**Week 11 Assignment**

Michael K. Zoucha

College of Science & Information Technology, Bellevue University

DSC 510: Introduction to Programming

Professor Eller

08 November 2021

**Week 11 Assignment**

Exercise 13-1:

Function CleanData given by author. The function replaces 9998 and 9999 with nan in the cmdivorcx variable, and addes four new variables for notdivorced, duration, durationsofar, and decade for grouping.

def CleanData(resp):  
  
 resp.cmdivorcx.replace([9998, 9999], np.nan, inplace=True)  
  
 resp['notdivorced'] = resp.cmdivorcx.isnull().astype(int)  
 resp['duration'] = (resp.cmdivorcx - resp.cmmarrhx) / 12.0  
 resp['durationsofar'] = (resp.cmintvw - resp.cmmarrhx) / 12.0  
  
 month0 = pd.to\_datetime('1899-12-15')  
 dates = [month0 + pd.DateOffset(months=cm)   
 for cm in resp.cmbirth]  
 resp['decade'] = (pd.DatetimeIndex(dates).year - 1900) // 10

Function ResampleDivorceCurveByDecade creates decade groups to pass to EstimateSurvivalByDecade function and sets up the configuration for the final plot including labels.

def ResampleDivorceCurveByDecade(resps):  
  
 for i in range(41):  
 samples = [thinkstats2.ResampleRowsWeighted(resp)  
 for resp in resps]  
 sample = pd.concat(samples, ignore\_index=True)  
 groups = sample.groupby('decade')  
 if i == 0:  
 survival.AddLabelsByDecade(groups, alpha=0.7)  
  
 EstimateSurvivalByDecade(groups, alpha=0.1)  
  
 thinkplot.Config(xlabel='Years',  
 ylabel='Fraction undivorced',  
 axis=[0, 28, 0, 1])

Function EstimateSurvivalByDecade groups the data frame by newly created decades column and passes the groups to EstimateSurvival for the survival and hazard functions

def EstimateSurvivalByDecade(groups, \*\*options):  
 thinkplot.PrePlot(len(groups))  
 for name, group in groups:  
 \_, sf = EstimateSurvival(group)  
 thinkplot.Plot(sf, \*\*options)

Function EstimateSurvival takes each data frame, splits them into divorced and not divorced then estimates the survival curve, returning a hazard and survival function

def EstimateSurvival(resp):  
complete = resp[resp.notdivorced == 0].duration.dropna()  
 ongoing = resp[resp.notdivorced == 1].durationsofar.dropna()  
  
 hf = survival.EstimateHazardFunction(complete, ongoing)  
 sf = hf.MakeSurvival()  
  
 return hf, sf

Taking the data from FemResp2002 and FemResp2010, we clean the data using the above function and create subsets of data using only respondents that have ever been married.

resp6 = survival.ReadFemResp2002()  
resp7 = survival.ReadFemResp2010()  
  
CleanData(resp6)  
married6 = resp6[resp6.evrmarry==1]  
  
CleanData(resp7)  
married7 = resp7[resp7.evrmarry==1]

Using the now cleaned and subset data, we can pass the data frames to the ResampleDivorceCurveByDecade to create the final plot.

ResampleDivorceCurveByDecade([married6, married7])

Chart, line chart

Description automatically generated

Results:  
As we can see, all age groups follow similar patters until about 5 years of marriage, when the more recent decades tend to drop off more than older decades. There is a complete drop off the 90s-decade group after 20 years because they have not been alive long enough to be married longer than that, so no one can be ‘undivorced’ if no one is married, though I would have expected it to drop off sooner.