

# Policy Report Workshop

# Missing Data

- Missing Cases
  - Who did not make it into your sample and why?
- Missing Variable Values
  - Proxy Variable
  - Compare those with missing values to those without
  - Sensitivity analysis
  - Substitute mean value
  - Selection bias model
- Why are data missing?
  - What does the missing data mean for your results?

# Common Data Issues

- Many categorical values
- Many zero values
- Missing values for different cases
- Know how your variables relate to each other
- What does your data say descriptively?

# Many Categorical Values

2005 HOUSEHOLD INCOME RECODES

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1. \$0-\$4,999	166	1.4	1.4	1.4
2. \$5,000-\$14,999	787	6.7	6.7	8.1
3. \$15,000-\$24,999	889	7.5	7.5	15.6
4. \$25,000-\$34,999	1104	9.3	9.3	24.9
5. \$35,000-\$49,999	1703	14.4	14.4	39.3
6. \$50,000-\$74,999	2702	22.8	22.8	62.1
7. \$75,000-\$99,999	1813	15.3	15.3	77.5
8. \$100,000-\$149,999	1748	14.8	14.8	92.2
9. \$150,000 AND OVER	920	7.8	7.8	100.0
Total	11832	100.0	100.0	

Use the uncategorized variable

Or

Recode at the midpoints

Or

Create dummy variables

HIGHEST LEVEL OF SCHOOL PERSON COMPLETED

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.<9TH GRADE	220	1.9	2.4	2.4
2.9-12TH, NO GED	759	6.4	8.4	10.8
3.HIGH SCHOOL GRAD	2095	17.7	23.1	33.8
4.GED	164	1.4	1.8	35.6
5.VOCATIONAL	278	2.3	3.1	38.7
6.SOME COL, NO DEGR	1523	12.9	16.8	55.4
7.ASSOCIATE DEGR	856	7.2	9.4	64.9
8.BACHELOR'S DEGR	1967	16.6	21.6	86.5
9.MASTER'S DEGR	925	7.8	10.2	96.7
10.PROF DEGR	109	.9	1.2	97.9
11.DOCTORATE DEGR	192	1.6	2.1	100.0
Total	9088	76.8	100.0	
Missing System	2744	23.2		
Total	11832	100.0		

# Many Zero Values

VISITED A ZOO, FAIR OR AMUSEMENT PARK

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid DID NOT PARTICIPATE	1104	9.3	47.0	47.0
1 time	493	4.2	21.0	68.0
2 times	328	2.8	14.0	81.9
3	145	1.2	6.2	88.1
4	77	.7	3.3	91.4
5	87	.7	3.7	95.1
6	23	.2	1.0	96.1
7	10	.1	.4	96.5
8	3	.0	.1	96.6
9	2	.0	.1	96.7
10	37	.3	1.6	98.3
12	11	.1	.5	98.8
15	4	.0	.2	98.9
16	1	.0	.0	99.0
18	1	.0	.0	99.0
20	14	.1	.6	99.6
24	1	.0	.0	99.7
25	2	.0	.1	99.7
30	1	.0	.0	99.8
45	1	.0	.0	99.8
50	1	.0	.0	99.9
52	1	.0	.0	99.9
60	1	.0	.0	100.0
200	1	.0	.0	100.0
Total	2349	19.9	100.0	
Missing System	9483	80.1		
Total	11832	100.0		

47% did not participate

52% visited a fair, zoo, or amusement park

Should this be a 0,1 variable?

# Missing Values for Different Cases

2005 HOUSEHOLD INCOME RECODES

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1. \$0-\$4,999	166	1.4	1.4	1.4
2. \$5,000-\$14,999	787	6.7	6.7	8.1
3. \$15,000-\$24,999	889	7.5	7.5	15.6
4. \$25,000-\$34,999	1104	9.3	9.3	24.9
5. \$35,000-\$49,999	1703	14.4	14.4	39.3
6. \$50,000-\$74,999	2702	22.8	22.8	62.1
7. \$75,000-\$99,999	1813	15.3	15.3	77.5
8. \$100,000-\$149,999	1748	14.8	14.8	92.2
9. \$150,000 AND OVER	920	7.8	7.8	100.0
Total	11832	100.0	100.0	

VISITED A ZOO, FAIR OR AMUSEMENT PARK

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid DID NOT PARTICIPATE	1104	9.3	47.0	47.0
1 time	493	4.2	21.0	68.0
2 times	328	2.8	14.0	81.9
3	145	1.2	6.2	88.1
4	77	.7	3.3	91.4
5	87	.7	3.7	95.1
6	23	.2	1.0	96.1
7	10	.1	.4	96.5
8	3	.0	.1	96.6
9	2	.0	.1	96.7
10	37	.3	1.6	98.3
12	11	.1	.5	98.8
15	4	.0	.2	98.9
16	1	.0	.0	99.0
18	1	.0	.0	99.0
20	14	.1	.6	99.6
24	1	.0	.0	99.7
25	2	.0	.1	99.7
30	1	.0	.0	99.8
45	1	.0	.0	99.8
50	1	.0	.0	99.9
52	1	.0	.0	99.9
60	1	.0	.0	100.0
200	1	.0	.0	100.0
Total	2349	19.9	100.0	
Missing System	9483	80.1		
Total	11832	100.0		

# Writing Up Results

- Abstract or Executive Summary
- Problem Statement
- Methods and Procedures
- Error Structure
- Findings
- Implications