Jacob Hobbs
CS 591.003
Project proposal

Who is the Most Superstitious of Them All?

Abstract.

Using publicly available marriage record data extending back into the Renaissance, I intend to find seasonal and monthly deviations from uniform frequency. I will correlate this with information on holy days and lunar cycles to determine which subsets of the early British and British Colonial (and possibly other) populations deferred important events to more "lucky" days than the rest.

Introduction

Before the rise of rationalism in the 18th century, the common European outlook was deeply embedded in pre-scientific thought. The stars and planets affected everyday events on earth, whether through their own ethereal influence or because God had organized them to reveal his will to the careful observer (see Genesis 1:14, where they are given as "signs"). Because of this, it was prudent to plan important events for days when the heavens indicated that efforts begun on those days would prosper.

Although historians have found plenty of anecdotal evidence for superstitious belief, I have not been able to discover good quantitative analysis of any of the records we have to verify the extent of it. Not all such beliefs can be quantified, but a few of them can. In particular, events that were regularly recorded, like marriages and births, can be checked for interesting patterns. For instance, it has been said that common folk in England preferred to marry while the moon was waxing, since it indicated growth and harmony for the marriage. Another pattern to look for is the attention people paid to holy days on their calendars. Holy days were often marked with red ink, and they are said to have been lucky days for doing things in the minds of ordinary people.

Description of the data

The calendar system used by western Europe is known as the Gregorian calendar. It was adopted by the Catholic church in 1582, but rejected in Protestant countries until, for the most part, some differing dates in the 18th century. This means that the holy days between countries will not always be in sync with each other. They will, however, use either the Gregorian or the Julian calendars. There is a python package called jdcal that allows for converting between Gregorian calendar and Julian calendar dates, which will make it easier to determine moon cycles by converting Julian calendar dates to Gregorian dates. The exact moon cycle also depends upon longitude, but at most it will affect the boundaries of moon cycles by a single day. I plan on using the moon cycles as determined by the Greenwich meridian for European dates, and GMT-5 for early American dates, since most of the data will center on those longitudes. If I stick to the British Isles and Colonies, which I might, then I only have to worry about the Julian calendar for any dates through 1752, at which time they all switched to the Gregorian.

So-called "moveable feasts" are holidays that were celebrated on different days each year. These holidays were celebrated a fixed number of days surrounding the Easter celebration for a given year. The date of Easter depended on which calendar, Julian or Gregorian, a particular country used.

Fortunately, there is a web site, http://www.smart.net/~mmontes/ec-cal.html, which allows for calculating the Julian and Gregorian dates for major moveable feasts in any given year. For years after 1583, the Gregorian dates can be checked against http://aa.usno.navy.mil/data/docs/easter.php for consistency. I will either use a package like Beautiful Soup to scrape the dates from these websites, or use code that has been posted for calculating the dates myself (such as the code at https://www.assa.org.au/edm for the Gregorian Easters, and the code at http://www.smart.net/~mmontes/ortheast.html for Julian Easters).

Marriage data is generally found from parish records. Most parish record info can be found in disparate formats across a number of websites for individual parishes across England. There are way too many individual sites to search, and often their databases are only available for querying specific names. Similar information for the US can be found in vital records; however, the US does not store such data so acquiring it requires the nightmare of going to individual states for their data. The familysearch.org and ancestry.com websites hosts the largest online collection of extracted record data; unfortunately, access to this data requires pay for ancestry.com so we are limited to familysearch.org.

FamilySearch has an enormous collection of data. Marriage records can be retrieved by knowing individual batch numbers or searching through specific collections. Useful for my task is to search through specific collections. I will be scraping data from the following collections: Massachusetts Marriages 1695-1910, Maryland Marriages 1666-1970, New Hampshire Marriage Records 1637-1947, England Marriages 1538-1973, England and Wales Non-Conformist Record Indexes, Scotland Marriages 1561-1910, Wales Marriages 1541-1900. With the option to search by marriage year, I can leave the name information blank and retrieve an exhaustive list of the marriage collections. I can manipulate the search URL's in order to look them up using Python's urllib2 library, and then scrape the results using Beautiful Soup. The results page gives names, marriage dates and locations without having to search through a deeper level in the records. The formatting for the results are pretty consistent among all the records, so it won't be too hard to gather and format the 30 million or so marriage records into a local file for analytics. I plan on gathering all available marriage records for the years 1538-1850 from the above collections.

The Massachusetts Vital Records Project (http://ma-vitalrecords.org/) has an exhaustive collection of early marriage information for the state. Unfortunately, the website is formatted to make it difficult to scrape that information, so I will only look into it if I have the time.

If I am able to pull marriage data from other countries, I will have to also add a column to my data to indicate which calendar the date is using.

Hypotheses

I predict that early English marriages happened more frequently on superstitiously lucky days. I believe there will be a strong enough correlation with moon cycles or calendar dates to justify the idea that many people married with the heavens or holy days in mind. I believe the correlation will be stronger in more rural areas. I also believe it will be stronger for Scotland than for England, up through

the 18th century. This is based on assumptions about the rurality of Scotland until a later date. I predict that marriages were more superstitious in 17th century England than in 17th century Puritan colonies. I base this on the fact that New England Puritans refused to celebrate common Christian holidays and removed the "red numbers" from their calendars. Their strict views should have had some effect on their superstitions. By the mid-18th century the influx of German and other settlers may increase the correlation with superstitious days to some degree, in certain areas. I predict that by the early 19th century, such correlations will die off everywhere and marriages will occur more uniformly during lunar cycles.

I also hypothesize that there will be major seasonal differences in overall marriage counts, but that these will not affect the distributions when measured by moon phases or holy calendars. I can split the data by season to verify this.

If I have time to scrape additional marriage data for other countries, like France, Spain and Germany, I will be able to check for variations between the countries. Since I have studied little of the early Renaissance folk culture of these other countries, I can only give a complete guess and hypothesize that they too will show lunar or holy calendar correlation amongst their marriage dates. I think it is very possible that the correlations in other regions could be *different*, however. Maybe the peasants of France thought it luckier to marry during a different part of the lunar cycle, for instance.

So far as I have been able to find, no one has done a project quite like this. I don't believe the topic is important enough for historians or computer scientists to have cared about. I'm really hoping for some correlations to be detected because that would make the results presentable in a historical or sociological journal, as further verification of the results of literary studies my hypothesis is currently based on (see the books in the references section for more info).

If my hypotheses don't pan out, and there isn't a significant correlation between marriage dates and holy days or between marriage dates and lunar cycles, then I predict that there will at least be a correlation between marriage dates and seasons. In particular, that there will be far more marriages occurring in the spring and fall than at other times of the year. This I believe will be due to the timing of work seasons and the favorableness of the weather, but I will not have a way verifying the causes without more data.

Proposed Method

The way I plan on measuring for statistically significant patterns is via the transformation of marriage dates to their time in the lunar cycle, as well as to their timing in the liturgical calendar for the given year. I plan on producing graphs of distributions and visually ascertaining the presence of any interesting patterns. To check for the presence of interesting patterns with respect to holy days I will measure the means and variances of marriage frequency across and within the different groupings of days. I will also look for patterns for marrying more frequently on particular days of the week and use the weekday distributions as the base for the null hypothesis of whether liturgical days and moon cycle moments are important.

Jacob Hobbs
CS 591.003
Project proposal

When comparing the 17th century New England Puritans to their English counterparts, I am inclined to believe that any apparent presence of correlation between holy calendar dates and marriages among New Englanders will be an indicator of noise in the distribution (variation from a uniform norm) rather than New England allegiance to the concept of lucky days. My reason for believing this is that the men who controlled the distribution of literature, especially the almanacks which were the source of figuring out holy days, did not include such superstitious material in their almanacks until the end of the century, unlike their English counterparts. Thus, such "noise" will be useful and I may be able to treat the distribution of apparent holy day observance among the New England natives as a base for the null hypothesis when trying to decide on the presence of superstition in other places and at other times.

For variation among marriage locations in the British Isles, I will determine if any individual locations display significant correlations with holy/lunar days across time, and which are more strongly correlated than others. I can use those locations to produce a heatmap of levels of correlation by geography and see whether higher correlations correspond to more rural areas. I can estimate the density of population in different locations also by looking at the number of marriages occurring in a particular location, and thus produce "population density" heatmaps as well to overlay with the superstition correlation heatmap, and see how correlated or inversely correlated they are. I can also consider the number of marriages in a particular location and the level of holy / lunar day correlation as two random variable outcomes and perform a chi-squared test to see numerically the strength of the correlation with respect to urban population centers versus rural areas.

The basic tools I'll be using for my analysis will be Python, Scipy and Matplotlib, since they are easy to use and I am familiar with them. They are well-suited for the tasks I'm planning on, especially for the size of the data I'll be analyzing.

Conclusion

There are plenty of modern superstitions that many people have, but not many of them are astronomically based anymore. Many people today are unaware of just how superstitious and focused on heavenly events their forebears used to be. I am hoping to find patterns between lunar cycles and the dates early Europeans and British colonists chose to marry as evidence for those superstitions. I hope to identify a temporal trend of the correlations decreasing as time moves forward into the industrial age. I will also be looking for correlations between marriage dates and Saints days and other holy days in their liturgical calendars. It will be fun to have a measure for just how superstitious they really were.

References

David D. Hall, Worlds of Wonder, Days of Judgment: Popular Religious Belief in Early New England, (Alfred A. Knopf Inc: New York, 1989)

David Stevenson, The Origins of Freemasonry: Scotland's century, 1590-1710, (Cambridge University Press: Cambridge, 1988)

Keith Thomas, <u>Religion and the Decline of Magic</u>, (Penguin Books: London, 1991)

Wikipedia, "Gregorian calendar", http://en.wikipedia.org/wiki/Gregorian_calendar

Wikipedia, "Old Style and New Style dates",
http://en.wikipedia.org/wiki/Old_Style_and_New_Style_dates