APPLIED STATISTICAL ANALYSIS LAB

NAME: ADITI KULKARNI
ROLL NO: 55
YEAR: SY
DIVISION: E (E3)
SRN NO: 202201893

ASSIGNMENT 1

STATEMENT: *Read the data from data set drinks.csv and compute standardized z-score values for the numeric variables.*

THEORY:

$$Z = (x - \mu)/ \sigma$$

where,

x *is variable*
μ *is mean*
σ *is standard deviation*

Z-score or standard score is used for standardizing scores on the same scale by dividing score deviation by the standard deviation in a data set.

Z-score can be positive or negative based on results. A negative score indicates that the value is less than the mean and a positive score indicates that the value is greater than the mean.

The average of every z-score for a data is zero.

Steps to find z-score using R:

1. Read the file
2. Find the mean
3. Find standard deviation
4. Conclude z-score
5. Validate the standardization

A z-score close to zero suggests that the data point is close to the mean, while a larger absolute z-score indicates greater deviation from the mean. Z-scores can be used to identify outliers in a dataset. Typically, data points with z-scores above 3 or below -3 are considered

outliners. Z-scores are commonly used in hypothesis testing and to set thresholds for statistical significance.

SOURCE CODE:

```
## Read the dataset drinks.csv

drinks<-read.csv(file.choose())

View(drinks)

## Find mean & standard deviation of a variable z = ( x - mean ) / sd

mean_bs<-mean(drinks$beer_servings)

mean_bs

sd_bs<-sd(drinks$beer_servings)

sd_bs

## Find and save standardized z-scores of a variable into another variable

drinks$S_beer_servings<-(drinks$beer_servings - mean_bs) / (sd_bs)

View(drinks)

## Validate standardization

round(mean(drinks$S_beer_servings),digits=0)       ## be zero or close to zero

sd(drinks$S_beer_servings) ## equal to 1

write.csv(drinks," D:\\Programming\\R\\z-drinks.csv",row.names=FALSE)
```

OUTPUT:

```
> mean_bs<-mean(drinks$beer_servings)
> mean_bs
[1] 106.1606
> sd_bs<-sd(drinks$beer_servings)
> sd_bs
[1] 101.1431
> drinks$S_beer_servings<-(drinks$beer_servings - mean_bs) / (sd_bs)
> View(drinks)
> round(mean(drinks$S_beer_servings),digits=0)
[1] 0
> sd(drinks$S_beer_servings) ## equal to 1
[1] 1
> write.csv(drinks,"D:\\Programming\\R\\Z-drinks.csv",row.names=FALSE)
```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent	S_beer_servings
70	Guatemala	53	69	2	2.2	*NA*	-0.525598092
71	Guinea	9	0	2	0.2	AF	-0.960625286
72	Guinea-Bissau	28	31	21	2.5	AF	-0.772772634
73	Guyana	93	302	1	7.1	SA	-0.130118826
74	Haiti	1	326	1	5.9	*NA*	-1.039721139
75	Honduras	69	98	2	3.0	*NA*	-0.367406386
76	Hungary	234	215	185	11.3	EU	1.263945588
77	Iceland	233	61	78	6.6	EU	1.254058607
78	India	9	114	0	2.2	AS	-0.960625286
79	Indonesia	5	1	0	0.1	AS	-1.000173212
80	Iran	0	0	0	0.0	AS	-1.049608121
81	Iraq	9	3	0	0.2	AS	-0.960625286
82	Ireland	313	118	165	11.4	EU	2.045017140
83	Israel	63	69	9	2.5	AS	-0.426728276
84	Italy	85	42	237	6.5	EU	-0.209214679
85	Jamaica	82	97	9	3.4	*NA*	-0.238875624
86	Japan	77	202	16	7.0	AS	-0.288310533
87	Jordan	6	21	1	0.5	AS	-0.990286231
88	Kazakhstan	124	246	12	6.8	AS	0.176377605
89	Kenya	58	22	2	1.8	AF	-0.476163184
90	Kiribati	21	34	1	1.0	OC	-0.841981506
91	Kuwait	0	0	0	0.0	AS	-1.049608121
92	Kyrgyzstan	31	97	6	2.4	AS	-0.743111689
93	Laos	62	0	123	6.2	AS	-0.436615258
94	Latvia	281	216	62	10.5	EU	1.728633726
95	Lebanon	20	55	31	1.9	AS	-0.851868487

FINAL EXCEL SHEET (csv):

	A	B	C	D	E	F	G	H
1	country	beer_servi	spirit_serv	wine_servi	total_litres	continent	S_beer_servings	
2	Afghanista	0	0	0	0	AS	-1.04961	
3	Albania	89	132	54	4.9	EU	-0.16967	
4	Algeria	25	0	14	0.7	AF	-0.80243	
5	Andorra	245	138	312	12.4	EU	1.372702	
6	Angola	217	57	45	5.9	AF	1.095867	
7	Antigua &	102	128	45	4.9	NA	-0.04114	
8	Argentina	193	25	221	8.3	SA	0.858579	
9	Armenia	21	179	11	3.8	EU	-0.84198	
10	Australia	261	72	212	10.4	OC	1.530894	
11	Austria	279	75	191	9.7	EU	1.70886	
12	Azerbaijan	21	46	5	1.3	EU	-0.84198	
13	Bahamas	122	176	51	6.3	NA	0.156604	
14	Bahrain	42	63	7	2	AS	-0.63435	
15	Bangladesl	0	0	0	0	AS	-1.04961	
16	Barbados	143	173	36	6.3	NA	0.36423	
17	Belarus	142	373	42	14.4	EU	0.354343	
18	Belgium	295	84	212	10.5	EU	1.867051	
19	Belize	263	114	8	6.8	NA	1.550668	
20	Benin	34	4	13	1.1	AF	-0.71345	
21	Bhutan	23	0	0	0.4	AS	-0.82221	
22	Bolivia	167	41	8	3.8	SA	0.601518	
23	Bosnia-He	76	173	8	4.6	EU	-0.2982	
24	Botswana	173	35	35	5.4	AF	0.66084	
25	Brazil	245	145	16	7.2	SA	1.372702	
26	Brunei	31	2	1	0.6	AS	-0.74311	

	A	B	C	D	E	F	G	H
27	Bulgaria	231	252	94	10.3	EU	1.234285	
28	Burkina Fa	25	7	7	4.3	AF	-0.80243	
29	Burundi	88	0	0	6.3	AF	-0.17955	
30	Cote d'Ivo	37	1	7	4	AF	-0.68379	
31	Cabo Verd	144	56	16	4	AF	0.374117	
32	Cambodia	57	65	1	2.2	AS	-0.48605	
33	Cameroon	147	1	4	5.8	AF	0.403778	
34	Canada	240	122	100	8.2	NA	1.323267	
35	Central Afı	17	2	1	1.8	AF	-0.88153	
36	Chad	15	1	1	0.4	AF	-0.9013	
37	Chile	130	124	172	7.6	SA	0.235699	
38	China	79	192	8	5	AS	-0.26854	
39	Colombia	159	76	3	4.2	SA	0.522422	
40	Comoros	1	3	1	0.1	AF	-1.03972	
41	Congo	76	1	9	1.7	AF	-0.2982	
42	Cook Islan	0	254	74	5.9	OC	-1.04961	
43	Costa Rica	149	87	11	4.4	NA	0.423552	
44	Croatia	230	87	254	10.2	EU	1.224398	
45	Cuba	93	137	5	4.2	NA	-0.13012	
46	Cyprus	192	154	113	8.2	EU	0.848692	
47	Czech Rep	361	170	134	11.8	EU	2.519592	
48	North Kore	0	0	0	0	AS	-1.04961	
49	DR Congo	32	3	1	2.3	AF	-0.73322	
50	Denmark	224	81	278	10.4	EU	1.165076	
51	Djibouti	15	44	3	1.1	AF	-0.9013	
52	Dominica	52	286	26	6.6	NA	-0.53549	

	A	B	C	D	E	F	G	H
53	Dominican	193	147	9	6.2	NA	0.858579	
54	Ecuador	162	74	3	4.2	SA	0.552083	
55	Egypt	6	4	1	0.2	AF	-0.99029	
56	El Salvado	52	69	2	2.2	NA	-0.53549	
57	Equatorial	92	0	233	5.8	AF	-0.14001	
58	Eritrea	18	0	0	0.5	AF	-0.87164	
59	Estonia	224	194	59	9.5	EU	1.165076	
60	Ethiopia	20	3	0	0.7	AF	-0.85187	
61	Fiji	77	35	1	2	OC	-0.28831	
62	Finland	263	133	97	10	EU	1.550668	
63	France	127	151	370	11.8	EU	0.206039	
64	Gabon	347	98	59	8.9	AF	2.381175	
65	Gambia	8	0	1	2.4	AF	-0.97051	
66	Georgia	52	100	149	5.4	EU	-0.53549	
67	Germany	346	117	175	11.3	EU	2.371288	
68	Ghana	31	3	10	1.8	AF	-0.74311	
69	Greece	133	112	218	8.3	EU	0.26536	
70	Grenada	199	438	28	11.9	NA	0.917901	
71	Guatemala	53	69	2	2.2	NA	-0.5256	
72	Guinea	9	0	2	0.2	AF	-0.96063	
73	Guinea-Bis	28	31	21	2.5	AF	-0.77277	
74	Guyana	93	302	1	7.1	SA	-0.13012	
75	Haiti	1	326	1	5.9	NA	-1.03972	
76	Honduras	69	98	2	3	NA	-0.36741	
77	Hungary	234	215	185	11.3	EU	1.263946	
78	Iceland	233	61	78	6.6	EU	1.254059	

	A	B	C	D	E	F	G	H
79	India	9	114	0	2.2	AS	-0.96063	
80	Indonesia	5	1	0	0.1	AS	-1.00017	
81	Iran	0	0	0	0	AS	-1.04961	
82	Iraq	9	3	0	0.2	AS	-0.96063	
83	Ireland	313	118	165	11.4	EU	2.045017	
84	Israel	63	69	9	2.5	AS	-0.42673	
85	Italy	85	42	237	6.5	EU	-0.20921	
86	Jamaica	82	97	9	3.4	NA	-0.23888	
87	Japan	77	202	16	7	AS	-0.28831	
88	Jordan	6	21	1	0.5	AS	-0.99029	
89	Kazakhstar	124	246	12	6.8	AS	0.176378	
90	Kenya	58	22	2	1.8	AF	-0.47616	
91	Kiribati	21	34	1	1	OC	-0.84198	
92	Kuwait	0	0	0	0	AS	-1.04961	
93	Kyrgyzstan	31	97	6	2.4	AS	-0.74311	
94	Laos	62	0	123	6.2	AS	-0.43662	
95	Latvia	281	216	62	10.5	EU	1.728634	
96	Lebanon	20	55	31	1.9	AS	-0.85187	
97	Lesotho	82	29	0	2.8	AF	-0.23888	
98	Liberia	19	152	2	3.1	AF	-0.86176	
99	Libya	0	0	0	0	AF	-1.04961	
100	Lithuania	343	244	56	12.9	EU	2.341627	
101	Luxembou	236	133	271	11.4	EU	1.28372	
102	Madagasc	26	15	4	0.8	AF	-0.79255	
103	Malawi	8	11	1	1.5	AF	-0.97051	
104	Malaysia	13	4	0	0.3	AS	-0.92108	

	A	B	C	D	E	F	G	H
105	Maldives	0	0	0	0	AS	-1.04961	
106	Mali	5	1	1	0.6	AF	-1.00017	
107	Malta	149	100	120	6.6	EU	0.423552	
108	Marshall Is	0	0	0	0	OC	-1.04961	
109	Mauritania	0	0	0	0	AF	-1.04961	
110	Mauritius	98	31	18	2.6	AF	-0.08068	
111	Mexico	238	68	5	5.5	NA	1.303494	
112	Micronesia	62	50	18	2.3	OC	-0.43662	
113	Monaco	0	0	0	0	EU	-1.04961	
114	Mongolia	77	189	8	4.9	AS	-0.28831	
115	Monteneg	31	114	128	4.9	EU	-0.74311	
116	Morocco	12	6	10	0.5	AF	-0.93096	
117	Mozambiq	47	18	5	1.3	AF	-0.58492	
118	Myanmar	5	1	0	0.1	AS	-1.00017	
119	Namibia	376	3	1	6.8	AF	2.667897	
120	Nauru	49	0	8	1	OC	-0.56515	
121	Nepal	5	6	0	0.2	AS	-1.00017	
122	Netherlan	251	88	190	9.4	EU	1.432024	
123	New Zeala	203	79	175	9.3	OC	0.957449	
124	Nicaragua	78	118	1	3.5	NA	-0.27842	
125	Niger	3	2	1	0.1	AF	-1.01995	
126	Nigeria	42	5	2	9.1	AF	-0.63435	
127	Niue	188	200	7	7	OC	0.809144	
128	Norway	169	71	129	6.7	EU	0.621292	
129	Oman	22	16	1	0.7	AS	-0.83209	
130	Pakistan	0	0	0	0	AS	-1.04961	

	A	B	C	D	E	F	G	H
131	Palau	306	63	23	6.9	OC	1.975808	
132	Panama	285	104	18	7.2	NA	1.768182	
133	Papua Nev	44	39	1	1.5	OC	-0.61458	
134	Paraguay	213	117	74	7.3	SA	1.056319	
135	Peru	163	160	21	6.1	SA	0.56197	
136	Philippines	71	186	1	4.6	AS	-0.34763	
137	Poland	343	215	56	10.9	EU	2.341627	
138	Portugal	194	67	339	11	EU	0.868466	
139	Qatar	1	42	7	0.9	AS	-1.03972	
140	South Kore	140	16	9	9.8	AS	0.334569	
141	Moldova	109	226	18	6.3	EU	0.028073	
142	Romania	297	122	167	10.4	EU	1.886825	
143	Russian Fe	247	326	73	11.5	AS	1.392476	
144	Rwanda	43	2	0	6.8	AF	-0.62447	
145	St. Kitts &	194	205	32	7.7	NA	0.868466	
146	St. Lucia	171	315	71	10.1	NA	0.641066	
147	St. Vincent	120	221	11	6.3	NA	0.13683	
148	Samoa	105	18	24	2.6	OC	-0.01148	
149	San Marin	0	0	0	0	EU	-1.04961	
150	Sao Tome	56	38	140	4.2	AF	-0.49594	
151	Saudi Arab	0	5	0	0.1	AS	-1.04961	
152	Senegal	9	1	7	0.3	AF	-0.96063	
153	Serbia	283	131	127	9.6	EU	1.748408	
154	Seychelles	157	25	51	4.1	AF	0.502648	
155	Sierra Leo	25	3	2	6.7	AF	-0.80243	
156	Singapore	60	12	11	1.5	AS	-0.45639	

	A	B	C	D	E	F	G	H
157	Slovakia	196	293	116	11.4	EU	0.88824	
158	Slovenia	270	51	276	10.6	EU	1.619877	
159	Solomon I	56	11	1	1.2	OC	-0.49594	
160	Somalia	0	0	0	0	AF	-1.04961	
161	South Afric	225	76	81	8.2	AF	1.174963	
162	Spain	284	157	112	10	EU	1.758295	
163	Sri Lanka	16	104	0	2.2	AS	-0.89142	
164	Sudan	8	13	0	1.7	AF	-0.97051	
165	Suriname	128	178	7	5.6	SA	0.215926	
166	Swaziland	90	2	2	4.7	AF	-0.15978	
167	Sweden	152	60	186	7.2	EU	0.453213	
168	Switzerlan	185	100	280	10.2	EU	0.779483	
169	Syria	5	35	16	1	AS	-1.00017	
170	Tajikistan	2	15	0	0.3	AS	-1.02983	
171	Thailand	99	258	1	6.4	AS	-0.0708	
172	Macedoni	106	27	86	3.9	EU	-0.00159	
173	Timor-Lest	1	1	4	0.1	AS	-1.03972	
174	Togo	36	2	19	1.3	AF	-0.69368	
175	Tonga	36	21	5	1.1	OC	-0.69368	
176	Trinidad &	197	156	7	6.4	NA	0.898127	
177	Tunisia	51	3	20	1.3	AF	-0.54537	
178	Turkey	51	22	7	1.4	AS	-0.54537	
179	Turkmenis	19	71	32	2.2	AS	-0.86176	
180	Tuvalu	6	41	9	1	OC	-0.99029	
181	Uganda	45	9	0	8.3	AF	-0.60469	
182	Ukraine	206	237	45	8.9	EU	0.98711	

183	United Ara	16	135	5	2.8	AS	-0.89142
184	United Kin	219	126	195	10.4	EU	1.115641
185	Tanzania	36	6	1	5.7	AF	-0.69368
186	USA	249	158	84	8.7	NA	1.41225
187	Uruguay	115	35	220	6.6	SA	0.087395
188	Uzbekistar	25	101	8	2.4	AS	-0.80243
189	Vanuatu	21	18	11	0.9	OC	-0.84198
190	Venezuela	333	100	3	7.7	SA	2.242757
191	Vietnam	111	2	1	2	AS	0.047847
192	Yemen	6	0	0	0.1	AS	-0.99029
193	Zambia	32	19	4	2.5	AF	-0.73322
194	Zimbabwe	64	18	4	4.7	AF	-0.41684
195							

CONCLUSION:

In summary, the provided R code accomplishes the following tasks:

1. Reads a dataset named "drinks.csv" by prompting the user to select the file.

2. Calculates the mean and standard deviation of the "beer_servings" variable.

3. Computes standardized z-score values for the "beer_servings" variable and saves them in a new column named "S_beer_servings" in the "drinks" data frame.

4. Validates the standardization by checking if the mean of the standardized values is close to zero and the standard deviation is approximately equal to 1.

5. Finally, it writes the modified data frame with the added standardized z-scores to a new CSV file named "Z-drinks.csv" located at "D:\\Programming\\R".

Note: The standardized z-scores are obtained by subtracting the mean of the "beer_servings" variable and dividing by its standard deviation. This process ensures that all the values in the "S_beer_servings" column have a mean close to zero and a standard deviation equal to 1. Standardization is often used to compare and analyze variables on a common scale.