Aula 2

Teste binomial pequenas amostras

O teste binomial é aplicado a populações cujos elementos possuem dois atributos de interesse (dados dicotômicos), como por exemplo: masculino e feminino; fértil e não fértil; casado ou não casado; peça defeituosa ou não defeituosa, etc.... O teste é usado para testar a proporção p de um dos atributos da população em estudo.

Estatística de teste

Visto que o interesse é testar o parâmetro p (probabilidade de ocorrência do atributo A), definiremos a estatística de teste como sendo o número T de vezes que o atributo A ocorreu nas n realizações da amostra, ou seja,

$$T = \sum_{i=1}^{n} X_i$$

que tem distribuição Binomial de parâmetros n e p.

Indicando por T^* o valor observado da variável aleatória T, a hipótese H_0 será rejeitada se $(T^* \leq t_1)$ ou $(T^* \geq t_2)$.

Teste Bilateral $P(T \leq t_1) = \alpha_1$ e $P(T \geq t_2) = \alpha_2$, $\alpha_1 + \alpha_2 = \alpha$

Teste Binomial

- *Null hypothesis*:

$$H_0 : \text{The machine is operating properly}$$

- *Alternative hypothesis*:

$$H_1 : \text{The machine needs attention}$$

- $H_0 : p \leq .05$ and $H_1 : p > .05$
- T: Total number of defective items.
- $T \sim \text{Binomial}(10, p)$
- Under H_0, $P(T \leq 2) \geq 0.9885$
- Critical region: $T > 2$ Ou seja, T>=3
- Suppose a random sample consisting of 10 machined parts is observed and 4 of the parts are found to be defective.
- Then $T = 4$ and the null hypothesis is rejected. We conclude that the machine needs attention. $\square$

Como resolve?

- Para o teste unilateral a direita obter t1 tal que
- P(T>=t1) seja aproximadamente igual ao nível de significância.
- Para T~bin(n=10 e p=0.05), y=P(T>2)=P(T>=3), logo t1=3 para nível de significância entre 0.012 e 0.085

y=P(T>x)

```
1-pbinom(0:10,10,0.05)
```

x	y
0	0.401
1	0.086
2	0.012
3	0.001
4	0.000
5	0.000
6	0.000
7	0.000
8	0.000
9	0.000
10	0.000

Poder do teste

- P(rejeitar H0|H0 é falsa), variamos o valor do parâmetro e observamos a probabilidade de rejeitar H0.
- $P(\text{reject } H_0) = \sum_{i=3}^{10} \binom{10}{i} p^i (1-p)^{10-i} = 1 - \sum_{i=0}^{2} \binom{10}{i} p^i (1-p)^{10-i}$

```
1-pbinom(2,10,seq(0,1,0.05))
```

p	$P(\text{reject } H_0)$	p	$P(\text{reject } H_0)$
0	0.0000	0.50	0.9453
0.05	0.0115	0.55	0.9726
0.10	0.0702	0.60	0.9877
0.15	0.1798	0.65	0.9952
0.20	0.3222	0.70	0.9984
0.25	0.4744	0.75	0.9996
0.30	0.6172	0.80	0.9999
0.35	0.7384	0.85	1.0000
0.40	0.8327	0.90	1.0000
0.45	0.9004	1.00	1.0000

Exercício

Obtenha o poder do teste binomial para

- H0: p <= 0.5 e n=5, alpha = 0.05

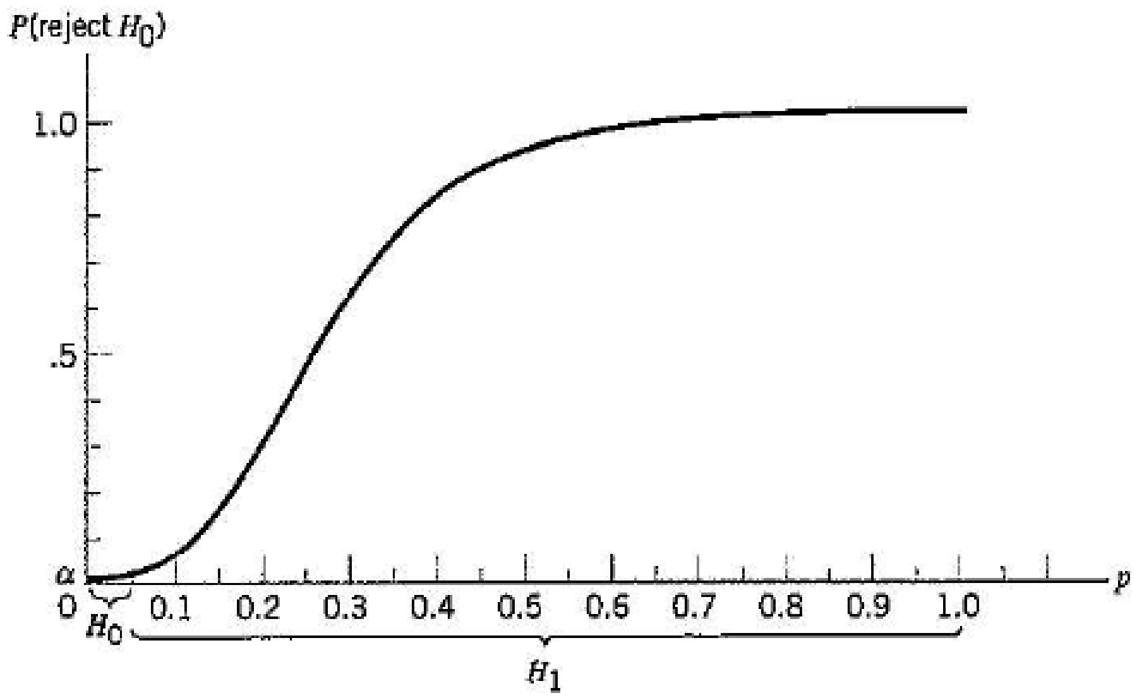
```
1-pbinom(0,5,seq(0,1,0.05))
```

Avalie o poder do teste variando n=5, 10, 50 e 100. Plote num gráfico, eixo x os valores de p e eixo y as probabilidades obtidas.

```
1-pbinom(0,n,seq(0,1,0.05))
```

Unbiased Test

- Um teste não tendencioso é aquele em que o poder é sempre pelo menos tão grande quanto o nível de significância



Exercício

- Produza com auxilio do R o gráfico do slide anterior. De acordo com o gráfico, a partir de que proporção o teste é não tendencioso?

```
x=seq(0,1,0.05)
y=round(1-pbinom(2,10,x),4)
plot(x,y, type="b")
```

A partir de p=0.09 o poder do teste é superior a 0.05 (poder = 0.054).

Exercício

Sabe-se que 20% de uma certa espécime de inseto exibem uma particular característica A. Dezoito insetos desta espécie foram estudados em uma usual experiência e nenhum deles apresentou a característica A . É razoável supor que os insetos observados têm origem na população de insetos com 20% de incidência da característica A?

Trata-se de um teste bilateral com hipóteses:

$H0 : p = 0.20$

$H1 : p \neq 0.20$

Sob a hipótese H0 , a variável aleatória T tem distribuição Binomial (18;0.20).

Conclua o teste com nível de significância de 3%.

Resposta

- Teremos t1=0 e t2=8 o que gera $\alpha 1=0.018$ e $\alpha 2=0.0162$, logo $\alpha=0.0342$.
- Não aceita-se a hipótese nula ao nível de significância de 0.03
- Não há evidência de que os insetos da amostra provenham da população especificada no enunciado.
- O p-valor é obtido por $P(T<=0)+P(T>=8)=0.018+0.016=0.034$

x	y	yacum
0	0.0180	0.0180
1	0.0811	0.0991
2	0.1723	0.2713
3	0.2297	0.5010
4	0.2153	0.7164
5	0.1507	0.8671
6	0.0816	0.9487
7	0.0350	0.9837
8	0.0120	0.9957
9	0.0033	0.9991
10	0.0008	0.9998
11	0.0001	1.0000
12	0.0000	1.0000
13	0.0000	1.0000
14	0.0000	1.0000
15	0.0000	1.0000
16	0.0000	1.0000
17	0.0000	1.0000
18	0.0000	1.0000

Teste Binomial (grandes amostras)

$$T = \sum_{i=1}^{n} X_i$$

- Se T tem distribuição Binomial de parâmetros n e p, e se n é suficientemente grande, usaremos a aproximação normal para a binomial, isto é, assumiremos que T é assintoticamente normal com média np e variância np(1-p).

Tamanho da amostra >=20

$$H_0: \; p = p^* \qquad H_1: \; p \neq p^*$$

$$P(Y \leq t_{\text{obs}}) \approx P\left(Z \leq \frac{t_{\text{obs}} - np^* + 0.5}{\sqrt{np^*(1-p^*)}}\right)$$

$$P(Y \geq t_{\text{obs}}) \approx 1 - P\left(Z \leq \frac{t_{\text{obs}} - np^* - 0.5}{\sqrt{np^*(1-p^*)}}\right)$$

$$H_0: \; p \geq p^* \qquad H_1: \; p < p^*$$

$$P(Y \leq t_{\text{obs}}) \approx P\left(Z \leq \frac{t_{\text{obs}} - np^* + 0.5}{\sqrt{np^*(1-p^*)}}\right)$$

$$H_0: \; p \leq p^* \qquad H_1: \; p > p^*$$

$$P(Y \geq t_{\text{obs}}) \approx 1 - P\left(Z \leq \frac{t_{\text{obs}} - np^* - 0.5}{\sqrt{np^*(1-p^*)}}\right)$$

Exemplo

- Sob herança mendeliana simples um cruzamento entre plantas de dois genótipos específicos pode produzir descendentes um quarto dos quais são anão e três quartos dos quais são normais. Conduza um teste para verificar se plantas seguem essa herança, considerando-se que de 925 plantas, 682 são normais.

$$H_0: \; p = 3/4 \quad \text{v.s.} \quad H_1: \; p \neq 3/4$$

$$\alpha = 0.05, \; n = 925 \text{ and } p^* = 3/4.$$

$$P(Y \leq 682) \approx P\left(Z \leq \tfrac{682-693.75+0.5}{13.17}\right) = P(Z \leq -0.8542) = 0.196$$

$$p\text{-value: } 2P(Y \leq 682) = 0.392$$

Se usarmos a distribuição binomial, o p-valor será obtido por
P(T<=682)+P(T>=705)=0.3825

n*p=693.75, logo a média está entre 693 e 694,
assim 693-682=11 e T2 = 694+11=705

Conclusão: O teste indica que as plantas seguem o padrão da herança mendeliana (p-valor =0.392)

Exercícios para avaliação

- 1)Acredita-se que 30% das pessoas tenham problemas de audição. Numa amostra de 80 indivíduos 20 apresentaram o problema. Execute um teste binomial não paramétrico. Avalie o poder do teste.
- 2)Um industrial afirma que seu processo de fabricação produz 90% de peças dentro das especificações. Deseja-se investigar se este processo de fabricação ainda está sob controle. Uma amostra de 15 peças foi analisada e foram constatadas 10 peças dentro das especificações. Ao nível de 5% de significância, podemos dizer ser verdadeira essa afirmação?

- Data de entrega a ser definida em tarefas do conexão uff.

Fim da aula 2