Frequency Distributions

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

Learning objectives:

Construct an ungrouped frequency distribution using RStudio

Construct a grouped frequency distribution using RStudio

Frequency Distributions

Frequency distributions provide us with a method for arranging and viewing data sets. This allows for easier interpretation and analysis of the data.

Ungrouped vs Grouped Frequency Distributions

Use ungrouped when there are fewer than 20 unique data values in the data set

Use grouped when there are more than 20 unique data values in the data set

Ungrouped Frequency Distribution

Using the same fan data as we employed for the run chart:

Fans 1-10:	68	72	72	74	72	69	75	75	72	73
Fans 10-20:	70	71	71	72	73	72	70	72	73	74

Ungrouped Frequency Distribution Example

	value	freq	rel.freq	cum.up	cum.down
1	68	1	0.05	0.05	1.00
2	69	1	0.05	0.10	0.95
3	70	2	0.10	0.20	0.90
4	71	2	0.10	0.30	0.80
5	72	7	0.35	0.65	0.70
6	73	3	0.15	0.80	0.35
7	74	2	0.10	0.90	0.20
8	75	2	0.10	1.00	0.10

Where:

- value = Score, Value, or Observation
- freq = Frequency
- rel.freq = Relative Frequency
- cum.up / cum.down = Cumulative

Ungrouped Frequency Distribution Example

	value	freq	rel.freq	cum.up	cum.down
1	68	1	0.05	0.05	1.00
2	69	1	0.05	0.10	0.95
3	70	2	0.10	0.20	0.90
4	71	2	0.10	0.30	0.80
5	72	7	0.35	0.65	0.70
6	73	3	0.15	0.80	0.35
7	74	2	0.10	0.90	0.20
8	75	2	0.10	1.00	0.10

Frequency distributions are considered ‘ungrouped’ when each row, or ‘class interval’, consists of only one score, value, or observation.

Ungrouped Frequency Distribution in R

```
> frequency.dist.ungrouped(fans$cfm)
```

Grouped Frequency Distribution

Ungrouped frequency distributions have one value for each class interval. Where the Range ($X_H - X_L$) of the data set is large, however, constructing a functional ungrouped frequency distribution becomes untenable.

Grouped Frequency Distribution

In these cases, we use a Grouped Frequency Distribution.

Grouped frequency distributions have a range of values associated with each interval.

- Example interval: 5 – 9
- Example interval: 1.230 - 1.234

Grouped Frequency Distribution Example

Forty (40) castings for use in a machining process have been randomly selected from an incoming lot from a supplier.

Grouped Frequency Distribution Example

Descriptive Statistics

Variable	Sample Size (n)	Mean	Std. Dev.	Low	High	Range
Weight	40	134.75	14.75	109	170	61

The data are initially arranged in an ungrouped frequency distribution:

Ungrouped Frequency Distribution

Too Many Intervals

	value	freq	rel.freq	cum.up	cum.down
1	109	1	0.025	0.025	1.000
2	111	1	0.025	0.050	0.975
3	117	1	0.025	0.075	0.950
4	118	1	0.025	0.100	0.925
5	120	1	0.025	0.125	0.900
6	121	1	0.025	0.150	0.875
7	122	2	0.050	0.200	0.850
8	124	2	0.050	0.250	0.800
9	125	1	0.025	0.275	0.750
10	126	2	0.050	0.325	0.725
11	128	2	0.050	0.375	0.675
12	129	3	0.075	0.450	0.625
13	130	1	0.025	0.475	0.550
14	131	2	0.050	0.525	0.525
15	132	1	0.025	0.550	0.475
16	133	1	0.025	0.575	0.450
17	134	1	0.025	0.600	0.425
18	135	2	0.050	0.650	0.400
19	137	1	0.025	0.675	0.350
20	139	1	0.025	0.700	0.325
21	143	2	0.050	0.750	0.300
22	146	1	0.025	0.775	0.250
23	148	2	0.050	0.825	0.225
24	152	1	0.025	0.850	0.175
25	155	2	0.050	0.900	0.150
26	158	1	0.025	0.925	0.100
27	162	1	0.025	0.950	0.075
28	165	1	0.025	0.975	0.050
29	170	1	0.025	1.000	0.025

Grouped Frequency Distribution

The data are then reorganized in a Grouped Frequency distribution

	l	min	midpoint	max	u	freq	rel.freq	cum.up	cum.down
1	[	105	107.5	110	)	1	0.025	0.025	1.000
2	[	110	112.5	115	)	1	0.025	0.050	0.975
3	[	115	117.5	120	)	2	0.050	0.100	0.950
4	[	120	122.5	125	)	6	0.150	0.250	0.900
5	[	125	127.5	130	)	8	0.200	0.450	0.750
6	[	130	132.5	135	)	6	0.150	0.600	0.550
7	[	135	137.5	140	)	4	0.100	0.700	0.400
8	[	140	142.5	145	)	2	0.050	0.750	0.300
9	[	145	147.5	150	)	3	0.075	0.825	0.250
10	[	150	152.5	155	)	1	0.025	0.850	0.175
11	[	155	157.5	160	)	3	0.075	0.925	0.150
12	[	160	162.5	165	)	1	0.025	0.950	0.075
13	[	165	167.5	170	)	1	0.025	0.975	0.050
14	[	170	172.5	175	)	1	0.025	1.000	0.025

Grouped Frequency Distribution in R

```
> frequency.dist.grouped(castings$weight)
```

Grouped Frequency Distribution

Important questions to answer:

- How many class intervals, optimally, should the frequency distribution have? How many is too few? Too many?

Grouped Frequency Distribution

- What class interval size is best for the data set we are attempting to portray in a frequency distribution?
- At what class interval should we start the grouped frequency distribution?

Constructing a Grouped Frequency Distribution

- Generate a frequency distribution with as close as you can get to 10 class intervals, without going under (divide the Range by 10 for an estimate of the class interval size you'll need);

Constructing a Grouped Frequency Distribution

- Use one of the following class interval sizes: 1, 2, 3, or 5; increasing the sizes in multiples of 10 where required (e.g. 10, 20, 30, 50, 100…)

Constructing a Grouped Frequency Distribution

- Start the first class interval with a number that is a multiple of the class interval size
- The first class interval must contain the lowest score in the data set (X_L)

Constructing a Grouped Frequency Distribution

- lolcat::freq.dist.grouped considers all of these rules to give an optimal result

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005