Information about the CCT data for the take-home exam

(Bernd Figner: March 16, 2014)

You get 2 data sets, (1) the data set with the risky choice data and (2) the data set with the demographics information.

These data are from the first lab session of the mixed models course, from the two risky choice tasks that you did (the "hot" and "cold" Columbia Card Task, CCT).

The following paper is NOT mandatory reading (and not really necessary for the take-home exam at all); it's for the curious and interested ones who want to know more about the risky choice task (or if you feel you need some more information to understand the data).

> Figner, B., & Weber, E. U. (2011). Who takes risk when and why? Determinants of risk-taking. *Current Directions in Psychological Science, 20*, 211-216.

If you want to refresh your memory how the risky choice task worked, here you can try out the demo versions:

Hot version: https://vlab.decisionsciences.columbia.edu/cct/game2-instructions1.php
Cold version: https://vlab.decisionsciences.columbia.edu/cct/game2-instructions1.php

OK, so here are some explanations regarding the data files and the variables it contains

(1) Risky choice data

I uploaded the file both in csv and xls format (both are the same, so use whichever is more convenient for you):

- CCTdata_12March2014a.csv
- CCTdata_12March2014a.xls

Some explanations how the task works

Each of the two CCT versions consists of 24 game rounds (=trials). In each game round, you choose how many cards you want to turn over; each game round starts with 32 cards shown from the back. Across trials, the following factors are systematically varied:

- Probability: 1 or 3 loss cards (out of the 32 cards)
- Gain amount: for each gain card that you turn over, you win either 10 or 30 points
- Loss amount: if you turn over a loss card, either 250 or 750 points are subtracted from the number of points you had accrued in the current game round (each new game round starts with a score of 0)

This gives 2 x 2 x 2 = 8 different combinations. Each of these combinations is presented 3 times in total: In each of 3 blocks, all 8 different combinations are presented (in random order). In other words: Each block is made up of 2 (probability: 1 or 3 loss cards) x 2 (gain amount: 10 or 30 points) x 2 (loss amount: 250 or 750 points) = 8 game rounds. Each combination is presented 3 times (thus 3 blocks), resulting in a total of 24 game rounds for each the hot and the cold CCT.

The main variable of interest is how many cards participants decide to turn over in each of the game rounds and whether they adjust their level of risk-taking (=number of cards chosen) as a function of probability, gain amount, and loss amount. An optimal solution should always adjust the number of cards depending on the levels of probability, gain amount, and loss amount in the given round.

Further, it could be that participants adjust their strategy over blocks. For example, they might start out taking too many or too few cards and then learn to get better and adjust their risk-taking levels in the course of the 3 blocks.

In the lab session where we collected these data, I asked you to do both the hot and the cold CCT: The difference between these two versions is that in the hot CCT, the participant gets immediate information. I.e., as soon as the participant clicks on the card, it is turned over and revealed whether it is a "good" win card or a "bad" loss card. In the cold CCT, the participant doesn't get any feedback, but s/he only indicates how many cards should be turned over.

One important effect this difference has is that in the hot CCT, there might be **"censored" trials**: Let's say you wanted to turn over 10 cards, but the 2nd card you turned over a loss card. The game round stops immediately and I as experimenter have no way to find out that you wanted to turn over 10 cards.
In the cold CCT, the task always saves in the data how many cards you wanted to turn over; i.e., the data are not censored for the cold CCT.

Further below, after the variable explanations, I will show you how you should model this task difference in your analyses.

Explanations for the variables in `CCTdata_12March2014a.csv` and `CCTdata_12March2014a.xls`

- p_serial: you already know that variable from the valuation ratings analyses. You can use it to extract the participant code and the order information.
- r_tasktype: "c" stands for the cold CCT; "h" for the hot CCT
- r_block: 1 is the first block, 2 the second block, 3 the third block (each new CCT starts with block = 1; i.e., if somebody did first the cold CCT, it will be block 1, 2, 3; then they did the hot, it will start again with 1, then 2, 3)
- r_trialnum: 1 is the first game round, 2 the 2nd etc (as for the block variable, each new CCT starts with trial number 1)
- r_lossnum: The number of loss cards in that game round (either 1 or 3); i.e., this is what I called above "Probability"
- r_winvalue: Gain amount (10 or 30 points per good card turned over)
- r_lossvalue: Loss amount (250 or 750 loss if a loss card is turned over)
- r_cardschosen: That is the main DV of interest: How many cards did the participant turn over (NOTE: in the hot CCT, this is how many cards they really turned over. In the cold CCT: If, for example, the participant said they wanted to turn over 10 cards, the entry in r_cardschosen is 10, irrespective of whether (when the computer later played out these game rounds for real to determine payment, there was a loss card of not among these 10 cards).
- r_censored: If no loss card was encountered, the value is 0. A value larger than 0 indicates that there was a loss card encountered (which matters only for the hot CCT).
- r_score: The final "score" (i.e., outcome) in that game round (e.g., -720 means that they ended up with a loss of 720 points in that game round).
- r_timespent: The duration of that game round (in seconds)
- p_totaltime: The duration of that whole (hot or cold) CCT (i.e., for all entries for the same CCT for the same participant, all entries are the same and do NOT change from game round to game round)

- payout: How much the person gained/lost in that CCT (the number of points divided by 100, as we ususally use the 'conversion rate' 1 point is worth 1 cent)
- r_payout: Whether that game round was randomly chosen to be paid out (=1) or not (=0)
- num_incorrect: At the end of the instructions, there was a short quizz containing 4 questions: The participants had to enter answers and we can use them to see how well they understood how the CCT worked. This variable indicates how many of these 4 questions were answered incorrectly.
- r_id: Another ID variable that is created by the data base that saves the data from the online CCTs (sometimes, this can be useful to merge with other data; you probably won't need it)
- r_pid: same as for r_id

(2) Demographics data

Explanations for the variables in
`demographics_cleaned_12March2014a.csv` and `demographics_cleaned_12March2014a.xls`

- d_id: same as for r_id and r_pid (i.e., some kind of ID variable created by the data base)
- d_pid: same again
- d_date: date and time of participation (note that this is New York time, as the server running the tasks and saving the data is in New York)
- d_gender: Gender of the participant: F = female; M = male
- d_age: Age of the participant in years
- d_status: Marital status: 0 = Single; 1 = Living together; 2 = Married; 3 = Divorced or living separated; 4 = Widowed
- d_children: number of children
- d_education: 0=No degree; 1=High school diploma; 2=Associate degree, occupational; 3=Associate degree, academic; 4=Bachelor's degree; 5=Master's degree; 6=Professional degree; 7=Doctoral degree
- d_employment: 0=No job/Unemployed; 1=Working in household; 2=Student; 3=Worker/Farmer; 4=Civil servant/employee; 5=Manager; 6=Entrepreneur; 7=Other
- d_income: 0=less than $19,999; 1=$20,000 - $34,999; 2=$35,000 - $49,999; 3=$50,000 - $99,999; 4=$100,000 - $199,999
  5=greater than $200,000; 6=Prefer not to answer
- d_smoke: Do you smoke? 1=yes; 0=no
- d_quantitysmoke: How much do you smoke? 0=Less than a pack a week; 1=1 to 2 packs a week; 2=3 to 6 packs a week; 3=A pack a day; 4=Two packs a day; 5=More than two packs a day
- d_lastsmoke: When did you last smoke? 0=Within the last hour; 1=Between 1 and 6 hours ago; 2=Between 7 and 12 hours ago; 3=Between 13 and 24 hours ago; 4=Between 1 and 2 days ago; 5=Between 3 and 7 days ago; 6=More than a week ago
- d_howlongsmoke: For how long have you been smoking? 0=Less than a month; 1=1 to 6 months; 2=7 to 12 months; 3=1 to 2 years; 4=3 to 10 years; 5=11 to 20 years; 6=More than 20 years
- p_id: same as r_id in the CCT data frame
- p_serial: same as p_serial in the CCT data frame

- p_totalpayout: Participants typically get a flat rate payment of Euro 10 for participating; the outcome of the 3 randomly chosen game rounds (i.e., the number in the variable "payout" in the CCT data frame) is then added (subtracted if it's a negative number) to these Euro 10, resulting in the number in "p_totalpayout"

HINT: There might be some missing data in some of the demographics variables; thus:
(a) Check whether this is the case for the demographic variable(s) that you plan to use in your analyses.
(b) The data base writes "NULL" as a missing value; R, however, needs NA as value for a missing value.
(c) If the demographics variable you plan to use has missing entries, it might be a good idea to create a new data frame from which the rows that have missing values have been removed (if you don't do that, you might encounter some error messages in some approaches to get p values, e.g., when using drop1).

(3) Some additional information regarding the risky choice data: Modelling the effect of "censored" game rounds (which can occur in the hot, but not the cold CCT)

To account for that difference in the task version, we have to do the following when we analyze the data:

We create a variable (e.g., called "censored_1Yes0No") that indicates whether a hot CCT trial was censored or not. We could for example code it with a rule like this (you could for example use the ifelse() command to create such a variable in R):
If this row of data belongs to the hot CCT AND if r_censored is greater than 0, then the variable censored_1Yes0No should have the value 1, otherwise it should have the value 0.

In other words: For all cold CCT game rounds, the value in the variable "censored_1Yes0No" should be 0 (irrespective of what the value in r_censored is). For the hot CCT, the game rounds in which the variable r_censored is 0, the value in the variable "censored_1Yes0No" should be also 0; only if the variable r_censored is greater than 0, the value in the variable "censored_1Yes0No" should be 1.