

Stat 218 - Day 4
Five-number summary, Boxplots

Five-number summary (FNS):

Minimum, Lower Quartile (Q1), Median, Upper Quartile (Q3), Maximum

Quartiles are calculated as the median of the values below/above the location of the actual median

Example: Natural selection (Bumpus)

Humerus lengths (inches) of sparrows who *survived* storm ($n=35$):

0.687	0.703	0.709	0.715	0.721	0.723	0.723	0.726	0.728
0.728	0.728	0.729	0.730	0.730	0.733	0.733	0.735	0.736
0.739	0.741	0.741	0.741	0.741	0.743	0.749	0.751	0.752
0.752	0.755	0.756	0.766	0.767	0.769	0.770	0.780	

Humerus lengths of sparrows who *perished* in storm ($n=24$):

0.659	0.689	0.702	0.703	0.709	0.713	0.720	0.720	0.726
0.726	0.729	0.731	0.736	0.737	0.738	0.738	0.739	0.743
0.744	0.745	0.752	0.752	0.754	0.765			

(a) Find FNS of humerus lengths for sparrows who survived the storm.

(b) Find FNS of humerus lengths for sparrows who did not survive the storm.

Boxplot: visual representation of FNS

(c) Produce (unmodified) boxplots to compare these distributions.

(d) Compare and contrast the two distributions of humerus lengths.

Modified boxplot: indicates outliers separately, extends “whisker” to most extreme non-outlier.

Outliers are values more than $1.5*(Q3-Q1)$ from nearer quartile

(e) Check for outliers in both groups.

(f) Create modified boxplots of humerus lengths for the two groups.

(g) Use Minitab (`bumpus.mtw`) to reproduce this analysis. [*Notes:* The data are presented in “unstacked” format in `c1` and `c2`, meaning that each group has its own column. The data are presented in “stacked” format in `c4` and `c5`, meaning that all of the data are in one column with a separate column indicating the group variable. You can use `Graph> Boxplot` and then choose “one Y, with groups” with the stacked data or “multiple Y’s, simple” with the unstacked data. Also, be aware that Minitab may calculate quartiles a bit differently.]

Example: Draft lottery

The following data are the draft numbers (1-366) assigned to birthdates in the 1970 draft lottery. Men born on the date assigned a draft number of 1 were the first to be drafted, followed by those born on the date assigned draft number 2, and so on.

Date	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	305	86	108	32	330	249	93	111	225	359	19	129
2	159	144	29	271	298	228	350	45	161	125	34	328
3	251	297	267	83	40	301	115	261	49	244	348	157
4	215	210	275	81	276	20	279	145	232	202	266	165
5	101	214	293	269	364	28	188	54	82	24	310	56
6	224	347	139	253	155	110	327	114	6	87	76	10
7	306	91	122	147	35	85	50	168	8	234	51	12
8	199	181	213	312	321	366	13	48	184	283	97	105
9	194	338	317	219	197	335	277	106	263	342	80	43
10	325	216	323	218	65	206	284	21	71	220	282	41
11	329	150	136	14	37	134	248	324	158	237	46	39
12	221	68	300	346	133	272	15	142	242	72	66	314
13	318	152	259	124	295	69	42	307	175	138	126	163
14	238	4	354	231	178	356	331	198	1	294	127	26
15	17	89	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	55	274	120	44	207	254	107	96
17	235	189	33	260	112	73	98	154	255	288	143	304
18	140	292	332	90	278	341	190	141	246	5	146	128
19	58	25	200	336	75	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	63	192	185	135
21	186	363	334	62	250	60	27	291	204	243	156	70
22	337	290	265	316	326	247	153	339	160	117	9	53
23	118	57	256	252	319	109	172	116	119	201	182	162
24	59	236	258	2	31	358	23	36	195	196	230	95
25	52	179	343	351	361	137	67	286	149	176	132	84
26	92	365	170	340	357	22	303	245	18	7	309	173
27	355	205	268	74	296	64	289	352	233	264	47	78
28	77	299	223	262	308	222	88	167	257	94	281	123
29	349	285	362	191	226	353	270	61	151	229	99	16
30	164		217	208	103	209	287	333	315	38	174	3
31	211		30		313		193	11		79		100

(a) What draft number was assigned to your birthday? Is this draft number in the top third, middle third, or last third of the draft order?

The following table arranges in order the draft numbers for each month:

Rank	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	17	4	29	2	31	20	13	11	1	5	9	3
2	52	25	30	14	35	22	15	21	6	7	19	10
3	58	57	33	32	37	28	23	36	8	24	34	12
4	59	68	108	62	40	60	27	44	18	38	46	16
5	77	86	122	74	55	64	42	45	49	72	47	26
6	92	89	136	81	65	69	50	48	63	79	51	39
7	101	91	139	83	75	73	67	54	71	87	66	41
8	118	144	166	90	103	85	88	61	82	94	76	43
9	121	150	169	124	112	104	93	102	113	117	80	53
10	140	152	170	147	130	109	98	106	119	125	97	56
11	159	179	200	148	133	110	115	111	149	138	99	70
12	164	181	213	191	155	134	120	114	151	171	107	78
13	186	189	217	208	178	137	153	116	158	176	126	84
14	194	205	223	218	183	180	172	141	160	192	127	95
15	199	210	239	219	197	206	187	142	161	196	131	96
16	211	212	256	231	226	209	188	145	175	201	132	100
17	215	214	258	252	250	222	190	154	177	202	143	105
18	221	216	259	253	276	228	193	167	184	220	146	123
19	224	236	265	260	278	247	227	168	195	229	156	128
20	235	285	267	262	295	249	248	198	204	234	174	129
21	238	290	268	269	296	272	270	245	207	237	182	135
22	251	292	275	271	298	274	277	261	225	241	185	157
23	280	297	293	273	308	301	279	286	232	243	203	162
24	305	299	300	312	313	335	284	291	233	244	230	163
25	306	302	317	316	319	341	287	307	242	254	266	165
26	318	338	323	336	321	353	289	311	246	264	281	173
27	325	347	332	340	326	356	303	324	255	283	282	240
28	329	363	334	345	330	358	322	333	257	288	309	304
29	337	365	343	346	357	360	327	339	263	294	310	314
30	349		354	351	361	366	331	344	315	342	348	320
31	355		362		364		350	352		359		328

(b) Use this information to calculate the median draft number for *your* birth month. Is this number in the top half or bottom half of the draft order?

(c) Pool the findings of the class and record the median draft number for each month. Do you notice any tendency in these median draft numbers over time?

(d) Create boxplots of the draft numbers for each month (`draft70.mtw`; `Graph>` `Boxplot; One Y with groups; draft num, month`). Do they reveal any tendency over time? Explain.