

Example 1
Baseball Data

Player	Batting	Player	Batting
Smith	0.300	Anderson	0.330
Smith	0.400	Anderson	0.280
Smith	0.325	Anderson	0.240
Smith	0.400	Anderson	0.300
Smith	0.325	Anderson	0.320
Smith	0.400	Mitchell	0.340
Smith	0.325	Mitchell	0.400
Smith	0.400	Mitchell	0.300
Smith	0.325	Mitchell	0.200
Smith	0.400	Mitchell	0.300
Jones	0.150	Mitchell	0.400
Jones	0.225	Rodriguez	0.600
Jones	0.220	Rodriguez	0.500
Jones	0.240	Rodriguez	0.550
Jones	0.180	Rodriguez	0.600
Jones	0.160	Rodriguez	0.500
Jones	0.200	Rodriguez	0.550
Jones	0.210	Suarez	0.600
Jones	0.230	Suarez	0.500
Jones	0.210	Suarez	0.550
Jones	0.200		
Anderson	0.300		

1. What kind of data is Baseball data? Categorical or Numerical?
2. Provide a histogram of the variable Batting? What do we need to do in JMP to change the number of classes (number of bar of the Histogram)?
3. Provide the Summary statistics of the variable Batting?
4. **Provide** a Bar chart or a Pie chart or a Pareto chart for the categorical variable.

Problem 2:
Bigclass Data

name	age	sex	height	weight	name	age	sex	height	weight
KATIE	12	F	59	95	FREDRICK	14	M	63	93
LOUISE	12	F	61	123	ALFRED	14	M	64	99
JANE	12	F	55	74	HENRY	14	M	65	119
JACLYN	12	F	66	145	LEWIS	14	M	64	92
LILLIE	12	F	52	64	EDWARD	14	M	68	112
TIM	12	M	60	84	CHRIS	14	M	64	99
JAMES	12	M	61	128	JEFFERY	14	M	69	113
ROBERT	12	M	51	79	MARY	15	F	62	92
BARBARA	13	F	60	112	AMY	15	F	64	112
ALICE	13	F	61	107	ROBERT	15	M	67	128
SUSAN	13	F	56	67	WILLIAM	15	M	65	111
JOHN	13	M	65	98	CLAY	15	M	66	105
JOE	13	M	63	105	MARK	15	M	62	104
MICHAEL	13	M	58	95	DANNY	15	M	66	106
DAVID	13	M	59	79	MARTHA	16	F	65	112
JUDY	14	F	61	81	MARIAN	16	F	60	115
ELIZABETH	14	F	62	91	PHILLIP	16	M	68	128
LESLIE	14	F	65	142	LINDA	17	F	62	116
CAROL	14	F	63	84	KIRK	17	M	68	134
PATTY	14	F	62	85	LAWRENCE	17	M	70	172

1. Provide the Summary statistics of all variables
 - a) Bar, Pie and Pareto for the categorical
 - b) Histogram for the numerical
2. Provide the Summary about the Skewness of the histograms (Skewness of a normal distribution is 0)
3. Scatterplot
4. Scatterplot as Fig. 1 (using sex as grouping variable)
5. Scatterplot as Fig. 2 (using age as grouping variable)
6. Conclusion?

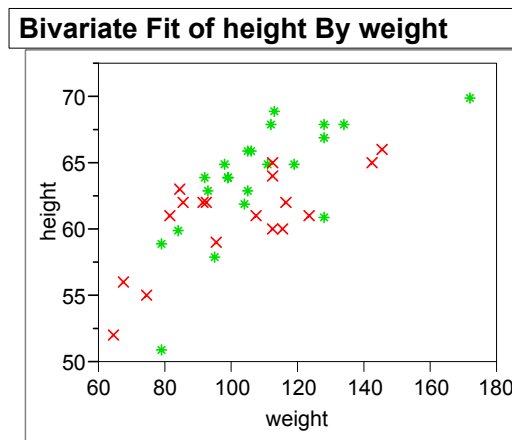


Fig 1

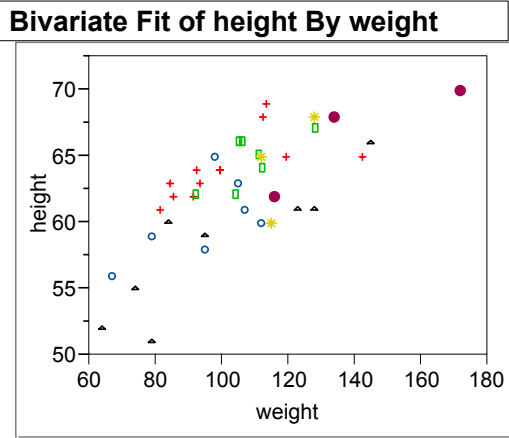


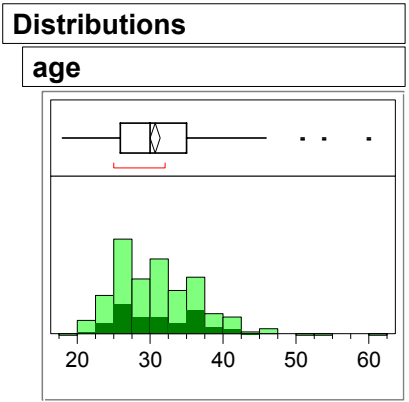
Fig. 2

Problem 3:

Data : CarPoll

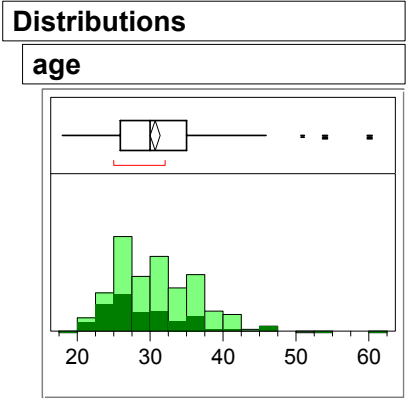
sex	marital status	age	country	size	type
Male	Married	34	American	Large	Family
Male	Single	36	Japanese	Small	Sporty
Male	Married	23	Japanese	Small	Family
Male	Single	29	American	Large	Family
Male	Married	39	American	Medium	Family
Male	Single	34	Japanese	Medium	Family
Female	Married	42	American	Large	Family
Female	Married	40	European	Medium	Family
Male	Married	28	American	Medium	Sporty
Female	Married	26	American	Medium	Family
Female	Married	26	European	Small	Sporty
Male	Single	26	European	Medium	Sporty
Female	Married	37	American	Medium	Sporty
Male	Single	25	Japanese	Small	Sporty
Female	Single	27	Japanese	Medium	Family
Male	Married	41	American	Large	Family
Male	Single	35	American	Small	Sporty
Male	Single	23	American	Medium	Work
Male	Married	28	Japanese	Small	Family
Male	Married	38	European	Medium	Family
Female	Married	25	Japanese	Small	Family
Male	Married	30	Japanese	Medium	Family
Male	Married	36	American	Large	Family
Female	Single	29	European	Small	Sporty
Female	Married	31	European	Medium	Family
Female	Married	26	European	Medium	Sporty
Male	Single	23	Japanese	Small	Sporty
Female	Married	31	American	Medium	Sporty
Female	Married	31	Japanese	Small	Sporty
Male	Married	35	American	Small	Sporty

How do we plot a histogram of a sample within the histogram of all data?



Example

Provide the histogram of the age of single people within the the histogram of the age the whole population (see below).



Example 4:
Data : Drug

Drug	x	y
a	11	6
a	8	0
a	5	2
a	14	8
a	19	11
a	6	4
a	10	13
a	6	1
a	11	8
a	3	0
d	6	0
d	6	2
d	7	3
d	8	1
d	18	18
d	8	4
d	19	14
d	8	9
d	5	1
d	15	9
f	16	13
f	13	10
f	11	18
f	9	5
f	21	23
f	16	12
f	12	5
f	12	16
f	7	1
f	12	20

1. make a scatterplot
2. Investigate the linear relationship between X and Y using graphical and numerical summary
3. Verify that the assumptions for fitting a linear regression are satisfied
4. Make the scatterplot and emphasize over the drug categories in the scatterplot

Example 5:

Data : Financial

1. Investigate Sale by Profit and discuss the shape of the scatterplot
2. Do you see any unusual point?
3. By removing the unusual observations, did we make any improvement to the model?