# Money, Control, and Prestige: A Study of the Determinants Associated with the Adoption Rates of Massive Open Online Courses (MOOCs) in American Postsecondary Education

[ Student name removed ]

## Abstract

Disruptive innovation in higher education has garnered significant attention from policymakers and business interests alike. However, how technological innovation diffuses across various institutional types over time is not well understood. This paper employs hazard modeling to assess when and what factors contributed to the early adoption of Massive Open Online Courses to begin to understand variations in institutional response to disruptive innovations. I use a newly created dataset that combines data from the National Center for Education Statistics' Integrated Postsecondary Education Data System (n=3,826) with a publicly available MOOCs dataset. Controlling for time, results suggest prestigious and wealthy colleges have an estimated risk about 2.3 times higher of adopting MOOCs compared to poor or low-prestige schools, while being public or private did not seem to make a significant difference. MOOCs were most likely to be adopted during 2013, about two years after the first MOOC was offered.

## Introduction

The merits of whether technological disruptors of higher education are an effective solution to improved efficiency and productivity have been fiercely debated in recent years (Bass, 2012; C. Christensen, 2015; C. M. Christensen & Eyring, 2011; Christensen, Horn, Caldera, & Soares, 2011; Hixon, 2014; Meisenhelder, 2014). This debate continues to influence the direction of education policy (Yuan & Powell, 2013), and the roles of the private sector in higher education and related products like Massive Open Online Courses (MOOCs), Competency Based Education (CBE), or microcredentialing in the delivery of higher education (Mintz, 2014).

A cadre of disruptors, or tools that are designed to transform the whole postsecondary education system such that education is produced more efficiently and at a lower cost, are an emergent phenomenon (C. Christensen, 2015; Stack, 2014). Technological disruptors have garnered substantial support from the White House (Carey, 2013; Ferenstein, 2013) to well known higher education officials at universities such as Yale and Princeton (Bowen, 2015; The Economist, 2015). Although the types of disruptors rapidly change (Butin, 2015), the volume of reports discussing any disruptor suggests that these types of technological innovations will continue and increasingly impact US higher education (Levine, 2015).

To date, research on the impacts of disruptors on higher education focuses mostly on the impacts on student learning (Anderson, Boyles, & Rainie, 2012; Jaggars, 2014; Markov, 2013; Porto, 2013; Selingo, 2013; Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011). There appears to be broad consensus that new innovations intended to transform traditional education from campus-based to digital-based do not positively impact student learners, particularly those who need greater academic support (Jaggars, 2014; Parr, 2013; Stein, 2013).

Despite the prevalence of studies of disruptive technologies on student learning outcomes, only a few studies examine the impact of disruptive technologies of institutional behavior, including which types of institutions adopt these technologies and their goals (Derousie, 2014; Hollands & Tirthali, 2014; Yuan & Powell, 2013). Other research rooted in policy and the politics of higher education demonstrates the need to focus on institutional behavior around adopting and implementing new reforms (Dougherty & Reddy, 2013; Mettler, 2014; Slaughter & Leslie, 1997; Slaughter & Rhoades, 2004). For example, a large body of research demonstrates that decreases in state appropriations have lead to the privatization of higher education, which has shown to have the following unintended impacts on institutional performance: a steady increase in a restriction of admissions of students who require remedial assistance (Zusman, 2005; Ehrenberg, 2006); narrowing of academic mission (Lieberwitz, 2005); redistribution of institutional funds from less marketable areas to more marketable areas (e.g. from humanities to the biosciences) (Slaughter & Leslie, 1997); increased supply of instructors to match increased student enrollment, but a decline in tenure track opportunities to cut costs (Ehrenberg & Zhang, 2007). These institutional decisions impact the public by diminishing the capacities of public universities and colleges to serve low and moderate income Americans (Ehrenberg, 2006; Mettler, 2014). Other research shows that state and federal policy designed to increase accountability results in restriction of student services, institutional mission, and student admissions (Lahr et al., 2014). And research estimating the impacts of political favoritism for for-profit colleges has shown to significantly disadvantage open-access institutions (Mettler, 2014).

In almost certain terms, higher education must innovate to address the ever growing demand for a college degree at a reduced cost and to extract itself from the “crisis” it is so often associated with (Carey, 2015; C. M. Christensen & Eyring, 2011; Dillon & Carey, 2009; Ebersole, 2014; Ferlie, Musselin, & Andresani, 2008). However, as prior research has demonstrated, adverse institutional responses have resulted from external pressures to address issues of accountability and decreased funding. This paper takes advantage of the rapid expansion of one recent technological innovation, Massive Open Online Courses (MOOCs) to start the empirical work needed to build the foundation for understanding how and when colleges and universities respond to the introduction of a disruptive innovation into its sphere. This paper employs a longitudinal quantitative analysis using a new administrative data set to investigate when colleges and universities first adopted MOOCs and what factors make certain colleges or universities more likely to adopt MOOCs.

## Putting MOOCs into Context

MOOCs, designed to increase efficiency and productivity, were introduced in 2011 to avail students of free, online courses often provided by elite, US research institutions (Pappano, 2012). More than 500 institutions offer MOOCs, which combined offer more than 4, 200 courses to more than 35 million students worldwide (Shah, 2015). As fully online courses that are usually free of charge, non-credit-bearing and open to the public, MOOCs present an alternative to location-bound, proprietary forms of campus-based learning (Mazoue, 2013). MOOCS saw exponential growth in enrollment since their beginnings (Shah, 2015). The founder of the MOOC aggregator site Class Central, Dhahwal Shah, noted that by winter 2011, Coursera, the largest MOOC provider, had enrolled over 1.7 million students in more than 200 courses from 30 plus universities. By fall of 2012, edX the joint start-up from MIT and Harvard registered more than 370,000 students in its first official course (Pappano, 2012; Roth, 2012). Dr. Anant Agarwal, the president of edX, predicted that in just one year, campuses would award students credit for edX certificates and eventually treat MOOC certificates as they currently treat AP coursework (ibid). At the close of 2015, more than 35 million students had signed up for at least one MOOC since 2011 and more than 550 colleges and universities world wide had begun to offer their own MOOCs via one of the emergent platform providers (Shah, 2015).

Of the works that examine institutional implementation or adoption of MOOCs, none to my knowledge have used empirical or theoretical frameworks to examine why or when colleges and universities decide to adopt MOOCs (Bowen, 2015; C. Christensen, 2015; Selingo, 2013). Only a few studies have explored why some institutions adopt MOOCs while others do not (Hollands & Tirthali, 2014; Yuan & Powell, 2013). Hollands and Tirthali (2014) found that institutions adopt MOOCs to maintain brand recognition, improve institutional efficiency, and increase revenue. A national survey by Allen and Seaman (2013) showed that public universities were more likely to offer MOOCs than were private and for-profit schools, which were more likely to be in the planning stages. Further, they show that large universities (with more than 15,000 students) were more likely to offer MOOCs; and that research/doctoral universities were almost twice as likely as other institutions to adopt MOOCs.

According to the popular and scholarly literature, MOOCs are a very controversial innovation in higher education (Rhoads, 2015; Wellen, 2013; CFHE, 2013). Lawmakers and political leaders want to control the spiraling public and private costs of enrolment increases, increased student debt, and anxiety related to the value of credentials (Cronin & Horton, 2009). Conversely, academics and theorists fear that online learning and its subsidiaries are manifestations of a neoliberal reform agenda of higher education where austerity and productivity are key (Jessop, 2012; Noble, 1998). Related to this last point, academics worry MOOCs as well as other online disruptors will further commoditize higher education, and become “digital diploma mills” or the online analog to for-profit colleges, which have recently been investigated as a deceptive and exploitive industry (Fain, 2013; Smith, 2015). Additionally, decreases in state appropriations has meant a shrinkage of programs that are not able to be self-sufficient (e.g. the humanities) and/or increased partnerships with laboratories or corporations that oftentimes fund research departments conditional on the corporation’s research agendas (Slaughter & Rhoades, 2004). Some fear the expansion of online learning could results in similar restriction or compromise important intangible qualities of higher education (Tilak, 2008), constrain faculty autonomy and research contributions from non-flagship universities (Benkler, 2008), or undermine the importance of academic community and strong governance needed in higher education (Bousquet, 2008; Delbanco, 2013).

## Theories of Innovation in Higher Education

Higher education institutions are often perceived as highly resistant to change (Ansell, 2008; Diamond, 2006; Woodhouse, 2016). Yet the rapid uptake of MOOCs signals that higher education, however, does in fact change. Still, extant research on diffusion of innovations and argues institutional change needs additional rigorous research (Kezar & Eckel, 2002; Kezar & Eckel, 2004). Given this lack of research, this paper draws largely from the vast literature on innovation and diffusion rooted in political science and sociology to frame the study approach.

Political scientists have identified two main principles that drive the adoption of a new program or policy: *internal determinants* and *diffusion* (Berry & Berry, 1990, 2014). Internal determinants are political, economic, and social factors that move an organization to innovate. Diffusion of an innovation occurs when it is “communicated through certain channels over time among the members of a social system” (Rogers, 1983, p. 5). Hence, mechanisms of diffusion are interorganizational. Examples of common mechanisms of diffusion are *normative pressures* (i.e., an institution adopts a policy because many organizations are adopting it); *imitation* (i.e., an institution adopts a policy because the perceived leader adopted it); and *competition* (i.e., an institution gains an economic advantage over competitors if it adopts a policy) (Berry & Berry, 2014; DiMaggio & Powell, 1983).

Diffusion scholars employ statistical models that use mechanisms of communication to justify why institutions emulate other institutions when adopting a policy (Berry & Berry, 2014). Often, they use event history analysis – a type of discrete-time hazard model – to identify when an even occurred and the factors associated with the first event as well as subsequent adoptions by neighboring states or districts (McLendon, Deaton, & Hearn, 2007; McLendon, Heller, & Young, 2005; Mintrom, 1997). The leader-laggard model assumes that certain institutions are leaders and that other institutions emulate them. This model incorporates multiple mechanisms; it makes the assumption that leadership is regional or national, if appropriate; and it assumes that the jurisdiction will take cues from one or more pioneers in its region when deciding to embrace a policy (Grupp & Richards, 1975; Rogers, 1983; Walker, 1969). This paper employs the leader-laggard model to frame the discussion and to select the internal determinants and associated mechanisms to estimate the odds that an institution will adopt a MOOC at a given point in time. Applying this theory, I focus on institutional prestige, institutional wealth, and whether the institution is publicly or privately controlled as the primary determinants because these are the elements that find the greatest support in the literature.

**Key Measures of Innovation in Higher Education**

*Institutional prestige*, also referred to as reputation, prestige, status, impression, stature and visibility, has not only been consistently linked to institutional change regardless of sector

(Cyert & March, 1963; DiMaggio & Powell, 1983; Rogers, 1983) but also to increased entrepreneurism in higher education (Clark, 2004; Goldman, Goldman, Gates, Brewer, & Brewer, 2004; Kezar, 2014; Marginson, 2006). Institutions with high levels of perceived prestige are those that command a global reputation for research, ability to produce PhD graduates, and an ingrained willingness to build on this research as a means for competing vigorously for more prestige and additional resources (Clark, 2004; Goldman et al., 2004)). Because organizations are resource dependent, they rely on improving or maintaining their perceived levels of prestige in order to secure critical sources of funding (Hearn, 1996). Measuring prestige, which is typically done using the Carnegie Classification or Barrron's selectivity index are highly controversial instruments (Kezar, 2013).

High levels of *institutional wealth* have also been found to be a strong predictor of institutional innovation (Damanpour, 1991; Mohr, 1969; Rogers, 1983), yet there is some discussion that the lack of resources also may lead organizations to change in order to become more efficient and cut costs (Schumpeter, 1934; Slaughter & Leslie, 1997). Everett Rogers the father of discourse on the diffusion of innovation, submits that organizational wealth may condition institutional change for either resource-rich or resource-poor institutions, generally less wealthy organizations will adopt an innovation last simply because they do not have the available resources to fund innovation or handle a loss (2003, p. 295).

A recent dissertation published in 2014, which examines the relationship between post secondary institutions and various innovations including MOOCs over a timespan of 25 years, does find evidence for a positive relationship between revenue and innovation (DeRousie, 2014). DeRousie finds that institutions that earned more revenue from tuition and fees compared to those who earned less had greater odds of adoption compared to institutions that charged less.

It is well known *innovation* strongly correlates with the *private sector* since commercial enterprises are required to be responsive to market pressures and to stay competitive (Cankar & Petkovsek, 2013; C. Christensen, 2015; West & Lu, 2009). In contrast, government agencies do not have to show a profit on their revenues and are not beholden to customers the way the commercial businesses are and therefore may be less likely to innovation. Additionally, because private entities are not regulated in the same ways as public organizations, they may enjoy higher degrees of autonomy, which is also strongly associated with innovation (Hartley, Sørensen, & Torfing, 2013).

Higher education institutions have been treated as if they were in the middle of the public-private continuum (Alpert, 1991; Berdahl, 1991; Birnbaum, 1991). Building on this, Kezar (2013) writes that higher education has benefitted from a degree of autonomy because no national ministry of higher education evolved and states have a tradition of giving the institutions their independence with very few local expectations. However, she goes on to note that different institutional types experience greater and lesser degrees of environmental independence (p. 65). For example, private institutions experience greater vulnerability to market forces, whereas public institutions tend be affected by state legislatures. Research on whether the noted emphasis on market forces has evened out public and private vulnerabilities to existing market forces.

## Data and Method

### Data Collection

Data for this study come from two independent data sets that I compiled. I used a subset of publicly available, administrative data from the federal National Center for Education Statistics, Institutional Post Secondary Education Data System (IPEDS) and combined these data with data from a website that aggregates key measures about MOOCs called Class Central (Class Central, 2016). IPEDS is a publicly available longitudinal data set with institutional data for more than 7,000 colleges and universities in the United States (NCES, 2016). I used the following surveys collected for IPEDS for academic years 2010-11 through 2014-15: admissions, institutional characteristics, fall enrollment data, and graduation data.

Class Central provides an index of most of the MOOCs offered worldwide since their first occurrence in 2011. Data used from this site included course name, course description, course subject, course dates, course instructor, length of course, university or college that provides the course, and name of the technical platform provider or LMS (such as Coursera, edX, canvas.net, etc.). Because data from Class Central are not compiled in the traditional sense the way public data may be available, data collection from Class Central required using a web crawler to comb through or “crawl” a website with the desired data (Thornton, 2014). I, as the researcher, select the data I wish to use by “training” the crawler and then save the output in the appropriate form to then be read into statistical software. I used a freely available crawler called IMPORT.io (Import.io via www.import.io, 2016) to collect my data and then analyzed it using SPSS statistical software.

Since data from this website are collected by a third party from the MOOC providers, I interviewed the founder and CEO of Class Central, Dhawal Shah, to ensure data collection and data reliability and took a random sampling of courses to cross-examine them with the provider and institution websites to ensure the data matched.

To create the final data set, I merged two data sets using institutional names as the key matching variable. There was a 100 percent match rate between MOOC institutions and the IPEDS university and college identifiers. The analytic sample includes 3,825 nonprofit, Title IV–eligible two- and four-year institutions in the United States. The academic years (AY) examined include 2010–11 through 2014–15. One of the major contributions of this study is that this is the first database of this kind to merge data for a higher education innovation with administrative postsecondary data; the emergent data set allows for many future analyses that can range in depth and breadth. Secondly, the data set can be expanded easily now that an algorithm exists for collecting MOOC data and because IPEDS data continue to be updated annually.

**Measures**

My primary dependent variable was a binary variable for MOOC adoption (1= yes; 0 = no). Since MOOCs are free and open courses that do not require students to have fulfilled any prerequisites to take the course, colleges and universities that offer online courses but also include admissions requirements or which are not free are not considered MOOC adopters. This study counts colleges and universities only as "adopters" if they began offering a MOOC starting from the beginning of the academic year 2011-2012 through the end of academic year 2014-15 (July 1, 2011-June 30, 2015), where academic years are those defined by IPEDS. In total, there were 140 colleges and universities that adopted MOOCs in this time period. Descriptive statistics are posted in Table 1 in the following section.

First, I use a log transformed measure of the college or university's yearly cash flow (*lncashflow*), which is calculated from total assets minus total liabilities and then log transformed in order to impose a normalized linear relationship to the outcome on this otherwise very skewed metric (Mertler & Vannatta, 2013).

Secondly, I included a variable that is a proxy measure for prestige based on the Carnegie rankings provided by IPEDs. Often researchers use Barron's selectivity index to estimate college and university prestige, which is a restricted NCES data set. I chose not to use this measure and instead relied on the publicly provided Carnegie ranking for two reasons. First, the Barron's index does not provide a rank for all colleges and universities in my study and secondly the data are restricted such that I cannot share my data with reviewers without prior consent from NCES.

The basic Carnegie ranking includes more than 10 subgroups that I simplified into a binary variable coded 1 for colleges with high levels of prestige approximated by using the Carnegie rankings (*prestige)*.

The last variable of interest is *public*, which denotes whether a university is publicly or privately controlled.

Time-invariant variables include the size of institution (*instsize 2)*, which denotes whether an institution is larger than 10,000 students and whether or not the institution is a four-year or two-year institution (four-year).

The number of periods examined is four and the number of total events included equal 140. I excluded international colleges and universities, which are subject to different political and economic influences as well as university systems that adopted MOOCs because they have different administration concerns that cannot be captured in my data set. I also excluded for-profit colleges because as of March 2014, there was only one for-profit that had adopted a MOOC in the dataset. Time invariant control measures include whether the institution is a four-year institution (1 = four-year, 0 = 2year) and the institutional size (1 = more than 10,000 students, 0 = less than 10,000 students).

**Empirical Strategy: Discrete-Time Hazard Modeling of Institutional Adoption of MOOCs as a Time-Varying Event**

To address the question of whether prestige, public ownership, or wealth corresponded to an institution's increased risk of adopting a MOOC—I used discrete-time hazard models using logit regression (Singer & Willett, 2003; Allison, 1982). A review of the literature showed that a few political scientists have specifically targeted the diffusion of education policy solutions from state government to state government using discrete-time hazard modeling (McLendon, Deaton, Hearn 2007; Mokher 2008; Doyle, 2006; Mintrom, 1997; McLendon et al 2005), but to my knowledge none have examined the diffusion of innovation between institutions themselves or outside of the government's purview. The application of a large national study using discrete-time hazard to model diffusion of innovation in higher education is one of the major contributions of this study.

Discrete-time hazard modeling is commonly used by social scientists to ascertain when and why an individual or group does something while another does not (Allison, 1982; Singer & Willett, 2003). The data are in the form of an event history, which is a record of when events occurred to a sample (Allison, 1982; Tuma and Hannan, 19878). For example, if a sample consists of women of childbearing age, each woman's event history might consist of the birthdates of her children (Allison 1982, p. 62). In this study, data are reported in long-form such that each institution's data is spread across multiple records where each record describes a specific time period (Singer & Willet, 2003, p.352). This is often referred to as the "person-period" data set, but in the case of the current paper is an "institution-period" data set. Because we are only concerned about the first "event" happening, the data set only grows in length and not width. In particular, records increase for each year until an institution adopts a MOOC; at this point, data stops being recorded for the individual (or institution).

Discrete-time hazard models make it possible to incorporate both time-varying and time invariant explanatory variables and allows for the impact of categorical predictors to be estimated by using log-linear approaches that are not well estimated using traditional ordinary least squares (OLS) (Allison, 1982). In this study, the time-varying predictors that have been linked to institutional innovation are *prestige* and *cashflow*; the variable *public*, conversely, is time invariant. As controls for the model, I use the size of the institution (*instsize2)* and whether or not the institution is a four-year or two-year (*fouryear).* In addition to these types of variables, discrete-time hazard modeling also permits the inclusion of time effects within the equation. This means we can control for the discrete nature of time due to the fact that MOOCs data are only reported in large units of time (e.g. month and year) (Allison, 1982).

The discrete-time hazard model is useful for identifying the probability of an occurrence in a specific period of time, provided that the event has never occurred before (Singer & Willett, 2003). Specifically, the result called the *hazard* is the conditional probability that individual *i* will experience the event in time period *j*, given that he or she did not experience it in any earlier time period (Singer & Willett, 2003, p. 330). Conditionality is important because it ensures that the hazard represents the probability of event occurrence among those individuals *eligible* to experience the event occurrence in that time period -- the *risk set*. The hazard function results in maximum likelihood estimates of the discrete-time hazard function for the risk-set (Singer & Willett, 2003). In this study, the risk set comprises institutions that had never adopted MOOCs before and are therefore “at risk” of adoption.

One disadvantage of the discrete-time hazard modeling that is not an issue in OLS is bias introduced by censored data, which occurs when a researcher does not know an individual’s event time (Singer & Willet, 2003; Allison, 1982). If an event time is never observed this is referred to as *right censoring* and is not as concerning as when an event time is not known because the beginning of time is not observed, referred to as *left censoring* (see Singer & Willett, 2003, p. 316-318). For practical purposes, in this study the beginning of time is 2011 and no event occurred before this time, eliminating any concern about left censoring. However, as illustrated in the results section many institutions have still not adopted MOOCs highlighting the high degree of right censoring.

Discrete-time hazard modeling is a common strategy for political scientists who investigate the diffusion of policy across states (e.g. Shilpan & Volden, 2008; Berry & Berry 1990; Mintrom, 1997). This study examines the diffusion of a technological innovation across higher education institutions, and while states are not the same as institutions of higher education, discrete-time hazard modeling seems an appropriate match to model diffusion. One technical advantage of this model is that it allows for a dichotomous outcome variable, the appropriate handling of data through censoring, and the ability to include the effect of time in the equation (Bowers, 2010).

Drawing from the recommendations of scholars of longitudinal analyses, the institution-level data set was converted into an institution-level data set, with *event* defined as the first institutional adoption MOOCs from academic year 2011-12 until the time when the latest institutional data is available from IPEDS, AY 2014-15. Institutions that had not adopted MOOCs by the end of the AY 2014-15 were censored from the data. No institutions were left censored.

The basic equation presented below is estimated with a cross-sectional time series logit model, and the results are presented as maximum likelihood estimates.

$$ADOPT_{i,t} = a_1D_1 + a_2D_2 + a_3D_3 + a_4D_4 + B_1(FOURYEAR)_{i,t} + B_2(INSTSIZE2)_{i,t} + B_3(PUBLIC)_{i,t} + B_4(PRESTIGE)_{i,t} + B_5(LNCASHFLOW)_{i,t}$$

In this equation, $ADOPT_{i,t}$ is the conceptual dependent variable (hazard rate), which are the odds that an institution *i* will adopt a MOOC in year *t*, given that the state has not adopted a MOOC the year prior. A constant $a_1D_1 - a_4D_4$ denotes each of the four time periods (AY 2011-12, 2012-13, 2013-14, 2014-15). $FOURYEAR$ is a binary variable that indicates whether a the college or university is a four-year institution. *INSTSIZE2* is a binary indicator for whether a school has a total student enrollment that exceeds 10,000 students. *PUBLIC* is a binary variable that denotes whether an institution is public. *Prestige* is a binary measure based on Carnegie rankings for doctoral/research institution, where the top-ranked universities are research universities (Research I and Research II). *LNCASHFLOW* is a continuous measure of the fiscal health of an institution calculated by taking subtracting the total liabilities from total assets.

A discrete-time hazard model was fit to the data by estimating parameters for each time period and for each of the three variables using logistic regression (, 2010). As mentioned previously, because no institutions adopted MOOCs prior to 2011 making it so there were no intercept parameters for prior years, the beginning of time is 2011. As the literature suggests, the model begins by conducting a test of significance of multiple pseudointercepts for each time point, which effectively models the effect of time in the analysis of an institution's risk of adopting a MOOC (Singer & Willet, 2003; Bowers, 2010). Then the additional parameters are added to the model as β estimates, and then the model fit is assessed. Table 4 presents odds-ratios for seven discrete-time hazard models, as well as parameter estimates and significance levels, standard errors (in parentheses), the overall N of the institution-period data set at each time point and the tests of model goodness of fit, including -2 Log likelihood, chi square, and Cox & Snell pseudo $R^2$.

Finally, as mentioned before, in the time since MOOCs first launched, there have been 140 instances of adoption (i.e. 140 events). That is, of the institutions included in this study, 140 began offering at least one MOOC between fall 2011 and summer 2015. This number is comparatively small compared to the total population of eligible MOOC adopters. The recommended guidelines for determining the minimum sample size required for survival analysis, a priori power analysis calculations assume a power of 0.8 ($\alpha$ =.05); however, no prediction rate from the literature on the rate at which higher education institutions adopt a technological innovation has been established and so determining the power is beyond the scope of the paper.

## Results

### Descriptive Results

Table 1 presents the descriptive statistics for variables used in this study. On average, more colleges and universities have less than 10,000 students; there are more four-year than two year colleges; roughly half of the schools are public; fewer than 10 percent of the schools are highly prestigious, and there is very little variation about the mean in terms of available cash flow.[1]

**Table 1: Descriptive Statistics**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| instsize2 | 3,825 | 0 | 1 | .29 | .454 |
| fouryear | 3,825 | 0 | 1 | .65 | .477 |
| public | 3,825 | 0 | 1 | .49 | .500 |
| prestige | 3,825 | 0 | 1 | .07 | .255 |
| lncashflow | 3,825 | 0 | 24.36 | 17.62 | 1.83 |
| Valid N (listwise) | 3,825 | | | | |

A preliminary multiple regression was conducted to calculate the Mahalanobis distance to identify outliers and examine multiple collinearity among the model predictors (*instsize2, fouryear, public, prestige,* and *lncashflow*). Data screening led to the elimination of 4 cases, which exceeded the critical chi-square value ($df$ = 9, critical chi-square criterion = 27.877, $p$ < .001). Although not required in logistic regression, evaluation of linearity justified a natural log transformation of the only continuous variable in this study (*cashflow*) into *lncashflow*.

[1] Should the Ns be the same for this table? Some of these colleges do not report their financial data and so it is hard to get this total number. I'd have to go entirely with listwise deletion which would limit my sample. Please advise.

According to the results in the coefficients table, all tolerance statistics exceed .1, indicating there are no issues of multicollinearity (see Table 2).

**Table 2: Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | -.096 | .011 | | -9.063 | .000 | | |
| | d2 | .008 | .003 | .034 | 3.156 | .002 | .671 | 1.491 |
| | d3 | .018 | .003 | .074 | 6.781 | .000 | .671 | 1.489 |
| | d4 | .013 | .003 | .055 | 5.095 | .000 | .673 | 1.485 |
| | instsize2 | .003 | .003 | .014 | 1.236 | .217 | .586 | 1.707 |
| | fouryear | -.003 | .003 | -.015 | -1.255 | .209 | .561 | 1.783 |
| | public | -.008 | .003 | -.038 | -2.976 | .003 | .493 | 2.030 |
| | lncashflow | .005 | .001 | .095 | 8.747 | .000 | .673 | 1.486 |
| | prestige | .063 | .004 | .164 | 15.925 | .000 | .741 | 1.349 |

a. Dependent Variable: event

The Lifetable (Table 3) summarizes the sample distribution of event occurrence. The data represent study findings, which tracked the adoption of Massive Open Online Courses of 3,825 colleges and universities over four years (2011-12 to 2014-15). This is a full-population study and no colleges closed during the study; however, 3,685 institutions were right censored at the end of the study. At the start of the study 100 percent of the colleges and universities had not adopted MOOCs; by the end of the study, 96 percent of colleges and universities still had not adopted any MOOC. The hazard rate, which is the conditional probability that a college or university will experience the event in a given time period, given that it did not experience it in any earlier time period, was the highest during year three (1.6 percent). In comparison, the survival rate, the cumulative risk associated with event occurrence at each period, was the lowest point in year three (97.4 percent).

**Table 3: Life Table[a]**

| | | Number of colleges and universities | | | Proportion of colleges and universities | |
|---|---|---|---|---|---|---|
| Year | Time Interval | That did not adopt MOOCs | That adopted MOOCs | Censored at the end of the year | That adopted MOOCs during the year | That did not adopt MOOCs during the year |
| 0 | [0,1) | 3825 | 0 | 0 | - | 1.0000 |
| 2011 | [1,2) | 3825 | 7 | 0 | 0.0018 | 0.9982 |
| 2012 | [2,3) | 3818 | 32 | 0 | 0.0084 | 0.9898 |
| 2013 | [3,4) | 3786 | 59 | 0 | 0.0156 | 0.9744 |
| 2014 | [4,5) | 3727 | 42 | 3,685 | 0.0113 | 0.9634 |

a. Median survival time is 4.0

Figure 1 includes a graph of the hazard and survival curves for the sample. After the peak in year three, the figure on the left depicts a decline in the adoption of MOOCs since they initially entered the higher education landscape. Though we do not have enough years of data to confirm this, the trend in the data conjure Gartner's hype cycle – a five-step process of "initial trigger", "inflated expectations", and "trough of disillusionment" which precede more reportedly productive periods called the "slope of enlightenment" and "plateau of productivity", when the technology's broad applicability prompts widespread mainstream adoption (McPherson & Bacow, 2014; Gartner, 2015). In this case, it appears the initial period of innovation is analogous to the "innovation trigger." The following period characterized by a rapid increase is the "peak of inflated expectations," which is followed by a sizeable decrease that might be evidence of the beginning of the "trough of disillusionment."

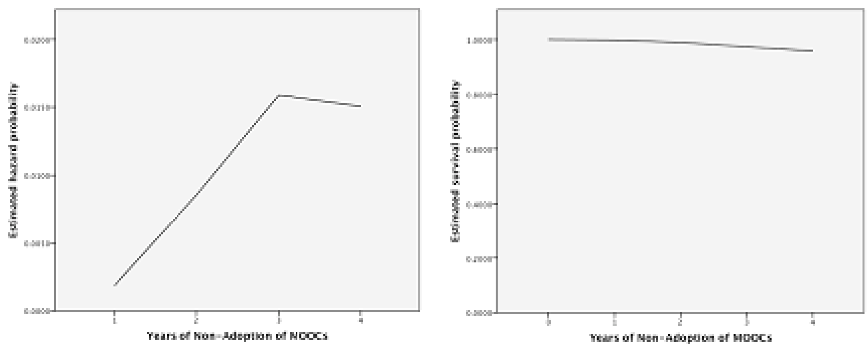



**Figure 1. Estimated hazard function (left) and estimated survival function (right) from study on the timing of adoption of post-secondary adoption of Massive Open Online Courses (MOOCs).**

**Analytic Results**

Table 4 reports the logit results converted to odds ratios for seven discrete-time hazard models for the year the institution first adopted a MOOC, where D1-D4 are a set of time indicators representing each year of the model. Model A includes only the main effect of time; Model B includes the main effects of time plus *instsize2*; Model C includes the main effects of time plus *fouryear*; Model D includes the main effects of time *public*; Model E includes the main effects of time plus *lncashflow*, a measure of the total assets minus total liabilities; Model F includes the main effects of time plus *prestige*; and, Model G includes the main effects of time plus all the substantive predictors: *instsize2*, *fouryear*, *public*, *lncashflow*, and *prestige*.

Not accounting for any differences caused by the substantive predictors, Model A reports that the likelihood that an institution will adopt a MOOC increases from year one to year three, but then begins to taper in year four. For example, the estimated odds institutions will adopt MOOCs in year one is approximately .1 times higher compared to previous years but rises to 1.9

time higher in year three and then falls slightly to 1.4 times more likely in year four. These values are similar to those presented in the column which reports survival rates in the Lifetable (Table 3).The baseline for Model A is the entire sample since it does not include any substantive predictors.

Model B reports on the risk of MOOC adoption after introducing the first exogenous predictor, institutional size (instsize2) and the main effect of time. Similar to Model A, the results suggest an increase in the risk of MOOC adoption until after year three when the risk begins to decline slightly. In this group, the baseline for the α's in this group is *Instsize2* = 0. Colleges and universities that are larger, therefore, have an increased risk of adopting MOOCs compared to smaller colleges and universities. The estimated odds large colleges and universities (enrollment >= 10,000 students) will adopt MOOCs is almost eight times greater than small colleges and universities.

Model C reports that the risk of adopting MOOCs after introducing the substantive predictor *fouryear* (baseline *fouryear* = 0) to the main effect of time illustrates again a constant increase in the risk of adopting MOOCs until the end of year three but a decrease in year four. Controlling for the effects of time, the odds four year institutions adopt MOOCs is almost six times higher than two-year institutions to adopt MOOCs.

Model D shows the risk of adopting MOOCs after introducing the variable *public*, an indicator for whether the school is public versus private (baseline = 0), to the main effect of time. This variable is not statistically significant, suggesting there is no difference in the risk of adopting MOOCs between public and private colleges and universities.

Model E introduces the measure *prestige* to the main effect of time and reports that very prestigious universities (Research I and Research II colleges according to the Carnegie ratings) have a risk 20 times higher than non-prestigious colleges and universities of adopt MOOCs, controlling for time. Again the risk for adoption increases with time until year four, at which point it stars to decrease.

Model F introduces the last substantive predictor, the logged cashflow variable (*lncashflow*) to the main effect of time. This model illustrates the same pattern of changes over time, but also shows that the estimated odds of first adopting a MOOC are approximately 6 times higher for colleges and universities with more positive cash flow than colleges and universities with presumably less disposable income.

Model G indicates that when all predictors are included that in addition to time being significant, the risk of adopting MOOCs is significantly higher for prestigious and financially well off colleges and universities compared to those that are not. Specifically, after controlling for time, prestigious colleges and universities that benefit from a higher, positive cash flow have an estimated risk of about 2.3 times higher of adopting MOOCs than less prestigious and poorer colleges and universities.

In addition to the coefficients presented in Table 4, there are a number of other values that can be used to analyze and understand how well the model fits the data. The log-likelihood is not really useful on its own but is useful for computing the deviance statistic, which is commonly used to assess model fit (Singer & Willett, 2003, p. 398). However, better modeling does result in increasing LL values for each added parameter, which is the case in my table. Deviance measures how much worse the model is in comparison to the best possible model that could be fit. That is, the better the fit of the model the smaller the deviance will be. In this case, Model G has the most parameters (6) and the smallest deviance (1016.63), while Model A has the fewest parameters (4) and the greatest deviance. However, Model D, which focuses on whether a university is public is only marginally smaller (1405.41). When the difference in deviance is small compared to the critical chi-square value, researchers fail to reject the null hypothesis, which indicates the reduced model is not substantively worse than the less parsimonious model (i.e., the model with more parameters). Both the deviance-based hypothesis tests and the Wald-Chi Square statistics ($df = 1$, critical chi-square at $p < .001 = 10.83$) demonstrate the null model should be rejected in favor of the more parameterized models (see p. 398-399).

Because the LL statistic inevitably increases with each addition new parameter, we use the AIC and BIC to correct for this effect. The AIC and BIC essentially penalize the LL statistic for the number of parameters present in the model. When we cannot calculate the deviance for non-nested parameters we also use the AIC and BIC to assess the benefit of a fully saturated model. Model G has the smallest AIC and BIC values, suggesting that again Model G has the best fit of any of the other models even though it has the greater number of parameters (AIC = 1028.63; BIC = 1045.97).

Finally, the Cox & Snell $R^2$ is a rough indicator of the amount of variance explained by the model (Bowers, 2010). In logistic equations, it is well known that the $R^2$ is not a true approximation of the variance explained by the equation (Aldrich & Nelson, 1984). However, Model G appears to explain substantially more variance (97.1 percent) in the probability of an institution adopting MOOCs compared to other models which hover around 70 percent, providing further evidence that Model G is the best model on which to explain the patterns of MOOC adoption in the early years.

**Table 4: Discrete-Time Hazard Models Fitted to the Year of First Adoption of Massive Open Online Courses (*n* =3,825, *n events* = 140)**

| | Model A | Model B | Model C | Model D | Model E | Model F | Model G |
|---|---|---|---|---|---|---|---|
| Parameter Estimates and Asymptotic Standard Errors | | | | | | | |
| $D_1$ | 0.002*** | 0.001*** | 0.000*** | 0.001*** | 0.001*** | 0.000*** | 0.000*** |
| | (0.448) | (0.476) | (0.552) | (0.457) | (0.465) | (1.427) | (1.715) |
| $D_2$ | 0.010*** | 0.003*** | 0.002*** | 0.009*** | 0.004*** | 0.000*** | 0.000*** |
| | (0.183) | (0.244) | (0.371) | (0.205) | (0.220) | (1.318) | (1.631) |
| $D_3$ | 0.019*** | 0.006*** | 0.004*** | 0.018*** | 0.008*** | 0.000*** | 0.000*** |
| | (0.134) | (0.207) | (0.349) | (0.161) | (0.174) | (1.284) | (1.610) |
| $D_4$ | 0.014*** | 0.005*** | 0.003*** | 0.013*** | 0.006*** | 0.000*** | 0.000*** |
| | (0.157) | (0.222) | (0.359) | (0.181) | (0.191) | (1.285) | (1.609) |
| *instsize2* | | 7.737*** | | | | | 1.501 |
| | | (0.206) | | | | | (0.298) |
| *fouryear* | | | 5.972*** | | | | 1.246 |
| | | | (0.346) | | | | (0.406) |
| *public* | | | | 0.824 | | | 0.885 |
| | | | | (0.175) | | | (0.216) |
| *prestige* | | | | | 20.146*** | | 2.342*** |
| | | | | | (0.183) | | (0.286) |
| *lncashflow* | | | | | | 2.962 | 2.323*** |
| | | | | | | (0.064) | (0.085) |
| | | | ○ | | | | |
| Goodness-of-fit | | | | | | | |
| LL | -703.32 | -642.21 | -681.08 | -702.71 | -579.39 | -519.04 | -508.32 |
| Deviance | 1406.63 | 1284.41 | 1362.15 | 1405.41 | 1158.77 | 1038.07 | 1016.63 |
| n parameters | 4 | 5 | 5 | 5 | 5 | 5 | 9 |
| AIC | 1414.63 | 1294.41 | 1372.15 | 1415.41 | 1168.77 | 1048.07 | 1028.63 |
| BIC | 1426.19 | 1308.86 | 1386.60 | 1429.86 | 1183.22 | 1062.52 | 1045.97 |
| Cox & Snell $R^2$ | 0.719 | 0.722 | 0.720 | 0.719 | 0.725 | 0.727 | 0.971 |
| | | | | | | | |
| Deviance-based Hypothesis Tests | | | | | | | |
| $H_0$: $\beta_{instsz.} = 0$ | | 122.22*** | | | | | - |
| $H_0$: $\beta_{\text{foury.}} = 0$ | | | 44.48*** | | | | - |
| $H_0$: $\beta_{\text{public}} = 0$ | | | | 1.22 | | | - |
| $H_0$: $\beta_{\text{prest.}} = 0$ | | | | | 247.86*** | | - |
| $H_0$: $\beta_{\text{lncas.}} = 0$ | | | | | | 368.56*** | - |
| | | | ○ | | | | |
| Wald Hypothesis Tests | | | | | | | |
| $H_0$: $\beta_{instsz.} = 0$ | | 98.23*** | | | | | 1.855 |
| $H_0$: $\beta_{\text{foury.}} = 0$ | | | 26.69*** | | | | 0.29 |
| $H_0$: $\beta_{\text{public}} = 0$ | | | | 1.221 | | | 0.32 |
| $H_0$: $\beta_{\text{prest.}} = 0$ | | | | | 270.05*** | | 8.88 |
| $H_0$: $\beta_{\text{lncas.}} = 0$ | | | | | | 290.38*** | 97.96*** |

~*p* <.10, **p* <.05; ***p* <.01; *** *p* <.001

**Discussion and Concluding Remarks**

The results from Table 4 demonstrate that the most prestigious and wealthiest of universities are the most likely to adopt MOOCs. Specifically, universities with the highest distinction of research according to Carnegie rankings and those with positive cash flow are both associated with higher risks of adopting MOOCs compared to colleges and universities that are not Research I or Research II institutions or which are more financially constrained. There is no evidence supporting the claim that public institutions may be more likely to innovate than private colleges and universities, calling into question whether financial motivation plays a significant role in innovation to increase revenue. I also find that once cash flow and prestige are introduced into the model, the size of an institution and whether or not it is a four-year institution become non-significant. Finally, I also find that universities had the highest risk of adopting MOOCs in year three, after which the risk of adoption began to decline.

There are several implications for these findings. For one, that wealthier institutions have higher hazard rates than poorer institutions is not surprising; as the literature points out, it is these institutions that are the least risk averse because of their greater financial cushion. Additionally, because MOOCs are quite expensive to produce, it may be that these wealthier institutions are the only organizations capable of producing them – even if poorer institutions were inclined to adopt MOOCs and market themselves to a global audience. Regarding the null finding for public universities, it is surprising indeed that there is no statistically significant result, either positive or negative. The privatization literature suggests that more market-oriented education organizations will be more innovative than those more heavily oriented to the public service mission. Extrapolating out from this theory, we would expect to find either lesser resourced public institutions have a higher rate of adoption compared to their private counterparts; or the opposite, private colleges and universities that have less regulatory oversight and more autonomy to innovate.

These findings help to clarify what really happened in the initial frenzy of MOOCs. While the hype of MOOCs has receded somewhat in the media and critics and advocates have become more tempered in their reviews over the last year or so, the findings of this study suggest the early attention paid to their rapid rise and disruptive force in higher education may have been accelerated or exaggerated. The scale at which the innovation expanded is in fact quite small as is illustrated by the hazard rate of 1.6 percent during year three, the year with the highest risk associated with adoption. It is important to keep in mind that the majority of colleges and universities in the United States have not adopted MOOCs and the risk for adoption remains very small. Nonetheless, there was a noticeable and statistically significant trend associated with the diffusion of MOOCs at a very early point in their history. Some theorists suggest this may be an early indication of influential incremental change that could lead to sustained, transformative change in the future (Thelen & Mahoney, 2010).

Because of the nature of MOOCs and their novelty in higher education as well as our limited general understanding about transformative versus incremental change in higher education, there are many questions still unanswered by this study. What factors distinguish early adopters from late adopters? To what extent do these actions reflect institutional response to political and economic pressures environmental pressures to change? Or alternatively, were there inter-institutional politics that catalyzed or suppressed institutional adoption? What about the impacts of the external but rapidly developing higher education market on the decisions to adopt MOOCs or not? For example, as the market for STEM degrees continues to flourish, are colleges and universities more likely to offer STEM courses over humanities? With regards to looking toward the market for revenue replacement, what relationships do we see between for-profit providers and different types of colleges and universities?

Establishing the initial rates of adoption as well as the influence of some substantive political and economic predictors of institutional change are first steps to understanding higher education response to the emergence of a technological innovation. Next steps in the study will be to interact these terms to generate better nuanced explanations of difference in innovation across institution type.

.

**Comment [AJB1]:** REVISION:
This is an excellent revision.

There's a few formatting issues still throughout, with a few missing citations and some footnotes, etc that appear to be notes to me.

In looking at Table 2, I may have missed this in the last version, but I'm not sure what the standardized coefficients are. These should be odds ratios. So I'm not sure how the "beta" was calculated. The unstandardized are logits, or at least should be.

Also, Table 2 is not in APA format. Please see previous DTHM papers for correct APA format.

Table 4 has much of what you need though, and it's well formatted. A few issues though. One should not have 0 as a coefficient as you do here. I know that the probability is really low, but it's significant and so there's no way it's zero, which is what you're reporting with 0.000.

Overall though, this is an excellent revision.

Final grade on Midterm A+ (99%).

**APPENDIX A: TABLES AND FIGURES**

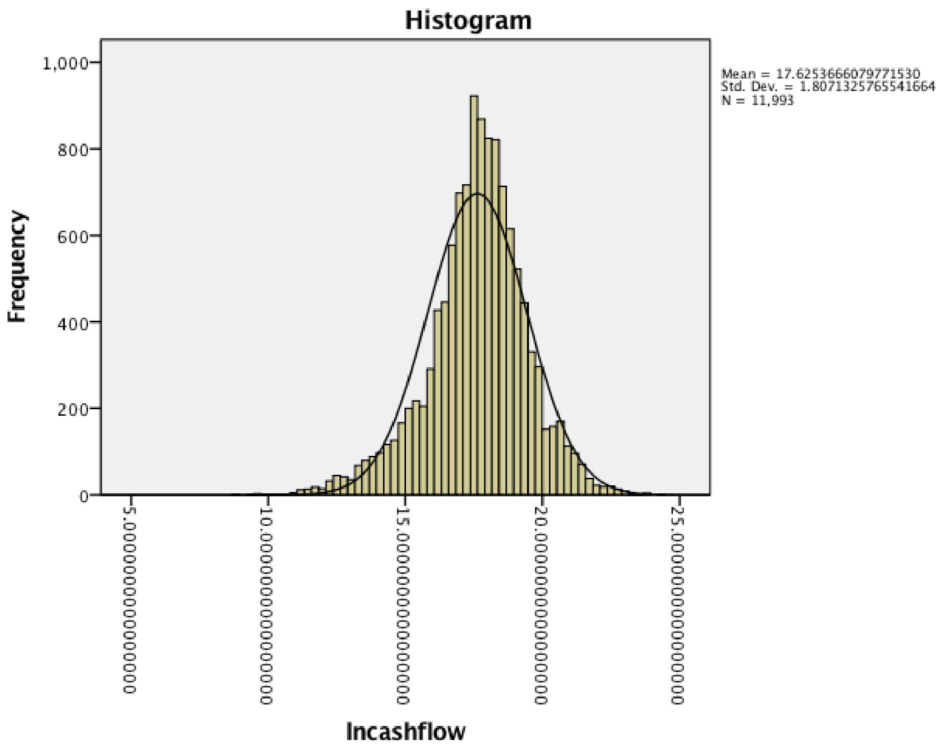



**Figure 2 . Histogram demonstrating normality of logged cash flow (lncashflow) variable.**

**APPENDIX B: Syntax & Output**

**Syntax**

```
* Encoding: UTF-8.
** Lara Pheatt
** Purpose: Preliminary MOOCs Hazard Modeling for Bowers Midterm, Spring 2016
** Data used: Derived/hazmod_institution_period_clean.xlsx, originally cleaned in STATA

** Life Table, p. 327 of S & W, 2003. Chapter 10.

GET DATA /TYPE=TXT
  /FILE='/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean.txt'
  /DELCASE=LINE
  /DELIMITERS=","
```

```
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=2
  /IMPORTCASE=ALL
  /VARIABLES=
  id F3.0
  time F1.0
  censor F1.0.
CACHE.
Execute.

** Makes Life Table

procedure output outfile= '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean1.txt'.
survival table = time
 /interval = thru 4 by 1
 /status = censor(0)
 /write=tables.

** AUTHOR IMPORTANT TO READ & DO: edit hazard model in .txt to get it into spss
acceptable format --> hazmod_institution_level_clean2.csv

get data
 /type = txt
 /file = '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean2.csv'
 /delimiters = ","
 /arrangement = delimited
 /variables =
 v1 f5.2
 v2 f7.2
 v3 f6.2
 v4 f6.2
 v5 f6.2.
execute.

save outfile  '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav'
 /drop v4
 /rename v1=year v2=begin v3=censored v5=left.
```

```
get file '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav'.
compute p1 = left/begin.
do if $casenum = 1.
compute p2 = 1.
else.
compute temp = lag(p2).
compute p2 = (1-p1)*temp.
format p1 p2 (f6.4) year begin left (f4.0) censored (f4.0).
end if.
if $casenum = 1 p1 =$sysmis.
list year begin left censored p1 p2.
save outfile '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean4.sav'.


GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=year
MEAN(p1)[name="MEAN_p1"] MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: year=col(source(s), name("year"), unit.category())
  DATA: MEAN_p1=col(source(s), name("MEAN_p1"))
  GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))
  GUIDE: axis(dim(2), label("Estimated hazard probability"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: line(position(year*MEAN_p1), missing.wings())
END GPL.


GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=year
MEAN(p2)[name="MEAN_p2"] MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
```

```
  DATA: year=col(source(s), name("year"), unit.category())
  DATA: MEAN_p2=col(source(s), name("MEAN_p2"))
  GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))
  GUIDE: axis(dim(2), label("Estimated survival probability"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: line(position(year*MEAN_p2), missing.wings())
END GPL.
```

```
** Begin Chap 11 of S&W, 2003

GET DATA /TYPE=XLSX
  /FILE='/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_period_clean.xlsx'
  /SHEET=name 'Sheet1'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.

FREQUENCIES
 VARIABLES= id
 /FORMAT=limit(10)
 /statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW
KURTOSIS SEKURT
 /HISTOGRAM NORMAL
 /order= analysis .
execute.


FREQUENCIES
 VARIABLES= lncashflow
 /FORMAT=limit(10)
 /statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW
KURTOSIS SEKURT
```

```
 /HISTOGRAM NORMAL
 /order= analysis .
execute.

descriptives event instsize2 fouryear public prestige lncashflow.
** note that degree granting institutions are eliminated when I use listwise deletion.
** also note that I transformed cashflow to lncashflow when cleaning data in stata because it was not normally distributed.

crosstabs
/tables=period by event
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cell.

** descriptives for event by lncashflow, binary outcome & continuous predictor = ttest

T-test
Groups= event (0,1)
 /missing=Analysis
 /variables=lncashflow
 /criteria=CI(.95).

** descriptives for event by lncashflow, binary outcome & binary predictors = chisquare

crosstabs
/tables=event by instsize2
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.

crosstabs
/tables=event by fouryear
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.
```

```
crosstabs
/tables=event by public
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.

crosstabs
/tables=event by prestige
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.

 ** check for outliers

REGRESSION
  /MISSING PAIRWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT id
  /METHOD=ENTER lncashflow
  /SAVE MAHAL.

EXAMINE VARIABLES=MAH_1
  /PLOT BOXPLOT STEMLEAF
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

** Drop outlier cases: # 12789, 12797, 13204, 14481

USE ALL.
COMPUTE filter_$=(MAH_1 <= 27.877).
VARIABLE LABELS filter_$ 'MAH_1 <= 27.877 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
```

```
FILTER BY filter_$.
EXECUTE.
```

** Check for initial predictor significance, NEED HELP HERE: I don't understand why my values are significant in STATA but not in SPSS.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV SIG N
  /MISSING PAIRWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT event
  /METHOD=ENTER d1 d2 d3 d4 instsize2 fouryear public lncashflow prestige
  /SAVE  MAHAL.
```

** Begin analysis for time using survival modeling

```
logistic regression var = event
    /method = enter d1 d2 d3 d4
    /method = enter d1 d2 d3 d4 instsize2
    /method = enter d1 d2 d3 d4 fouryear
    /method = enter d1 d2 d3 d4 public
    /method = enter d1 d2 d3 d4 prestige
    /method = enter d1 d2 d3 d4 lncashflow
    /method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow
    /origin.
```

** ALTERNATIVE CODE from UCLA website

** model a

```
logistic regression var = event
 /method = enter d1 d2 d3 d4
 /origin.
```

** model b

```
logistic regression var = event
```

```
 /method = enter d1 d2 d3 d4 instsize2
 /origin.

** model c

logistic regression var = event
 /method = enter d1 d2 d3 d4 fouryear
 /origin.

** model d

logistic regression var = event
 /method = enter d1 d2 d3 d4  public
 /origin.

** model e

logistic regression var = event
 /method = enter d1 d2 d3 d4  prestige
 /origin.

** model f

logistic regression var = event
 /method = enter d1 d2 d3 d4  lncashflow
 /origin.

** model g

logistic regression var = event
 /method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow
 /origin.
```

**Output**

```
** note that degree granting institutions are eliminated when I use listwise deletion.
* Encoding: UTF-8.
** Lara Pheatt
** Purpose: Preliminary MOOCs Hazard Modeling for Bowers Midterm, Spring 2016
** Data used: Derived/hazmod_institution_period_clean.xlsx, originally cleaned in STATA
```

```
** Life Table, p. 327 of S & W, 2003. Chapter 10.

GET DATA /TYPE=TXT
  /FILE='/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean.txt'
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=2
  /IMPORTCASE=ALL
  /VARIABLES=
  id F3.0
  time F1.0
  censor F1.0.
CACHE.
Execute.

** Makes Life Table

procedure output outfile= '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean1.txt'.
survival table = time
 /interval = thru 4 by 1
 /status = censor(0)
 /write=tables.
```

**Survival Analysis**

**Notes**

| | |
|---|---|
| Output Created | 28-MAR-2016 00:09:21 |
| Comments | |

| | | |
|---|---|---|
| Input | Data | /Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean.txt |
| | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 3825 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Cases with missing values on a variable are excluded from any calculation involving that variable. |
| Syntax | | survival table = time /interval = thru 4 by 1 /status = censor(0) /write=tables. |
| Resources | Processor Time | 00:00:00.01 |
| | Elapsed Time | 00:00:00.00 |
| File Saved | Life Table Output File | /Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean1.txt |

**Survival Variable: time**

**Life Table[a]**

| Interval Start Time | Number Entering Interval | Number Withdrawing during Interval | Number Exposed to Risk | Number of Terminal Events | Proportion Terminating | Proportion Surviving |
|---|---|---|---|---|---|---|
| 0 | 3825 | 0 | 3825.000 | 0 | .00 | 1.00 |
| 1 | 3825 | 0 | 3825.000 | 7 | .00 | 1.00 |
| 2 | 3818 | 0 | 3818.000 | 32 | .01 | .99 |
| 3 | 3786 | 0 | 3786.000 | 59 | .02 | .98 |
| 4 | 3727 | 3685 | 1884.500 | 42 | .02 | .98 |

**Life Table[a]**

| Interval Start Time | Cumulative Proportion Surviving at End of Interval | Std. Error of Cumulative Proportion Surviving at End of Interval | Probability Density | Std. Error of Probability Density | Hazard Rate | Std. Error of Hazard Rate |
|---|---|---|---|---|---|---|
| 0 | 1.00 | .00 | .000 | .000 | .00 | .00 |
| 1 | 1.00 | .00 | .002 | .001 | .00 | .00 |
| 2 | .99 | .00 | .008 | .001 | .01 | .00 |
| 3 | .97 | .00 | .015 | .002 | .02 | .00 |
| 4 | .95 | .00 | .000 | .000 | .00 | .00 |

a. The median survival time is 4.00

** AUTHOR IMPORTANT TO READ & DO: edit hazard model in .txt to get it into spss acceptable format --> hazmod_institution_level_clean2.csv

```
get data
 /type = txt
 /file = '/Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean2.csv'
```

```
 /delimiters = ","
 /arrangement = delimited
 /variables =
 v1 f5.2
 v2 f7.2
 v3 f6.2
 v4 f6.2
 v5 f6.2.
execute.

save outfile  '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav'
 /drop v4
 /rename v1=year v2=begin v3=censored v5=left.

get file '/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav'.
compute p1 = left/begin.
do if $casenum = 1.
compute p2 = 1.
else.
compute temp = lag(p2).
compute p2 = (1-p1)*temp.
format p1 p2 (f6.4) year begin left (f4.0) censored (f4.0).
end if.
if $casenum = 1 p1 =$sysmis.
list year begin left censored p1 p2.
```

**List**

| Notes | |
|---|---|
| Output Created | 28-MAR-2016 00:09:22 |
| Comments | |

| | | |
|---|---|---|
| Input | Data | /Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav |
| | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 5 |
| Syntax | | list year begin left censored p1 p2. |
| Resources | Processor Time | 00:00:00.00 |
| | Elapsed Time | 00:00:00.00 |

/Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean3.sav

| year | begin | left | censored | p1 | p2 |
|---|---|---|---|---|---|
| 0 | 3825 | 0 | 0 | . | 1.0000 |
| 1 | 3825 | 7 | 0 | .0018 | .9982 |
| 2 | 3818 | 32 | 0 | .0084 | .9898 |
| 3 | 3786 | 59 | 0 | .0156 | .9744 |
| 4 | 3727 | 42 | 3685 | .0113 | .9634 |

Number of cases read: 5 Number of cases listed: 5

save outfile '/Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean4.sav'.

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=year
MEAN(p1)[name="MEAN_p1"] MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: year=col(source(s), name("year"), unit.category())
  DATA: MEAN_p1=col(source(s), name("MEAN_p1"))
  GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))
  GUIDE: axis(dim(2), label("Estimated hazard probability"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: line(position(year*MEAN_p1), missing.wings())
END GPL.
```

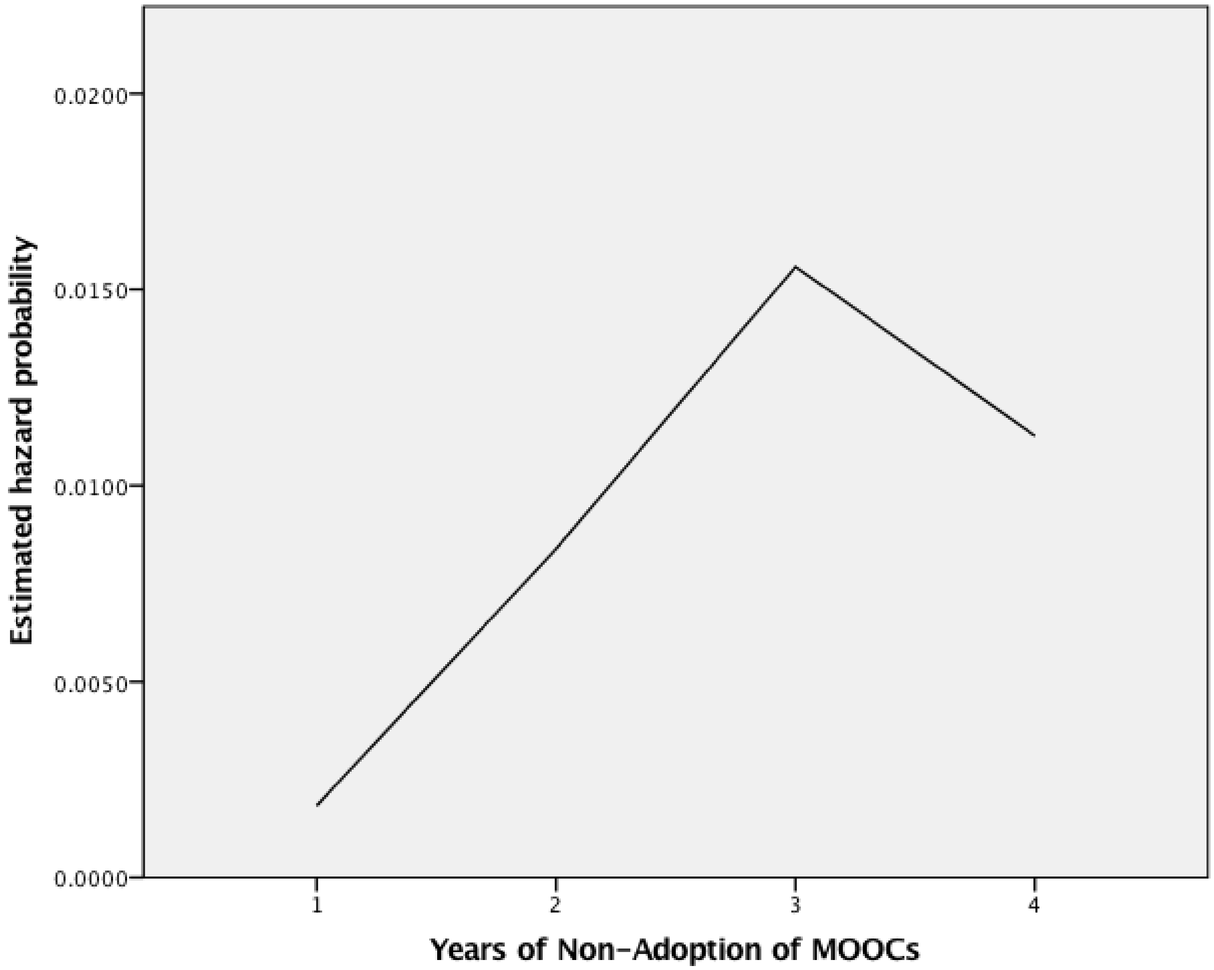
**GGraph**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:22 |
| Comments | | |
| Input | Data | /Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean4.sav |
| | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 5 |

| | | |
|---|---|---|
| Syntax | | GGRAPH<br>/GRAPHDATASET NAME="graphdataset" VARIABLES=year MEAN(p1)[name="MEAN_p1"] MISSING=LISTWISE<br>REPORTMISSING=NO<br>/GRAPHSPEC SOURCE=INLINE.<br>BEGIN GPL<br>SOURCE: s=userSource(id("graphdataset"))<br>DATA: year=col(source(s), name("year"), unit.category())<br>DATA: MEAN_p1=col(source(s), name("MEAN_p1"))<br>GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))<br>GUIDE: axis(dim(2), label("Estimated hazard probability"))<br>SCALE: linear(dim(2), include(0))<br>ELEMENT: line(position(year*MEAN_p1), missing.wings())<br>END GPL. |
| Resources | Processor Time | 00:00:00.16 |
| | Elapsed Time | 00:00:00.00 |

/Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean4.sav




```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=year
MEAN(p2)[name="MEAN_p2"] MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
```

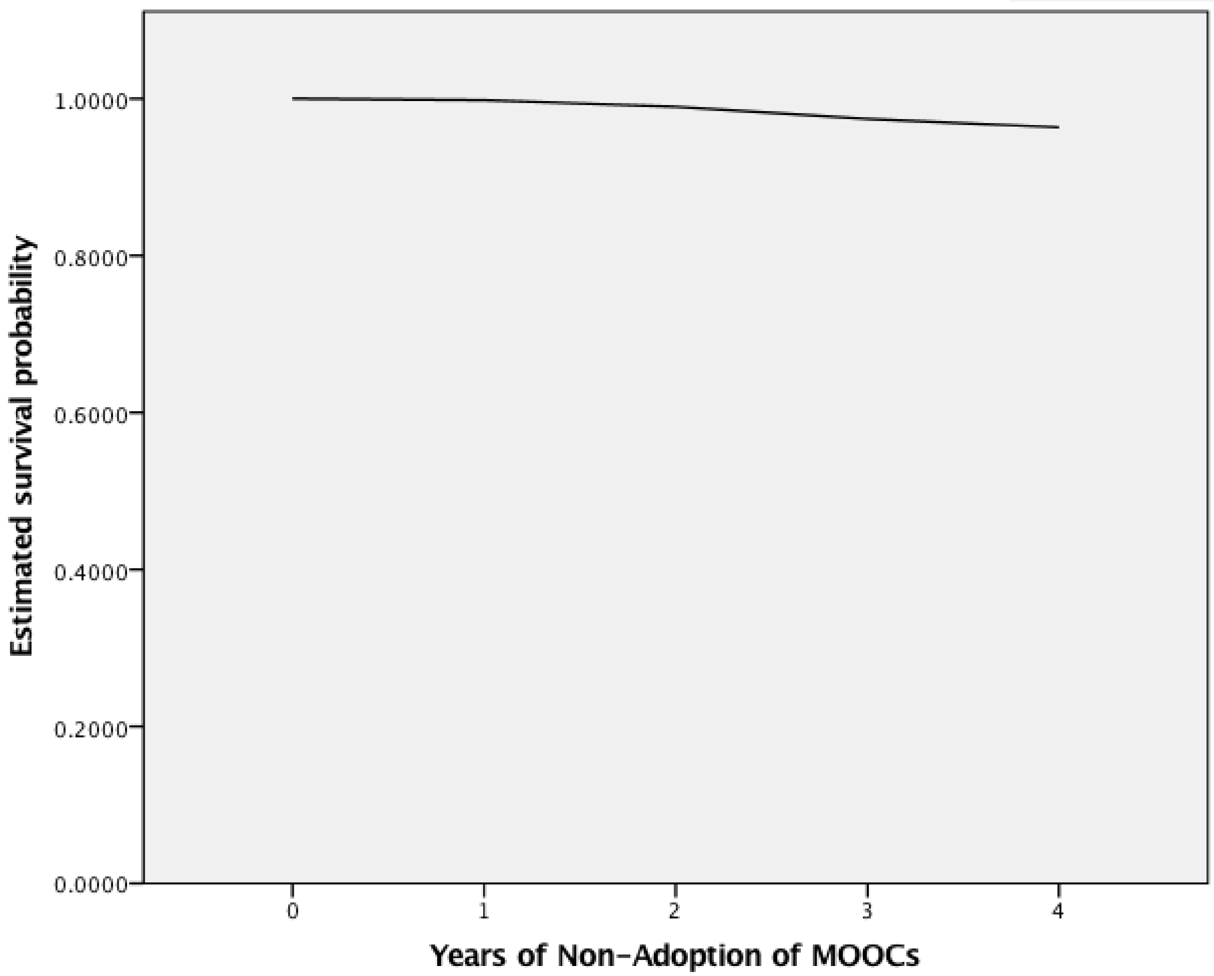
```
  DATA: year=col(source(s), name("year"), unit.category())
  DATA: MEAN_p2=col(source(s), name("MEAN_p2"))
  GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))
  GUIDE: axis(dim(2), label("Estimated survival probability"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: line(position(year*MEAN_p2), missing.wings())
END GPL.
```

**GGraph**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:22 |
| Comments | | |
| Input | Data | /Users/Lara/Dropbox/TC - General/PhD Ideas/Quant/Data/Derived/hazmod_institution_level_clean4.sav |
| | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 5 |

| | | |
|---|---|---|
| Syntax | | GGRAPH<br>/GRAPHDATASET NAME="graphdataset" VARIABLES=year MEAN(p2)[name="MEAN_p2"] MISSING=LISTWISE<br>REPORTMISSING=NO<br>/GRAPHSPEC SOURCE=INLINE.<br>BEGIN GPL<br>SOURCE: s=userSource(id("graphdataset"))<br>DATA: year=col(source(s), name("year"), unit.category())<br>DATA: MEAN_p2=col(source(s), name("MEAN_p2"))<br>GUIDE: axis(dim(1), label("Years of Non-Adoption of MOOCs"))<br>GUIDE: axis(dim(2), label("Estimated survival probability"))<br>SCALE: linear(dim(2), include(0))<br>ELEMENT: line(position(year*MEAN_p2), missing.wings())<br>END GPL. |
| Resources | Processor Time | 00:00:00.16 |
| | Elapsed Time | 00:00:00.00 |




** Begin Chap 11 of S&W, 2003

```
GET DATA /TYPE=XLSX
  /FILE='/Users/Lara/Dropbox/TC - General/PhD
Ideas/Quant/Data/Derived/hazmod_institution_period_clean.xlsx'
  /SHEET=name 'Sheet1'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.

FREQUENCIES
 VARIABLES= id
 /FORMAT=limit(10)
 /statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW
KURTOSIS SEKURT
 /HISTOGRAM NORMAL
 /order= analysis .
```

**Frequencies**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:23 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics are based on all cases with valid data. |

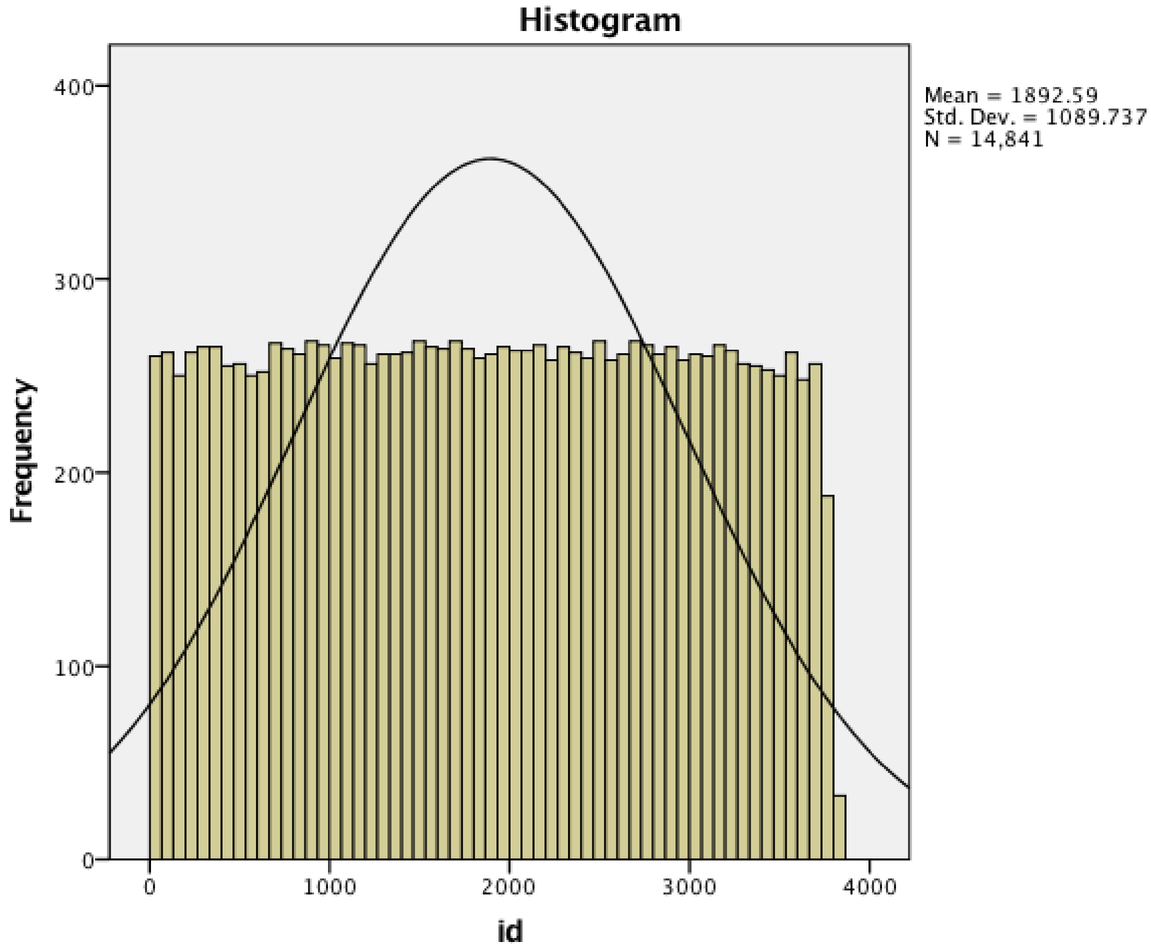
| | | |
|---|---|---|
| Syntax | | FREQUENCIES<br>VARIABLES= id<br>/FORMAT=limit(10)<br>/statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW KURTOSIS SEKURT<br>/HISTOGRAM NORMAL<br>/order= analysis . |
| Resources | Processor Time | 00:00:00.25 |
| | Elapsed Time | 00:00:00.00 |

**Statistics**

id

| | | |
|---|---|---|
| N | Valid | 14841 |
| | Missing | 0 |
| Mean | | 1892.59 |
| Median | | 1891.00 |
| Mode | | 2[a] |
| Std. Deviation | | 1089.737 |
| Skewness | | .002 |
| Std. Error of Skewness | | .020 |
| Kurtosis | | -1.190 |
| Std. Error of Kurtosis | | .040 |
| Minimum | | 1 |
| Maximum | | 3826 |

a. Multiple modes exist. The smallest value is shown




execute.

```
FREQUENCIES
 VARIABLES= lncashflow
 /FORMAT=limit(10)
 /statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW
KURTOSIS SEKURT
 /HISTOGRAM NORMAL
```

/order= analysis .

**Frequencies**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:23 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics are based on all cases with valid data. |
| Syntax | | FREQUENCIES VARIABLES= lncashflow /FORMAT=limit(10) /statistics=stddev MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW KURTOSIS SEKURT /HISTOGRAM NORMAL /order= analysis . |
| Resources | Processor Time | 00:00:00.47 |
| | Elapsed Time | 00:00:01.00 |

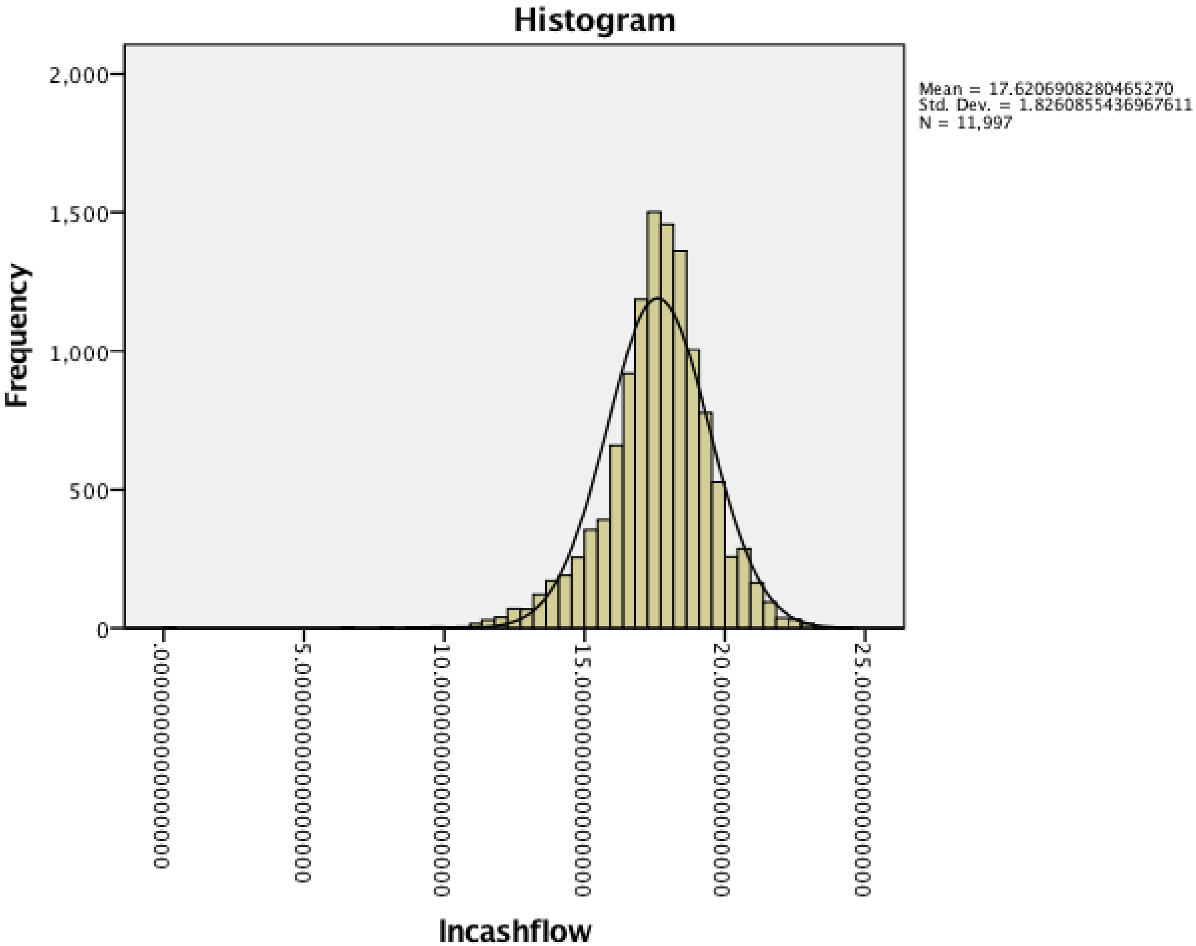
**Statistics**

lncashflow

| | | |
|---|---|---|
| N | Valid | 11997 |

| | |
|---|---|
| Missing | 2844 |
| Mean | 17.620690828046570 |
| Median | 17.732652664184570 |
| Mode | 19.1252670288085940[a] |
| Std. Deviation | 1.826085543696751 |
| Skewness | -.630 |
| Std. Error of Skewness | .022 |
| Kurtosis | 2.542 |
| Std. Error of Kurtosis | .045 |
| Minimum | .0000000000000000 |
| Maximum | 24.3572483062744140 |

a. Multiple modes exist. The smallest value is shown




execute.

descriptives event instsize2 fouryear public prestige lncashflow.

**Descriptives**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User defined missing values are treated as missing. |
| | Cases Used | All non-missing data are used. |
| Syntax | | descriptives event instsize2 fouryear public prestige lncashflow. |
| Resources | Processor Time | 00:00:00.01 |
| | Elapsed Time | 00:00:00.00 |

**Descriptive Statistics**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| event | 14841 | 0 | 1 | .01 | .097 |
| instsize2 | 14483 | 0 | 1 | .29 | .454 |
| fouryear | 14841 | 0 | 1 | .65 | .477 |
| public | 14841 | 0 | 1 | .49 | .500 |
| prestige | 14841 | 0 | 1 | .07 | .255 |
| lncashflow | 11997 | .0000000000000000 | 24.35724830627 44140 | 17.62069082804 6556 | 1.826085543696 740 |
| Valid N (listwise) | 11993 | | | | |

** note that degree granting institutions are eliminated when I use listwise deletion.
** also note that I transformed cashflow to lncashflow when cleaning data in stata because it was not normally distributed.

crosstabs
/tables=period by event

/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cell.

**Crosstabs**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table. |
| Syntax | | crosstabs<br>/tables=period by event<br>/format=avalue tables<br>/statistics=chisq<br>/cells=count row column total<br>/count round cell. |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |
| | Dimensions Requested | 2 |
| | Cells Available | 524245 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| period * event | 14841 | 100.0% | 0 | 0.0% | 14841 | 100.0% |

**period * event Crosstabulation**

| | | | event | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | Total |
| period | 1 | Count | 3702 | 7 | 3709 |
| | | % within period | 99.8% | 0.2% | 100.0% |
| | | % within event | 25.2% | 5.0% | 25.0% |
| | | % of Total | 24.9% | 0.0% | 25.0% |
| | 2 | Count | 3678 | 32 | 3710 |
| | | % within period | 99.1% | 0.9% | 100.0% |
| | | % within event | 25.0% | 22.9% | 25.0% |
| | | % of Total | 24.8% | 0.2% | 25.0% |
| | 3 | Count | 3666 | 59 | 3725 |
| | | % within period | 98.4% | 1.6% | 100.0% |
| | | % within event | 24.9% | 42.1% | 25.1% |
| | | % of Total | 24.7% | 0.4% | 25.1% |
| | 4 | Count | 3655 | 42 | 3697 |
| | | % within period | 98.9% | 1.1% | 100.0% |
| | | % within event | 24.9% | 30.0% | 24.9% |
| | | % of Total | 24.6% | 0.3% | 24.9% |
| Total | | Count | 14701 | 140 | 14841 |
| | | % within period | 99.1% | 0.9% | 100.0% |
| | | % within event | 100.0% | 100.0% | 100.0% |
| | | % of Total | 99.1% | 0.9% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 40.687[a] | 3 | .000 |
| Likelihood Ratio | 48.891 | 3 | .000 |
| Linear-by-Linear Association | 25.242 | 1 | .000 |
| N of Valid Cases | 14841 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 34.88.

** descriptives for event by lncashflow, binary outcome & continuous predictor = ttest

T-test
Groups= event (0,1)
/missing=Analysis
/variables=lncashflow
/criteria=CI(.95).

**T-Test**

**Notes**

| | | | |
|---|---|---|---|
| Output Created | | | 28-MAR-2016 00:09:24 |
| Comments | | | |
| Input | Filter | <none> | |
| | Weight | <none> | |
| | Split File | <none> | |
| | N of Rows in Working Data File | | 14841 |

| | | |
|---|---|---|
| Missing Value Handling | Definition of Missing | User defined missing values are treated as missing. |
| | Cases Used | Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis. |
| Syntax | | T-test<br>Groups= event (0,1)<br>/missing=Analysis<br>/variables=lncashflow<br>/criteria=CI(.95). |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |

**Group Statistics**

| | event | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| lncashflow | 0 | 11864 | 17.589827243090480 | 1.802417131790307 | .016547779971854 |
| | 1 | 133 | 20.373815428941770 | 1.831758347603016 | .158833592496560 |

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | | | |
| lncashflow | Equal variances assumed | 1.801 | .180 | -17.711 | 11995 | .000 | -2.783988185851293 | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal variances not assumed | -17.433 | 134.881 | .000 | -2.783988185851293 | | | |

**Independent Samples Test**

| | | t-test for Equality of Means | | |
|---|---|---|---|---|
| | | | 95% Confidence Interval of the Difference | |
| | | Std. Error Difference | Lower | Upper |
| lncashflow | Equal variances assumed | .157191357996785 | -3.092108677287715 | -2.475867694414870 |
| | Equal variances not assumed | .159693265754572 | -3.099814847053851 | -2.468161524648735 |

** descriptives for event by lncashflow, binary outcome & binary predictors = chisquare

```
crosstabs
/tables=event by instsize2
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.
```

**Crosstabs**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |

| | | |
|---|---|---|
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table. |
| Syntax | | crosstabs<br>/tables=event by instsize2<br>/format=avalue tables<br>/statistics=chisq<br>/cells=count row column total<br>/count round cel. |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |
| | Dimensions Requested | 2 |
| | Cells Available | 524245 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| event * instsize2 | 14483 | 97.6% | 358 | 2.4% | 14841 | 100.0% |

**event * instsize2 Crosstabulation**

| | | | instsize2 | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | Total |
| event | 0 | Count | 10241 | 4102 | 14343 |
| | | % within event | 71.4% | 28.6% | 100.0% |
| | | % within instsize2 | 99.7% | 97.4% | 99.0% |
| | | % of Total | 70.7% | 28.3% | 99.0% |

| | | | | |
|---|---|---|---|---|
| 1 | Count | 32 | 108 | 140 |
| | % within event | 22.9% | 77.1% | 100.0% |
| | % within instsize2 | 0.3% | 2.6% | 1.0% |
| | % of Total | 0.2% | 0.7% | 1.0% |
| Total | Count | 10273 | 4210 | 14483 |
| | % within event | 70.9% | 29.1% | 100.0% |
| | % within instsize2 | 100.0% | 100.0% | 100.0% |
| | % of Total | 70.9% | 29.1% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 158.457[a] | 1 | .000 | | |
| Continuity Correction[b] | 156.111 | 1 | .000 | | |
| Likelihood Ratio | 139.876 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 158.446 | 1 | .000 | | |
| N of Valid Cases | 14483 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 40.70.
b. Computed only for a 2x2 table

```
crosstabs
/tables=event by fouryear
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.
```

**Crosstabs**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table. |
| Syntax | | crosstabs /tables=event by fouryear /format=avalue tables /statistics=chisq /cells=count row column total /count round cel. |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |
| | Dimensions Requested | 2 |
| | Cells Available | 524245 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| event * fouryear | 14841 | 100.0% | 0 | 0.0% | 14841 | 100.0% |

**event * fouryear Crosstabulation**

| | | | fouryear | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | Total |
| event | 0 | Count | 5195 | 9506 | 14701 |
| | | % within event | 35.3% | 64.7% | 100.0% |
| | | % within fouryear | 99.8% | 98.7% | 99.1% |
| | | % of Total | 35.0% | 64.1% | 99.1% |
| | 1 | Count | 10 | 130 | 140 |
| | | % within event | 7.1% | 92.9% | 100.0% |
| | | % within fouryear | 0.2% | 1.3% | 0.9% |
| | | % of Total | 0.1% | 0.9% | 0.9% |
| Total | | Count | 5205 | 9636 | 14841 |
| | | % within event | 35.1% | 64.9% | 100.0% |
| | | % within fouryear | 100.0% | 100.0% | 100.0% |
| | | % of Total | 35.1% | 64.9% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 48.413[a] | 1 | .000 | | |
| Continuity Correction[b] | 47.183 | 1 | .000 | | |
| Likelihood Ratio | 61.653 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 48.410 | 1 | .000 | | |
| N of Valid Cases | 14841 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 49.10.

b. Computed only for a 2x2 table

```
crosstabs
/tables=event by public
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.
```

**Crosstabs**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table. |
| Syntax | | crosstabs /tables=event by public /format=avalue tables /statistics=chisq /cells=count row column total /count round cel. |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |

| | | |
|---|---|---|
| | Dimensions Requested | 2 |
| | Cells Available | 524245 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| event * public | 14841 | 100.0% | 0 | 0.0% | 14841 | 100.0% |

**event * public Crosstabulation**

| | | | public | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | Total |
| event | 0 | Count | 7559 | 7142 | 14701 |
| | | % within event | 51.4% | 48.6% | 100.0% |
| | | % within public | 99.2% | 98.9% | 99.1% |
| | | % of Total | 50.9% | 48.1% | 99.1% |
| | 1 | Count | 64 | 76 | 140 |
| | | % within event | 45.7% | 54.3% | 100.0% |
| | | % within public | 0.8% | 1.1% | 0.9% |
| | | % of Total | 0.4% | 0.5% | 0.9% |
| Total | | Count | 7623 | 7218 | 14841 |
| | | % within event | 51.4% | 48.6% | 100.0% |
| | | % within public | 100.0% | 100.0% | 100.0% |
| | | % of Total | 51.4% | 48.6% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 1.806[a] | 1 | .179 | | |

| | | | | | |
|---|---|---|---|---|---|
| Continuity Correction[b] | 1.585 | 1 | .208 | | |
| Likelihood Ratio | 1.806 | 1 | .179 | | |
| Fisher's Exact Test | | | | .202 | .104 |
| Linear-by-Linear Association | 1.806 | 1 | .179 | | |
| N of Valid Cases | 14841 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 68.09.
b. Computed only for a 2x2 table

```
crosstabs
/tables=event by prestige
/format=avalue tables
/statistics=chisq
/cells=count row column total
/count round cel.
```

**Crosstabs**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |

| | | |
|---|---|---|
| | Cases Used | Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table. |
| Syntax | | crosstabs<br>/tables=event by prestige<br>/format=avalue tables<br>/statistics=chisq<br>/cells=count row column total<br>/count round cel. |
| Resources | Processor Time | 00:00:00.02 |
| | Elapsed Time | 00:00:00.00 |
| | Dimensions Requested | 2 |
| | Cells Available | 524245 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| event * prestige | 14841 | 100.0% | 0 | 0.0% | 14841 | 100.0% |

**event * prestige Crosstabulation**

| | | | prestige | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | Total |
| event | 0 | Count | 13746 | 955 | 14701 |
| | | % within event | 93.5% | 6.5% | 100.0% |
| | | % within prestige | 99.6% | 91.8% | 99.1% |
| | | % of Total | 92.6% | 6.4% | 99.1% |
| | 1 | Count | 55 | 85 | 140 |
| | | % within event | 39.3% | 60.7% | 100.0% |
| | | % within prestige | 0.4% | 8.2% | 0.9% |
| | | % of Total | 0.4% | 0.6% | 0.9% |

| | | | | |
|---|---|---|---|---|
| Total | Count | 13801 | 1040 | 14841 |
| | % within event | 93.0% | 7.0% | 100.0% |
| | % within prestige | 100.0% | 100.0% | 100.0% |
| | % of Total | 93.0% | 7.0% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 625.581[a] | 1 | .000 | | |
| Continuity Correction[b] | 617.288 | 1 | .000 | | |
| Likelihood Ratio | 278.318 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 625.538 | 1 | .000 | | |
| N of Valid Cases | 14841 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.81.
b. Computed only for a 2x2 table

** check for outliers

```
REGRESSION
 /MISSING PAIRWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT id
 /METHOD=ENTER lncashflow
 /SAVE MAHAL.
```

**Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| | Cases Used | Correlation coefficients for each pair of variables are based on all the cases with valid data for that pair. Regression statistics are based on these correlations. |
| Syntax | | REGRESSION /MISSING PAIRWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT id /METHOD=ENTER lncashflow /SAVE MAHAL. |
| Resources | Processor Time | 00:00:00.03 |
| | Elapsed Time | 00:00:00.00 |
| | Memory Required | 2832 bytes |
| | Additional Memory Required for Residual Plots | 0 bytes |
| Variables Created or Modified | MAH_1 | Mahalanobis Distance |

**Variables Entered/Removed[a]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | lncashflow[b] | . | Enter |

a. Dependent Variable: id
b. All requested variables entered.

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .205[a] | .042 | .042 | 1066.541 |

a. Predictors: (Constant), lncashflow
b. Dependent Variable: id

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 601142982.085 | 1 | 601142982.085 | 528.473 | .000[b] |
| | Residual | 13644430532.473 | 11995 | 1137509.840 | | |
| | Total | 14245573514.559 | 11996 | | | |

a. Dependent Variable: id
b. Predictors: (Constant), lncashflow

**Coefficients[a]**

| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | (Constant) | 4052.680 | 94.467 | | 42.900 | .000 |

| | | | | | |
|---|---|---|---|---|---|
| lncashflow | -122.588 | 5.333 | -.205 | -22.989 | .000 |

a. Dependent Variable: id

**Residuals Statistics[a]**

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | 1066.76 | 4052.68 | 1892.59 | 223.857 | 11997 |
| Std. Predicted Value | -3.689 | 9.649 | .000 | 1.000 | 11997 |
| Standard Error of Predicted Value | 9.737 | 94.467 | 12.987 | 4.580 | 11997 |
| Adjusted Predicted Value | 1066.51 | 4059.00 | 1892.58 | 223.855 | 11997 |
| Residual | -2529.958 | 2200.886 | -72.272 | 1028.823 | 11997 |
| Std. Residual | -2.372 | 2.064 | -.068 | .965 | 11997 |
| Stud. Residual | -2.373 | 2.064 | -.068 | .965 | 11997 |
| Deleted Residual | -2532.260 | 2201.396 | -72.266 | 1029.003 | 11997 |
| Stud. Deleted Residual | -2.374 | 2.064 | -.068 | .965 | 11997 |
| Mahal. Distance | .000 | 93.112 | 1.000 | 2.131 | 11997 |
| Cook's Distance | .000 | .003 | .000 | .000 | 11997 |
| Centered Leverage Value | .000 | .008 | .000 | .000 | 11997 |

a. Dependent Variable: id

```
EXAMINE VARIABLES=MAH_1
 /PLOT BOXPLOT STEMLEAF
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

**Explore**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:24 |
| Comments | | |
| Input | Filter | <none> |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 14841 |
| Missing Value Handling | Definition of Missing | User-defined missing values for dependent variables are treated as missing. |
| | Cases Used | Statistics are based on cases with no missing values for any dependent variable or factor used. |
| Syntax | | EXAMINE VARIABLES=MAH_1 /PLOT BOXPLOT STEMLEAF /COMPARE GROUPS /STATISTICS DESCRIPTIVES EXTREME /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL. |
| Resources | Processor Time | 00:00:00.63 |
| | Elapsed Time | 00:00:01.00 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| Mahalanobis Distance | 11997 | 80.8% | 2844 | 19.2% | 14841 | 100.0% |

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| Mahalanobis Distance | Mean | | .9999166 | .01945434 |
| | 95% Confidence Interval for Mean | Lower Bound | .9617830 | |
| | | Upper Bound | 1.0380503 | |
| | 5% Trimmed Mean | | .7094967 | |
| | Median | | .3122464 | |
| | Variance | | 4.541 | |
| | Std. Deviation | | 2.13085011 | |
| | Minimum | | .00000 | |
| | Maximum | | 93.11156 | |
| | Range | | 93.11156 | |
| | Interquartile Range | | 1.00130 | |
| | Skewness | | 15.986 | .022 |
| | Kurtosis | | 598.328 | .045 |

**Extreme Values**

| | | | Case Number | Value |
|---|---|---|---|---|
| Mahalanobis Distance | Highest | 1 | 12789 | 93.11156 |
| | | 2 | 12797 | 93.11156 |
| | | 3 | 13204 | 36.66228 |
| | | 4 | 14481 | 28.67441 |
| | | 5 | 9083 | 23.35196 |
| | Lowest | 1 | 6357 | .00000 |
| | | 2 | 1087 | .00000 |
| | | 3 | 14366 | .00000 |
| | | 4 | 583 | .00000 |
| | | 5 | 11225 | .00000 |

**Mahalanobis Distance**

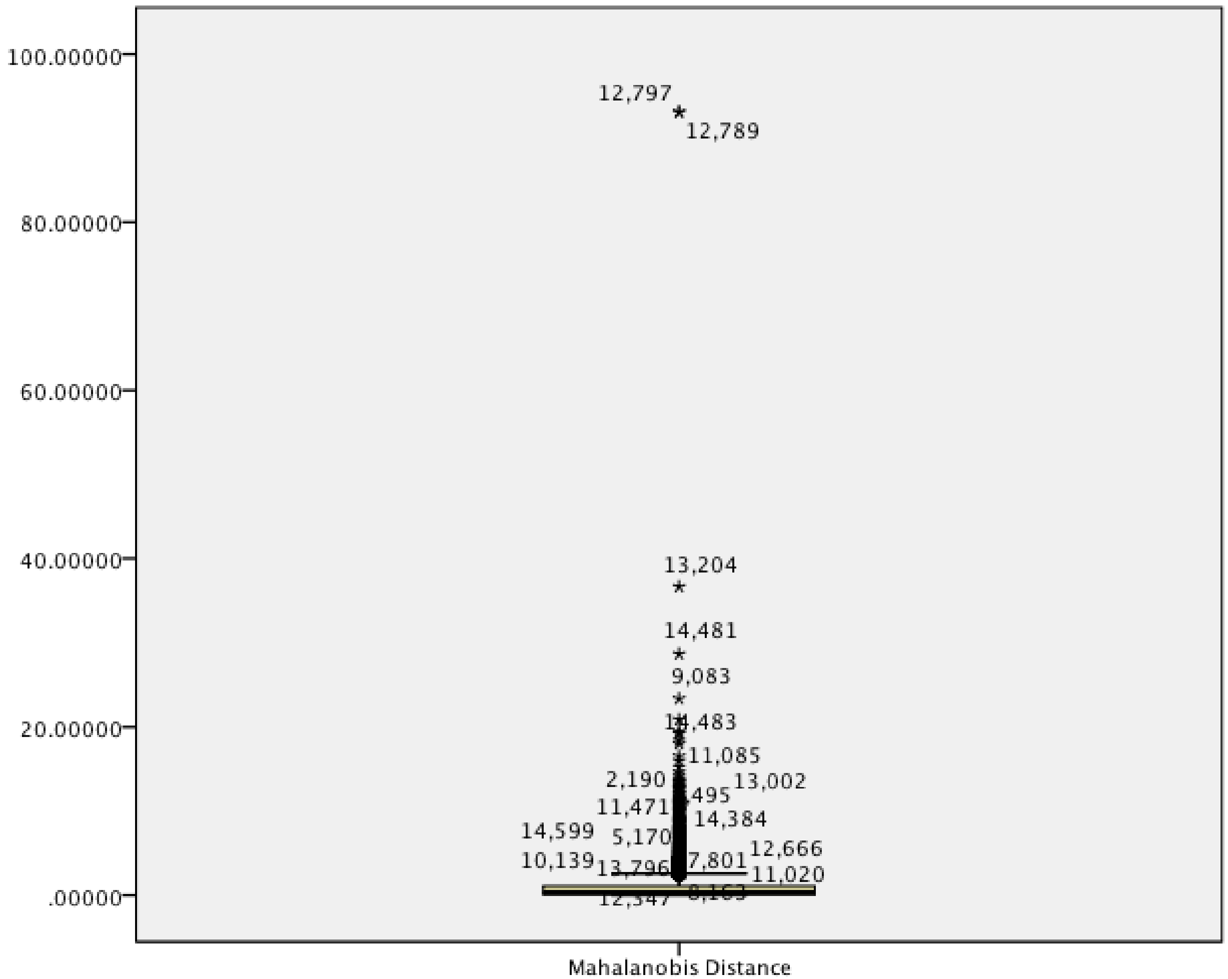
Mahalanobis Distance Stem-and-Leaf Plot

```
 Frequency    Stem &  Leaf

  3663.00        0 .
00000000000000000000000000000001111111111111222222222333333344444455555666667
77778888899999
  1313.00        1 .  00001111222333444555666777888999
   929.00        2 .  00011122334445566778899
   678.00        3 .  0011223344556677889
   528.00        4 .  00123456789
   457.00        5 .  0123456789
   397.00        6 .  0123456789
   348.00        7 .  0123456789
   263.00        8 .  0123456789
   258.00        9 .  0123456789
   233.00       10 .  125679&&
   163.00       11 .  458&&
   149.00       12 .  16&&&
   158.00       13 .  56&&&
   145.00       14 .  1&&&
   147.00       15 .  &&&&
   123.00       16 .  &&&
   121.00       17 .  &&&
    84.00       18 .  &&
    73.00       19 .  &&
    81.00       20 .  &&
    69.00       21 .  &&
    84.00       22 .  &&
    66.00       23 .  &&
    69.00       24 .  &&
    51.00       25 .  &
  1347.00 Extremes    (>=2.57)

 Stem width:    .10000
 Each leaf:       40 case(s)
```

& denotes fractional leaves.




** Drop outlier cases: # 12789, 12797, 13204, 14481

```
USE ALL.
COMPUTE filter_$=(MAH_1 <= 27.877).
VARIABLE LABELS filter_$ 'MAH_1 <= 27.877 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

** Check for initial predictor signficance, NEED HELP HERE: I don't understand why my values are signficant in STATA but not in SPSS.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV SIG N
  /MISSING PAIRWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT event
  /METHOD=ENTER d1 d2 d3 d4 instsize2 fouryear public lncashflow prestige
  /SAVE  MAHAL.
```

**Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |

| | | |
|---|---|---|
| | Cases Used | Correlation coefficients for each pair of variables are based on all the cases with valid data for that pair. Regression statistics are based on these correlations. |
| Syntax | | REGRESSION<br>/DESCRIPTIVES MEAN STDDEV SIG N<br>/MISSING PAIRWISE<br>/STATISTICS COEFF OUTS R ANOVA COLLIN TOL<br>/CRITERIA=PIN(.05) POUT(.10)<br>/NOORIGIN<br>/DEPENDENT event<br>/METHOD=ENTER d1 d2 d3 d4 instsize2 fouryear public lncashflow prestige<br>/SAVE MAHAL. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |
| | Memory Required | 9824 bytes |
| | Additional Memory Required for Residual Plots | 0 bytes |
| Variables Created or Modified | MAH_2 | Mahalanobis Distance |

**Descriptive Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| event | .01 | .105 | 11993 |
| d1 | .25 | .435 | 11993 |
| d2 | .25 | .434 | 11993 |
| d3 | .25 | .433 | 11993 |
| d4 | .24 | .429 | 11993 |
| instsize2 | .31 | .463 | 11989 |

| | | | |
|---|---|---|---|
| fouryear | .70 | .458 | 11993 |
| public | .47 | .499 | 11993 |
| lncashflow | 17.62536660797 7096 | 1.807132576554 185 | 11993 |
| prestige | .08 | .275 | 11993 |

**Correlations**

| | | event | d1 | d2 | d3 | d4 | instsize2 | fouryear | public | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sig. (1-tailed) | event | . | .000 | .238 | .000 | .041 | .000 | .000 | .141 | | |
| | d1 | .000 | . | .000 | .000 | .000 | .026 | .243 | .358 | | |
| | d2 | .238 | .000 | . | .000 | .000 | .209 | .439 | .350 | | |
| | d3 | .000 | .000 | .000 | . | .000 | .343 | .398 | .414 | | |
| | d4 | .041 | .000 | .000 | .000 | . | .009 | .385 | .296 | | |
| | instsize2 | .000 | .026 | .209 | .343 | .009 | . | .000 | .000 | | |
| | fouryear | .000 | .243 | .439 | .398 | .385 | .000 | . | .000 | | |
| | public | .141 | .358 | .350 | .414 | .296 | .000 | .000 | . | | |
| | lncashflow | .000 | .278 | .409 | .433 | .423 | .000 | .000 | .000 | | |
| | prestige | .000 | .059 | .098 | .390 | .005 | .000 | .000 | .000 | | |
| N | event | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | d1 | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | d2 | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | d3 | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | d4 | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | instsize2 | 11989 | 11989 | 11989 | 11989 | 11989 | 11989 | 11989 | 11989 | | |
| | fouryear | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | public | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | lncashflow | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |
| | prestige | 11993 | 11993 | 11993 | 11993 | 11993 | 11989 | 11993 | 11993 | | |

**Correlations**

| | | lncashflow | prestige |
|---|---|---|---|
| Sig. (1-tailed) | event | .000 | .000 |

| | | | |
|---|---|---|---|
| | d1 | .278 | .059 |
| | d2 | .409 | .098 |
| | d3 | .433 | .390 |
| | d4 | .423 | .005 |
| | instsize2 | .000 | .000 |
| | fouryear | .000 | .000 |
| | public | .000 | .000 |
| | lncashflow | . | .000 |
| | prestige | .000 | . |
| N | event | 11993 | 11993 |
| | d1 | 11993 | 11993 |
| | d2 | 11993 | 11993 |
| | d3 | 11993 | 11993 |
| | d4 | 11993 | 11993 |
| | instsize2 | 11989 | 11989 |
| | fouryear | 11993 | 11993 |
| | public | 11993 | 11993 |
| | lncashflow | 11993 | 11993 |
| | prestige | 11993 | 11993 |

**Variables Entered/Removed[a]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | prestige, d3, public, d4, lncashflow, d2, instsize2, fouryear[b] | . | Enter |

a. Dependent Variable: event

b. Tolerance = .000 limit reached.

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .230[a] | .053 | .052 | .102 |

a. Predictors: (Constant), prestige, d3, public, d4, lncashflow, d2, instsize2, fouryear
b. Dependent Variable: event

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 6.938 | 8 | .867 | 83.417 | .000[b] |
| | Residual | 124.544 | 11980 | .010 | | |
| | Total | 131.481 | 11988 | | | |

a. Dependent Variable: event
b. Predictors: (Constant), prestige, d3, public, d4, lncashflow, d2, instsize2, fouryear

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | -.096 | .011 | | -9.063 | .000 | | |
| | d2 | .008 | .003 | .034 | 3.156 | .002 | .671 | 1.491 |
| | d3 | .018 | .003 | .074 | 6.781 | .000 | .671 | 1.489 |
| | d4 | .013 | .003 | .055 | 5.095 | .000 | .673 | 1.485 |
| | instsize2 | .003 | .003 | .014 | 1.236 | .217 | .586 | 1.707 |
| | fouryear | -.003 | .003 | -.015 | -1.255 | .209 | .561 | 1.783 |
| | public | -.008 | .003 | -.038 | -2.976 | .003 | .493 | 2.030 |
| | lncashflow | .005 | .001 | .095 | 8.747 | .000 | .673 | 1.486 |
| | prestige | .063 | .004 | .164 | 15.925 | .000 | .741 | 1.349 |

a. Dependent Variable: event

**Excluded Variables[a]**

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics: Tolerance | VIF | Minimum Tolerance |
|---|---|---|---|---|---|---|---|---|
| 1 | d1 | .[b] | . | . | . | -1.781E-12 | -561356964934.887 | -1.781E-12 |

a. Dependent Variable: event
b. Predictors in the Model: (Constant), prestige, d3, public, d4, lncashflow, d2, instsize2, fouryear

**Collinearity Diagnostics[a]**

| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions (Constant) | d2 | d3 | d4 | instsize2 | fouryear |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4.602 | 1.000 | .00 | .01 | .01 | .01 | .01 | .01 |
| | 2 | 1.032 | 2.112 | .00 | .03 | .00 | .16 | .06 | .01 |
| | 3 | 1.000 | 2.145 | .00 | .12 | .33 | .05 | .00 | .00 |
| | 4 | .983 | 2.164 | .00 | .19 | .01 | .14 | .04 | .00 |
| | 5 | .781 | 2.428 | .00 | .00 | .00 | .00 | .05 | .05 |
| | 6 | .313 | 3.837 | .00 | .03 | .03 | .03 | .65 | .06 |
| | 7 | .209 | 4.697 | .00 | .54 | .54 | .54 | .09 | .08 |
| | 8 | .077 | 7.728 | .02 | .07 | .07 | .07 | .01 | .78 |
| | 9 | .004 | 34.953 | .97 | .00 | .00 | .00 | .10 | .02 |

**Collinearity Diagnostics[a]**

| Model | Dimension | Variance Proportions: public | lncashflow | prestige |
|---|---|---|---|---|
| 1 | 1 | .01 | .00 | .01 |
| | 2 | .01 | .00 | .22 |
| | 3 | .00 | .00 | .00 |
| | 4 | .00 | .00 | .11 |
| | 5 | .12 | .00 | .28 |

| | | | |
|---|---|---|---|
| 6 | .16 | .00 | .29 |
| 7 | .13 | .00 | .00 |
| 8 | .56 | .02 | .02 |
| 9 | .00 | .98 | .06 |

a. Dependent Variable: event

**Residuals Statistics**[a]

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | -.05 | .12 | .01 | .024 | 11989 |
| Std. Predicted Value | -2.395 | 4.342 | .000 | 1.000 | 11989 |
| Standard Error of Predicted Value | .002 | .006 | .003 | .001 | 11989 |
| Adjusted Predicted Value | -.05 | .11 | .01 | .024 | 11989 |
| Residual | -.110 | 1.003 | .000 | .102 | 11989 |
| Std. Residual | -1.079 | 9.838 | .000 | 1.000 | 11989 |
| Stud. Residual | -1.080 | 9.841 | .000 | 1.000 | 11989 |
| Deleted Residual | -.110 | 1.004 | .000 | .102 | 11989 |
| Stud. Deleted Residual | -1.080 | 9.880 | .000 | 1.004 | 11989 |
| Mahal. Distance | 4.004 | 39.079 | 7.999 | 3.781 | 11989 |
| Cook's Distance | .000 | .019 | .000 | .001 | 11989 |
| Centered Leverage Value | .000 | .003 | .001 | .000 | 11989 |

a. Dependent Variable: event

** Begin analysis for time using survival modeling

```
logistic regression var = event
   /method = enter d1 d2 d3 d4
   /method = enter d1 d2 d3 d4 instsize2
   /method = enter d1 d2 d3 d4 fouryear
   /method = enter d1 d2 d3 d4 public
   /method = enter d1 d2 d3 d4 prestige
```

/method = enter d1 d2 d3 d4 lncashflow
/method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow
/origin.

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 /method = enter d1 d2 d3 d4 instsize2 /method = enter d1 d2 d3 d4 fouryear /method = enter d1 d2 d3 d4 public /method = enter d1 d2 d3 d4 prestige /method = enter d1 d2 d3 d4 lncashflow /method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow /origin. |
| Resources | Processor Time | 00:00:00.31 |

| | |
|---|---|
| Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11989 | 100.0 |
| | Missing Cases | 4 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11856 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.

b. Initial Log-likelihood Function: -2 Log Likelihood = 16620.283
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3022.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2772.338 | 1 | .000 |
| | | d4 | 2763.299 | 1 | .000 |
| | Overall Statistics | | 11464.859 | 4 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15213.723 | 4 | .000 |
| | Block | 15213.723 | 4 | .000 |
| | Model | 15213.723 | 4 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1406.560[a] | .719 | .959 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -6.409 | .448 | 205.051 | 1 | .000 | .002 |
| | d2 | -4.604 | .183 | 629.555 | 1 | .000 | .010 |
| | d3 | -3.943 | .134 | 869.233 | 1 | .000 | .019 |
| | d4 | -4.253 | .157 | 731.337 | 1 | .000 | .014 |

a. Variable(s) entered on step 1: d1, d2, d3, d4.

**Block 2: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 122.154 | 1 | .000 |
| | Block | 122.154 | 1 | .000 |
| | Model | 15335.877 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 1 | 1284.406[a] | .722 | .962 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -7.566 | .476 | 252.163 | 1 | .000 | .001 |
| | d2 | -5.735 | .244 | 552.329 | 1 | .000 | .003 |
| | d3 | -5.045 | .207 | 594.460 | 1 | .000 | .006 |
| | d4 | -5.328 | .222 | 578.109 | 1 | .000 | .005 |
| | instsize2 | 2.046 | .206 | 98.230 | 1 | .000 | 7.739 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, instsize2.

**Block 3: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | Chi-square | df | Sig. |
|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Step 1 | Step | 67.477 | 1 | .000 |
| | Block | 67.477 | 1 | .000 |
| | Model | 15403.354 | 6 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1216.929[a] | .723 | .964 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| Observed | | | Predicted event 0 | Predicted event 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -9.453 | .583 | 262.659 | 1 | .000 | .000 |
| | d2 | -7.628 | .415 | 337.045 | 1 | .000 | .000 |
| | d3 | -6.919 | .392 | 310.783 | 1 | .000 | .001 |
| | d4 | -7.200 | .401 | 322.963 | 1 | .000 | .001 |
| | instsize2 | 2.240 | .207 | 116.583 | 1 | .000 | 9.397 |
| | fouryear | 2.134 | .348 | 37.687 | 1 | .000 | 8.448 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, fouryear.

**Block 4: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 5.516 | 1 | .019 |
| | Block | 5.516 | 1 | .019 |
| | Model | 15408.869 | 7 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1211.414[a] | .723 | .965 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -9.216 | .594 | 240.999 | 1 | .000 | .000 |
| | d2 | -7.386 | .430 | 294.387 | 1 | .000 | .001 |
| | d3 | -6.675 | .408 | 267.006 | 1 | .000 | .001 |
| | d4 | -6.953 | .417 | 278.602 | 1 | .000 | .001 |
| | instsize2 | 2.511 | .235 | 114.440 | 1 | .000 | 12.316 |
| | fouryear | 1.947 | .359 | 29.455 | 1 | .000 | 7.004 |
| | public | -.491 | .207 | 5.624 | 1 | .018 | .612 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, public.

**Block 5: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 83.640 | 1 | .000 |
| | Block | 83.640 | 1 | .000 |
| | Model | 15492.509 | 8 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1127.774[a] | .725 | .967 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| Observed | Predicted |
|---|---|

| | | | event 0 | event 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -8.458 | .598 | 200.309 | 1 | .000 | .000 |
| | d2 | -6.599 | .436 | 229.322 | 1 | .000 | .001 |
| | d3 | -5.839 | .414 | 198.489 | 1 | .000 | .003 |
| | d4 | -6.077 | .422 | 207.124 | 1 | .000 | .002 |
| | instsize2 | 1.404 | .282 | 24.769 | 1 | .000 | 4.073 |
| | fouryear | .959 | .389 | 6.068 | 1 | .014 | 2.608 |
| | public | -.427 | .209 | 4.157 | 1 | .041 | .653 |
| | prestige | 2.029 | .242 | 70.150 | 1 | .000 | 7.603 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, prestige.

**Block 6: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 111.147 | 1 | .000 |
| | Block | 111.147 | 1 | .000 |
| | Model | 15603.656 | 9 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1016.627[a] | .728 | .971 |

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

**Classification Table[a]**

| Observed | | | Predicted event 0 | Predicted event 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11851 | 5 | 100.0 |
| | | 1 | 128 | 5 | 3.8 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -23.405 | 1.715 | 186.194 | 1 | .000 | .000 |
| | d2 | -21.387 | 1.631 | 171.938 | 1 | .000 | .000 |
| | d3 | -20.502 | 1.606 | 162.948 | 1 | .000 | .000 |
| | d4 | -20.743 | 1.609 | 166.272 | 1 | .000 | .000 |
| | instsize2 | .406 | .298 | 1.855 | 1 | .173 | 1.501 |
| | fouryear | .220 | .406 | .293 | 1 | .588 | 1.245 |
| | public | -.122 | .216 | .316 | 1 | .574 | .886 |
| | prestige | .851 | .286 | 8.881 | 1 | .003 | 2.343 |
| | lncashflow | .843 | .085 | 97.964 | 1 | .000 | 2.323 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, lncashflow.

**Block 7: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Model | 15603.656 | 9 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1016.627[a] | .728 | .971 |

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted event 0 | Predicted event 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11851 | 5 | 100.0 |
| | | 1 | 128 | 5 | 3.8 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -23.405 | 1.715 | 186.194 | 1 | .000 | .000 |
| | d2 | -21.387 | 1.631 | 171.938 | 1 | .000 | .000 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | d3 | -20.502 | 1.606 | 162.948 | 1 | .000 | .000 |
| | d4 | -20.743 | 1.609 | 166.272 | 1 | .000 | .000 |
| | instsize2 | .406 | .298 | 1.855 | 1 | .173 | 1.501 |
| | fouryear | .220 | .406 | .293 | 1 | .588 | 1.245 |
| | public | -.122 | .216 | .316 | 1 | .574 | .886 |
| | prestige | .851 | .286 | 8.881 | 1 | .003 | 2.343 |
| | lncashflow | .843 | .085 | 97.964 | 1 | .000 | 2.323 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, instsize2, fouryear, public, prestige, lncashflow.

** ALTERNATIVE CODE from UCLA website

** model a

```
logistic regression var = event
 /method = enter d1 d2 d3 d4
 /origin.
```

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |

| | | |
|---|---|---|
| Syntax | | logistic regression var = event<br>/method = enter d1 d2 d3 d4<br>/origin. |
| Resources | Processor Time | 00:00:00.06 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11993 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11860 | .0 |

| | | | | |
|---|---|---|---|---|
| | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16625.828
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3024.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2773.336 | 1 | .000 |
| | | d4 | 2764.298 | 1 | .000 |
| | Overall Statistics | | 11468.857 | 4 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15219.195 | 4 | .000 |
| | Block | 15219.195 | 4 | .000 |
| | Model | 15219.195 | 4 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1406.633[a] | .719 | .959 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11860 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -6.410 | .448 | 205.093 | 1 | .000 | .002 |
| | d2 | -4.604 | .183 | 629.555 | 1 | .000 | .010 |
| | d3 | -3.943 | .134 | 869.389 | 1 | .000 | .019 |
| | d4 | -4.254 | .157 | 731.460 | 1 | .000 | .014 |

a. Variable(s) entered on step 1: d1, d2, d3, d4.

```
** model b

logistic regression var = event
 /method = enter d1 d2 d3 d4 instsize2
 /origin.
```

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 instsize2 /origin. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11989 | 100.0 |
| | Missing Cases | 4 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| Observed | | | Predicted event 0 | Predicted event 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 0 | event | 0 | 0 | 11856 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16620.283
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3022.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2772.338 | 1 | .000 |
| | | d4 | 2763.299 | 1 | .000 |
| | | instsize2 | 3324.184 | 1 | .000 |
| | Overall Statistics | | 11470.734 | 5 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15335.877 | 5 | .000 |
| | Block | 15335.877 | 5 | .000 |
| | Model | 15335.877 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1284.406[a] | .722 | .962 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted: event 0 | Predicted: event 1 | Predicted: Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11856 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -7.566 | .476 | 252.163 | 1 | .000 | .001 |
| | d2 | -5.735 | .244 | 552.329 | 1 | .000 | .003 |
| | d3 | -5.045 | .207 | 594.460 | 1 | .000 | .006 |
| | d4 | -5.328 | .222 | 578.109 | 1 | .000 | .005 |
| | instsize2 | 2.046 | .206 | 98.230 | 1 | .000 | 7.739 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, instsize2.

** model c

```
logistic regression var = event
 /method = enter d1 d2 d3 d4 fouryear
 /origin.
```

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 fouryear /origin. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11993 | 100.0 |

| | | | |
|---|---|---|---|
| | Missing Cases | 0 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11860 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16625.828
c. The cut value is .500

**Variables not in the Equation**

| | Score | df | Sig. |
|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3024.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2773.336 | 1 | .000 |
| | | d4 | 2764.298 | 1 | .000 |
| | | fouryear | 7900.332 | 1 | .000 |
| | Overall Statistics | | 11470.359 | 5 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15263.677 | 5 | .000 |
| | Block | 15263.677 | 5 | .000 |
| | Model | 15263.677 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1362.151[a] | .720 | .960 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | | | 0 | 1 | |
| Step 1 | event | 0 | 11860 | 0 | 100.0 |

| | | | | |
|---|---|---|---|---|
| | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -7.903 | .552 | 205.321 | 1 | .000 | .000 |
| | d2 | -6.102 | .371 | 270.389 | 1 | .000 | .002 |
| | d3 | -5.440 | .349 | 243.172 | 1 | .000 | .004 |
| | d4 | -5.752 | .359 | 257.031 | 1 | .000 | .003 |
| | fouryear | 1.787 | .346 | 26.688 | 1 | .000 | 5.974 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, fouryear.

** model d

logistic regression var = event
/method = enter d1 d2 d3 d4 public
/origin.

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |

| | | |
|---|---|---|
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 public /origin. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11993 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11860 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16625.828
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3024.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2773.336 | 1 | .000 |
| | | d4 | 2764.298 | 1 | .000 |
| | | public | 5394.361 | 1 | .000 |
| | Overall Statistics | | 11468.910 | 5 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15220.417 | 5 | .000 |
| | Block | 15220.417 | 5 | .000 |
| | Model | 15220.417 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1405.411[a] | .719 | .959 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted: event 0 | Predicted: event 1 | Predicted: Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | event | 0 | 11860 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -6.506 | .457 | 202.830 | 1 | .000 | .001 |
| | d2 | -4.700 | .205 | 525.671 | 1 | .000 | .009 |
| | d3 | -4.039 | .161 | 625.728 | 1 | .000 | .018 |
| | d4 | -4.349 | .181 | 575.650 | 1 | .000 | .013 |
| | public | .193 | .175 | 1.221 | 1 | .269 | 1.213 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, public.

** model e

```
logistic regression var = event
 /method = enter d1 d2 d3 d4  prestige
 /origin.
```

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:25 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 prestige /origin. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11993 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11860 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.

b. Initial Log-likelihood Function: -2 Log Likelihood = 16625.828

c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3024.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2773.336 | 1 | .000 |
| | | d4 | 2764.298 | 1 | .000 |
| | | prestige | 693.482 | 1 | .000 |
| | Overall Statistics | | 11490.648 | 5 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15467.058 | 5 | .000 |
| | Block | 15467.058 | 5 | .000 |
| | Model | 15467.058 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1158.770[a] | .725 | .966 |

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11860 | 0 | 100.0 |
| | | 1 | 133 | 0 | .0 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -7.401 | .465 | 253.369 | 1 | .000 | .001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | d2 | -5.553 | .220 | 639.619 | 1 | .000 | .004 |
| | d3 | -4.804 | .174 | 764.082 | 1 | .000 | .008 |
| | d4 | -5.053 | .191 | 702.867 | 1 | .000 | .006 |
| | prestige | 3.003 | .183 | 270.050 | 1 | .000 | 20.155 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, prestige.

** model f

```
logistic regression var = event
 /method = enter d1 d2 d3 d4  lncashflow
 /origin.
```

**Logistic Regression**

**Notes**

| | | |
|---|---|---|
| Output Created | | 28-MAR-2016 00:09:26 |
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 lncashflow /origin. |
| Resources | Processor Time | 00:00:00.07 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11993 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage |
| | Observed | | 0 | 1 | Correct |
| Step 0 | event | 0 | 0 | 11860 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16625.828
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3024.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2773.336 | 1 | .000 |
| | | d4 | 2764.298 | 1 | .000 |
| | | lncashflow | 11267.486 | 1 | .000 |
| | Overall Statistics | | 11482.459 | 5 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15587.762 | 5 | .000 |
| | Block | 15587.762 | 5 | .000 |
| | Model | 15587.762 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1038.066[a] | .727 | .970 |

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage |
| | Observed | | 0 | 1 | Correct |
| Step 1 | event | 0 | 11856 | 4 | 100.0 |
| | | 1 | 128 | 5 | 3.8 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -27.464 | 1.427 | 370.626 | 1 | .000 | .000 |
| | d2 | -25.413 | 1.318 | 371.571 | 1 | .000 | .000 |
| | d3 | -24.556 | 1.284 | 365.621 | 1 | .000 | .000 |
| | d4 | -24.835 | 1.285 | 373.725 | 1 | .000 | .000 |
| | lncashflow | 1.086 | .064 | 290.382 | 1 | .000 | 2.962 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, lncashflow.

** model g

logistic regression var = event
/method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow
/origin.

**Logistic Regression**

**Notes**

| | |
|---|---|
| Output Created | 28-MAR-2016 00:09:26 |

| | | |
|---|---|---|
| Comments | | |
| Input | Filter | MAH_1 <= 27.877 (FILTER) |
| | Weight | <none> |
| | Split File | <none> |
| | N of Rows in Working Data File | 11993 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing |
| Syntax | | logistic regression var = event /method = enter d1 d2 d3 d4 instsize2 fouryear public prestige lncashflow /origin. |
| Resources | Processor Time | 00:00:00.08 |
| | Elapsed Time | 00:00:00.00 |

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 11989 | 100.0 |
| | Missing Cases | 4 | .0 |
| | Total | 11993 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 11993 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

**Block 0: Beginning Block**

**Classification Table[a,b,c]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| Observed | | | 0 | 1 | |
| Step 0 | event | 0 | 0 | 11856 | .0 |
| | | 1 | 0 | 133 | 100.0 |
| | Overall Percentage | | | | 1.1 |

a. No terms in the model.
b. Initial Log-likelihood Function: -2 Log Likelihood = 16620.283
c. The cut value is .500

**Variables not in the Equation**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | d1 | 3022.033 | 1 | .000 |
| | | d2 | 2907.190 | 1 | .000 |
| | | d3 | 2772.338 | 1 | .000 |
| | | d4 | 2763.299 | 1 | .000 |
| | | instsize2 | 3324.184 | 1 | .000 |
| | | fouryear | 7896.335 | 1 | .000 |
| | | public | 5394.361 | 1 | .000 |
| | | prestige | 693.482 | 1 | .000 |
| | | lncashflow | 11263.596 | 1 | .000 |
| | Overall Statistics | | 11490.663 | 9 | .000 |

**Block 1: Method = Enter**

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15603.656 | 9 | .000 |
| | Block | 15603.656 | 9 | .000 |
| | Model | 15603.656 | 9 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 1016.627[a] | .728 | .971 |

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | | | Predicted | | |
|---|---|---|---|---|---|
| | | | event | | Percentage Correct |
| | Observed | | 0 | 1 | |
| Step 1 | event | 0 | 11851 | 5 | 100.0 |
| | | 1 | 128 | 5 | 3.8 |
| | Overall Percentage | | | | 98.9 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | d1 | -23.405 | 1.715 | 186.194 | 1 | .000 | .000 |
| | d2 | -21.387 | 1.631 | 171.938 | 1 | .000 | .000 |
| | d3 | -20.502 | 1.606 | 162.948 | 1 | .000 | .000 |
| | d4 | -20.743 | 1.609 | 166.272 | 1 | .000 | .000 |
| | instsize2 | .406 | .298 | 1.855 | 1 | .173 | 1.501 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | fouryear | .220 | .406 | .293 | 1 | .588 | 1.245 |
| | public | -.122 | .216 | .316 | 1 | .574 | .886 |
| | prestige | .851 | .286 | 8.881 | 1 | .003 | 2.343 |
| | lncashflow | .843 | .085 | 97.964 | 1 | .000 | 2.323 |

a. Variable(s) entered on step 1: d1, d2, d3, d4, instsize2, fouryear, public, prestige, lncashflow.

## References

Comment [AJB2]: This should go before the appendix.

Alpert, D. (1991). Performance and paralysis: The organizational context of the American research university. In M. W. Peterson, E. E. Chaffee, and T. H. White (Eds.), ASHE Reader on Organization and Governance in Higher Education (4th ed.). Needham Heights, MA.: Ginn Press.

Anderson, J. Q., Boyles, J. L., & Rainie, L. (2012). The Future Impact of the Internet on Higher Education: Experts Expect More Efficient Collaborative Environments and New Grading Schemes; They Worry about Massive Online Courses, the Shift Away from On-Campus Life. *Pew Internet & American Life Project*. Retrieved from http://eric.ed.gov/?id=ED534048

Ansell, B. W. (2008). University Challenges: Explaining Institutional Change in Higher Education. *World Politics*, *60*(2), 189–230.

Bass, R. (2012). Disrupting ourselves: The problem of learning in higher education. *Educause Review*, *47*(2), 23–33.

Berry, F., & Berry, W. (2014). Innovation and diffusion models in policy research. In *Theories of the Policy Process* (Third Edition, Third Edition edition). Boulder, CO: Westview Press.

Berry, F., & Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review*, *84*(2), 395–415.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, *103*(3), 191-207.

Bowen, W. G. (2015). *Higher education in the digital age*. Princeton University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=26MpBQAAQBAJ&oi=fnd&pg=PP1&dq=bowen+digital+age&ots=5CuDAzbeTb&sig=iGjCcqE64NgEiItPdXuRUto6hbU

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, *103*(3), 191–207.

Butin, D. (2015). The Future of The Future of Higher Education | Inside Higher Ed. Retrieved April 19, 2016, from https://www.insidehighered.com/blogs/higher-ed-beta/future-future-higher-education

Cankar, S. S., & Petkovsek, V. (2013). Private and public sector innovation and the importance of cross-sector collaboration. *Journal of Applied Business Research*, *29*(6), 1597.

Carey, K. (2013, February 13). Obama, Rubio Agree on One Thing: Technology Could Fix the Higher Ed Mess. *Slate*. Retrieved from http://www.slate.com/blogs/future_tense/2013/02/13/state_of_the_union_moocs_obama_rubio_agree_on_using_tech_to_fix_higher_ed.html

Carey, K. (2015). *The End of College: Creating the Future of Learning and the University of Everywhere*. Penguin.

Christensen, C. (2015). Disruptive Innovation. Retrieved from http://www.claytonchristensen.com/key-concepts/

Christensen, C. M., & Eyring, H. J. (2011). *The Innovative University: Changing the DNA of Higher Education from the Inside Out* (1 edition). San Francisco: Jossey-Bass.

Christensen, C. M., Horn, M. B., Caldera, L., & Soares, L. (2011). *Disrupting College: How Disruptive Innovation Can Deliver Quality and Affordability to Postsecondary Education*. Innosight Institute. Retrieved from http://eric.ed.gov/?id=ED535182

Clark, B. R. (2004). Delineating the character of the entrepreneurial university. *Higher Education Policy*, *17*(4), 355–370.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm* (Vol. 2). Englewood Cliffs, NJ. Retrieved from https://books.google.com/books?hl=en&lr=&id=qqZ_FDFoDcMC&oi=fnd&pg=PA60&dq=Cyert+%26+march+1963&ots=9V0PLgxv6P&sig=y57Oe52K9IoVyh5nEwSu1uJBZ4A

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal*, *34*(3), 555–590.

Derousie, J. (2014). *An Exploration of the Diffusion and Adoption of Four Innovations in Higher Education*.

Diamond, R. (2006, September 8). Why Colleges Are So Hard to Change | Inside Higher Ed. Retrieved April 20, 2016, from https://www.insidehighered.com/views/2006/09/08/diamond

Dillon, E., & Carey, K. (2009). Drowning in Debt: The Emerging Student Loan Crisis. *Higher Education*.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 147–160.

Dougherty, K. J., & Reddy, V. (2013). *Performance Funding for Higher Education: What Are the Mechanisms What Are the Impacts: ASHE Higher Education Report, 39:2*. John Wiley & Sons.

Ebersole, J. (2014, January 13). Top Issues Facing Higher Education In 2014. Retrieved September 10, 2015, from http://www.forbes.com/sites/johnebersole/2014/01/13/top-issues-facing-higher-education-in-2014/

Ehrenberg, R. G. (2006). The perfect storm and the privatization of public higher education. *Change: The Magazine of Higher Learning*, *38*(1), 46–53.

Fain, P. (2013, January 16). California looks at MOOCs in online push | Inside Higher Ed. Retrieved January 21, 2013, from http://www.insidehighered.com/news/2013/01/16/california-looks-moocs-online-push

Ferenstein, G. (2013). Why Obama’s Radical Education Plan Could Finally Disrupt College. Retrieved from http://social.techcrunch.com/2013/08/22/why-obamas-radical-education-plan-could-finally-disrupt-higher-education/

Ferlie, E., Musselin, C., & Andresani, G. (2008). The Steering of Higher Education Systems: A Public Management Perspective. *Higher Education*, *56*(3), 325–348.

Goldman, C. A., Goldman, C., Gates, S. M., Brewer, A., & Brewer, D. J. (2004). *In pursuit of prestige: Strategy and competition in US higher education*. Transaction Publishers. Retrieved from https://books.google.com/books?hl=en&lr=&id=NUUe4XUsd3gC&oi=fnd&pg=PR8&dq=Brewer+Gates+and+Goldman&ots=Dzm0uH5y6N&sig=gT-gs4Pwpnvb69pmDbVs4aPryFo

Grupp, F. W., & Richards, A. R. (1975). Variations in elite perceptions of American states as referents for public policy making. *American Political Science Review*, *69*(3), 850–858.

Hartley, J., Sørensen, E., & Torfing, J. (2013). Collaborative Innovation: A Viable Alternative to Market Competition and Organizational Entrepreneurship. *Public Administration Review*, *73*(6), 821–830. http://doi.org/10.1111/puar.12136

Hearn, J. C. (1996). Transforming US higher education: An organizational perspective. *Innovative Higher Education*, *21*(2), 141–154.

Hixon, T. (2014, January 6). Higher Education Is Now Ground Zero For Disruption. Retrieved April 19, 2016, from http://www.forbes.com/sites/toddhixon/2014/01/06/higher-education-is-now-ground-zero-for-disruption/

Hollands, F., & Tirthali, D. (2014). *MOOCs: Expectations and reality. Full Report for the Center for Benefit-Cost Studies of Education*. Teachers College, Columbia University.

Jaggars, S. S. (2014). Choosing between online and face-to-face courses: Community college student voices. *American Journal of Distance Education*, *28*(1), 27–38.

Kezar, A. J. (2014). *How colleges change: understanding, leading, and enacting change*. New York: Routledge.

Lahr, H., Pheatt, L., Dougherty, K. J., Jones, S., Natow, R. S., & Reddy, V. (2014). Unintended impacts of performance funding on community colleges and universities in three states. Retrieved from http://academiccommons.columbia.edu/catalog/ac:180622

Levine, A. (2015, September 14). Time Is Right for Colleges to Shift  From Assembly-Line Education. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Time-Is-Right-for-Colleges-to/233057/

Marginson, S. (2006). Dynamics of National and Global Competition in. *Higher Education*, *52*(1), 1–39. http://doi.org/10.1007/s10734-004-7649-x

Markov, J. (2013, January 17). Measuring the Success of Online Education. Retrieved January 20, 2013, from http://bits.blogs.nytimes.com/2013/01/17/measuring-the-success-of-online-education/

McLendon, M. K., Deaton, S. B., & Hearn, J. C. (2007). The Enactment of Reforms in State Governance of Higher Education: Testing the Political Instability Hypothesis. *The Journal of Higher Education*, *78*(6), 645–675.

McLendon, M. K., Heller, D. E., & Young, S. P. (2005). State postsecondary policy innovation: Politics, competition, and the interstate migration of policy ideas. *The Journal of Higher Education*, *76*(4), 363–400.

Meisenhelder, S. (2014, April 16). Online Higher Ed a Cost Savings for Students, Universities? Think Again. Retrieved April 19, 2016, from http://www.huffingtonpost.com/susan-meisenhelder/online-higher-ed-a-cost-s_b_4782301.html

Mettler, S. (2014). *Degrees of Inequality: How the Politics of Higher Education Sabotaged the American Dream*.

Mintrom, M. (1997). Policy entrepreneurs and the diffusion of innovation. *American Journal of Political Science*, 738–770.

Mintz, S. (2014, September 30). The Future of Higher Education | Inside Higher Ed. Retrieved April 20, 2016, from https://www.insidehighered.com/blogs/higher-ed-beta/future-higher-education

Mohr, L. B. (1969). Determinants of innovation in organizations. *American Political Science Review*, *63*(1), 111–126.

Not classy enough. (2015, March 28). *The Economist*. Retrieved from http://www.economist.com/news/special-report/21646986-online-learning-could-disrupt-higher-education-many-universities-are-resisting-it-not

Pappano, L. (2012, November 2). Massive Open Online Courses Are Multiplying at a Rapid Pace. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Parr, C. (2013, March 10). New study of low MOOC completion rates. Retrieved December 1, 2015, from https://www.insidehighered.com/news/2013/05/10/new-study-low-mooc-completion-rates

Porto, S. (2013, May 31). The Impact of Disruptive Technology-Based Innovations in Higher Education. Retrieved from http://evolllution.com/opinions/impact-disruptive-technology-based-innovations-higher-education/

Rogers, E. M. (1983). Difiusion of innovations. *New York: Free Press*, *18*(20), 271.

Roth, C. (2012). *Are the sleeping giants awake? Non-profit universities enter online education at scale.* Parthenon Perspectives.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle* (Vol. 55). Transaction Publishers. Retrieved from http://books.google.com/books?hl=en&lr=&id=-OZwWcOGeOwC&oi=fnd&pg=PR6&dq=Schumpeter,+J.+A.+(1934).+The+theory+of+economic+development:+An+inquiry+into+profits,+capital,+credit,+interest,+and+the+business+cycle+(Vol.+55).+Transaction+Publishers.&ots=iM3_q2uaB8&sig=Nnbrs5ideLQnO0e5HgGNTHuMOek

Selingo, J. J. (2013). *College (Un)Bound: The Future of Higher Education and What It Means for Students*. Houghton Mifflin Harcourt.

Shah, D. (2015, December 21). By The Numbers: MOOCS in 2015. Retrieved March 29, 2016, from https://www.class-central.com/report/moocs-2015-stats/

Slaughter, S., & Leslie, L. L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University.* The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, MD 21218-4319 ($39.95). Retrieved from http://eric.ed.gov/?id=ED409816

Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, state, and higher education*. JHU Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=Y-mISmAUa38C&oi=fnd&pg=PR9&dq=slaughter+%26+Rhoades+2004&ots=E3s1nk8njc&sig=CV3mBvzhjjiOhRwVNpj6beomWM4

Smith, A. (2015, July 15). The for-profit industry is struggling, but has not reached the end of the road | Inside Higher Ed. Retrieved December 1, 2015, from https://www.insidehighered.com/news/2015/07/15/profit-industry-struggling-has-not-reached-end-road

Stack, B. (2014, July 1). Competency Education: The Next Great Disruptor in Education. Retrieved from http://connectedprincipals.com/archives/10582

Stein, K. (2013, December 5). Penn GSE Study Shows MOOCs Have Relatively Few Active Users, With Only a Few Persisting to Course End | Penn GSE Press Room. Retrieved December 1, 2015, from http://www.gse.upenn.edu/pressroom/press-releases/2013/12/penn-gse-study-shows-moocs-have-relatively-few-active-users-only-few-persisti

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research*, *81*(1), 4–28. http://doi.org/10.3102/0034654310393361

Walker, J. L. (1969). The diffusion of innovations among the American states. *American Political Science Review*, *63*(3), 880–899.

West, D. M., & Lu, J. (2009). *Comparing technology innovation in the private and public sectors*. Governance Studies at Brookings. Retrieved from http://www.brookings.edu/~/media/Research/Files/Papers/2009/6/technology%20west/06_technology_west.pdf

Woodhouse, K. (2016, February 16). Author discusses new book on changes in higher education marketplace. Retrieved April 20, 2016, from https://www.insidehighered.com/news/2016/02/16/author-discusses-new-book-changes-higher-education-marketplace

Yuan, L., & Powell, S. (2013). MOOCs and disruptive innovation: Implications for higher education. *eLearning Papers, In-Depth*, *33*(2), 1–7.