Causal Inference and Its Limits

Professor Allen Hicken

جامعة قطر
QATAR UNIVERSITY

SESRI
معهد البحوث الاجتماعية والاقتصادية المسحية
Social & Economic Survey Research Institute

What is Causality?

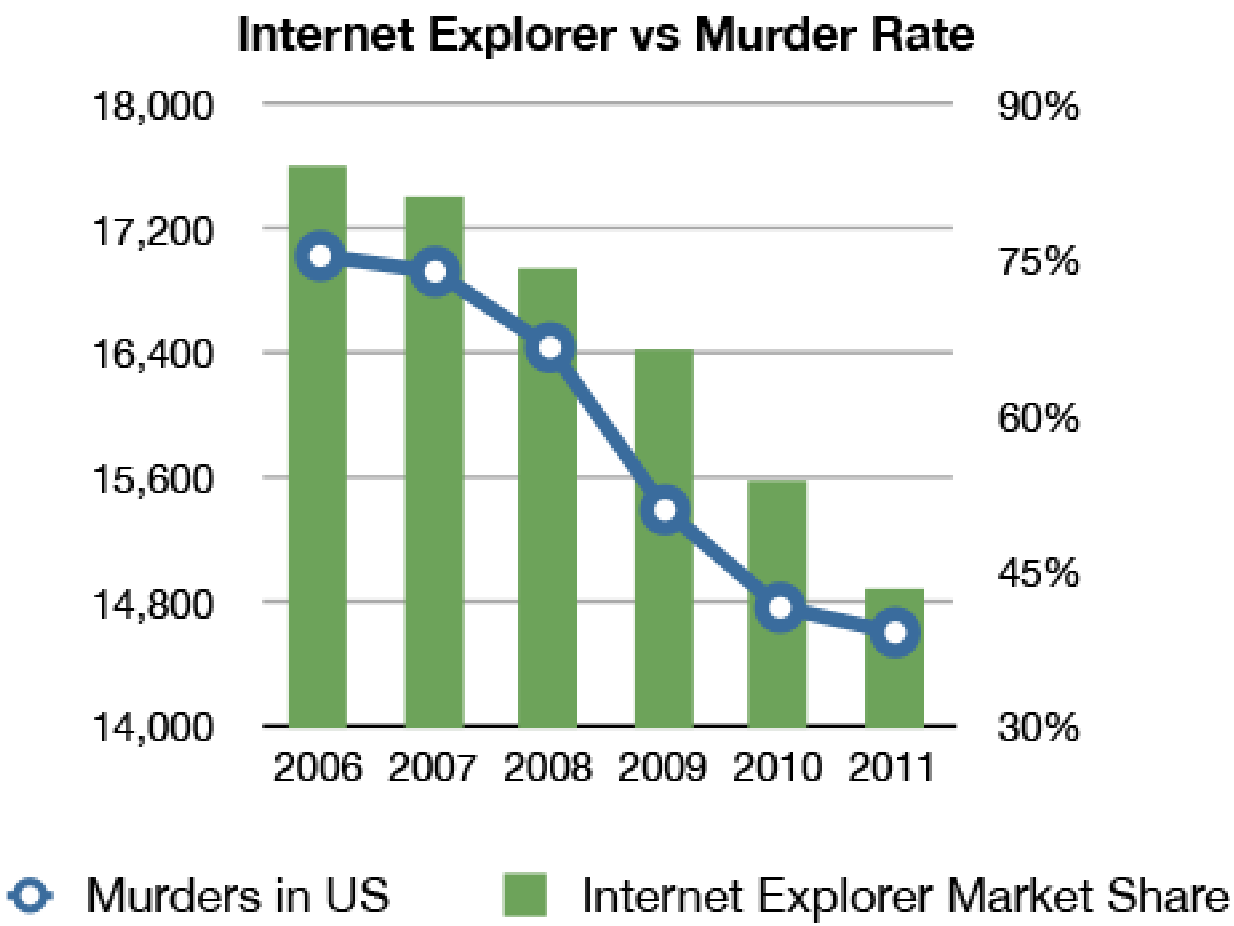
- Causality versus causal inference.
- Causality: X is a cause of Y; Y is an effect of X (X is the *treatment*, Y is the effect)
- Causal Inference: Can we infer X→Y from our sample?

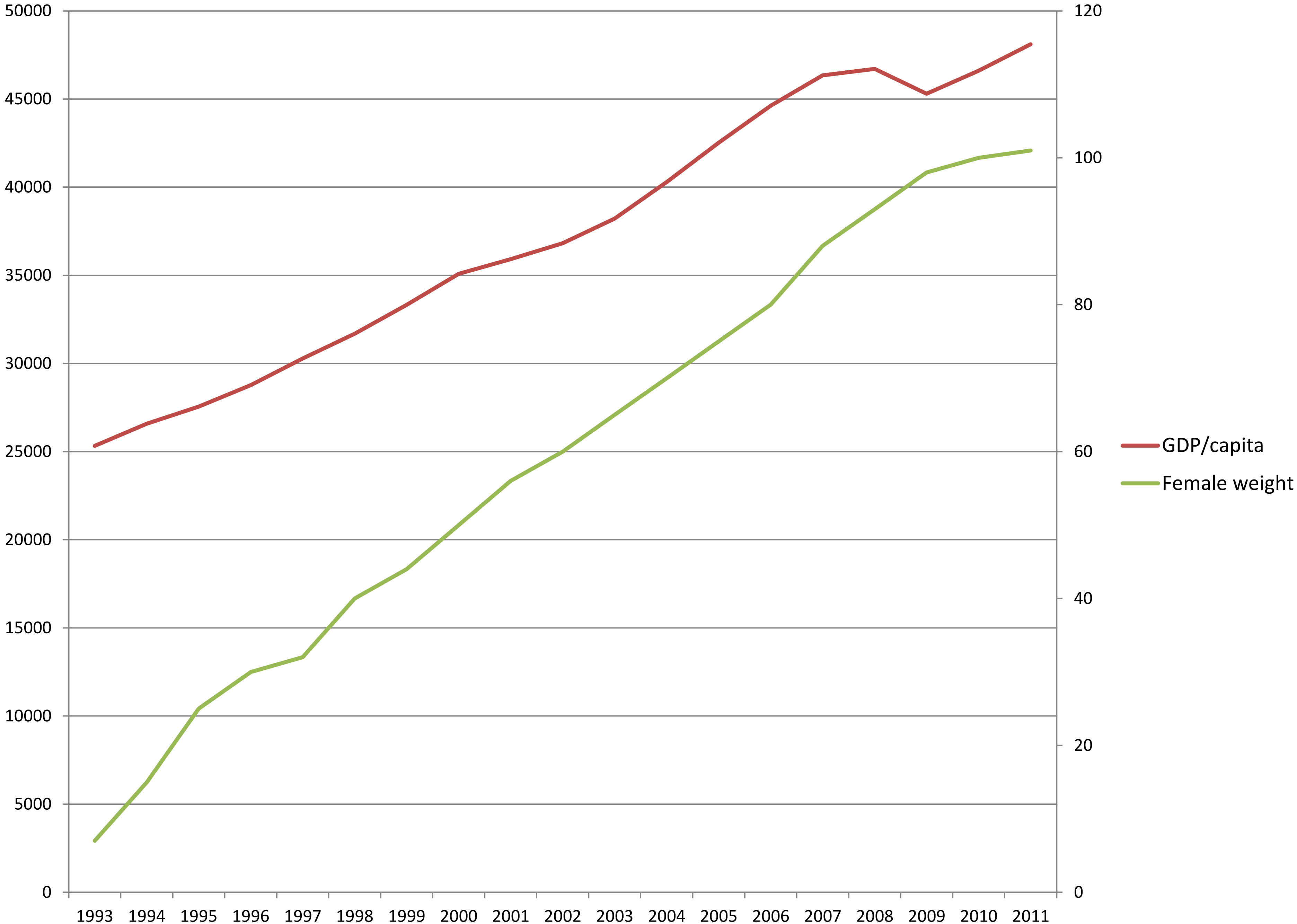
Criteria for Establishing Causality

1. Correlation (Association)
2. Temporal Ordering
3. Theory (Causal Mechanisms and Counterfactuals)
4. Isolation (Rule out Confounds and Alternative Explanations)

Criterion #1. Correlation

- Two variables are “correlated” when changes in one variable occur together with changes in the other (Louise White)
  - Correlation is roughly synonymous with association and co-variance.
  - A correlation between two variables can be positive or negative.


Internet Explorer vs Murder Rate
18,000
17,200
16,400
15,600
14,800
14,000
90%
75%
60%
45%
30%
2006
2007
2008
2009
2010
2011
Murders in US
Internet Explorer Market Share



50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
120
100
80
60
40
20
0
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
GDP/capita
Female weight

60000
50000
40000
30000
20000
10000
0
40
35
30
25
20
15
10
5
0
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
GDP/capita
Median % Obese in U.S.

Establishing Causality

- We observe: X correlates with Y

1. Causation is not involved at all
2. There is a causal link
3. Confounding (omitted) variable (Z) causes both X and Y

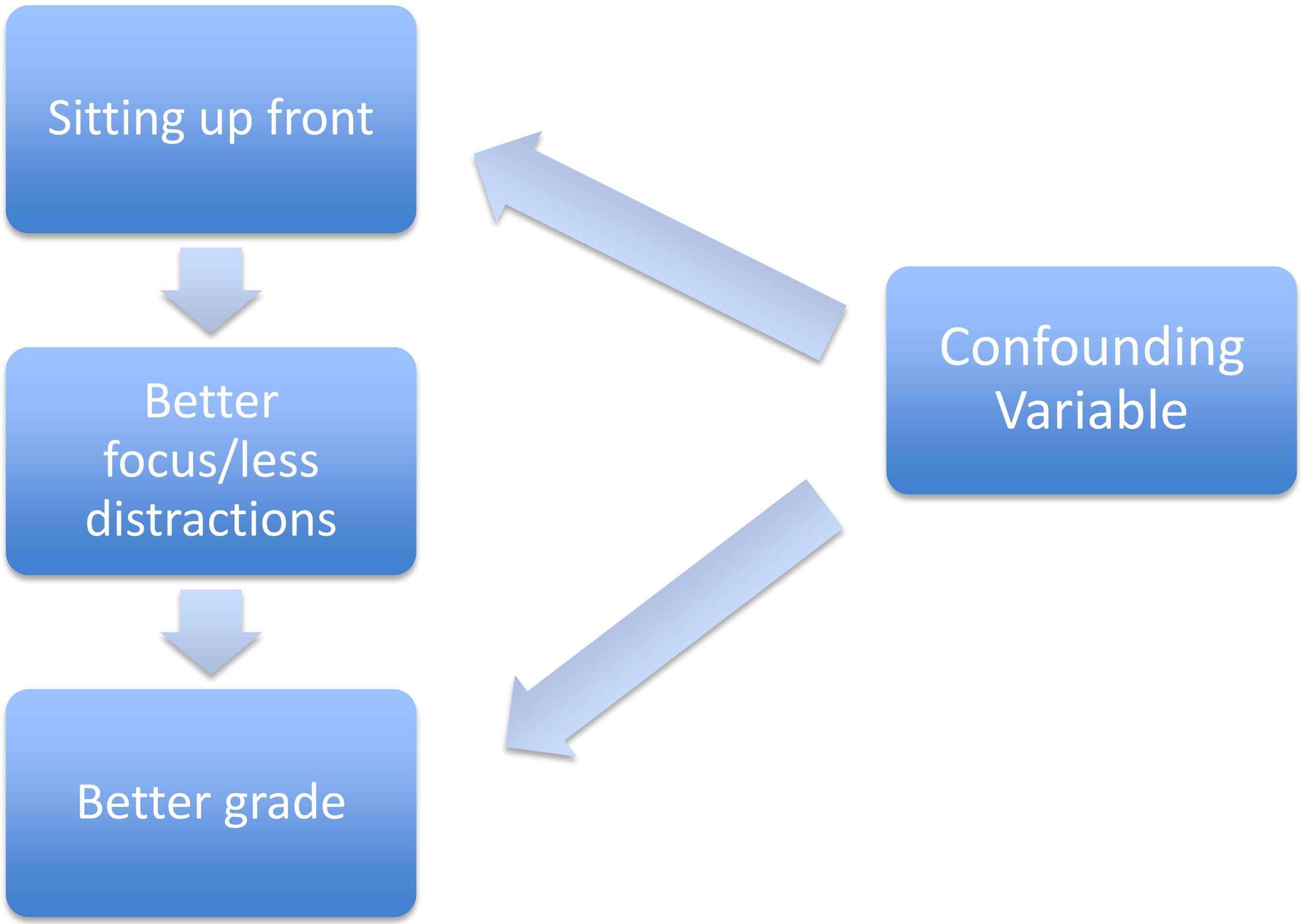
Look for Alternative Explanations

Criterion #2 Temporal Ordering

- The hypothesized cause (IV) must come before the effect (DV).
  - Rise in GDP/capita precedes rise in obesity in U.S.
  - Students decide whether or not to sit in the front of class before the get their final grade.
    - Or do they?
  - Social science has lots of tricky “chicken-and-egg” situations.

Criterion #3 Causal Mechanism

- You have to be able to tell a plausible story that connects the cause (IV) to the effect (DV)
  - This story often includes an "intervening variable" that gets us from the cause to the effect
  - Students who sit up front are able to hear better, see better, better comprehend the lecture, and are less tempted by distractions (plausible story)
  - Students who sit up front of the class bask in my aura and absorb more of my genius by just being close to me (not plausible)


Sitting up front
Better focus/less distractions
Better grade
Confounding Variable

Criterion #4 Isolation (Rule Out Alternative Explanations and Confounds)

- If there is a confounding variable that is causally prior to both a cause (IV) and an effect (DV), then the correlation we observe between the cause and the effect may be spurious.

Criterion #4 Isolation (Rule Out Confounds)

- If there is a confounding variable that is causally prior to both an cause and an effect, then the correlation we observe between the cause and the effect may be spurious.

- When it comes to causal inference this is perhaps the biggest challenge for non-experimental researchers.

The Fundamental Problem of Causal Inference

- Problem. We cannot rerun history to see whether changing the value of an independent variable would have changed the value of the dependent variable.

- Solution #1. Give up.

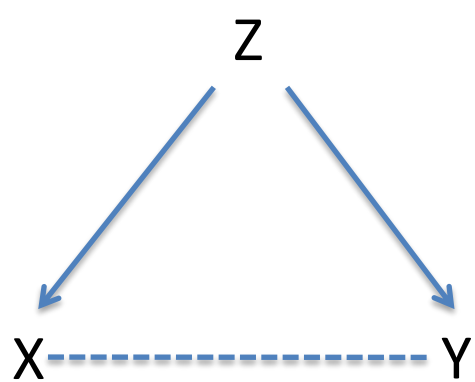
The Fundamental Problem of Causal Inference

- Solution #2. Design your research in a way that comes as close as possible to rerunning history.
  - Observe the effects of changes in one independent variable when all other independent variables remain the same, or
  - Measure other independent variables, then use statistical techniques to hold them constant.

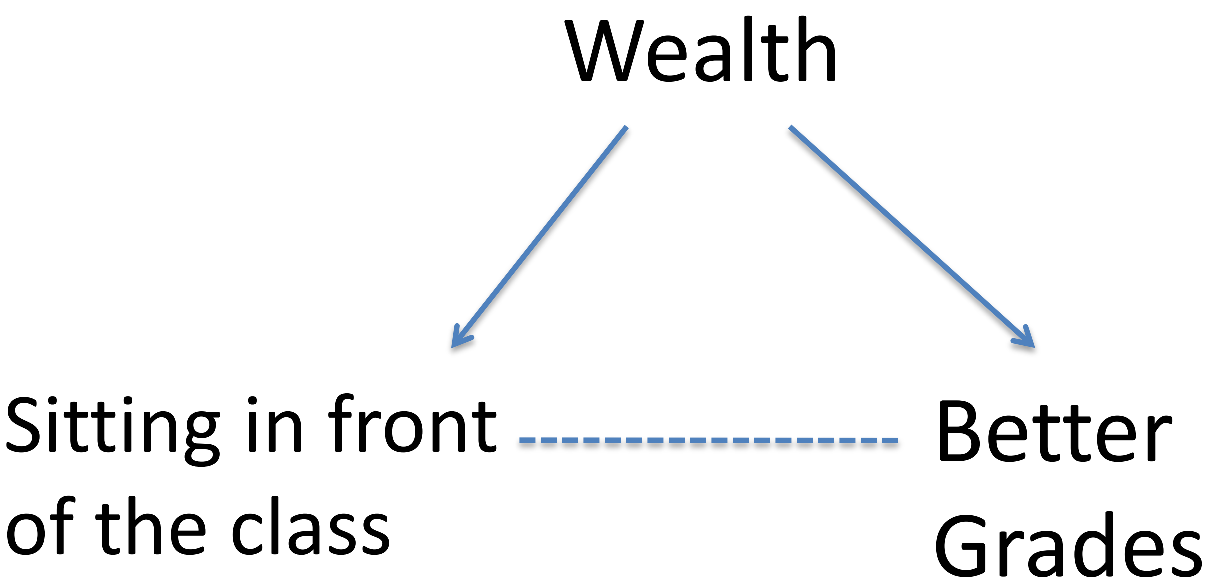
Establishing Causality

- We observe: X correlates with Y

1. Causation is not involved at all
2. There is a causal link
3. Confounding variable causes both X and Y

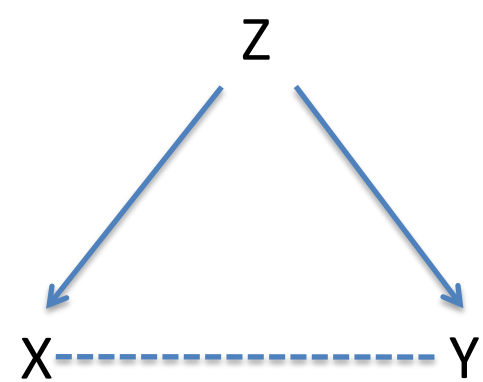


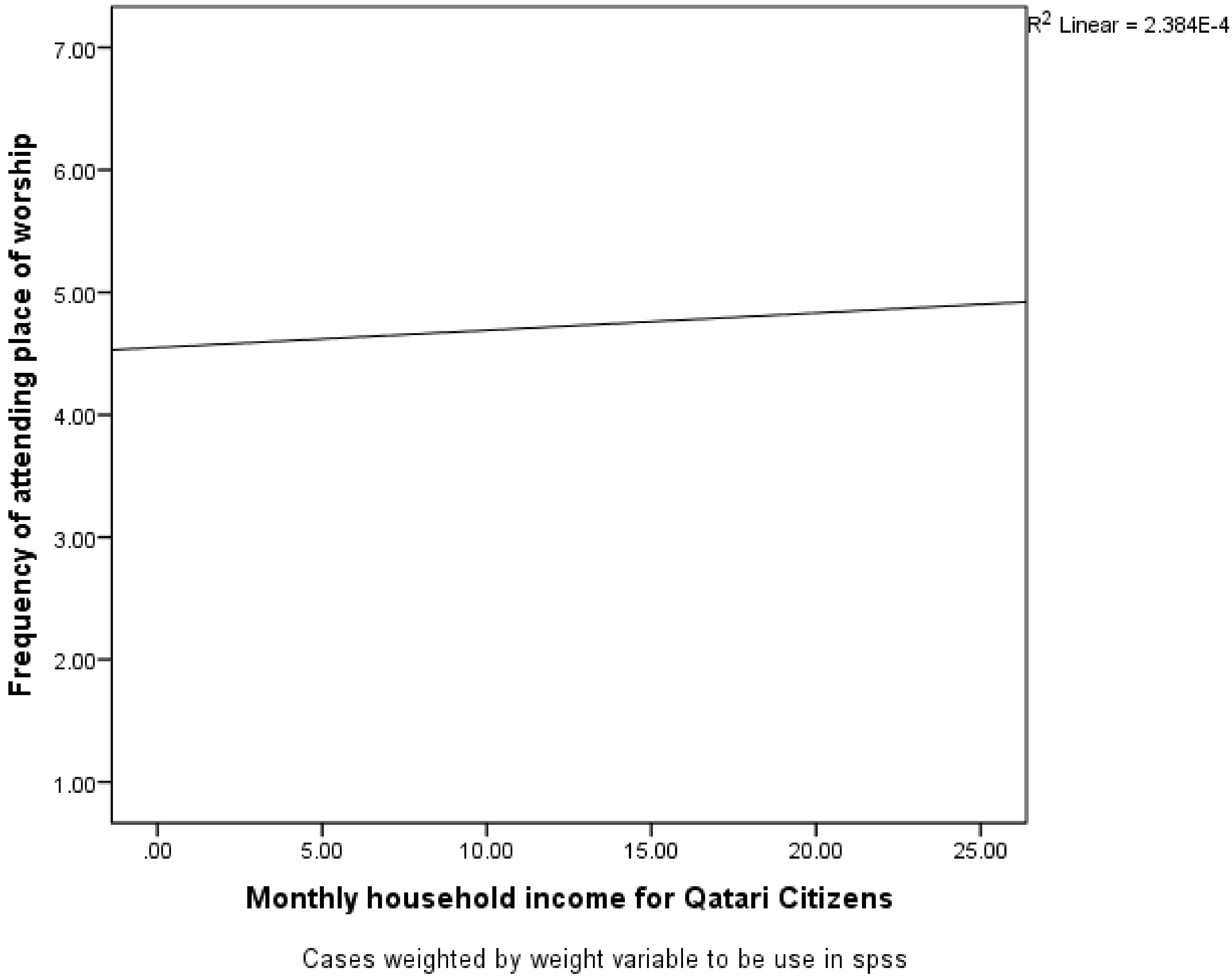
Example



Dealing with Confounding Variables

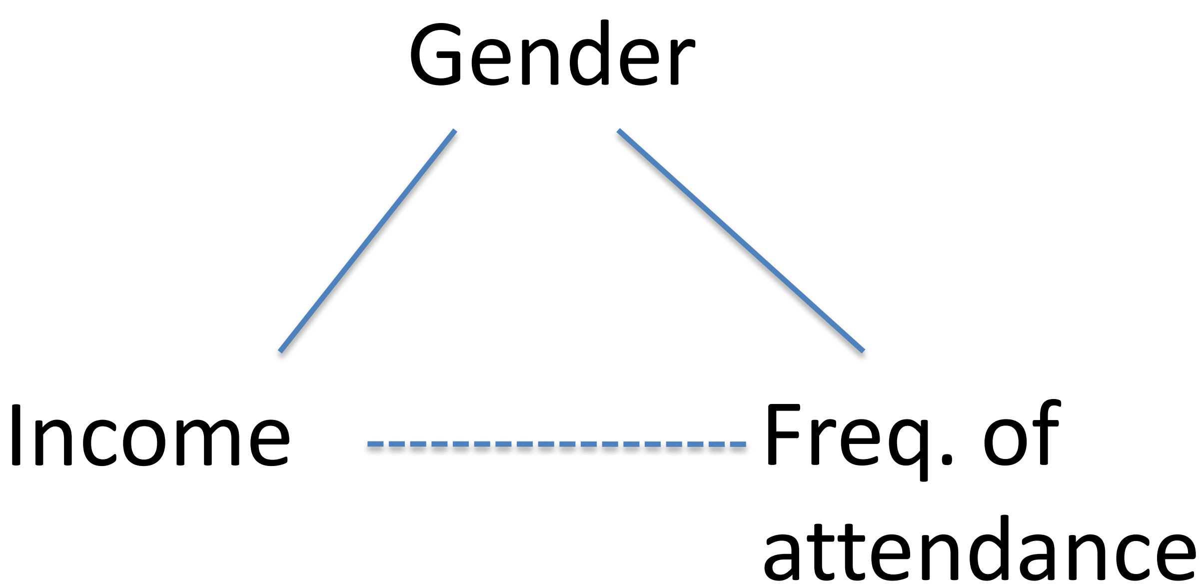
- Control variables
  - Holding potential confounding variables constant
- 3 possible outcomes when control for Z
  - Spurious relationship
  - Additive relationship
  - Interactive relationship




R² Linear = 2.384E-4
7.00
6.00
5.00
4.00
3.00
2.00
1.00
Frequency of attending place of worship
.00
5.00
10.00
15.00
20.00
25.00
Monthly household income for Qatari Citizens
Cases weighted by weight variable to be use in spss

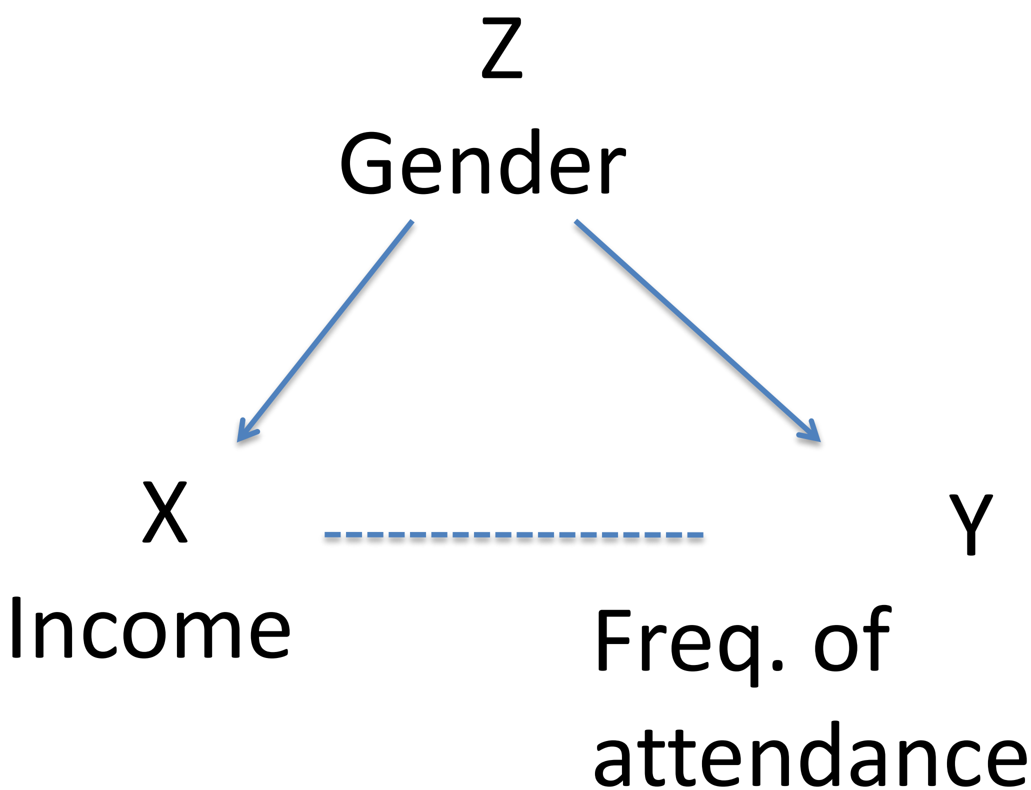
Examples

- Relationship between income and religiosity
  - Income → higher attendance at religious services
  - What could be a confounding or control variable?

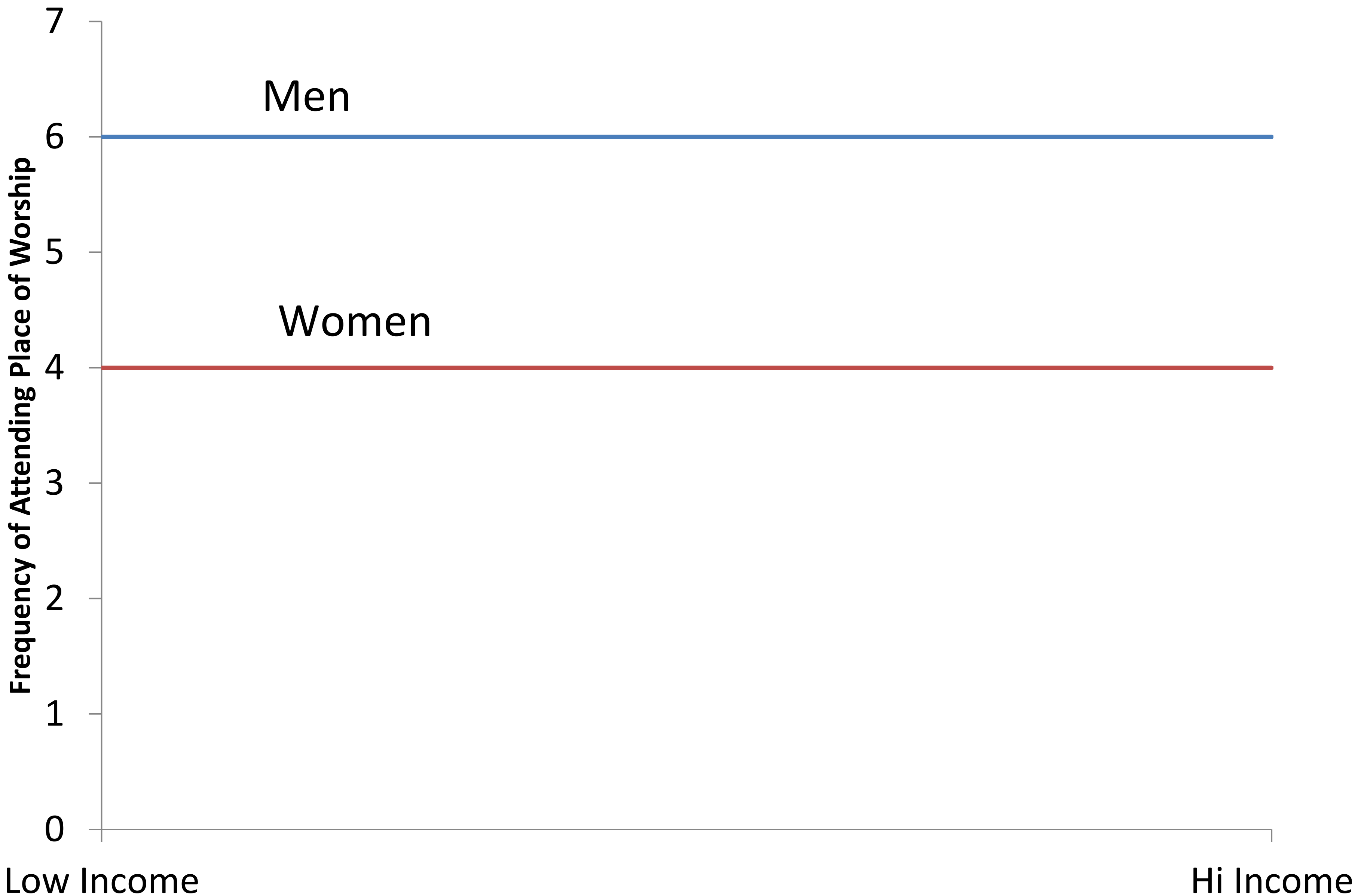


Spurious relationship

- After holding Z constant the causal connection between X and Y disappears

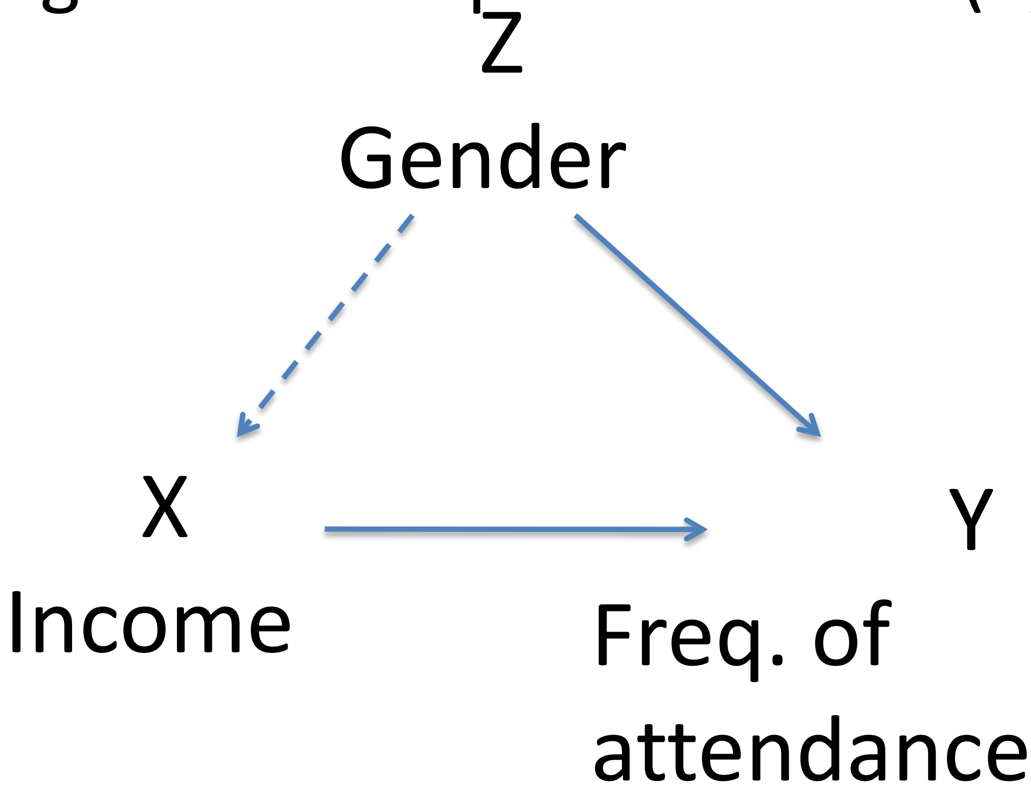


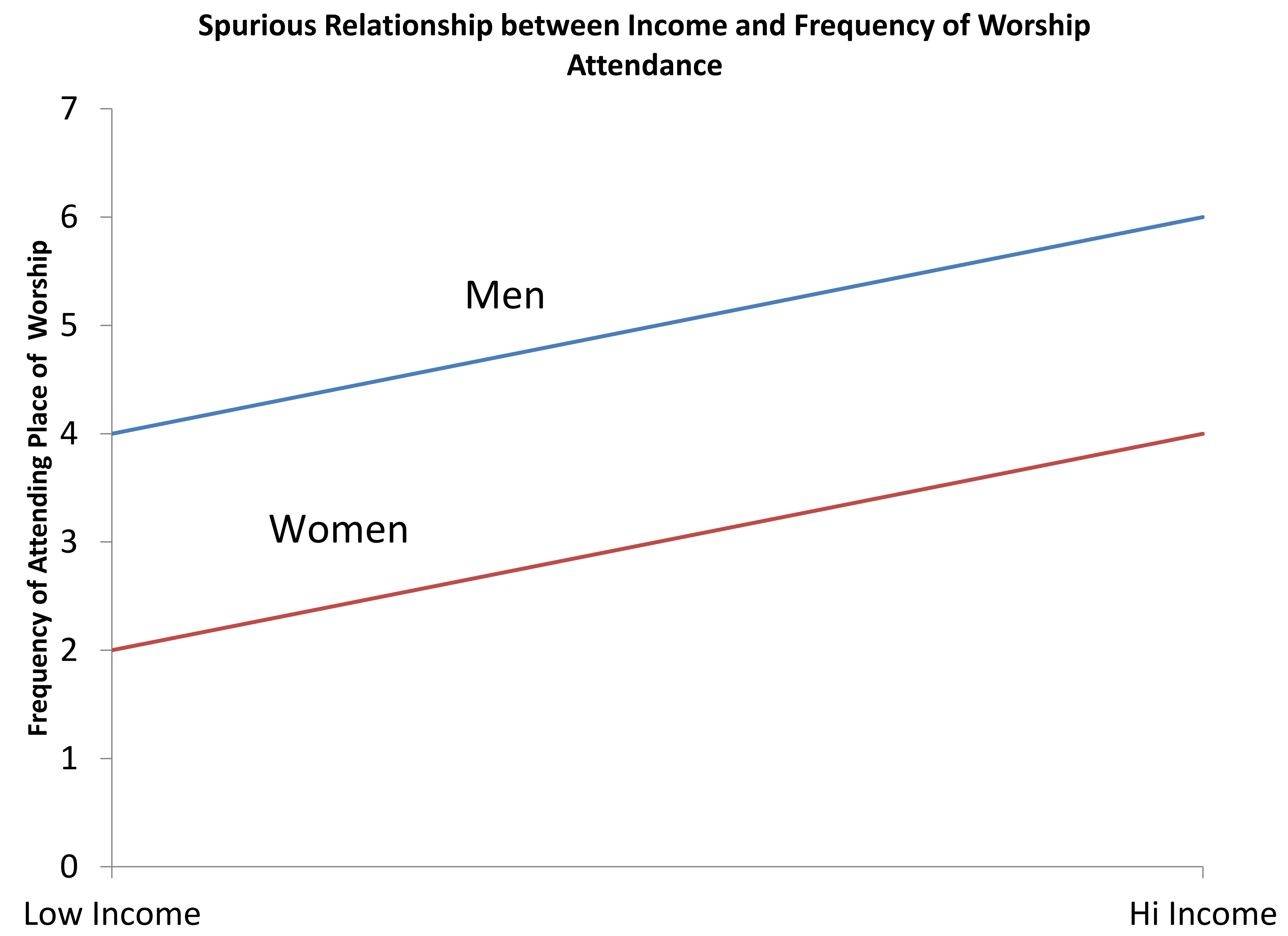
Spurious Relationship between Income and Frequency of Worship Attendance



Additive Relationship

- The control variable (Z) has a weak or non-existent relationship with the IV (X) and a strong relationship with the DV (Z).

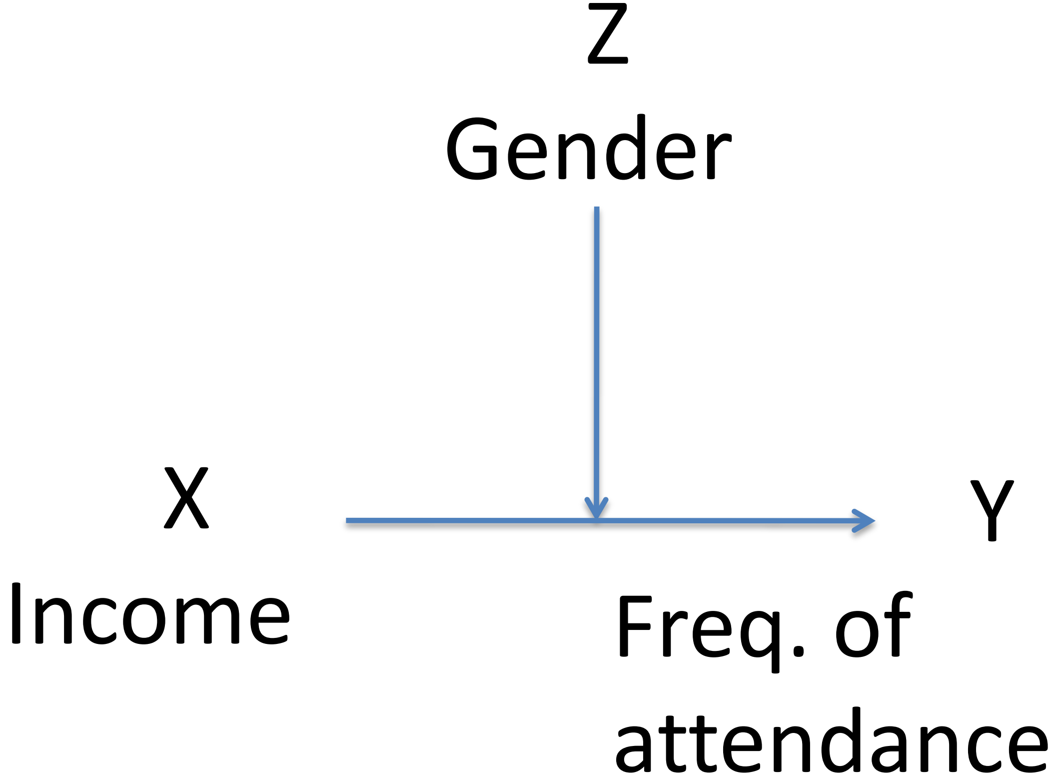


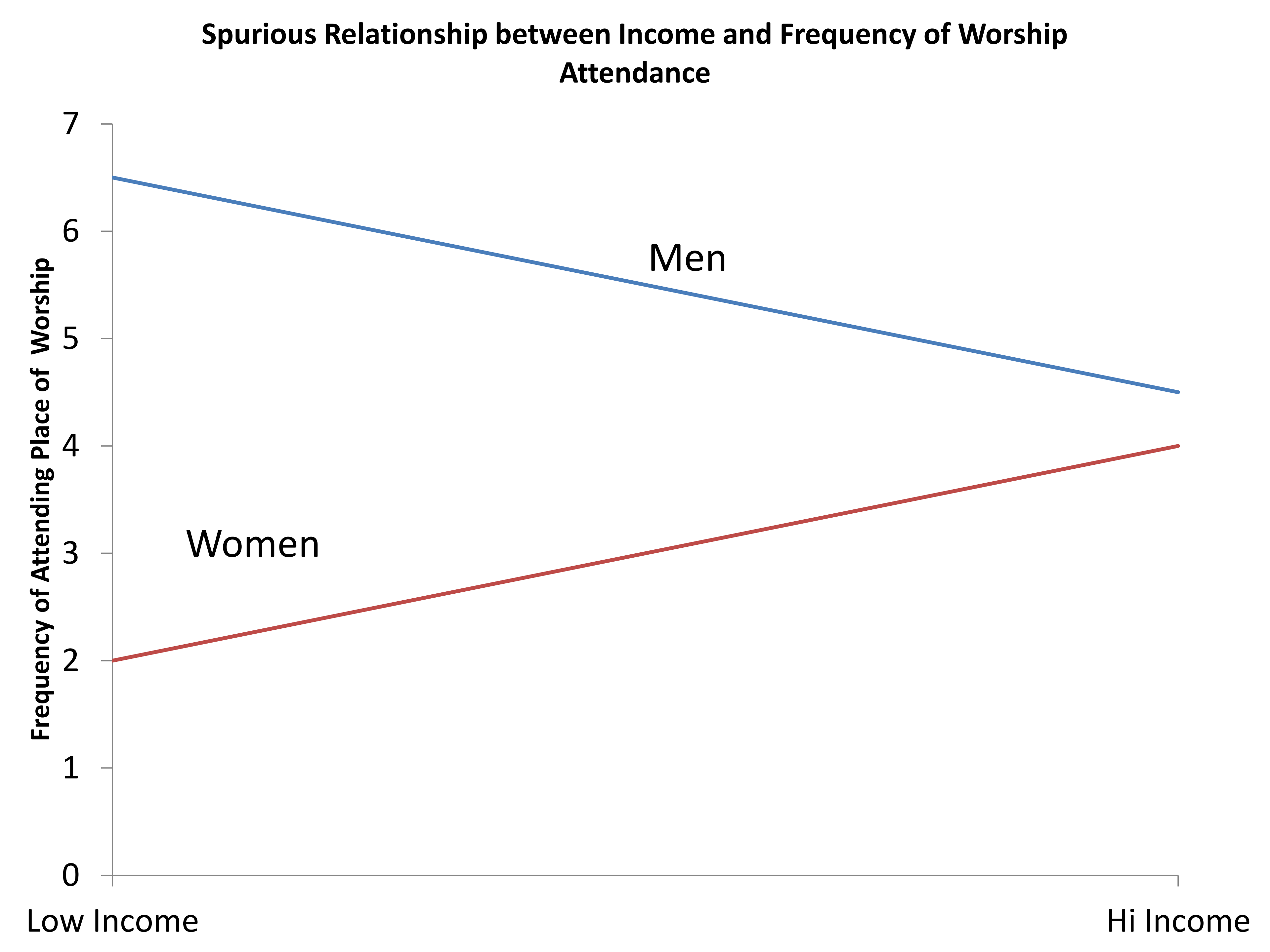


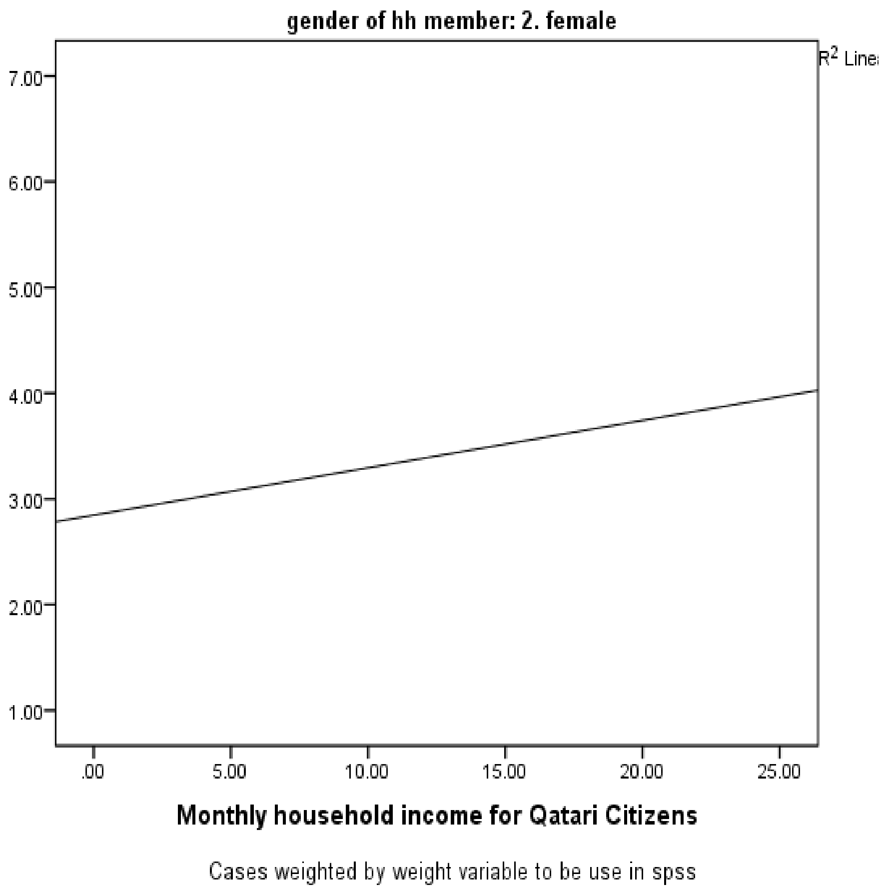
Spurious Relationship between Income and Frequency of Worship Attendance
Frequency of Attending Place of Worship
7
6
5
4
3
2
1
0
Men
Women
Low Income
Hi Income

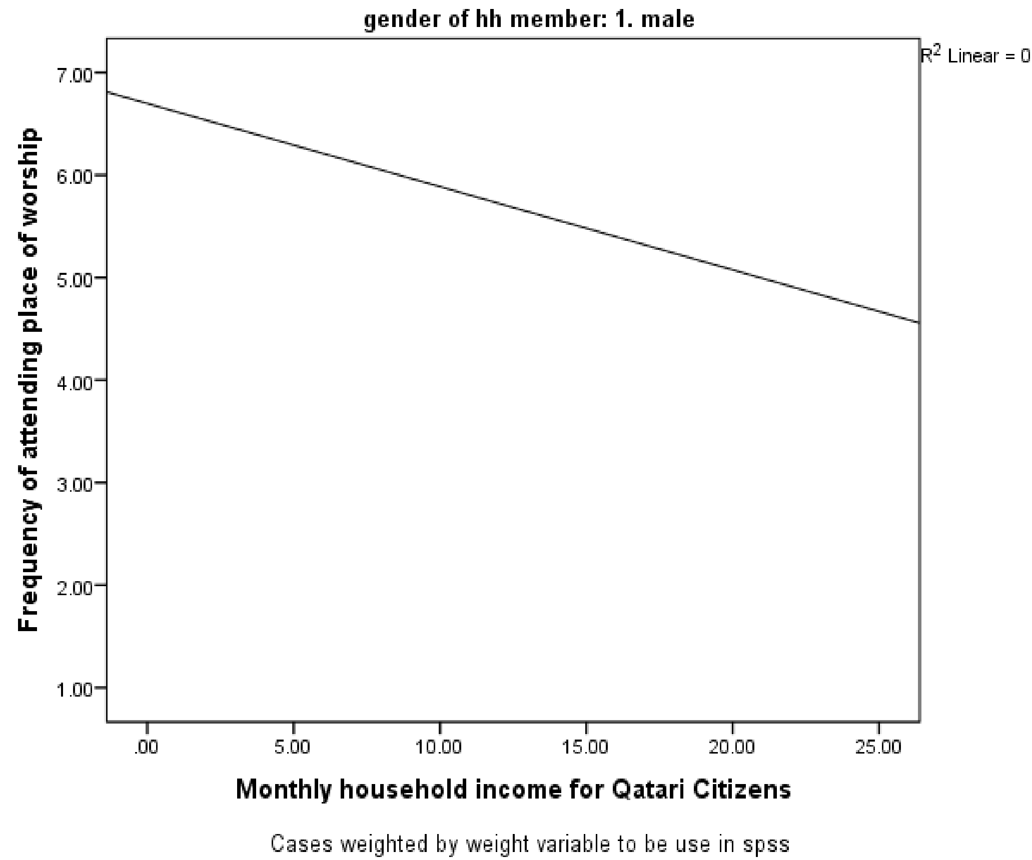
Interactive Relationships

- The relationship between the IV (X) and DV (Y) depends on the value of the control variable (Z)




Spurious Relationship between Income and Frequency of Worship Attendance
Frequency of Attending Place of Worship
7
6
5
4
3
2
1
0
Men
Women
Low Income
Hi Income


gender of hh member: 2. female
R2 Line
7.00
6.00
5.00
4.00
3.00
2.00
1.00
.00
5.00
10.00
15.00
20.00
25.00
Monthly household income for Qatari Citizens
Cases weighted by weight variable to be use in spss


gender of hh member: 1. male
R2 Linear = 0
Frequency of attending place of worship
7.00
6.00
5.00
4.00
3.00
2.00
1.00
.00
5.00
10.00
15.00
20.00
25.00
Monthly household income for Qatari Citizens
Cases weighted by weight variable to be use in spss

Exercises

Each of following conclusions is based on a relationship between X and Y that could be spurious. For each one: (i) identify a plausible confounding variable (Z) for which you would ideally control, (ii) Briefly describe how Z might be affecting the relationship between X and Y.

1. In Great Britain, the level of ice cream sales (X) and drowning deaths (Y) are strongly related; as sales go up, so do deaths from drowning. Conclusion: To save lives we should prohibit ice cream sales.
2. Car color (X) and accident rates (Y) are linked: Red cars are more likely to be involved in accidents than are non-red cars. Conclusion: If red cars are banned, the accident rate will drop.
3. Women's education (X) and divorce rates (Y) are correlated: more educated women have a higher divorce rate than less-educated women. Conclusion: Education causes divorce.

Adapted from Pollock 2009

Exercise

- In groups of 3-4 review your hypotheses from yesterday
- What potential confounding variables can you identify as a group?
- How might they effect the relationship between your "X" and "Y" variables?

Sources

- Philip H. Pollock III. 2009. *The Essentials of Political Analysis.* CQ Press.
- W. Philips Shively. *The Craft of Political Research*. Pearson Prentice Hall.