Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 24, 2021

2021 Spring

Chapter 14. Nonparametric Statistics

Plan

The Kruskal-Wallis Test

What is the nonparametric counterpart for the one-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Kruskal-Wallis Test

What is the nonparametric counterpart for the one-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Kruskal-Wallis Test

What is the nonparametric counterpart for the one-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Kruskal-Wallis Test

What is the nonparametric counterpart for the one-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

The Kruskal-Wallis Test

What is the nonparametric counterpart for the one-way ANOVA?

Setup Suppose that $k \geq 2$ independent sample of size $n_1, \cdots, n_k$ are drawn from k

identically shaped and scaled pdfs,
except for possibly different medians.

Let $\widetilde{\mu}_1, \cdots, \widetilde{\mu}_k$ be the medians.

Test $H_0 : \widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_k$ vs. H_1 : not all the $\widetilde{\mu}_i$'s are equal.

Remark This is the test for median not mean, but if pdfs are symmetric, they are the same.

Kruskal-Wallis statistic B

$$B = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{\cdot j}^2}{n_j} - 3(n+1)$$

where

Kruskal-Wallis statistic B

$$B = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1)$$

where

Table 14.4.1 Notation for Kruskal-Wallis Procedure

	Treatment Level			
	1	2	$\cdots$	k
	$Y_{11}(R_{11})$	$Y_{12}(R_{12})$		$Y_{1k}(R_{1k})$
	$Y_{21}(R_{21})$			
	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	$Y_{n_1 1}(R_{n_1 1})$	$Y_{n_2 2}(R_{n_2 2})$		$Y_{n_k k}(R_{n_k k})$
Totals	$R_{.1}$	$R_{.2}$		$R_{.k}$

Theorem Under the above setup and under H_0, then

$$B = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1) \overset{approx}{\sim} \chi^2_{k-1}.$$

H_0 should be rejected at the α level of significance if $b > \chi^2_{1-\alpha,k-1}$.

Theorem Under the above setup and under H_0, then

$$B = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1) \overset{approx}{\sim} \chi^2_{k-1}.$$

H_0 should be rejected at the α level of significance if $b > \chi^2_{1-\alpha,k-1}$.

E.g. Lottery over the year 1969; Whether lottery is random?

Test if $H_0 : \widetilde{\mu}_{\mathrm{Jan}} = \widetilde{\mu}_{\mathrm{Feb}} = \cdots = \widetilde{\mu}_{\mathrm{Dec}}$ at $\alpha = 0.01$

E.g. Lottery over the year 1969; Whether lottery is random?

Test if $H_0 : \widetilde{\mu}_{\text{Jan}} = \widetilde{\mu}_{\text{Feb}} = \cdots = \widetilde{\mu}_{\text{Dec}}$ at $\alpha = 0.01$

Table 14.4.2 1969 Draft Lottery, Highest Priority (001) to Lowest Priority (366)

Date	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1	305	086	108	032	330	249	093	111	225	359	019	129
2	159	144	029	271	298	228	350	045	161	125	034	328
3	251	297	267	083	040	301	115	261	049	244	348	157
4	215	210	275	081	276	020	279	145	232	202	266	165
5	101	214	293	269	364	028	188	054	082	024	310	056
6	224	347	139	253	155	110	327	114	006	087	076	010
7	306	091	122	147	035	085	050	168	008	234	051	012
8	199	181	213	312	321	366	013	048	184	283	097	105
9	194	338	317	219	197	335	277	106	263	342	080	043
10	325	216	323	218	065	206	284	021	071	220	282	041
11	329	150	136	014	037	134	248	324	158	237	046	039
12	221	068	300	346	133	272	015	142	242	072	066	314
13	318	152	259	124	295	069	042	307	175	138	126	163
14	238	004	354	231	178	356	331	198	001	294	127	026
15	017	089	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	055	274	120	044	207	254	107	096
17	235	189	033	260	112	073	098	154	255	288	143	304
18	140	292	332	090	278	341	190	141	246	005	146	128
19	058	025	200	336	075	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	063	192	185	135
21	186	363	334	062	250	060	027	291	204	243	156	070
22	337	290	265	316	326	247	153	339	160	117	009	053
23	118	057	256	252	319	109	172	116	119	201	182	162
24	059	236	258	002	031	358	023	036	195	196	230	095
25	052	179	343	351	361	137	067	286	149	176	132	084
26	092	365	170	340	357	022	303	245	018	007	309	173
27	355	205	268	074	296	064	289	352	233	264	047	078
28	077	299	223	262	308	222	088	167	257	094	281	123
29	349	285	362	191	226	353	270	061	151	229	099	016
30	164		217	208	103	209	287	333	315	038	174	003
31	211		030		313		193	011		079		100
Totals:	6236	5886	7000	6110	6447	5872	5628	5377	4719	5656	4462	3768

Sol. Rank the lottery for the year (see the previous table).

Compute b using the formula:

$$b = \frac{12}{366 \times 367}\left[\frac{6236^2}{31} + \frac{5886^2}{29} + \cdots + \frac{3768^2}{31}\right] - 3 \times 367$$
$$= 25.95.$$

Critical region is $C = \{b : b \geq \chi^2_{0.99,11} = 24.725\}$.

Conclusion: Reject (Lottery is *NOT* random). ■

Sol. Rank the lottery for the year (see the previous table).

Compute b using the formula:

$$b = \frac{12}{366 \times 367}\left[\frac{6236^2}{31} + \frac{5886^2}{29} + \cdots + \frac{3768^2}{31}\right] - 3 \times 367$$
$$= 25.95.$$

Critical region is $C = \{b : b \geq \chi^2_{0.99,11} = 24.725\}$.

Conclusion: Reject (Lottery is *NOT* random). ∎

Sol. Rank the lottery for the year (see the previous table).

Compute b using the formula:

$$b = \frac{12}{366 \times 367}\left[\frac{6236^2}{31} + \frac{5886^2}{29} + \cdots + \frac{3768^2}{31}\right] - 3 \times 367$$
$$= 25.95.$$

Critical region is $C = \left\{b : b \geq \chi^2_{0.99,11} = 24.725\right\}$.

Conclusion: Reject (Lottery is *NOT* random). ∎

Sol. Rank the lottery for the year (see the previous table).

Compute b using the formula:

$$b = \frac{12}{366 \times 367}\left[\frac{6236^2}{31} + \frac{5886^2}{29} + \cdots + \frac{3768^2}{31}\right] - 3 \times 367$$
$$= 25.95.$$

Critical region is $C = \left\{b : b \geq \chi^2_{0.99,11} = 24.725\right\}$.

Conclusion: Reject (Lottery is *NOT* random). ■