# Missing Data

Author: Nicholas G Reich

*This material is part of the* **statsTeachR** *project*

# Today's Lecture

- Types of missing data
- Describing your missing data
- Multiple imputation

# Types of Missing Data

Missing Completely at Random (MCAR)

- 

Missing at Random (MAR)

- 

Missing Not at Random (MNAR)

-

# Multiple imputation results

**Regression coefficients from five imputed data sets**

| Data set | Estimated parameter | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ |
|---|---|---|---|---|---|---|---|
| 1 | Coefficient | -11.535 | -2.780 | 1.029 | -.031 | -0.359 | 0.572 |
| | Variance | 43.204 | 3.323 | 0.013 | 0.013 | 0.013 | 0.012 |
| 2 | Coefficient | -11.501 | -4.149 | 1.040 | -0.093 | -0.583 | 0.876 |
| | Variance | 40.488 | 2.680 | 0.010 | 0.009 | 0.009 | 0.007 |
| 3 | Coefficient | -10.141 | -5.038 | 0.766 | 0.123 | -0.252 | 0.625 |
| | Variance. | 42.055 | 3.301 | 0.010 | 0.010 | 0.010 | 0.009 |
| 4 | Coefficient | -11.533 | -6.920 | 0.870 | 0.084 | -0.458 | 0.815 |
| | Variance | 28.751 | 1.796 | 0.081 | 0.007 | 0.007 | 0.007 |
| 5 | Coefficient | -14.586 | -1.115 | 0.718 | 0.050 | -0.373 | 0.814 |
| | Variance | 32.856 | 2.362 | 0.009 | 0.009 | 0.009 | 0.008 |
| | Mean $b_i$ | -11.859 | -4.000 | 0.885 | 0.027 | -0.405 | 0.740 |
| | Mean Var.($\overline{W}$) | 37.471 | 2.692 | 0.025 | 0.010 | 0.010 | 0.009 |
| | Var. of $b_i$ (B) | 2.682 | 4.859 | 0.022 | 0.008 | 0.015 | 0.018 |
| | T | | | | | | |
| | $\sqrt{T}$ | 40.69 | 8.523 | 0.051 | 0.020 | 0.028 | 0.031 |
| | | 6.379 | 2.919 | 0.226 | 0.141 | 0.167 | 0.176 |
| | $t$ | -1.859 | -1.370 | 3.916* | 0.191 | 2.425* | 4.204* |

\* $p < .05$  "Var." refers to the squared standard error of the coefficient.

DC Howell, Treatment of Missing Data – Part II.

# Multiple imputation results

**Regression coefficients from five imputed data sets**

| Data set | Estimated parameter | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ |
|---|---|---|---|---|---|---|---|
| 1 | Coefficient | -11.535 | -2.780 | 1.029 | -.031 | -0.359 | 0.572 |
|   | Variance | 43.204 | 3.323 | 0.013 | 0.013 | 0.013 | 0.012 |
| 2 | Coefficient | -11.501 | -4.149 | 1.040 | -0.093 | -0.583 | 0.876 |
|   | Variance | 40.488 | 2.680 | 0.010 | 0.009 | 0.009 | 0.007 |
| 3 | Coefficient | -10.141 | -5.038 | 0.766 | 0.123 | -0.252 | 0.625 |
|   | Variance. | 42.055 | 3.301 | 0.010 | 0.010 | 0.010 | 0.009 |
| 4 | Coefficient | -11.533 | -6.920 | 0.870 | 0.084 | -0.458 | 0.815 |
|   | Variance | 28.751 | 1.796 | 0.081 | 0.007 | 0.007 | 0.007 |
| 5 | Coefficient | -14.586 | -1.115 | 0.718 | 0.050 | -0.373 | 0.814 |
|   | Variance | 32.856 | 2.362 | 0.009 | 0.009 | 0.009 | 0.008 |
|   | Mean $b_i$ | -11.859 | -4.000 | 0.885 | 0.027 | -0.405 | 0.740 |
|   | Mean Var.($\bar{W}$) | 37.471 | 2.692 | 0.025 | 0.010 | 0.010 | 0.009 |
|   | Var. of $b_i$ (B) | 2.682 | 4.859 | 0.022 | 0.008 | 0.015 | 0.018 |
|   | T | 40.69 | 8.523 | 0.051 | 0.020 | 0.028 | 0.031 |
|   | $\sqrt{T}$ | 6.379 | 2.919 | 0.226 | 0.141 | 0.167 | 0.176 |
|   | t | -1.859 | -1.370 | 3.916* | 0.191 | 2.425* | 4.204* |

* $p < .05$  "Var." refers to the squared standard error of the coefficient.

DC Howell, Treatment of Missing Data – Part II.