathbf{X}$, o problema pode ser expresso como testando:

$$H_0 : f(\mathbf{x}) = f_0(\mathbf{x}) \text{ versus} H_1 : f(\mathbf{x}) = f_1(\mathbf{x})$$

onde f_0 e f_1 são funções de densidade específicas (não dependem de nenhum parâmetro).

Definição:

Erro tipo I: Rejeitamos H_0 quando ela é verdadeira;

Erro tipo II: Não rejeitamos H_0 quando ela é falsa.

Os **tamanhos dos erros** são definidos como:

$$\begin{aligned} \alpha &= \text{“tamanho do erro tipo I”} = \mathbb{P}_{H_0}(\text{ rejeitar } H_0) \\ \beta &= \text{“tamanho do erro tipo II”} = \mathbb{P}_{H_1}(\text{ não rejeitar } H_0). \end{aligned}$$

O erro tipo I pode ter probabilidade 0, escolhendo a região crítica vazia, isto é, adotando a regra de nunca rejeitar H_0 não importando o que os dados podem dizer. Entretanto neste caso, $\beta = 1$. Vemos que, as probabilidades de erro estão negativamente correlacionadas, se diminuimos α crescemos β e vice-versa. Já estudamos cuidadosamente este tipo de teste no exemplo dos parafusos.

Em vista da perversidade inerente nas probabilidades de erro - que reduzindo um deles aumentamos o outro - não é claro como se escolher um teste. Um grande número de métodos podem, e são, utilizados.

Um método muito utilizado é baseado no fato que escolhemos H_0 de modo a ter o erro tipo I como mais grave. Um nível aceitável α é prescrito e o teste é escolhido a fim de minimizar β sem deixar a probabilidade de erro tipo I ultrapassar o valor préestabelecido α. Mesmo este método é subjetivo, pois o investigador tem que escolher o nível α. Ele tem que de alguma forma determinar se, digamos 5%, é uma escolha razoável para a probabilidade de rejeitar H_0 quando ela é verdadeira.

Deve-se notar que embora não se pode minimizar ambos as probabilidades de erro quando o tamanho da amostra é fixado, isto pode ser feito aumentando-se o tamanho da amostra.

Definição: Uma *hipótese estatística* é uma afirmação ou conjectura sobre o modelo probabilístico que rege o fenômeno que estamos interessados em estudar. Isto é, hipóteses sobre a distribuição das v.a.'s que compõem a nossa amostra aleatória. Se a hipótese estatística especifica completamente a distribuição, então ela é chamada *simples*; caso contrário ela é dita ser *composta*.

Definição: Um *teste de hipóteses* é uma regra de decisão.

Um teste é baseado em uma certa estatística, $T = t(X_1, \ldots, X_n)$, chamada *estatística do teste*. O conjunto de valores de T que, de acordo com um dado teste, fará com que rejeitemos a hipótese sendo testada é chamada *região crítica* do teste. A região crítica define o teste e vice-versa.

Exemplo: No exemplo dos parafusos a estatística do teste é $\bar{R} = (1/n)\sum_{i=1}^{n} R_i$ e a região crítica $\Gamma = \{y \in [0, \infty); y \leq 150\}$.

Exemplo: Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $\exp(\theta)$.
$H_0 : \theta = 0.001$ é hipótese simples;
$H_1 : \theta > 0.001$ é hipótese composta.

Um teste possível é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeitar } H_0 & \text{se } \bar{X} < 1100 \\ \text{Não rejeitar } H_0 & \text{se } \bar{X} \leq 1100 \end{cases}$$

Definição: Seja δ um teste para a hipótese estatística H_0 definido por: Rejeitar H_0 se, e somente se, $(X_1, \ldots, X_n) \in \Gamma_\delta$; então Γ_δ é chamada *região crítica* do teste δ.

Função Poder: Seja um teste δ para testar H_0, hipótese sobre um parâmetro θ. Definimos

$$\begin{aligned}\beta(\theta,\delta) &= \mathbb{P}(\text{rejeitar } H_0|\theta \text{ é o verdadeiro valor do parâmetro}) \\ &= \mathbb{P}_\theta(\text{rejeitar } H_0) \\ &= \mathbb{P}_\theta((X_1,\ldots,X_n)\in\Gamma_\delta)\end{aligned}$$

A função poder, no contexto de testes de hipóteses, tem o mesmo papel que o EQM no contexto de estimação. Em geral, utilizamos o poder para comparar entre os testes.

Exemplo: Seja $X_1,\ldots,X_{15}$ uma amostra aleatória de uma distribuição $\exp(\theta)$. Queremos testar:
$H_0 : \theta = 0.001$ versus $H_1 : \theta > 0.001$.

Um teste possível é:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeita } H_0 & \text{se } \bar{X} < 1100 \\ \text{Não rejeita } H_0 & \text{se } \bar{X} \le 1100 \end{cases}$$

O poder do teste é:

$$\begin{aligned}\beta(\theta,\delta) &= \mathbb{P}_\theta(\text{rejeitar } H_0) \\ &= \mathbb{P}_\theta(\bar{X} < 1100) \\ &= \mathbb{P}_\theta(\sum_{i=1}^{15} X_i < 15\times 1100)\end{aligned}$$

o que pode ser obtido de uma tabela gama (ver MINITAB).

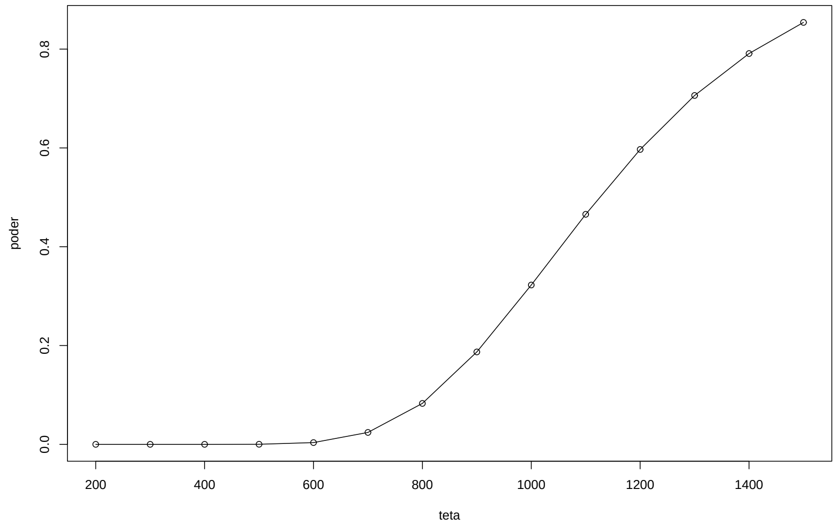

Figura 4.2.1: Função poder

θ	200	300	400	500	600
$\beta(\theta, \delta)$	0.000e+00	4.598e-11	8.581e-07	1.631e-04	3.549e-03
θ	700	800	900	1000	1100
$\beta(\theta, \delta)$	2.409e-02	8.283e-02	5.969e-01	7.061e-01	7.910e-01
θ	1200	1300	1400	1500	
$\beta(\theta, \delta)$	8.540e-01	7.061e-01	3.225e-01	4.656e-01	

Outra estatística do teste que pode ser utilizada é $X_{(1)}$ o mínimo da amostra. (Fazer como exercício).

Exemplo: 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:

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

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

Tabela 4.2.1: Função poder

Portanto,

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

A função poder é útil para nos informar quão bom é um teste em particular. Neste caso, se θ é maior que 20 é quase certo que rejeitamos H_0. Se θ é menor que 16 é quase certo não rejeitarmos H_0. Mas se $17 < \theta < 18$ (assim H_0 é falsa) o teste δ tem menos de 50% de chance de rejeitar H_0.

Definição: 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 difinido como:

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

Muitas vezes $t(\delta)$ é chamado o *nível de significância* do teste δ.

Exemplo: No exemplo anterior:

$$\begin{aligned}t(\delta) &= \sup_{\theta\leq 17} \beta(\theta,\delta)\\ &= \sup_{\theta\leq 17}(1-\Phi(18-\theta))\\ &= 1-\Phi(18-17)\\ &= 1-\Phi(1) \approx .159\end{aligned}$$

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

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.

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

Função poder:

$$
\begin{aligned}
\beta(\theta, \delta) &= \mathbb{P}_\theta(\text{rejeitar } H_0) \\
&= \mathbb{P}_\theta(\sum_{i=1}^{10} X_i \geq 6) \\
&= \mathbb{P}_\theta(Y = 6) + \ldots + \mathbb{P}_\theta(Y = 10), Y \sim b(10, \theta)
\end{aligned}
$$

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, $t(\delta) = .377$.

Definição: Após a coleta dos dados, o menor nível de significância para o qual rejeitamos H_0 é chamado o *p-valor*.

4.3 Relação entre testes 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:

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

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

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

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

4.4 Exemplos utilizando a distribuição normal

4.4.1 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:

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

Uma estatística que altamente correlacionada à hipótese a ser testada é $T(\mathcal{X}) = \bar{X}$. Neste caso, se a estatística do teste é grande temos evidências a favor de H_1. Um teste a ser considerado seria:

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

onde k^* é escolhido de forma a que o teste tenha nível de significância α. Neste caso, é fácil ver que

$$\mathbb{P}_\mu(\bar{X} \geq k^*) \ = \ \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{k^* - \mu}{\sigma/\sqrt{n}}\right)$$

é uma função crescente de μ e portanto atinge seu supremo em μ_0 quando restrita a H_0 e neste caso se equacionamos

$$\mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{k^* - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

temos que

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

e o teste 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 encontrar uma estatística que se comporta diferentemente sob as duas hipóteses e basear nosso teste nela. Tal estatística (analogamente ao caso anterior) 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

$$\begin{aligned} \alpha &= \sup_{\mu \in H_0} \mathbb{P}_\mu[T \geq k] \\ &= \mathbb{P}_{\mu_0}[T \geq k] \\ &= 1 - \mathbb{P}_{\mu_0}[T < k] \end{aligned}$$

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 4.4.1. *A Secretaria de Saúde de Campinas deseja determinar se a contagem média de bactérias por unidade de volume de água na lagoa do Taquaral está abaixo do limite de segurança de 200. Um pesquisador coletou 50 amostras de 10 locações da lagoa e encontrou os valores médios como sendo: 175, 190, 215, 198, 184, 207, 210, 193, 196 e 180 bactérias por litro. Você acha que com base nestes dados deve-se acreditar que o valor médio na lagoa está dentro dos limites de segurança?*

Considere μ como sendo a contagem média de bactérias por litro na Lagoa do Taquaral. Sendo assim, dizer que estamos fora dos limites de sugurança significa dizer $\mu \geq 200$. Considerando o problema de saúde pública, o erro mais grave é dizer que a água está dentro dos limites quando na realidade não está. Portanto, queremos testar:

$$H_0 : \mu \geq 200 \qquad \text{vs.} \qquad H_1 : \mu < 200.$$

Como cada uma das 10 observações que temos é uma média de 50 amostras de água, é razoável supor que a distribuição normal é apropriada. Sendo assim, o teste de hipóteses com nível de significância de 1% é dado por: rejeitamos H_0 quando

$$\frac{\bar{X} - 200}{S/\sqrt{10}} < t_{9;0.01} = -2.821.$$

Neste caso, $\bar{x} = 194,8$, $s = 13,14$ e temos $T(x) = -1,25$. Sendo assim, a hipótese nula não é rejeitada a um nível de 1%. A Secretaria de Saúde deveria se preocupar com a água da Lagoa.

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 $\alpha = 1 - \gamma$.

Exemplo 4.4.2. *Um experimentador deseja saber se o peso médio de tomates orgânicos pode ser considerado como 20 gramas. Ele tira uma amostra de 18 tomates e os pesa. Assumindo que o peso de tomates seja uma v.a.a normalmente distribída, ele contrói o IC de nível 95% e obtém*

$$\bar{x} \pm 2.11s/\sqrt{18} = [17,1; 29,3]$$

. Sendo assim, se ele deseja testar $H_0 : \mu = 20$ versus $H_1 : \mu \neq 20$, ele não deve rejeitar a hipótese nula a um nível de significância de 5%.

Exercício: 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 $\mu > \mu_2$?

4.4.2 Testes a respeito da 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:

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

e a hipótese alternativa é unilateral. Neste caso sabemos que a estatística $T(\mathcal{X}) = \sum_{i=1}^n (X_i - \mu)^2$ nos dá informação sobre as hipóteses. Neste caso, se $T(\mathcal{X})$ é grande temos mais evidência a favor de H_1 e um teste a ser proposto é:

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

onde k é tal que

$$\sup_{\sigma \in H_0} \mathbb{P}_{\sigma_0^2}[\sum_{i=1}^n (X_i - \mu)^2 \geq k] = \mathbb{P}_{\sigma_0^2}[\sum_{i=1}^n (X_i - \mu)^2 \geq k] = \alpha,$$

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

(2) 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:

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

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

(2) Para μ desconhecido, 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:

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

4.4.3 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 isto 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. Este problema é, em geral, referido como um problema de *análise de variância de um fator.*

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. Em geral, milhos híbridos produzem mais do que variedades puras de milho, a Embrapa está estudando uma nova combinação de variedades de milho e gostaria de saber se esta variedade híbrida produz mais do que a variedade mãe. Para isto plantamos alguns hectares com a nova variedade e alguns hectares com a variedade mãe e com base na produção média destas plantações desejamos comparar os dois tipos de milho.

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.

Exemplo 4.4.3. *Um pesquisador da área médica deseja determinar se a pílula anticoncepcional tem o efeito colateral de reduzir a pressão arterial de suas usuárias. O estudo se conduz medindo a pressão arterial de 15 usuárias de pílula com idades entre 25 e 27 anos. Após o uso regular do medicamento por 6 meses, suas pressões arteriais são novamente medidas e registradas.*

Indivíduo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Antes (x)	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
Depois (y)	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74
d = (x - y)	2	8	10	6	18	10	4	26	18	-8	0	32	0	-4	10

Assim, se X = pressão antes de tomar o medicamento e Y = pressão pós tomar o medicamento, temos a v.a. $D = X - Y$ com suas estatísticas:

$$\bar{d} = 8,80 \quad \text{e} \quad s_D = 10,98.$$

Portanto se assumimos que tanto X como Y são normalmente distribuídas temos

$$\frac{\bar{D} - (\mu_1 - \mu_2)}{S_D/\sqrt{15}} \sim N(0,1).$$

Para testarmos $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 > \mu_2$ podemos usar como estatística do teste:

$$T = \frac{\bar{D}}{S_D/\sqrt{15}}$$

a qual sob H_0 tem distribuição normal padrão. Sendo assim, nosso teste de hipóteses é dado por:

$$\delta(\mathbf{X}) = \begin{cases} \text{Rejeitar } H_0 & \text{se } \frac{\bar{D}}{S_D/\sqrt{15}} > t_{14,0.05} \\ \text{Não rejeitar } H_0 & \text{caso contrário} \end{cases}$$

Neste caso, $t_{14,0.05} = 1.761$ e

$$t = \frac{\bar{d}}{s_D/\sqrt{15}} = 3.10 > 1.761.$$

Portanto, rejeitamos a hipótese nula a um nível de significância de 5%.

p-valor (Menor nível de significância tal que a hipótese nula é rejeitada. Neste caso, p-valor $= \mathbb{P}(T > 3.10) = 0.0003317958$. Sendo assim, temos fortes evidências a favor da hipótese alternativa. Para não rejeitar a hipótese nula teríamos que ter um nível de significância menor que 0.00033.

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. Para todos os testes que propusermos neste caso, a probabilidade de erro tipo I dependerá dos valores desconhecidos de σ_1 e σ_2 e isso torna impossível calcular exatamente o teste. Veremos depois que podemos achar testes assintóticos neste caso e assim resolver pelo menos parcilamente este problema.

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

$$T = \frac{\sqrt{n_1 n_2/(n_1+n_2)}(\bar{X} - \bar{Y} - \mu_1 - \mu_2)}{\sqrt{[\sum(X_i - \bar{X})^2 + \sum(Y_i - \bar{Y}^2]/(n_1+n_2-2)}}$$

a qual tem uma distribuição $t_{n_1+n_2-2}$ e assim o IC de nível $\gamma = 1-\alpha$ é dado por $IC = [T_1; T_2]$ onde:

$$T_1 = \sqrt{\frac{n_1 n_2}{(n_1+n_2)}}\left(\bar{X} - \bar{Y} - t_{\alpha/2;n_1+n_2-2}\sqrt{\frac{\sum(X_i-\bar{X})^2 + \sum(Y_i - \bar{Y}^2)}{(n_1+n_2-2)}}\right)$$

e

$$T_2\sqrt{\frac{n_1 n_2}{(n_1+n_2)}}\left(\bar{X} - \bar{Y} + t_{\alpha/2;n_1+n_2-2}\sqrt{\frac{\sum(X_i-\bar{X})^2 + \sum(Y_i - \bar{Y}^2)}{(n_1+n_2-2)}}\right)$$

e o teste de tamanho α é:

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

Exemplo 4.4.4. *Um teste de alimentação foi realizado com 25 vacas a fim de se comparar dois tipos de ração na produção de leite. Das 25 vacas, foram selecionadas aleatoriamente 12 vacas para receber a ração A e as 13 restantes receberam a ração B. As vacas receberam a alimentação durante 3 semanas (21 dias) e foi registrado a produção média obtida por cada vaca (em litros).*

	Produção de leite (em litros)
Ração A	35, 47, 55, 29, 40, 39, 32, 41, 42, 57, 51, 39
Ração B	44, 44, 56, 46, 47, 38, 58, 53, 49, 35, 46, 30, 41

Assumindo que as variâncias são iguais, temos que a estatística do teste é dada por:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}S}$$

onde

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Neste caso, $\bar{x} = 42,25$, $\sum(x - \bar{x})^2 = 840,25$, $s_1^2 = 76,39$, $\bar{y} = 45,15$, $\sum(y - \bar{y})^2 = 767,69$ e $s_Y^2 = 63,97$. Assim,

$$s^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2} = 69,9.$$

Como o valor tabelado é: $t_{23,0.025} = -2.069$ e o valor observado da estatística é:

$$t = \frac{45,15 - 42,25}{8.63\sqrt{\frac{1}{13} + \frac{1}{12}}} = 0,84,$$

não rejeitamos H_0 a um nível de significância de 5%. De fato, temos **p-valor** $\mathbb{P}(|T| > 0,84) = 0,41$. Isto é, deveríamos crescer nosso nível de significância para 0,41 se queremos rejeitar H_0. Isto mostra que os dados trazem muito pouca evidência a favor de H_1.

Teste de igualdade de 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. Se não pudermos garantir a igualdade das variâncias, estas precisam ser testadas antes das médias.

Considere as hipóteses $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$. Neste caso, é razoável utilizar como estatística do teste a razão entre as variâncias amostrais. Sabemos que:

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$$

e

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

são independentes e sob H_0

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1).$$

Também sabemos que quanto mais distante de 1 estiver F mais evidência a favor da hipótese alternativa. Neste caso, nosso teste de hipóteses deveria ser:

$$\delta(\mathbf{X}, \mathbf{Y}) = \begin{cases} \text{Rejeito } H_0, & \text{se } F \leq c_1 \text{ ou } F \geq c_2 \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

onde $c_1 = F_{n_1-1,n_2-1,\alpha/2}$ e $c_2 = F_{n_1-1,n_2-1,1-\alpha/2}$. Se fixarmos $\alpha = 0,05$ temos $c_1 = 0,301$ e $c_2 = 3,43$. Portanto, não rejeitamos H_0 a um nível de significância de 5%. De fato, p-valor= 0,76.

4.5 Exemplos utilizando a distribuição binomial

4.5.1 Pequenas amostras

Exemplo 4.5.1. *Durante muitos anos um medicamento padrão para a cura de alergia tem sido usado e a taxa de sucesso é de 40%. Um novo medicamento vai ser introduzido no mercado e se antecipa que sua taxa de cura será melhor que o medicamento padrão. Suponha que o novo medicamento será testado em 20 pacientes e que será registrado a v.a. X número de pacientes curados.*

Como podemos utilizar os dados coletados no experimento para responder a questão: "*Há evidências substanciais de que o novo medicamento tem uma taxa de cura maior que o medicamento padrão.*

A taxa de cura do novo medicamento p é um valor que não pode ser exatamente encontrado, exceto após um longo tempo de aplicação em um grande número de pacientes. Sendo assim, temos que usar os dados disponíveis com a aplicação do medicamento aos 20 pacientes para testar as seguintes hipóteses:

- O novo medicamento é mais eficiente que o padrão: $p > .4$;
- O novo medicamento não é mais eficiente que o padrão: $p \leq .4$

Como é mais grave colocar no mercado um medicamento sem que ele seja eficiente, temos que o erro tipo I deveria ser "dizer que o nove medicamento é mais eficiente que o padrão quando na realida não é". Portante temos:

- H_0: O novo medicamento não é mais eficiente que o padrão: $p \leq .4$;
- H_1: O novo medicamento é mais eficiente que o padrão: $p > .4$;

Estatística do teste: X = número de pacientes curados dentre os 20 pacientes testados.

Se o experimento foi desenhado adequadamente temos $X \sim bin(20, p)$.

Portanto, quanto **maior** X mais evidência a favor de H_1 e podemos utilizar como regra de decisão:

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

Para encontrar o valor de c temos que fixar a probabilidade de erro tipo I. Note que:

$\mathbb{P}$(erro tipo I)	c=11	c=12	c=13
$\mathbb{P}(X > c)$	.128	.057	.0.021

Portanto, se quisermos construir um teste que tenha nível de significância de 10% temos que utilizar:

$$\delta(X) = \begin{cases} \text{Rejeito } H_0, & \text{se } X \geq 12 \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

O tamanho do teste é: .057.

Por outro lado, se queremos que nosso nível de significância seja 5% temos que utilizar:

$$\delta(X) = \begin{cases} \text{Rejeito } H_0, & \text{se } X \geq 13 \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

O tamanho do teste será: .021.

Não é possível construir um teste com tamanho exatamente 5% ou 10%.

Utilizando o teste que rejeita H_0 se $X \geq 12$ podemos construir a função poder:

$$\beta(\theta, \delta) = \mathbb{P}_\theta(\text{rejeitar } H_0) = \mathbb{P}_\theta(X \geq 12).$$

Neste caso,

θ	$\beta(\theta, \delta)$
0.1	3.92e-09
0.2	1.51e-05
0.3	1.27e-03
0.4	2.10e-02
0.5	1.31e-01
0.6	4.15e-01
0.7	7.72e-01
0.8	9.67e-01
0.9	9.99e-01

Suponha agora que observamos os pacientes e obtivemos $X = 14$, neste caso rejeitamos a hipótese nula a um nível de 5%, a um nível de 2%. Qual o menor nível de significância que rejeitamos H_0?

p-valor $= \mathbb{P}_{.4}(X \geq 15) = 0.00161.$

4.5.2 Grandes amostras

Seja $X_1, \ldots, X_n$ i.i.d. $b(1, p)$ onde n é grande. Neste caso, $X = \sum_{i=1}^n X_i \sim bin(n, p)$. Queremos utilizar a amostra para testar $H_0 : p = p_0$ versus $H_1 : p \neq p_0$.

Estatística do teste:

$$\hat{P} = \frac{X}{n} = \bar{X}.$$

Utilizando o Teorema Central do Limite sabemos que, sob a hipótese nula, a distribuição assintótica da estatística:

$$T(X) = \frac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

é $N(0,1)$.

Mais ainda, sabemos que quanto **maior** $|T(X)|$, mais evidência a favor da hipótese alternativa e um teste poderia ser da forma:

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

onde a constante c é escolhida de forma que o teste tenha nível de significância (probabilidade de erro tipo I) fixado α.

Neste caso, $c = z_{\alpha/2}$.

Exemplo 4.5.2. *O censo de 1990 registrou que no distrito de Barão Geraldo 20% das pessoas viviam abaixo do nível da pobreza. Para determinar se esta porcentagem mudou ao longo dos últimos anos, uma amostra aleatória de 400 famílias foi retirada e encontrou-se que 100 delas viviam abaixo do nível de pobreza.*

Seja p o valor atual da proporção de famílias que vivem abaixo do nível de pobreza em Barão Geraldo. Desejamos procurar evidência de que p seja maior que 20%. Portanto, queremos testar:

$$H_0 : p = 0.20 \qquad H_1 : p > .20.$$

Como o tamanho da amostra $n = 400$ é grande vamos usar como estatística do teste:

$$Z = \frac{\hat{P} - 0.20}{\sqrt{.2 \times .8/400}}.$$

Fazendo $\alpha = 0.05$ temos o teste

$$\delta(X) = \begin{cases} \text{Rejeito } H_0, & \text{se } Z > 1.64 \\ \text{Não rejeito } H_0, & \text{caso contrário} \end{cases}$$

O valor de Z calculado na amostra é:

$$z = \frac{(100/400) - 0.20}{\sqrt{.2 \times .8/400}} = 2.5.$$

Como $2,5 > 1.64$ rejeitamos H_0 a um nível de significância de 5%. Além disso, o p-valor é: $\mathbb{P}(Z > 2.5) = 0.00621$.