Testes de variância e Análise de Variância (ANOVA)

Prof. Marcos Vinicius Pó

Introdução à Inferência Estatística

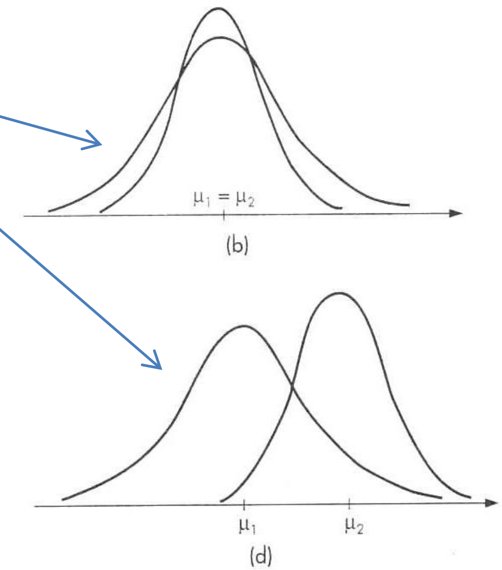
Introdução à Inferência Estatística

TESTE DE VARIÂNCIAS E DISTRIBUIÇÃO F

Prof. Marcos Vinicius Pó

Testes sobre variâncias

- Problema: queremos saber se há diferenças estatisticamente significativas entre os desvios-padrão de duas amostras. ou seja, se elas são ou não homocedásticas



- Em termos de teste de hipótese isso significa:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

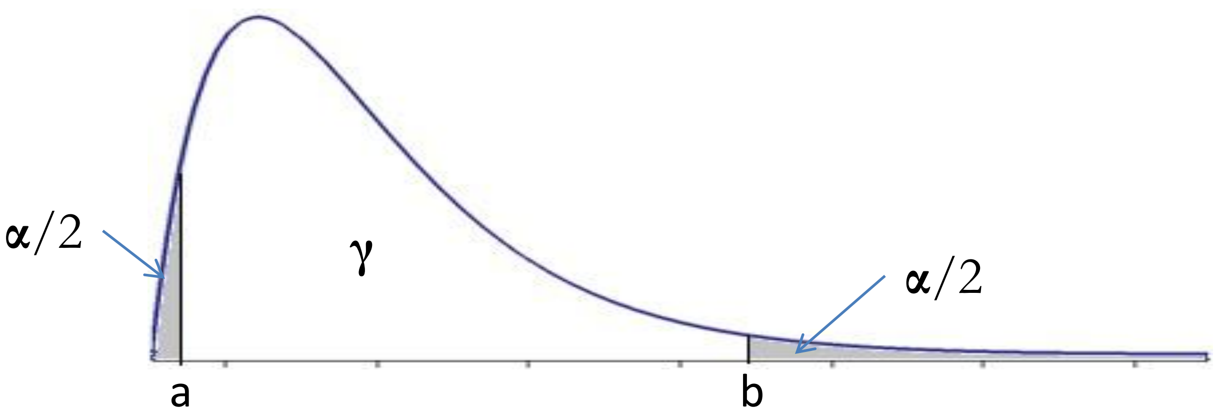
$$H_A : \sigma_1^2 \neq \sigma_2^2$$

Teste para variância usando distribuição Qui-quadrado ($\chi 2$)

- Pode-se testar uma variância nos mesmos moldes de teste de médias, utilizando-se a distribuição χ^2

- A medição de χ^2_{Obs} é determinada por: $$\chi^2_{Obs} = \frac{(n-1)S^2_{obs}}{\sigma^2_{ref}}$$

- A região crítica é bicaudal, determinada pelo **γ**, de acordo com os graus de liberdade (**ν**) determinados pela amostra.



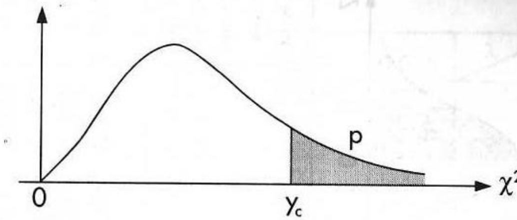
Distribuição qui-quadrado (χ^2)

Tabela IV — Distribuição Qui-quadrado

$Y \sim \chi^2(\nu)$

Corpo da tabela dá os valores y_c tais que $P(Y > y_c) = p$.

Para valores $\nu > 30$, use a aproximação normal dada no texto.



Graus de liberdade ν	p = 99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%	Graus de liberdade ν
1	$0{,}0^{3}16$	$0{,}0^{3}63$	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	4,218	5,024	5,412	6,635	9,550	10,827	1
2	0,020	0,040	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	6,438	7,378	7,824	9,210	12,429	13,815	2
3	0,115	0,185	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	8,311	9,348	9,837	11,345	14,796	16,266	3
4	0,297	0,429	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	10,026	11,143	11,668	13,277	16,924	18,467	4
5	0,554	0,752	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	11,644	12,832	13,388	15,086	18,907	20,515	5
6	0,872	1,134	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	13,198	14,449	15,033	16,812	20,791	22,457	6
7	1,239	1,564	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	14,703	16,013	16,622	18,475	22,601	24,322	7
8	1,646	2,032	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	16,171	17,534	18,168	20,090	24,352	26,125	8
9	2,088	2,532	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	17,608	19,023	19,679	21,666	26,056	27,877	9
10	2,558	3,059	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	19,021	20,483	21,161	23,209	27,722	29,588	10
11	3,053	3,609	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	20,412	21,920	22,618	24,725	29,354	31,264	11
12	3,571	4,178	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	21,785	23,337	24,054	26,217	30,957	32,909	12
13	4,107	4,765	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	23,142	24,736	25,472	27,688	32,535	34,528	13
14	4,660	5,368	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	24,485	26,119	26,873	29,141	34,091	36,123	14
15	5,229	5,985	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	25,816	27,488	28,259	30,578	35,628	37,697	15
16	5,812	6,614	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	27,136	28,845	29,633	32,000	37,146	39,252	16
17	6,408	7,255	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	28,445	30,191	30,995	33,409	38,648	40,790	17
18	7,015	7,906	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	29,745	31,526	32,346	34,805	40,136	42,312	18
19	7,633	8,567	8,906	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	31,037	32,852	33,687	36,191	41,610	43,820	19
20	8,260	9,237	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	32,321	34,170	35,020	37,566	43,072	45,315	20
21	8,897	9,915	10,283	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	33,597	35,479	36,343	38,932	44,522	46,797	21
22	9,542	10,600	10,982	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	34,867	36,781	37,659	40,289	45,962	48,268	22
23	10,196	11,293	11,688	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	36,131	38,076	38,968	41,638	47,391	49,728	23
24	10,856	11,992	12,401	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	37,389	39,364	40,270	42,980	48,812	51,179	24
25	11,524	12,697	13,120	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	38,642	40,646	41,566	44,314	50,223	52,620	25
26	12,198	13,409	13,844	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	39,889	41,923	42,856	45,642	51,627	54,052	26
27	12,879	14,125	14,573	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	41,132	43,194	44,140	46,963	53,022	55,476	27
28	13,565	14,847	15,308	16,928	18,939	21,588	23,647	27,336	31,319	34,027	37,916	41,337	42,370	44,461	45,419	48,278	54,411	56,893	28
29	14,258	15,574	16,047	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	43,604	45,722	46,693	49,588	55,792	58,302	29
30	14,953	16,306	16,791	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	44,834	46,979	47,962	50,892	57,167	59,703	30
	p = 99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%	

Teste para variância utilizando a estatística F (Fisher-Snedecor)

- Sejam S_1^2 e S_2^2 as variâncias de duas amostras, definimos **F** como:

$$\boxed{F = \frac{S_1^2}{S_2^2}}$$

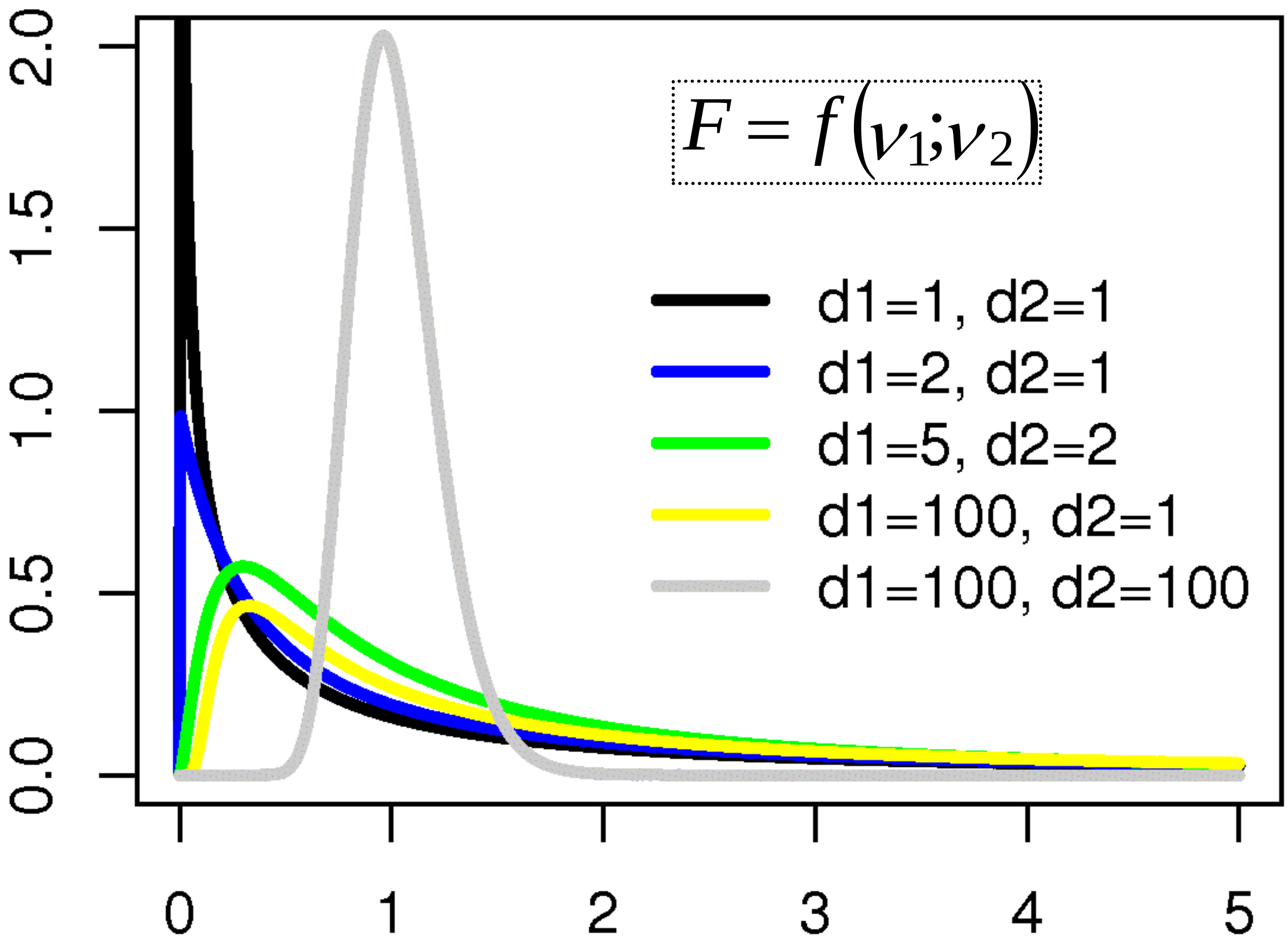
- Para um determinado coeficiente de confiança (), temos o seguinte intervalo de confiança para a razão entre duas variâncias

$$\boxed{IC\left(\sigma_2^2 / \sigma_1^2 ; \gamma\right) = f_1 \frac{s_2^2}{s_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < f_2 \frac{s_2^2}{s_1^2}}$$

Propriedades da distribuição F

- Família de curvas determinada pelos graus de liberdade no numerador e no denominador (ν_1 e ν_2);
- São representadas graficamente de forma positiva;
- A área total sob cada curva de uma distribuição F é 1;
- Valores F são sempre iguais ou maiores que zero;
- Para todas as distribuições F, o valor médio de F é aproximadamente 1.

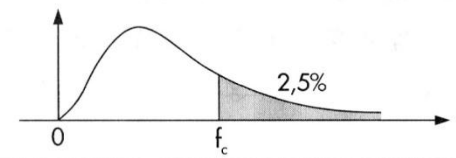
Exemplo de curvas da distribuição F



Fonte: Wikipédia

Exemplo de tabela F: p=2,5%

Tabela VI — Distribuição F (continuação)
Corpo da tabela dá os valores f_c tais que $P(F > f_c) = 0{,}025$.



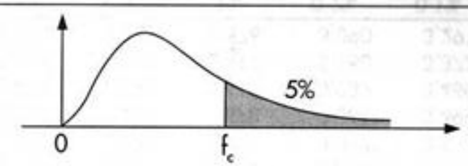
Graus de liberdade do denominador de F: ν_2	Grau de liberdade do numerador de F: ν_1 1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	Graus de liberdade do denominador de F: ν_2
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	963,3	976,7	984,9	993,1	997,2	1001	1006	1010	1014	1018	1
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50	2
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90	3
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	3,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26	4
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02	5
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85	6
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14	7
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67	8
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33	9
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08	10
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88	11
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72	12
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60	13
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49	14
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,80	2,76	2,70	2,64	2,59	2,52	2,46	2,40	15
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32	16
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25	17
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19	18
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13	19
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09	20
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04	21
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00	22
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97	23
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94	24
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91	25
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88	26
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85	27
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83	28
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81	29
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79	30
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64	40
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48	60
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31	120
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00	∞
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	

Pr

Exemplo de tabela F: p=5%

Tabela VI — Distribuição F

Corpo da tabela dá os valores f_c tais que $P(F > f_c) = 0{,}05$.

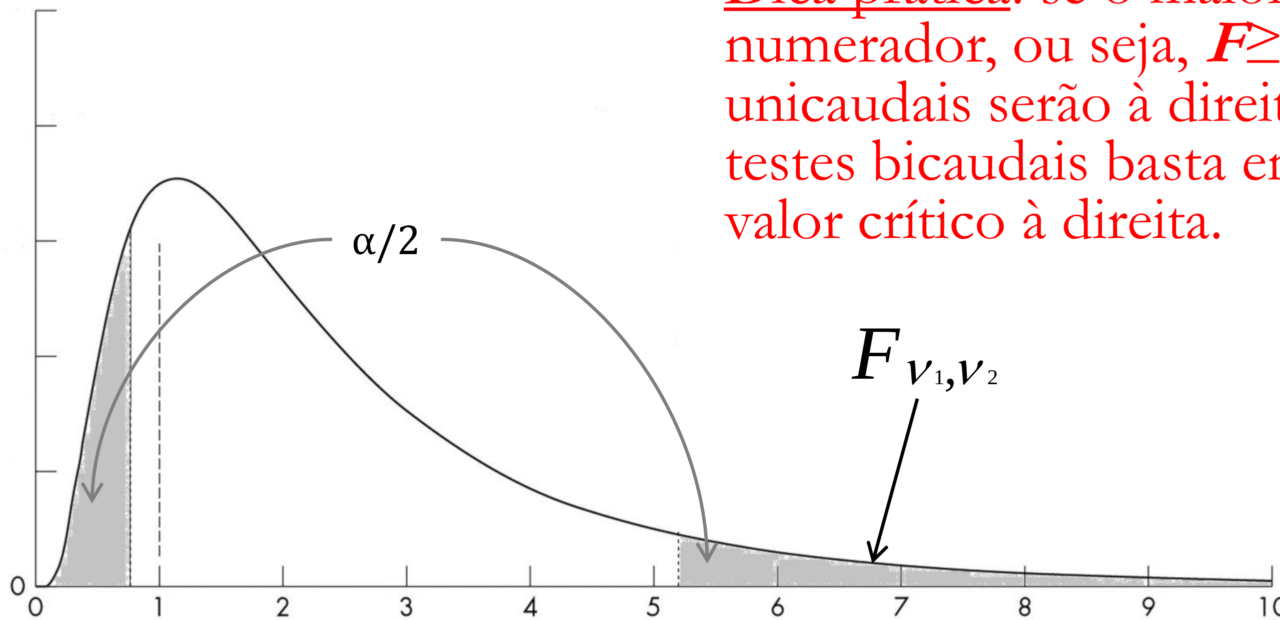


Grau de liberdade do numerador de F: ν_1

Graus de liberdade do denominador de F: ν_2	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	24	30	40	60	120	∞	Graus de liberdade do denominador de F: ν_2
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,4	245,9	246,5	247,3	248,0	249,1	250,1	251,1	252,2	253,3	254,3	1
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,43	19,44	19,45	19,45	19,46	19,47	19,48	19,49	19,50	2
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,72	8,70	8,69	8,67	8,66	8,64	8,62	8,59	8,57	8,55	8,53	3
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,86	5,84	5,82	5,80	5,77	5,75	5,72	5,69	5,66	5,63	4
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,62	4,60	4,58	4,56	4,53	4,50	4,46	4,43	4,40	4,36	5
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,94	3,92	3,90	3,87	3,84	3,81	3,77	3,74	3,70	3,67	6
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,51	3,49	3,47	3,44	3,41	3,38	3,34	3,30	3,27	3,23	7
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,22	3,20	3,17	3,15	3,12	3,08	3,04	3,01	2,97	2,93	8
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	3,01	2,99	2,96	2,94	2,90	2,86	2,83	2,79	2,75	2,71	9
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,87	2,85	2,83	2,80	2,77	2,74	2,70	2,66	2,62	2,58	2,54	10
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,72	2,70	2,67	2,65	2,61	2,57	2,53	2,49	2,45	2,40	11
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,62	2,60	2,57	2,54	2,51	2,47	2,43	2,38	2,34	2,30	12
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,53	2,52	2,48	2,46	2,42	2,38	2,34	2,30	2,25	2,21	13
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,46	2,44	2,41	2,39	2,35	2,31	2,27	2,22	2,18	2,13	14
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,40	2,39	2,35	2,33	2,29	2,25	2,20	2,16	2,11	2,07	15
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,35	2,33	2,30	2,28	2,24	2,19	2,15	2,11	2,06	2,01	16
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,34	2,31	2,29	2,26	2,23	2,19	2,15	2,10	2,06	2,01	1,96	17
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,27	2,25	2,22	2,19	2,15	2,11	2,06	2,02	1,97	1,92	18
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,23	2,22	2,18	2,16	2,11	2,07	2,03	1,98	1,93	1,88	19
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,20	2,18	2,15	2,12	2,08	2,04	1,99	1,95	1,90	1,84	20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,18	2,16	2,12	2,10	2,05	2,01	1,96	1,92	1,87	1,81	21
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,15	2,13	2,10	2,07	2,03	1,98	1,94	1,89	1,84	1,78	22
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,13	2,11	2,08	2,05	2,01	1,96	1,91	1,86	1,81	1,76	23
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,11	2,09	2,05	2,03	1,98	1,94	1,89	1,84	1,79	1,73	24
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,09	2,07	2,04	2,01	1,96	1,92	1,87	1,82	1,77	1,71	25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,07	2,05	2,02	1,99	1,95	1,90	1,85	1,80	1,75	1,69	26
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,06	2,04	2,00	1,97	1,93	1,88	1,84	1,79	1,73	1,67	27
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,04	2,02	1,99	1,96	1,91	1,87	1,82	1,77	1,71	1,65	28
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,03	2,01	1,97	1,94	1,90	1,85	1,81	1,75	1,70	1,64	29
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	2,01	1,99	1,96	1,93	1,89	1,84	1,79	1,74	1,68	1,62	30
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,92	1,90	1,87	1,84	1,79	1,74	1,69	1,64	1,58	1,51	40
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,84	1,81	1,78	1,75	1,70	1,65	1,59	1,53	1,47	1,39	60
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,77	1,75	1,72	1,69	1,66	1,61	1,55	1,50	1,43	1,35	1,25	120
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,69	1,67	1,63	1,60	1,57	1,52	1,46	1,39	1,32	1,22	1,00	∞
	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	24	30	40	60	120	∞	

Região Crítica na curva F

- <u>Dica prática</u>: se o maior valor ficar no numerador, ou seja, $\boldsymbol{F}\geq 1$, os testes unicaudais serão à direita e para os testes bicaudais basta encontrar o valor crítico à direita.



$f_1 = F_{\nu_1,\nu_2,\alpha/2}$ $\quad$ $f_2 : F_{\nu_1,\nu_2,\alpha/2}$

- Como as tabelas são limitadas, lembrar que:

$$F_{\nu_1,\nu_2,\alpha} = \frac{1}{F_{\nu_2,\nu_1,\alpha}}$$

Exemplos

1. Da população $X \sim N(50; 100)$ retirou-se uma amostra aleatória simples $n=12$. Da população $Y \sim N(60; 100)$ retirou-se uma amostra aleatória simples $m=8$. Obtemos respectivamente S_1^2 e S_2^2. (baseado em Bussab; Moretin, 2002: 361)
   a. Encontre o valor de a, tal que $P(S_1^2/S_2^2 < a) = 95\%$
   b. Encontre o valor de b, tal que $P(S_1^2/S_2^2 > b) = 95\%$

2. Estabeleça um intervalo de confiança de 95% (γ) para S_1^2/S_2^2 considerando duas amostras idênticas e independentes com os seguintes tamanhos: 10; 30; 120

3. Deseja-se comparar a uniformidade da produção de duas fábricas em relação ao comprimento dos produtos. Tomaram-se duas amostras aleatórias conforme tabela abaixo. Podemos afirmar que a uniformidade das fábricas é a mesma? (baseado em Bussab; Moretin, 2002: 361)

Estatísticas	Fábrica A	Fábrica B
Amostra	21	16
Média	21,15	21,12
Desvio-padrão	0,2030	0,4164

Introdução à Inferência Estatística

ANOVA

Prof. Marcos Vinicius Pó

Modelos explicativos estatísticos

- Modelos estatísticos visam descrever sinteticamente o comportamento de variáveis. Eles podem ser definidos como:

Observação = Previsível + Aleatório;

ou

Observação = Previsível x Aleatório

- Assim, um modelo estatístico para uma observação pode ser definido basicamente por uma equação do tipo:

$$y_i = \Theta + e_i$$

Onde:
$\boldsymbol{y_i}$: efeito verificado na i-ésima observação
$\boldsymbol{\theta}$: efeito fixo, comum a todos
$\boldsymbol{e_i}$: erro, devido à fatores não explícitos no modelo, com distribuição $e_i \sim N(0;\sigma^2)$

ANOVA e modelos estatísticos

- O objetivo dos nossos modelos explicativos estatísticos é diminuir o erro, ou seja, aquilo que não é explicado.
- Até agora os nossos modelos restringiam-se a apenas uma estimativa:

$$y_i = \mu + e_i \quad \text{ou} \quad y_i = p + e_i \quad \text{onde} \quad e_i = f(\sigma)$$

- Será que em alguns casos não diminuiremos a nossa variação (ou seja, aumentamos a precisão) e reduziremos o erro se usarmos várias médias, relacionadas a outras variáveis explicativas não tratadas no modelo simples?
- A ANOVA permite testar e mensurar isso

Utilidades da ANOVA

- **ANOVA** = **AN**alisys **O**f **Va**riance

- A ANOVA permite fazer a comparação global de diversas amostras ou subamostras, minimizando a probabilidade de erro amostral, já que, conforme aumenta o número de amostras, o total de comparações entre pares aumenta exponencialmente

Amostras/subamostras	Total de comparações
3	3
4	6
5	10
8	28
10	45

ANOVA de 1 fator (unidirecional)

- Objetivo: avaliar se várias médias populacionais são iguais ou se, pelo menos uma, é diferente
  - ▸ H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_n$

- Para isso, verificamos como se comporta a variação **entre** as médias das várias populações e a variação **dentro** dessas populações.

Pressupostos da ANOVA

- Amostras aleatórias simples
- Amostras independentes
- Populações normais
- As populações são homocedásticas (se tiver dúvida, teste!)

Estatística do teste ANOVA

- Estatística F (razão F)

- Essa estatística indica o tamanho da diferença entre as amostras, em função do tamanho da variação dentro de cada amostra.

$$F = \frac{MSe}{MSd}$$

Onde:

MSe = Variância entre amostras
MSd = Variância dentro das amostras

Tabela da ANOVA

- Para facilitar o manuseio dos dados, eles são organizados em uma tabela:

  n: número de amostras

  k: número de subpopulações

$$SQe = \sum_1^k n\left(\bar{\bar{x}} - \bar{x}_i\right)^2 \qquad SQd = \sum_1^k (n_i - 1)s_i^2 \qquad SQt = \frac{1}{(n-1)}\left(\sum_1^n \bar{x}_i^2 - n \bullet \bar{x}^2\right)$$

Variação	Soma dos Quadrados (SQ)	Graus de Liberdade	Quadrados das Médias (SQM)	F
Entre populações/ grupos	SQ_e	$gl_e = k-1$	$MQe = \frac{SQe}{gl_N}$	$\frac{MQe}{MQd}$
Dentro das populações/ grupos	SQ_d	$gl_d = n-k$	$MQd = \frac{SQd}{gl_D}$	
Total	SQ_t	$gl_t = n-1$		

Prof. Marcos Vinicius Pó

Exemplo: definição de modelo explicativo e uso da ANOVA

- Um psicólogo deseja avaliar explicações para o tempo de reação das pessoas a determinado estímulo visual.
- Para isso ele mediu o tempo de reação (***y***) de 20 pessoas e compilou outras variáveis que, com base nas teorias, podem afetar ***y***.
- Modelo básico (<u>sempre o mais simples – K.I.S.S.</u>): o tempo de reação dos indivíduos varia aleatoriamente na população, podendo ser explicado apenas pela média e a variância.

Indivíduo	*Tempo de reação (ms)*	*Gênero (M/F)*	*Idade (anos)*	*Acuidade Visual (%)*
i	***y***	***w***	***x***	***z***
1	96	M	20	90
2	92	F	20	100
3	106	M	20	80
4	100	F	20	90
5	98	F	25	100
6	104	M	25	90
7	110	M	25	80
8	101	F	25	90
9	116	F	30	70
10	106	M	30	90
11	109	M	30	90
12	100	F	30	80
13	112	F	35	90
14	105	F	35	80
15	118	M	35	70
16	108	M	35	90
17	113	F	40	90
18	112	F	40	90
19	127	M	40	60
20	117	M	40	80

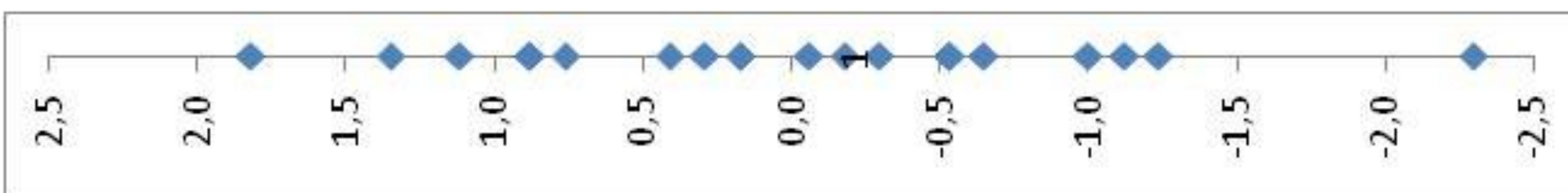
Dados tirados de Bussab, Wilton. Análise de Variância e Regressão. 2a. Ed. Editora Atual: São Paulo. 1988

Modelo I: média e desvio-padrão

- Obtenha média e desvio-padrão do tempo de reação;
- Calcule os resíduos normalizados $[(\bar{y}-y_i)/s]$ e analise-os.
- Calcule a soma dos erros quadráticos $[\Sigma(\bar{y}-y_i)^2]$

O que vocês acham? O modelo explica bem o fenômeno? Por que?

Será possível melhorar? Será que vale a pena melhorar esse modelo tornando-o mais complexo?



Modelo II: separando por Gênero (duas populações)

- Adicionamos uma discriminação nos nossos dados: j = Gênero (M/F2; M=1,F=2...)

$$y_{ij} = \Theta_i + e_{ij}$$

- Temos agora duas populações, Masculina e Feminina
  - ▸ Calcular média e desvio-padrão para ambas.
  - ▸ Calcule a soma dos erros quadráticos de ambas
  - ▸ São estatisticamente diferentes? Será que o modelo fica melhor adicionando essa variável?

$\bar{y}_M$= 110,1; σ^2_M = 74,54; SEQ_M = 566,9

$\bar{y}_F$ = 104,9; σ^2_F = 62,99; SEQ_F = 670,9

Fazendo a tabela de ANOVA

- Precisamos calcular:
- Variância dentro das populações (SQ_d)
  - Soma da soma de erros quadráticos de cada uma das populações
- Variância entre as populações (SQ_e)
  - $SQ_t = SQ_e + SQ_d \rightarrow = SQ_e = SQ_t - SQ_d$
- Determinar os graus de liberdade

Variação	Soma dos Quadrados (SQ)	Graus de Liberdade	Quadrados das Médias (SQM)	F
Entre populações/ grupos	SQ_e= 1373-1237,8 **=135,2**	1	$\frac{SSe}{gl_n} =$ 135,2	$\frac{MSe}{MSd} =$ 1,97
Dentro das populações/ grupos	SQ_d=566,9 + 670,9 **=1237,8**	n-2 = 18	$\frac{SSd}{gl_d} =$ 68,77	
Total	1373,0	n-1 = 19		

Medidas que a ANOVA permite

- **R^2:** coeficiente de explicação

$$R^2 = \frac{SQe}{SQt}$$

  - Significa a quantidade de informação que é explicada pelo modelo adotado
  - No nosso caso, $R^2 = 135{,}2/1373 = 9{,}85\%$, ou seja, a separação por gênero explica muito pouco do resultado
- **p-valor de F**: indica a possibilidade de generalização do modelo para a população
  - Igual ao p-valor de um teste de hipótese, ou seja, o nível em que podemos afirmar que o modelo é significativo
  - No nosso caso, $F_{(1,97;1;18)} = 0{,}177$, ou seja, o modelo é pouco significativo

Exercício para a próxima aula

- Separar as populações por idade

- Calcular para cada uma
  - Média e desvio-padrão (colocar em um quadro comparativo)
  - Soma dos quadrados dos erros

- Colocar na tabela de ANOVA

- O que parece? Esse modelo melhora a nossa previsão? Quanto?

Modelo III: Múltiplas populações (separação por idade)

- Resultado

<table>
<tr><th></th><th>Total</th><th>20</th><th>25</th><th>30</th><th>35</th><th>40</th></tr>
<tr><td>média</td><td>107,5</td><td>98,5</td><td>103,3</td><td>107,8</td><td>110,8</td><td>117,3</td></tr>
<tr><td>dpad</td><td>8,50</td><td>5,97</td><td>5,12</td><td>6,65</td><td>5,62</td><td>6,85</td></tr>
<tr><td rowspan="2">SQDesvios</td><td rowspan="2">1373,0</td><td>107,0</td><td>78,8</td><td>132,8</td><td>94,8</td><td>140,8</td></tr>
<tr><td colspan="5">554,0</td></tr>
</table>

ANOVA do Modelo III

- n: número de amostras
- k: número de populações

$Fc_{(5\%;4;15)} = 3,06$
p-valor $_{(5,54;4;14)} = 0,61\%$
$R^2 = 0,587$

Variação	Soma dos Quadrados (SQ)	Graus de Liberdade	Quadrados das Médias (SQM)	F
Entre populações/ grupos	$SS_e = 819$	k-1 = 4	$\frac{SSe}{gl_n} = 204,75$	$\frac{MSe}{MSd} = 5,54$
Dentro das populações/ grupos	$SS_d = 554,0$	n-k = 15	$\frac{SSd}{gl_d} = 36,93$	
Total	1373,0	n-1 = 19	72,26	

Conclusões do Modelo III

- É estatisticamente significativo (ao nível de menos de 1%)
- Possui um bom valor explicativo (57,9%)
- Portanto, o modelo III tem qualidades para ser adotado.
- Isso significa que a idade é um fator explicativo relevante para o fenômeno observado (tempo de reação).

Exercícios

1. Em um curso de extensão universitária pesquisaram-se os salários mensais (em unidades de referência) e a área de formação acadêmica dos estudantes, com base em uma amostra aleatória. Após eliminar-se os dados excessivamente destoantes, obteve-se o resultado abaixo. Podemos considerar que os salários de cada área são iguais?

	n	Média	Desvio-padrão^2
Sociais	21	30,9	19,2
Engenharia	15	34,2	28,2
Biológicas	7	38,1	22,3

Tabela da ANOVA

- Para facilitar o manuseio dos dados, eles são organizados em uma tabela:

n: número de amostras

k: número de subpopulações

$$MQe = \frac{SQe}{gl_e} = \frac{\sum_1^k n_i(\bar{\bar{x}} - \bar{x}_i)^2}{gl_e}$$

$$MQd = \frac{SQd}{gl_D} = \frac{\sum_1^k (n_i - 1)s_i^2}{gl_D}$$

Variação	Soma dos Quadrados (SQ)	Graus de Liberdade	Quadrados das Médias (SQM)	F
Entre populações/ grupos	SQ_e	$gl_e = k-1$	$MQe = \frac{SQe}{gl_N}$	$\frac{MQe}{MQd}$
Dentro das populações/ grupos	SQ_d	$gl_d = n-k$	$MQd = \frac{SQd}{gl_D}$	
Total	SQ_t	$gl_t = n-1$		

Exercícios

2. Um analista quer determinar se há diferença na média de vendas mensais de quatro regiões diferentes. É feita uma seleção aleatória de vendedores de cada região e cada um fornece os resultados (em R$ mil) do mês anterior. Com $\alpha = 5\%$ podemos concluir que há diferença na média de vendas de pelo menos uma das regiões?

	Norte	Leste	Sul	Oeste
Variância total = 68,10	34	47	40	21
	28	36	30	30
	18	30	41	24
	24	38	29	37
		44		23
Média	*26*	*39*	*35*	*27*
Variância	*45,33*	*45*	*40,67*	*42,5*