



ACADGILD

SESSION 8: Exploratory Data Analytics

Assignment 2

Submitted by: Munmun Ghosal

Login Id: munmun55@gmail.com

(M):+91-8007178659

Table of Contents

1. Problem Statement.....	3
2. Solution.....	3

1. Problem Statement

```
library(RcmdrPlugin.IPSUR)
data(RcmdrTestDrive)
```

Perform the below operations: -

- Compute the measures of central tendency for salary and reduction which variable has highest center?
- Which measure of center is more appropriate for before and after?

2. Solution

a. Compute the measures of central tendency for salary and reduction which variable has highest center?

The R-script for the given problem is as follows:

```
library(Rcmdr)
library(RcmdrPlugin.IPSUR)
data(RcmdrTestDrive)
RcmdrTestDrive
str(RcmdrTestDrive)

library(psych)

describe(RcmdrTestDrive$salary)
describe(RcmdrTestDrive$reduction)
```

The output of the R-Script (from Console window) is given as follows:

```
> library(Rcmdr)
> library(RcmdrPlugin.IPSUR)
> data(RcmdrTestDrive)
> RcmdrTestDrive
```

	order	smoking	gender	race	before	after	salary	reduction	parking
1	1	Nonsmoker	Female	Caucasian	72.6	75.2	618.65	9	2
2	2	Nonsmoker	Male	AfricanAmer	75.3	73.2	544.56	62	1
3	3	Nonsmoker	Female	Caucasian	75.5	74.5	550.24	19	4
4	4	Nonsmoker	Female	Caucasian	71.3	74.6	616.16	30	1
5	5	Nonsmoker	Female	Hispanic	74.3	73.8	543.39	105	1
6	6	Nonsmoker	Male	Caucasian	73.0	73.6	692.09	43	1
7	7	Smoker	Male	Hispanic	72.4	70.7	800.48	229	5
8	8	Nonsmoker	Male	Hispanic	73.6	74.0	703.79	40	1
9	9	Nonsmoker	Female	Caucasian	73.7	75.9	540.06	101	2
10	10	Nonsmoker	Female	Hispanic	74.6	74.8	522.28	440	1
11	11	Nonsmoker	Female	AfricanAmer	75.8	73.1	377.17	213	1
12	12	Nonsmoker	Female	Caucasian	75.3	72.1	525.96	474	2

13	13	Nonsmoker	Female	Caucasian	75.0	72.5	548.88	144	1
14	14	Nonsmoker	Male	Asian	72.8	72.7	537.70	179	2
15	15	Nonsmoker	Male	Asian	74.4	75.7	500.20	63	3
16	16	Nonsmoker	Female	Hispanic	72.9	73.1	597.73	570	1
17	17	Nonsmoker	Female	Hispanic	72.3	74.0	578.95	437	4
18	18	Nonsmoker	Male	Caucasian	74.0	74.6	690.06	62	2
19	19	Nonsmoker	Male	Caucasian	73.1	72.8	748.98	437	2
20	20	Nonsmoker	Male	AfricanAmer	74.0	76.1	811.71	60	1
21	21	Nonsmoker	Male	Other	73.6	74.5	660.58	255	1
22	22	Nonsmoker	Male	Hispanic	73.4	75.0	586.29	133	4
23	23	Nonsmoker	Male	AfricanAmer	73.9	74.0	387.59	88	1
24	24	Nonsmoker	Male	Caucasian	73.0	73.9	524.54	116	1
25	25	Nonsmoker	Female	Hispanic	74.2	75.7	536.87	48	3
26	26	Nonsmoker	Male	Caucasian	73.6	75.4	503.64	365	1
27	27	Smoker	Male	AfricanAmer	74.6	68.1	496.09	73	1
28	28	Nonsmoker	Male	AfricanAmer	74.5	72.6	701.91	306	5
29	29	Nonsmoker	Female	Caucasian	72.6	73.2	595.70	497	1
30	30	Nonsmoker	Male	Asian	72.6	74.1	759.30	32	1
31	31	Nonsmoker	Female	Hispanic	72.1	73.7	717.91	497	1
32	32	Nonsmoker	Male	Asian	73.2	73.5	808.63	21	2
33	33	Smoker	Male	Caucasian	73.2	70.0	682.60	291	1
34	34	Nonsmoker	Male	Asian	74.3	75.2	623.09	83	1
35	35	Smoker	Male	AfricanAmer	74.0	68.7	550.28	55	2
36	36	Nonsmoker	Male	AfricanAmer	75.5	72.9	646.25	100	8
37	37	Nonsmoker	Female	AfricanAmer	75.4	72.6	635.43	439	4
38	38	Nonsmoker	Male	Caucasian	75.5	72.5	437.19	419	1
39	39	Nonsmoker	Female	Caucasian	74.4	73.6	619.29	23	2
40	40	Nonsmoker	Male	Caucasian	73.7	75.0	593.68	71	1
41	41	Nonsmoker	Male	AfricanAmer	75.8	73.1	546.26	109	4
42	42	Nonsmoker	Female	Caucasian	74.3	72.2	704.83	98	1
43	43	Nonsmoker	Male	Caucasian	74.7	73.1	764.15	78	1
44	44	Nonsmoker	Female	Caucasian	74.9	72.0	859.67	257	3
45	45	Nonsmoker	Female	AfricanAmer	75.3	76.2	724.25	487	1
46	46	Nonsmoker	Male	AfricanAmer	75.6	75.0	631.62	213	3
47	47	Nonsmoker	Female	Hispanic	72.7	73.4	478.39	383	1
48	48	Nonsmoker	Female	Caucasian	75.6	74.9	652.79	116	1
49	49	Nonsmoker	Male	Caucasian	73.8	71.9	545.66	1632	2
50	50	Nonsmoker	Male	Caucasian	74.7	75.8	515.95	151	1
51	51	Nonsmoker	Male	AfricanAmer	75.4	74.8	612.27	152	3
52	52	Nonsmoker	Female	Hispanic	74.3	73.8	633.12	390	2
53	53	Nonsmoker	Male	AfricanAmer	75.0	73.2	671.35	64	1
54	54	Nonsmoker	Female	AfricanAmer	75.3	73.8	643.83	85	1
55	55	Nonsmoker	Male	Hispanic	74.8	73.6	794.66	71	2
56	56	Smoker	Female	Asian	73.2	70.6	888.00	37	1
57	57	Nonsmoker	Female	Caucasian	74.0	75.8	602.94	89	2
58	58	Smoker	Male	Caucasian	75.5	74.3	716.78	172	1
59	59	Nonsmoker	Male	Caucasian	75.3	72.8	606.12	3	1
60	60	Nonsmoker	Male	AfricanAmer	73.9	73.7	704.90	247	5
61	61	Nonsmoker	Male	Caucasian	71.7	72.5	620.32	127	2
62	62	Nonsmoker	Male	Caucasian	73.6	74.7	515.92	337	1
63	63	Nonsmoker	Female	AfricanAmer	72.1	73.7	655.72	123	1
64	64	Nonsmoker	Female	Hispanic	72.7	73.1	619.44	205	4
65	65	Nonsmoker	Female	Caucasian	74.5	71.9	640.48	61	1
66	66	Smoker	Male	Caucasian	73.2	72.8	844.32	119	2
67	67	Nonsmoker	Female	Caucasian	73.3	74.9	918.03	165	2
68	68	Nonsmoker	Female	Asian	74.2	75.1	933.49	480	6

69	69	Nonsmoker	Female	Hispanic	74.7	74.2	699.63	39	3
70	70	Nonsmoker	Female	Caucasian	74.4	74.2	593.27	434	4
71	71	Smoker	Male	Caucasian	74.5	69.7	634.24	147	1
72	72	Smoker	Female	Caucasian	73.0	69.3	686.98	270	2
73	73	Nonsmoker	Female	Hispanic	73.5	72.5	618.68	384	1
74	74	Smoker	Female	Hispanic	72.3	70.6	631.20	87	1
75	75	Nonsmoker	Female	Caucasian	75.7	73.8	608.88	291	3
76	76	Smoker	Female	Hispanic	75.6	69.1	686.28	31	2
77	77	Smoker	Female	AfricanAmer	75.4	70.0	715.44	549	1
78	78	Nonsmoker	Male	Hispanic	73.4	74.8	754.66	172	2
79	79	Nonsmoker	Male	AfricanAmer	72.9	74.6	865.89	251	1
80	80	Nonsmoker	Female	Caucasian	72.3	74.0	890.88	335	6
81	81	Smoker	Male	AfricanAmer	74.4	70.7	777.91	319	1
82	82	Smoker	Male	Caucasian	72.8	70.5	680.56	519	1
83	83	Nonsmoker	Male	Caucasian	75.1	73.5	594.61	94	2
84	84	Nonsmoker	Male	AfricanAmer	73.2	75.1	651.73	15	1
85	85	Smoker	Male	Caucasian	74.0	71.3	601.11	397	5
86	86	Nonsmoker	Female	Asian	73.8	72.9	626.71	95	2
87	87	Nonsmoker	Female	Caucasian	73.5	74.8	643.80	551	2
88	88	Smoker	Male	Hispanic	72.2	66.6	724.52	89	1
89	89	Nonsmoker	Female	AfricanAmer	74.4	75.3	745.57	121	2
90	90	Smoker	Male	Caucasian	75.2	72.5	842.05	319	1
91	91	Nonsmoker	Male	AfricanAmer	73.6	74.2	880.47	424	3
92	92	Nonsmoker	Female	Caucasian	73.1	72.6	1016.21	79	2
93	93	Nonsmoker	Male	AfricanAmer	73.9	73.3	726.13	372	5
94	94	Nonsmoker	Male	Caucasian	74.9	74.4	780.21	195	1
95	95	Nonsmoker	Female	Caucasian	72.5	75.0	704.08	324	1
96	96	Nonsmoker	Female	Other	75.0	73.4	785.89	532	3
97	97	Nonsmoker	Male	AfricanAmer	73.8	75.2	662.98	91	2
98	98	Nonsmoker	Male	Caucasian	73.6	75.2	621.30	32	1
99	99	Smoker	Male	Asian	74.8	71.3	521.17	94	2
100	100	Nonsmoker	Female	Caucasian	73.8	74.3	714.58	95	3
101	101	Nonsmoker	Male	Caucasian	75.8	74.6	728.94	99	5
102	102	Smoker	Male	Caucasian	75.5	71.1	812.26	275	1
103	103	Smoker	Male	Caucasian	72.4	71.7	924.78	203	1
104	104	Nonsmoker	Female	AfricanAmer	73.6	74.3	1001.31	131	3
105	105	Nonsmoker	Male	Hispanic	73.3	74.3	724.99	116	2
106	106	Nonsmoker	Male	Hispanic	72.9	73.3	822.35	66	1
107	107	Nonsmoker	Male	Hispanic	75.7	73.1	653.58	574	1
108	108	Nonsmoker	Female	Asian	72.6	73.3	642.28	87	1
109	109	Nonsmoker	Male	AfricanAmer	73.8	73.6	730.12	149	1
110	110	Smoker	Female	AfricanAmer	72.8	70.6	708.30	538	1
111	111	Nonsmoker	Male	Caucasian	73.9	71.9	629.17	419	2
112	112	Nonsmoker	Male	Caucasian	73.2	75.1	790.33	33	1
113	113	Nonsmoker	Male	AfricanAmer	75.5	73.8	788.05	213	1
114	114	Nonsmoker	Female	Caucasian	72.4	74.5	849.25	44	1
115	115	Nonsmoker	Male	AfricanAmer	72.8	74.5	1036.06	814	1
116	116	Nonsmoker	Male	Hispanic	74.8	75.2	1149.92	131	2
117	117	Smoker	Male	Caucasian	75.6	72.4	854.31	100	4
118	118	Nonsmoker	Female	Caucasian	74.1	74.2	768.94	688	4
119	119	Smoker	Male	Caucasian	75.3	69.6	666.74	83	1
120	120	Nonsmoker	Female	Hispanic	75.1	73.2	639.72	185	1
121	121	Smoker	Male	AfricanAmer	74.1	70.3	744.38	60	2
122	122	Nonsmoker	Female	Caucasian	74.6	74.1	584.08	6	1
123	123	Nonsmoker	Male	Caucasian	74.1	72.5	712.00	60	2
124	124	Nonsmoker	Female	AfricanAmer	73.9	72.7	789.76	282	1

125	125	Smoker	Male	Hispanic	73.0	67.3	719.06	31	1
126	126	Nonsmoker	Male	AfricanAmer	75.3	73.8	903.34	82	2
127	127	Nonsmoker	Male	Caucasian	73.5	75.3	1044.98	65	1
128	128	Nonsmoker	Male	Asian	72.3	74.8	1027.36	26	2
129	129	Nonsmoker	Female	AfricanAmer	73.5	73.7	855.36	117	1
130	130	Nonsmoker	Male	Caucasian	72.9	76.2	796.51	205	1
131	131	Smoker	Male	Caucasian	72.6	70.3	771.74	99	3
132	132	Nonsmoker	Male	Caucasian	76.3	74.2	780.27	401	1
133	133	Nonsmoker	Male	AfricanAmer	73.0	75.2	808.65	8	2
134	134	Nonsmoker	Female	Caucasian	74.7	74.7	632.05	469	4
135	135	Smoker	Female	Hispanic	74.5	67.5	681.58	116	4
136	136	Nonsmoker	Male	Caucasian	71.4	74.6	823.38	298	4
137	137	Nonsmoker	Male	Hispanic	74.4	73.9	754.55	115	2
138	138	Nonsmoker	Male	Asian	72.1	73.1	938.47	721	1
139	139	Nonsmoker	Male	Caucasian	73.1	76.4	1072.65	135	1
140	140	Nonsmoker	Male	AfricanAmer	73.7	73.3	1021.69	202	1
141	141	Nonsmoker	Female	Caucasian	73.0	73.3	785.75	642	1
142	142	Nonsmoker	Female	Hispanic	73.8	74.4	882.78	95	1
143	143	Nonsmoker	Female	Caucasian	73.6	72.0	762.43	262	2
144	144	Nonsmoker	Male	Hispanic	73.1	74.2	863.78	564	1
145	145	Nonsmoker	Male	Caucasian	73.4	73.9	745.97	258	3
146	146	Nonsmoker	Female	Hispanic	74.0	72.4	809.26	41	1
147	147	Nonsmoker	Female	AfricanAmer	75.8	72.9	668.26	77	3
148	148	Smoker	Female	Asian	74.2	67.8	780.61	429	2
149	149	Nonsmoker	Female	AfricanAmer	75.4	73.3	749.43	557	1
150	150	Nonsmoker	Male	Caucasian	75.1	72.9	889.55	89	1
151	151	Nonsmoker	Female	Caucasian	74.6	74.9	1025.09	59	1
152	152	Smoker	Male	Caucasian	75.5	69.8	1156.16	370	1
153	153	Nonsmoker	Male	AfricanAmer	74.9	74.3	777.93	202	2
154	154	Nonsmoker	Male	AfricanAmer	73.6	74.3	835.96	111	2
155	155	Nonsmoker	Female	Caucasian	74.5	72.6	668.69	598	3
156	156	Nonsmoker	Female	Caucasian	75.7	74.6	870.52	55	1
157	157	Nonsmoker	Male	AfricanAmer	72.6	73.8	827.18	750	1
158	158	Smoker	Male	Caucasian	74.1	70.8	689.23	83	2
159	159	Nonsmoker	Female	AfricanAmer	73.6	74.2	662.17	257	1
160	160	Smoker	Female	Caucasian	75.0	70.3	820.52	303	1
161	161	Nonsmoker	Female	AfricanAmer	73.1	74.8	780.51	79	2
162	162	Nonsmoker	Male	Hispanic	73.6	74.3	980.09	156	4
163	163	Nonsmoker	Male	AfricanAmer	73.6	75.1	1084.21	166	6
164	164	Smoker	Male	Hispanic	73.5	72.1	1073.50	9	1
165	165	Nonsmoker	Male	AfricanAmer	73.7	72.5	908.11	409	3
166	166	Nonsmoker	Male	Hispanic	73.1	73.4	793.42	424	2
167	167	Nonsmoker	Male	Hispanic	74.5	74.9	804.78	205	1
168	168	Nonsmoker	Male	AfricanAmer	73.7	74.1	790.82	47	2

```
> str(RcmdrTestDrive)
```

```
'data.frame': 168 obs. of 9 variables:
```

```
$ order : int 1 2 3 4 5 6 7 8 9 10 ...
```

```
$ smoking : Factor w/ 2 levels "Nonsmoker","Smoker": 1 1 1 1 1 1 2 1 1 1
```

```
...
```

```
$ gender : Factor w/ 2 levels "Female","Male": 1 2 1 1 1 2 2 2 1 1 ...
```

```
$ race : Factor w/ 5 levels "AfricanAmer",...: 3 1 3 3 4 3 4 4 3 4 ...
```

```
$ before : num 72.6 75.3 75.5 71.3 74.3 73 72.4 73.6 73.7 74.6 ...
```

```
$ after : num 75.2 73.2 74.5 74.6 73.8 73.6 70.7 74 75.9 74.8 ...
```

```
$ salary : num 619 545 550 616 543 ...
```

```
$ reduction: int 9 62 19 30 105 43 229 40 101 440 ...
```

```
$ parking : int 2 1 4 1 1 1 5 1 2 1 ...
```

```
>
> library(psych)
```

Attaching package: 'psych'

The following object is masked from 'package:car':

logit

The following objects are masked from 'package:ggplot2':

%+%, alpha

```
> describe(RcmdrTestDrive$salary)
  vars   n  mean    sd median trimmed  mad   min    max range skew
x1    1 168 724.52 148.86 710.15  715.53 134.6 377.17 1156.16 778.99 0.51
      kurtosis   se
x1         0.2 11.49
> describe(RcmdrTestDrive$reduction)
  vars   n  mean    sd median trimmed  mad min  max range skew kurtosis
x1    1 168 223.63 213.06  139.5  194.43 136.4   3 1632  1629 2.29    10.02
      se
x1 16.44
```

Conclusion/Interpretation:

Variable salary has the highest center

b. Which measure of center is more appropriate for before and after?

The R-script for the given problem is as follows:

```
describe(RcmdrTestDrive$before)
describe(RcmdrTestDrive$after)
```

The output of the R-Script (from Console window) is given as follows:

```
> describe(RcmdrTestDrive$before)
  vars   n  mean    sd median trimmed  mad   min  max range skew kurtosis
se
x1    1 168 73.97 1.08   73.9   73.98 1.19 71.3 76.3    5 -0.04    -0.8
0.08
> describe(RcmdrTestDrive$after)
  vars   n  mean    sd median trimmed  mad   min  max range skew kurtosis
se
x1    1 168 73.27 1.88   73.7   73.48 1.56 66.6 76.4   9.8 -1.16    1.32
0.15
```

Conclusion/Interpretation:

Both the variables 'before' & 'after' are negatively skewed, median is more appropriate measure of center.