

Descriptions of the Seasons Data Sets

The *Seasons* data come from a longitudinal study conducted by researchers at the University of Massachusetts Medical School (Merriam et al., 1999; Ockene et al., 2004, also see Chapter 26). Subjects were volunteers recruited from the membership of the Fallon Healthcare System, a large HMO in central Massachusetts. For some of the variables, subjects provided data during each season of the year. The number at the end of the variable name indicates the season: 1=winter; 2=spring; 3=summer; and 4=fall. Note that there are many missing data points in the Seasons file.

Of major interest in the study was whether, and how, certain physiological and psychological variables varied seasonally. Also, measures of activity, diet, and light exposure were collected (of these, only some of the light-exposure data are included here). Data were collected through a combination of quarterly questionnaires and three phone interviews each quarter for each subject.

Note that measures of physical activity are contained in the *Seasons physical activity* data set which has its own description file.

The variables in the main *Seasons* file are as follows:

id	Subject identification number
mstatus	Marital status: 1=single; 2=married; 3=living with partner; 4=separated; 5=divorced; 6=widowed
schoolyr	School completed: 1=no high school; 2=some high school; 3=high school diploma; 4=vocational/trade school; 5=some college; 6=associate degree; 7=bachelor's degree; 8=graduate school
employed	1=yes, employed full time; 2=yes, employed part time; 3=no occupation type of employment (not included in the .txt file)
age	Age of subject when entering the study
agegrp	Age category: 1=less than 40 years of age; 2=40-49 years; 3=50-59 years; 4=60 years old or older
sex	Sex: 0=male, 1=female
sayhlth	Self-rating of health — "In general, would you say your health is..." 1=Excellent; 2=Very good; 3=Good; 4=Fair; 5=Poor
wgtcomp	Self-rating of weight 1=Very overweight; 2=somewhat overweight; 3=About average; 4=Below average/lean
beck_a	1-4 scores on the Beck anxiety scale 1=winter; 2=spring; 3=summer; 4=fall
beck_d	1-4 scores on the Beck depression scale 1=winter; 2=spring; 3=summer; 4=fall. Larger scores indicate more depression. A score of 10 is the cutoff for dysphoria; 16 is the cutoff for depression
bmi	1-4 Body-mass index (a measure of body density – weight in kilograms divided by the square of height in meters)

The next set of 16 variables try to get at seasonal variation in the emotions hostility, anger, irritability, and anxiety, respectively. For each of these emotions, subjects are asked to indicate on a 9-point scale,

Compared to how you generally feel, how do you feel this time of year, where 1=much less, 5=about average, and 9=much more

host	1-4 hostility 1=winter; 2=spring; 3=summer; 4=fall
anger	1-4 anger
irrit	1-4 irritability
anxiety	1-4 anxiety
tc	1-4 total cholesterol level

The next 8 variables have to do with direct exposure to daylight (data obtained using three telephone interviews per quarter -- in which subjects were asked about their activities during the previous day). There is recent evidence that many older women hospitalized with acute hip fractures are low in vitamin D – a substance that helps the body absorb calcium and is readily synthesized through exposure to sunlight, so direct daylight exposure as a function of age and sex is of interest.

dirwdc	1-4 Direct exposure to daylight during weekdays (in hours per day)
dirwec	1-4 Direct exposure to daylight during weekend days (in hours per day)

The next four measures represent averages over the four seasons for subjects who had data on the measure in each season.

beck_a	Mean Beck anxiety score averaging over the four seasons (for subjects who had Beck anxiety data in all four seasons)
beck_d	Mean Beck depression score
bmi	Mean body-mass index
tc	Mean total cholesterol score
height	Height in cm
weight	Weight in kg (measured during winter)

All of the variables named above appear in the large *Seasons* data file. In addition, we have extracted several other files that are used in various examples, and in exercises at the ends of chapters. Most of these are smaller and were constructed to include only subsets of variables or of subjects who met certain criteria; however, a few contain extra variables. Descriptions of those files, and the variables included in them are as follows:

age_BMI

This is a copy of the Seasons data file to be used with the EL_x_age.sps SPSS syntax file that is available on the website.

Beck_D

Variables: id, schoolyr, agegrp, sex, beck_d1 - 4, mean_d. Includes only those subjects who had scores in both winter and spring seasons (beck_d1 and beck_d2); some d3 and d4 scores are missing; mean_d is average depression score for those subjects with all four seasonal depression scores.

Male_educ

Variables: schoolyr, mean_d. Includes only male subjects with beck_d scores in all four seasons who had one of four levels of education; schoolyr = 3,5,7, or 8.

Sayhlth

Variables: id, schoolyr, employed, agegrp, sex, sayhlth, beck_d1-4, d_mean. Excluded are 6 (of 641) subjects who had no sayhlth score or a sayhlth score of 5 (poor self-rating).

tc

Variables: schoolyr, agegrp, sex, sayhlth, tc1-4, tc_mean, and EL. Includes only subjects with tc scores in all four seasons. EL (educational level) = 1 if schoolyr = 1, 2, or 3; EL = 2 if schoolyr = 4 or 5; EL = 3 if schoolyr = 6, 7, or 8.

Male_D under 35

Variables: schoolyr, employed, age, agegrp, beck_d1-4, diff1_2, Psi_d, d_mean, sex. Diff1_2 is the difference between winter and spring depression scores and Psi_d is the difference between the winter beck_d score and the average of scores in the other three seasons. Only subjects under 36 years of age are included. For further explanation and illustrative calculations, see Chapter 6.

D_change data

Variables: sex, beck_d1, beck_d2, change. The “change” scores are the difference in depression scores between winter and spring (d1 and d2).

Seasons exercises

This is a version of the Seasons data set used by several exercises in Chapter 23. It contains the following variables from the main Seasons file: id, mstatus, schoolyr, employed, age, sex, wgtcomp, sayhlth, beck_a, beck_d, bmi, tc, height, and weight. The file also contains the following additional variables:

bmiplusnorm05	obtained by adding to each value of bmi a number randomly selected from a normal distribution with mean 0 and standard deviation 5.
bmiplusnorm020	obtained by adding to each value of bmi a number randomly selected from a normal distribution with mean 0 and standard deviation 20.
tcplusnorm040	obtained by adding to each value of tc a number randomly selected from a normal distribution with mean 0 and standard deviation 40.

tcformen

This is a subset of the Seasons data that contains only data for men.

tcforwomen

This is a subset of the seasons data file that contains only data for women.