Aula 9 - Tópicos Especiais

PhD. Wagner Hugo Bonat

Laboratório de Estatística e Geoinformação-LEG
Universidade Federal do Paraná

1/2017

Comparação de duas médias

- Comparar duas populações com relação às suas médias.
- Ao comparar parâmetros de duas populações precisamos verificar se estão ou não relacionadas.
- Se dependentes - Caso 1 - Teste t-pareado.
- Caso não ainda temos que verificar se estas tem a variância conhecida (Caso 2).
- Se a variância é desconhecida temos que verificar se são iguais (caso 3A) ou diferentes (caso 3b).

Teste t-pareado

- Exemplo 9.1 – 9.3 - Uma distribuidora de combustíveis deseja verificar se um novo tipo de gasolina é eficaz na revitalização de motores velhos. Com esse objetivo, seleciona 12 automóveis de um mesmo modelo com mais de 8 anos de uso e, após regulagem de seus motores, verifica o consumo de combustível. Em seguida, o carro é abastecido com o novo tipo de combustível durante 15 semanas, e uma nova aferição do consumo é feita. Defina as variáveis aleatórias X_i e Y_i como o rendimento do automóvel i respectivamente antes e após as 15 semanas. Os valores observados, em km/l, junto com as diferenças $D_i = Y_i - X_i$, para os 12 automóveis são apresentados na tabela a seguir.

Laboratório de Estatística e Geoinformação

Exemplo 9.1 – 9.3

```
> Antes <- c(8.1,7.9,6.8,7.8,7.6,7.9,5.7,8.4,8.0,9.5,8.0,6.8)
> Apos <- c(11.6,8.8,9.9,9.5,11.6,9.1,10.6,10.8,13.4,10.6,10.5,11.4)
> Diferenca <- Apos - Antes
> t(data.frame(Apos, Antes, Diferenca))
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
Apos	11.6	8.8	9.9	9.5	11.6	9.1	10.6	10.8	13.4	10.6	10.5	11.4
Antes	8.1	7.9	6.8	7.8	7.6	7.9	5.7	8.4	8.0	9.5	8.0	6.8
Diferenca	3.5	0.9	3.1	1.7	4.0	1.2	4.9	2.4	5.4	1.1	2.5	4.6

- $H_0 : \mu_D = 0$ (o novo combustível não aumenta o rendimento);
- $H_1 : \mu_D > 0$ (o novo combustível aumenta o rendimento).
- Estimador da variância $S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$
- Estatística T

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

que, sob H_0 segue uma distribuição t-Student com $n - 1$ graus de liberdade.

Laboratório de Estatística e Geoinformação

Exemplo 9.1 – 9.3

- Diferença média $\bar{D}$

```
> D.bar <- mean(Diferenca)
```

- Erro padrão S_D

```
> Sd <- sqrt(sum( (Diferenca - mean(Diferenca))^2)/(12-1))
```

- Estatística T

```
> (D.bar - 0) / (Sd/sqrt(12))
[1] 6.539586
```

- Valor T-tabelado $\alpha = 0.05$ temos $T_c = 1.796$.
- IMPORTANTE: Tabela T apresenta $P(-t_c \leq t \leq t_c)$.
- $T > T_c$, concluímos que o novo combustível é eficaz na melhora do rendimento.

Laboratório de Estatística e Geoinformação

Caso 2: Amostras independentes com variâncias conhecidas

- Exemplo 9.2 – 9.4 Um estudo envolve a avaliação de um novo sistema operacional de computador, desenvolvido para crianças com idades entre 8 e 12 anos. Afirma-se que o novo sistema é mais rápido do que o atual, líder de mercado. Para testar esta afirmação, foram selecionados em uma mesma escola dois grupos com 15 crianças casa. As crianças, sem conhecimento prévio relacionado ao uso de computadores, utilizaram máquinas de mesma configuração para realizar uma certa tarefa, que teve seu tempo anotado. O primeiro grupo, denominado Grupo A, trabalhou com o sistema operacional convencional ao passo que o segundo grupo, Grupo B, desenvolvey atividades no novo sistema. As final do experimento todas as 30 crianças haviam realizado a tarefa. Os dados obtidos estão na tabela a seguir.

Exemplo 9.2 – 9.4

- Dados

```
> GA <- c(182,185,193,175,184,192,175,173,186,178,162,179,164,182,186)
> GB <- c(92,76,76,90,97,90,86,93,100,115,85,80,90,86,94)
```

- H_0 : Tempo médio é igual para ambos os sistemas ($\mu_1 = \mu_2$);
- H_1 : Aprendizado do novo sistema é, em média, mais rápido ($\mu_1 > \mu_2$).
- Defina $\bar{D} = \bar{X} - \bar{Y}$
- $H_0 : \mu_D = 0$ contra $H_1 : \mu_D > 0$.
- $Var(\bar{D}) = Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y})$.
- Supondo que a variâncias das duas populações são iguais σ_o^2 temos

$$Var(\bar{D}) = \frac{\sigma_o^2}{n_1} + \frac{\sigma_o^2}{n_2}$$

Exemplo 9.2 – 9.4

- Assuma que informações adicionais fornecidas pelas empresas indicam que a v.a são independentes com $\sigma_o = 10$min.
- Estatística de teste
$$T = \frac{\bar{D} - 0}{\sqrt{Var \bar{D}}}$$
que tem distribuição Normal padrão, sob H_0.
- Médias

```
> mean(GA)

[1] 179.7333

> mean(GB)

[1] 90

> D.bar <- mean(GA) - mean(GB)
> D.bar

[1] 89.73333

> Var.D.bar <- (10^2)/15 + (10^2)/15
> Var.D.bar

[1] 13.33333

> TT <- D.bar/sqrt(Var.D.bar)
> TT

[1] 24.57449
```

Caso 3A: Amostras independentes com variâncias desconhecidas e iguais

- Exemplo 9.7: Digitadores são treinados em uma empresa em duas turmas distintas. Na primeira, denominada J, utiliza-se um método japonês de ensino, ao passo que na segunda turma, denominada A, utiliza-se um método alemão. Deseja-se comparar os dois métodos e para tanto 16 alunos de cada turma foram escolhidos aleatoriamente e uma mesma tarefa foi atribuída a cada um. Ao final do experimento, o tempo gasto na realização da tarefa, para cada aluno foi anotado. No processo, dois computadores utilizados pelos alunos selecionados da turma J e três da turma A apresentaram problemas que impediram a realização da tarefa. Apesar de não conhecidas, as variâncias populacionais para as duas turmas são consideradas iguais. Os dados obtidos foram:

```
> JJ <- c(10,13,9,10,14,13,10,15,12,10,9,10,13,14)
> AA <- c(15,12,18,16,15,17,17,15,16,17,11,17,14)
```

Laboratório de Estatística e Geoinformação

Caso 3A: Amostras independentes com variâncias desconhecidas e iguais

- Dados

```
> # X
> JJ <- c(10,13,9,10,14,13,10,15,12,10,9,10,13,14)
> # Y
> AA <- c(15,12,18,16,15,17,17,15,16,17,11,17,14)
```

- $H_0 : \mu_X = \mu_Y$ contra $H_1 : \mu_X \neq \mu_Y$.
- Estatística de teste

$$\frac{\bar{D} - (\mu_X - \mu_Y)}{S_C \sqrt{1/n_1 - 1/n_2}}$$

  sob H_0 tem distribuição t-Student com $n_1 + n_2 - 2$ graus de liberdade.
- Estimador para variância populacional

$$S_c^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{(n_1 - 1) + (n_2 - 1)}.$$

Laboratório de Estatística e Geoinformação

Caso 3A: Amostras independentes com variâncias desconhecidas e iguais

- Calculando as estatísticas

```
> # X
> mean(JJ)

[1] 11.57143

> var(JJ)

[1] 4.263736

> # Y
> mean(AA)

[1] 15.38462

> var(AA)

[1] 4.25641

> # Diferença
> D.bar <- mean(JJ) - mean(AA)
> D.bar

[1] -3.813187

> Sc <- ((14-1)*var(JJ) + (13-1)*var(AA))/((14-1) + (13-1))
> Sc

[1] 4.26022

> # Estatistica T
> t_obs <- D.bar/(sqrt(Sc)*sqrt(1/14 + 1/13))
> t_obs

[1] -4.796516
```

Caso 3B: Amostras independentes com variâncias desconhecidas e diferentes

- Para os mesmos dados do exemplo 9.7 considere que as variâncias populacionais são diferentes.
- Estatística de teste

$$T = \frac{\bar{D} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n_1 + S_Y^2/n_2}}.$$

que tem distribuição t-Student com

$$\nu = \frac{(S_X^2/n_1 + S_Y^2/n_2)^2}{\frac{(S_X^2/n_1)^2}{(n_1-1)} + \frac{(S_Y^2/n_2)^2}{(n_2-1)}}.$$

Exercícios recomendados

- Seção 9.2 - 1 a 6.