

Cohabitation and Divorce

INTRODUCTION

This module is designed to illustrate the effects of selection bias on the observed relationship between premarital cohabitation and later divorce. It also serves as a review of key methodological concepts introduced in the first part of the course.

LEARNING OBJECTIVES

Skills

- Learning about survey methodology and sampling methods
- Using software to access and analyze census data
- Employing control variables
- Learning how to construct, read, and interpret bivariate tables displaying frequencies and percentages
- Review the Following Methodological Concept:
 - Randomization vs. Random Sampling
 - Causality vs. Spuriousness
 - Self-Selection Bias
 - Social Measurement
 - Hypothesis Formation
 - Independent and Dependent Variables

Substance

- Review Substantive Material:
 - The Effects of Cohabitation on Marital Success
 - Selection vs. Experience Effects
 - Unconventionality Hypothesis
 - Risk Differences Hypothesis
 - Deviance Hypothesis

DATASET

The data sets you will be using - [OCEDPV9](#) - come from the U.S. Census 5% Public Use Micro-data Set (PUMS) and consist of a random sample of all individuals in the U.S. between the ages of 15 and 44 in 2000. The variables included in each of the data sets are as follows:

- MARSTUS: The current marital status of the respondent – Never Married (Nmarried), Living Together (Livingtg), Married (Married), Divorced (Divorced), Separated (Separatd), Widowed (Widowed)
- RACELAT: A combination of one's race and ethnicity – Non-Latino White (NLWhite), Black (Black), Latino/a (Latino), Asian (Asian), Native American (AmIndian), Non-Latino Other (NLOther)
- SEX: Household member's biological sex – Male (Male), Female (Female)

- AGE: Household member's age group - 16-24, 25-34, 35-44, 45-54, 55-64, 65+
- POV: Indicates whether or not the total family income was below the poverty line for the year (Poverty) or above the poverty line (NonPov).
- OCCUP: Classification of occupation into one of five social class groups – Higher Status White Collar (TopWC), Other White Collar (OtrWC), Service (Service), Higher Status Blue Collar (TopBC), and Other Blue Collar (OtrBC)
- EDUC: Classification of highest level of education completed into four categories – Less Than High School (LTHS), High School (HSgrad), Some College (SomeColl), and College Graduate (Collgrad)
- KID: Whether or not the respondent has a child who is living with him/her – No, Yes

METHODOLOGICAL AND THEORETICAL REVIEW

Before you begin, complete the following tasks:

- Review the following methodological concepts and conceptual distinctions:
 - Random Sampling vs. Randomization
 - Selection Bias
 - Causality vs. Spuriousness
 - Statistical Control
- Review the distinction between selection and experience as possible explanations for the negative relationship between cohabitation and marital success.
- Discuss the difference between the unconventionality hypothesis and the risk differences hypothesis and explain how these two hypotheses reflect the methodological effects of selection as a reason for the negative relationship between cohabitation and marital success. What is it about those who choose to cohabit that puts them at risk of divorce regardless of whether or not they cohabited before marriage?
- Discuss the deviance hypothesis and explain how this hypothesis reflects the causal effect of experience as a reason for the negative relationship between cohabitation and marital success. What is it about the experience of cohabitation that increases the risk of divorce regardless of one's prior situation?

THEORETICAL DEVELOPMENT

Having reviewed the necessary theories and concepts, we now need to develop the problem, identify measures to use in an analysis of the problem, and specify hypotheses to be tested in the analysis.

PROBLEM STATEMENT

In this exercise we will not be able to test whether it is selection or experience that accounts for the negative effect of cohabitation on marital success because we do not have longitudinal data to show what happens to cohabiters after they get married. We can, however, examine whether or not important elements of the logic of the selection explanation have empirical validity.

Recall from earlier in the course the logic of a spurious relationship. In a spurious relationship, two variables appear to be related because they are both being caused by a third variable. In the example being used here, people who cohabit tend to be the same people who have a lower chance of marital success. Is this relationship causal? That is, do people who have a lack of marital success do so because the experience of cohabiting changes them or the conditions under which they live their later marital lives? Or, is the relationship due to the fact that the individuals who cohabit and have low marital success also have some third set of characteristics that have caused them to cohabit and independently cause them to experience low marital success regardless of whether or not they cohabited? That is, is the relationship spurious?

For example, low SES individuals may choose to cohabit in greater numbers than high SES individuals because they can't afford to get married (i.e., they self-select into cohabitation and non-cohabitation groups in a biased way). Low SES individuals also have a lower likelihood of marital success because of higher levels of stress in their environments. Thus, you are likely to find that the people who cohabited before marriage are the same people who experience low marital success even though the cohabitation experience changes neither them nor the conditions of their lives in any causal manner. If this is the case, then the relationship is spurious. This is the risk differences hypothesis and reflects the effect of selection as an explanation for the negative effect between cohabitation and marital success. The difference between selection and experience as explanations for the negative relationship between cohabitation and marital success is that selection says that the relationship is spurious, whereas experience says that the relationship is causal.

In order for selection to be a source of spuriousness in the relationship between cohabitation and marital success, four conditions must be met: (1) the extraneous variable (in this case the selection variable) must come prior to both; (2) the extraneous variable must predict the likelihood of cohabitation; (3) the extraneous variable must decrease the likelihood of marital success; and (4) the effect of cohabitation on marital success must reduce or disappear when you control for the extraneous variable. Diagrammatically, the first three of these logical pre-requisites look as follows:

- Extraneous
- Variable
- Cohabitation Marital Success

Notice that in this diagram the causal arrows go from the extraneous variable to both cohabitation and marital success and the non-causal relationship line (i.e., the line without arrows) connects these latter two variables. In a spurious relationship the variables covary but do not exert any causal force on each other (thus the lack of an arrow). Rather, they are both the outcome of a common causal force being exerted by a third variable.

In this data analysis exercise, your task will be to examine the U.S. Census data to see if there is any reason to believe that selection could be a factor in explaining the negative relationship between cohabitation and marital success as spurious. To do so, you will examine the empirical validity of the assumption that the same characteristics that predict cohabitation also predict divorce (i.e., is there any evidence to suggest that the arrows in the above diagram are valid). We will not be able to assess the causal priority of the extraneous variables nor will we be able to examine the relationship between cohabitation and marital success under conditions of the extraneous variables (i.e., statistically control for selection). We will, however, be able to satisfy the other two requirements for demonstrating a spurious effect. Thus, you will be doing two sets of analyses – one to examine the predictors of being in a cohabiting vs. a married relationship, and one to examine the predictors of being separated or divorced vs. still married. The goal is to see if the predictors of cohabitation are the same as the predictors of divorce. If they are, then it is possible that the relationship between cohabitation and divorce is spurious.

MEASUREMENT

In order to test for the effects of unconventionality and risk exposure on cohabitation and divorce, we will first need to identify potential indicators for each of these concepts. Ideally we would identify or construct the best measures for each concept and then conduct our own survey using those measures. As is frequently the case, however, social scientists must rely on data that has already been collected to test their hypotheses. The data that we have to use for this analysis comes from the U.S. Census and was described above. Indicators have already been chosen for you with respect to cohabitation, divorce, and unconventionality. You will be asked at the end of this section to identify, specify, and provide a rationale for, an indicator to measure risk exposure.

MEASURING COHABITATION, MARRIAGE, AND DIVORCE

These concepts will be measured using two modified versions of the marital status variable (MARITAL). The variable MARITAL consists of responses to two items in the Census – one asking for one's current marital status (never

married, married, divorced, separated, or widowed) and one asking about one's relationship to the head of the household (with "living together" being a category that is distinguished from "roommate"). In the analysis of the predictors of cohabitation vs. marriage we are only interested in those who are making choices with respect to forming a relationship. Therefore, you will need to modify the marital status variable prior to these analyses to omit the categories of never married, divorced, separated, and widowed. This will leave you with a variable with only two categories: married vs. living together. For those analyses looking at predictors of divorce, we are only interested in those who are either currently married vs. those who have had a previous marriage that ended in divorce or separation. For these analyses you will need to modify the marital status variable differently. First you will have to omit the categories for never married and widowed. Next you will combine the categories of divorced and separated into a single category. As a result, you will be left with a variable with two categories: married vs. divorced/separated.

MEASURING UNCONVENTIONALITY

Everyone will use the same variable to test the validity of the unconventional hypothesis – race/ethnicity (RACELAT). The concept of unconventional is a difficult one to measure because the census data does not contain measures of peoples' attitudes. Instead, we have to infer peoples' attitudes about the institution of marriage and divorce based on background characteristics. Based on what we have learned about differences between race and ethnic groups in the United States, it could be argued that Non-Latino Whites and Asian-Americans are the most supportive of traditional family and therefore will be the least unconventional. African-Americans and Native-Americans, on the other hand, have demonstrated a much more flexible definition of the family that doesn't fit so clearly with the dominant institutional definition of the family as a heterosexual couple joined in a life-long legal marriage. Single parent forms are more prevalent among these two race/ethnic groups and due to both originating cultural differences as well as historical and contemporary patterns of discrimination they are more likely to live their lives outside of the dominant cultural norms. Thus, these two groups could be considered the most unconventional. Hispanics and other groups are likely to fall somewhere in the middle. Based on this reasoning, it is argued that race/ethnicity is a reasonable indicator of unconventional attitudes regarding marriage and family life. To examine the effects of unconventional on the likelihood of cohabitation and divorce you will first modify this variable to combine the six groups identified by the Census into three categories of unconventional as shown below.

- Unconventional Variable: Race/Ethnicity (RACELAT)
- Low Unconventional: Non-Latino White, Asian
- Medium Unconventional: Latino; Non-Latino Other
- High Unconventional: Black; Native American

MEASURING RISK EXPOSURE

- There are several variables in the data set that you can use to measure risk exposure to test the risk differences hypothesis. These include: Occupation, Education, Poverty Status, and Child Presence. Each individual in your group needs to choose a different one of these variables to use as an indicator of risk exposure. Indicate the variable that you will be using for your assignment.
- Now you need to provide a clear statement that indicates what categories under this variable represent low, medium and high levels of risk exposure. For example, if employment status was one of the variables you could use, you might specify that full-time employed are lowest risk, part-time employed are medium risk and unemployed are highest risk. If the variable you choose has only two categories (e.g., Poverty), then just indicate the high risk vs. low risk categories.
- Finally, provide a clear justification for why this is a reasonable measure of risk exposure to use in testing the risk differences hypothesis (similar to how I justified the unconventionality variable above).

SPECIFICATION OF HYPOTHESES

The final preliminary step that needs to be completed before doing the analysis is to specify what you expect to find in the analysis with respect to the variables that you will be using. This is called “formulating hypotheses”. Hypotheses always specify how you expect units (in this case, individuals) to vary together (positively or negatively) on two different variables (or sets of characteristics). For this exercise we will need two sets of hypotheses to test the assumptions of the selection explanations for the negative relationship between cohabitation and marital success. The first set is for the effect of unconventionality on the likelihood of cohabitation and the effect of unconventionality on marital success. The second set is for the effect of risk exposure on the likelihood of cohabitation and the effect of risk exposure on marital success. If you find that both effects are observed within a given set, then it is reasonable to think that the relationship between cohabitation and marital success could be due to selection and therefore spurious. To get you started, the unconventionality hypotheses are developed below. All you need to do is circle the appropriate word in the parentheses. You will then need to specify the risk differences hypotheses on your own.

Unconventionality Hypotheses: *Circle the appropriate direction for the change in the outcome variable in each hypothesis; and list the independent and dependent variables.*

1. Hypothesis 1 (Extraneous Variable Effects on the Likelihood of Cohabitation)
 - a. The more unconventional the racial/ethnic group one identifies with (with Non-Latino Whites and Asians being considered the least unconventional and Blacks and Native Americans being considered the most unconventional), the (greater / less) the likelihood of having a marital status of living together (vs. being married).
 - b. Independent Variable
 - c. Dependent Variable
 - d. Rationale: The more unconventional one's race/ethnic group is, the more likely it is that they will not follow the dominant cultural norms about marriage and family formation and be more willing to engage in cohabitation as an alternative to marriage.
2. Hypothesis 2 (Extraneous Variable Effects on the Likelihood of Divorce)
 - a. The more unconventional the racial/ethnic group one identifies with (with Non-Latino Whites and Asians being considered the least unconventional and Blacks and Native Americans being considered the most unconventional), the (greater / less) the likelihood of having a marital status of divorced/separated (vs. being married).
 - b. Independent Variable
 - c. Dependent Variable
 - d. Rationale: The more unconventional one's race/ethnic group is, the more likely they will be willing to violate dominant cultural norms against divorce and separation and therefore be more likely to get a divorce or separation regardless of whether or not they cohabited.

Notice that in both hypotheses I have stated the variable name as well as the values on each value that I expected to be related. Notice also that there is a subtle difference in the marital status comparisons being made in each hypothesis. In the first hypothesis we are trying to predict the likelihood of forming a cohabiting relationship vs. choosing to get married. As a result, those who are still in the process of making a choice (i.e., those who are currently single either because they have never been married or have chosen to remain divorced, separated, or widowed) are irrelevant to the hypothesis. In the second hypothesis we are trying to predict the likelihood of getting divorced after one gets married. As a result, we are interested in those who got married at some point and have maintained their marriage vs. those who got married at some point and chose to end their marriage. Those who never married, are currently living together, or whose marriage ended for other reasons than divorce or separation (i.e., widowed) are not relevant to testing this hypothesis. Note that some of those currently living together are doing so after a divorce but we have no way of knowing this given how the variable was constructed from the Census data files.

Risk Differences Hypotheses: *Using the hypotheses above as a guide, specify the two hypotheses using the risk exposure variable that you chose for your analysis. Fill in the name of the independent and dependent variables in each hypothesis and provide a reason why you expect the variables to be related. The first hypothesis involves a prediction of the likelihood of cohabiting and the second a prediction of the likelihood of being separated or divorced.*

- Hypothesis 1 (Extraneous Variable Effects on the Likelihood of Cohabitation)
 - Independent Variable
 - Dependent Variable
 - Rationale
- Hypothesis 2 (Extraneous Variable Effects on the Likelihood of Divorce)
 - Independent Variable
 - Dependent Variable
 - Rationale

EXPLORING THE DATA SET

As noted above, this analysis involves the use of a special data set – ocedpv2k – based on the 2000 census. To begin your analysis, follow these instructions:

- Go to <http://www.ssdan.net/datacounts>
- Click on the “Data” in the menu bar
- From there, click “Browse” on the left sidebar. Find “census2000” in the drop-down box and select it.
- Scroll down through the list of data sets until you find “ocedpv2k.dat” Highlight and click “submit.”
- You can also go to <http://ssdan.net/datacounts/webchip> to launch the dataset in WebCHIP.

MARGINAL DISTRIBUTIONS

Questions: *Get descriptive information on the sample for each of the variables included in the data set by creating a Marginals frequency distribution table.*

1. What is the unit of analysis and how many units are included in this sample?
2. What proportion of the sample is currently cohabiting?
 - a. According to lecture, about 10% of all couples in the United States are cohabiting. If this is true, then why is the percent cohabiting shown here so small? (HINT: What is the unit of Analysis?)
3. What proportion of the sample is currently divorced?

- a. Why does this number not reflect the total number of persons in the sample who have ever experienced a divorce?
4. Describe the sample in terms of the percent in each race/ethnicity group and the percent in each category of the risk exposure variable that you chose for analysis.
 - a. Unconventionality Variable (RACELAT):
 - i. NLOther
 - ii. Black
 - iii. Latinos
 - iv. Asian
 - v. American Indian
 - vi. Non-Latinos
 - vii. Others
 - viii. Summary
 - b. Risk Exposure Variable:
 - i. Percent
 - ii. Summary

TEST THE UNCONVENTIONALITY AND RISK DIFFERENCES HYPOTHESES FOR THE LIKELIHOOD OF COHABITATION

In this section you will create variables for and test two hypotheses related to the unconventionality and risk differences hypotheses for the effects of cohabitation on divorce. Specifically, you will test to see if those with unconventional attitudes and those with high divorce risks are more likely to cohabit. If they are, then this increases the likelihood that the cohabitation-divorce relationship is spurious. If, on the other hand, unconventionality and risk exposure do not predict cohabitation, then it will be impossible for these factors to be alternative explanations for the cohabitation-divorce relationship. As the diagram presented earlier shows, there must be a relationship between the extraneous variable (the unconventionality or risk differences variable) and the independent variable (cohabitation) in order for a condition of spuriousness to be true. This would leave one with the conclusion that the experience of cohabitation somehow changes individuals or relationships in such a way as to make them more vulnerable to later divorce.

Modify Variables:

- Before you can test these two hypotheses, you first need to modify the marital status variable so as to isolate those who cohabit vs. those who chose to marry. This modification involves omitting the categories for the never married, divorced, separated, and widowed. Omit the following categories from the "Marital" variable": (hold down the CTRL key on your keyboard and click) "Nmarried", "Divorced", "Separatd", and "Widowed".
- Next, you will have to modify the race/ethnicity variable to create a three category unconventionality variable. Combine the following categories in the "Racelat" variable: . (hold down the CTRL key on

your keyboard and click) "NLWhite", and "Asian".. When asked for a label, enter "LeastUnconv".

- To create the next most unconventional category, combine the following categories in the "Racelat" variable: (hold down the CTRL key on your keyboard and click) "Latino" and "NLOthr". When asked for a label, enter "MidUnconv".
- To create the most unconventional category, combine the following categories in the "Racelat" variable: (hold down the CTRL key on your keyboard and click) "Black" and "AmIndian". When asked for a label, enter "MostUnconv".
- Check that you did the modifications correctly by creating a Marginals frequency distribution table. Enter the percentage results.
 - Cohabitation
 - Married
 - LeastUnconventional
 - MidUnconventional
 - MostUnconventional

CROSSTABULATIONS

Once these variables are modified, create a Percent Down Table with "Marital" as the row variable, and "RaceLat" as the column variable. This will produce a table of percentages for you to use in testing your first unconventionality hypothesis. Fill in the table below with the appropriate percentages.

	LeastUnconventional	MidUnconventional	MostUnconventional
Married			
LivingTg			

Based on the results in the table above, what would you conclude with respect to your hypothesis?

Now test whether or not your Risk Differences variable predicts cohabitation. Create a Percent Down Table with "Marital" as the row variable and the Risk Differences variable you chose as the column variable. This will produce a table of percentages for you to use in testing your first risk differences hypothesis. Fill in the table below with the appropriate percentages:

	Low Risk	Medium Risk	High Risk
Married			
LivingTg			

Based on the results in the table above, what would you conclude with respect to your hypothesis?

TEST THE UNCONVENTIONALITY AND RISK DIFFERENCES HYPOTHESES FOR THE LIKELIHOOD OF DIVORCE

In this section you will create variables for and test two additional hypotheses related to the unconventionality and risk differences hypotheses for the effects of cohabitation on divorce. Specifically, you will test to see if those with unconventional attitudes and those with high divorce risks are more likely to divorce. If they are, then this increases the likelihood that the cohabitation-divorce relationship is spurious. If, on the other hand, unconventionality and risk exposure do not predict divorce, then it will be impossible for these factors to be alternative explanations for the cohabitation-divorce relationship. As the diagram presented earlier shows, a relationship between the extraneous variable (the unconventional or risk differences variable) and the independent variable (cohabitation) is not sufficient for establishing the plausibility of a spurious relationship. There must also be a relationship between the extraneous variable (the unconventionality or risk differences variable) and the dependent variable (divorce) in order for a condition of spuriousness to be true. If significant effects are NOT found between the extraneous variable (either the unconventionality or risk differences variable) and BOTH the independent and dependent variables (cohabitation and divorce, respectively), then you would not have grounds for concluding that the relationship between cohabitation and divorce is spurious. This would leave one with the conclusion that the experience of cohabitation somehow changes individuals or relationships in such a way as to make them more vulnerable to later divorce.

- **Modify Variables:** Before you can test these two hypotheses, you first need to modify the marital status variable so as to isolate those who are currently married vs. those who chose to separate or divorce. Since you already modified this variable differently above, you will need to open the data file again to return to the unmodified version.
 - To do this, go to <http://www.ssdan.net/datacounts>.
 - Click on the "Data" in the menu bar.
 - From there, click "Browse" on the left sidebar. Find "census2000" in the drop-down box and select it.
 - Scroll down through the list of data sets until you find "ocedpv2k.dat" Highlight and click "submit."
 - You can also go to <http://ssdan.net/webchip/webchip4/> to launch the dataset in WebCHIP.
- This modification will require that omit and combine certain variable categories. First, omit the following categories from the "Marital" variable: , (hold down the "CTRL" key on your keyboard and click) "Nmarried", "Livingtg" and "Widowed".

- Next, combine the following categories in the "Marital" variable (hold down the "CTRL" key on your keyboard and click) "Divorced" and "Separatd". Label this new category "DivSep".
- Unfortunately, since you reopened the data file you will now have to re-modify the RaceLat variable to recreate your unconventionality categories.
- Next, you will have to modify the race/ethnicity variable to create a three category unconventionality variable. Combine the following categories in the "Racelat" variable: (hold down the CTRL key on your keyboard and click) "NLWhite", and "Asian". When asked for a label, enter "LeastUnconv".
- To create the next most unconventional category, combine the following categories in the "Racelat" variable: (hold down the CTRL key on your keyboard and click) "Latino" and "NLOthr". When asked for a label, enter "MidUnconv".
- To create the most unconventional category, combine the following categories in the "Racelat" variable: (hold down the CTRL key on your keyboard and click) "Black" and "AmIndian". When asked for a label, enter "MostUnconv".
- Check that you did the modifications correctly by creating a Marginals frequency distribution table. Find the percentage results for the following:
 - Cohabitation
 - Married
 - DivSep
 - Unconventionality
 - LeastUnconventional
 - MidUnconventional
 - MostUnconventional
- Crosstabulations: Once these variables are modified, create a Percent Down table with "Marital" as the row variable and "RaceLat" as the column variable. This will produce a table of percentages for you to use in testing your first unconventionality hypothesis. Fill in the table below with the appropriate percentages:

	LeastUnconventional	MidUnconventional	MostUnconventional
Married			
DivSep			

- Based on the results in the table above, what would you conclude with respect to your hypothesis?
- Test whether or not your Risk Differences variable predicts cohabitation. Create a Percent Down table with "Marital" as the row variable and the Risk Differences variable you chose as the column variable. This will produce a table of percentages for you to use in

testing your first risk differences hypothesis. Fill in the table below with the appropriate percentages:

	Low Risk	Medium Risk	High Risk
Married			
DivSep			

- Based on the results in the table above, what would you conclude with respect to your hypothesis?

DISCUSSION AND CONCLUSIONS

Summarize what you learned from this exercise with respect to selection vs. experience factors as explanations for the effect of cohabitation on marital success. Based on your analyses, is it plausible that the relationship between cohabitation and divorce is spurious? If so, what extraneous variables did you identify that could be alternative explanations for the effect? Explain. If not, how did your tests help you decide this?

You were only able to partially test for selection factors as explanations for the cohabitation-divorce relationship. Explain why these results are only preliminary to a complete test of the spuriousness of the effect? To fully answer the question of whether the negative effect of cohabitation on marital success is causal or spurious due to selection, what additional analysis is needed?