UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

MAT02023 - INFERÊNCIA B - 2019/2

Plano Aula 26

Markus Stein

13 November 2019

. . . continuação TRV considerações finais

Distribuições discretas

- **Exemplo 1**: (*Equilíbrio de Hardy-Weinberg*) Seja $\boldsymbol{X} = (X_1, \dots, X_n)$ uma a. a. de $X \sim Multinomial(N, \pi_1, \pi_2, \pi_3)$. Use o TRV para testar $H_0 : \pi_1 = \pi_2 = \pi_3$.
- **Exemplo 2**: *(Tabelas $r \times c$)* Suponha que temos uma tabela de contingência $r \times c$ com n indivíduos independentemente selecionados, sendo n_{ij} o número de unidades classificadas na linha i e na coluna j, para todo $i = 1, \dots, r$ e $j = 1, \dots, c$. Seja π_{ij} a probabilidade de um indivíduo ser classificado na linha i e coluna j, tal que $\pi_{ij} \geq 0$ e $\sum_{i=1}^{r} \sum_{j=1}^{c} \pi_{ij} = 1$.

a. Encontre o TRV para testar $H_0 : \pi_{ij} = a_i b_j$, para algum $a_i > 0$ e $b_j > 0$ tais que $\sum_{i=1}^{r} a_i = 1$ e $\sum_{j=1}^{c} b_j = 1$, contra a alternativa $H_1 : \pi_{ij} \neq a_i b_j$ para pelo meno
b. Compare o teste do ítem (a) com o teste qui quadrado de independência, para tesar se a variável da linha e da coluna são independentes.

Teste Exato de Fisher

- **Exemplo 3**: *(Tabela 2×2 restrita)* Seja $S_1 \sim Binomial(n_1, \pi_1)$ independente de $S_2 \sim Binomial(n_2, \pi_2)$. Para testar as hipóteses $H_0 : \pi_1 = \pi_2$ contra $H_1 : \pi_1 > \pi_2$:

a. Mostre que sob H_0 temos que $S = S_1 + S_2$ é estatística suficiente e $S_1 | S = s \sim Hipergeométrica(n_1 + n_2, n_1, s)$.
b. Calcule o valor p (condicional) para esse teste?
c. Compare com os valores p do TRV e do teste qui quadrado do exercício 5.

Testes Qui Quadrado

- Pearson e o teste *Goodness-of-fit*
- ajustamento (homogeneidade) $\times$ independência

Testes Bayesianos

Leitura: Ler seções 8.2.2 e 8.3.5 do livro Casella e Berger.

Tarefa: Fazer lista 5 para entregar.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

MAT02023 - INFERÊNCIA B - 2019/2

Teste qui quadrado

Exemplo - Bombas em Londres

- Foram observadas 535 bombas lançadas em 576 áreas. Média de $\bar{x} = 535/576 = 0.93$ bombas por àrea.
- As àreas serão classificadas pela variável X: *nº de bombas em cada àrea.*
- H_0: "as bombas foram lançadas uniformemente?"
  - Assumimos que sob H_0 $X \sim Poisson(0.93)$

Com os dados do problema, primeiramente vamos calcular o valor da estatística $X^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$ para os dados observados.

```r
O <- c(229, 211, 93, 35, 7, 1)     # num. OBS. areas com 0, 1, 2, 3, 4 e 5+ bombas
n <- sum(O)                        # num. de areas
k <- length(O)                     # num. de categorias
xbarra <- 535 / n                  # num. medio de bombas por area
E <- 576 * c(dpois(0:4, xbarra), 1-ppois(4, xbarra))  # num. ESP. areas com ... bombas
```

O valor para os dados observados, temos que X^2 é igual a

```r
## funcao para calcular a estatistica qui quadrado `X2`
X2 <- function(O) sum((O - E)^2 / E)
X2(O)
```

```
## [1] 1.17238
```

Agora geraremos amostras Monte Carlo (MC) sob H_0 e calcularemos a estatística X^2 para cada amostra.

```r
## Geracao de possiveis amostras da `Poisson(xbarra)`
nsim <- 1000                                     # num. de simulacoes
amostrasMC <- replicate(nsim, rpois(n, xbarra))  # amostras simuladas
amostrasMC[amostrasMC > 5] <- 5                  # valores > 5 virarao 5

## calculando a estatistica `X2` para todas as amostras MC geradas
# for(ii in 1:nsim) X2(table(factor(amostrasMC[,ii], 0:5)))
X2MC <- apply(amostrasMC, 2, function(xx) X2(table(factor(xx, 0:5))))
```

Com base nas 1000 amostras MC geradas agora construiremos o histograma dos X^2 gerados. A distribuição dos X^2 se assemelha a uma distribuição qui quadrado χ^2?

```r
# histograma das estatisticas `X2`
par(mar=c(5, 4, 2, 2), cex=0.6)
hist(X2MC, main="", ylab="Proporção", xlab=expression(X^2), prob=TRUE)

# densidade teórica da dist. qui quadrado com k-1 g.l.
lines(X2MC[order(X2MC)], dchisq(X2MC[order(X2MC)], df=k-1), lwd=2, col="red")
```