Principles of Measurement Assignment 1

Andrés Araos

# Exercise 1

a)

Var(x1+x2+x3+x4) = var(x1+x2) + var(x3+x4) + 2cov(x1+x2, x3+x4)

= var(x1) + var(x2) + 2cov(x1, x2) + var(x3) + var(x4) + 2cov(x3, x4) + 2cov(x1+x2, x3+x4)

But:

cov(x1+x2, x3+x4) = E[(x1+x2)(x3+x4)] - E[x1+x2]E[x3+x4]

= E[x1x3 + x1x4 + x2x3 + x2x4] - (E[x1] + E[x2])(E[x3] + E[x4])

= E[x1x3] + E[x1x4] + E[x2x3] + E[x2x4] - (E[x1]E[x3] + E[x1]E[x4] + E[x2]E[x3] + E[x2]E[x4])

= E[x1x3] - E[x1]E[x3] + E[x1x4] - E[x1]E[x4] + E[x2x3] - E[x2]E[x3] + E[x2x4] - E[x2]E[x4]

Since cov(xi,xj)= E[xixj] - E[xi]E[xj]

Then,

cov(x1+x2, x3+x4) = cov(x1,x3) + cov(x1,x4) + cov(x2,x3) + cov(x2,x4)

Thus:

Var(x1+x2+x3+x4) =

= var(x1) + var(x2) + 2cov(x1, x2) + var(x3) + var(x4) + 2cov(x3, x4) + 2cov(x1+x2, x3+x4)

= var(x1) + var(x2) + var(x3) + var(x4) + 2cov(x1, x2) + 2cov(x1,x3) + 2cov(x1,x4) + 2cov(x2,x3) + 2cov(x2,x4) + 2cov(x3, x4)

= 0.25 + 0.77 + 3.68 + 1.18 + 2\*2.12 + 2\*0.36 + 2\*0.16 + 2\*0.66 + 2\*0.35 + 2\*0.74

= 5.88 + 8.68 = 14.56

b)

Sample Pearson correlation ρxixj= Cov (Xi,Xj)/SxSy

Thus:

| ρx1x1 | ρx2x1 | ρx3x2 | ρx4x1 |
| --- | --- | --- | --- |
| ρx1x2 | ρx2x2 | ρx3x2 | ρx4x2 |
| ρx1x3 | ρx2x3 | ρx3x3 | ρx4x3 |
| ρx1x4 | ρx2x4 | ρx3x4 | ρx4x4 |

=

| var(x1)/var(x1) | cov(x2,x1)  √var(x1)√var(x2) | cov(x3,x1)  √var(x3)√var(x2) | cov(x4,x1)  √var(x4)√var(x2) |
| --- | --- | --- | --- |
| cov(x1,x2)  √var(x1)√var(x2) | var(x2)/var(x2) | cov(x3,x2)  √var(x3√var(x2) | cov(x4,x2)  √var(x4)√var(x2) |
| cov(x1,x3)  √var(x1)√var(x3) | cov(x2,x3)  √var(x2)√var(x3) | var(x3)/var(x3) | cov(x4,x3)  √var(x4)√var(x3) |
| cov(x1,x4)  √var(x1)√var(x4) | cov(x2,x4)  √var(x2)√var(x4) | cov(x3,x4)  √var(x3)√var(x4) | var(x4)/var(x4) |

=

| 1.0 | - | - | - |
| --- | --- | --- | --- |
| 0.274 | 1.0 | - | - |
| 0.375 | 0.392 | 1.0 | - |
| 0.294 | 0.367 | 0.355 | 1.0 |

c) Based on the correlations calculated it is possible to say that the correlation between all the variables is weak.

d) Further information that would be useful for interpreting the results would be:

* Nature of the measurement, that is, what is being measured in each item. This information would help understand the underlying or theoretical relationship between the items.
* Scale used for measurement. This, together with the nature of measurement, would enable identifying what type of score we are dealing with, regardless of the construct being measured. In other words, whether we are dealing with nominal, ordinal or ratio scores. This would also enable checking for the assumptions of the statistics used.
* Sample size, since it is necessary information to know how significant the correlations obtained are, i.e. to conduct an hypothesis test on the relation of the variables using the Pearson correlation.

# Exercise 2

Each item can be seen as a Bernoulli distributed variable.

Thus, =

While the estimate of

Thus,

=

=

=

=

=

=

# Exercise 3

## a) Coding items:

**Part 1:** I scored the items in Part 1 using a 0 to 4 scale that I considered as an ordinal score because we can’t know exactly what is the distance between the different answers (for example at least once a week could be from 1 to 6) and also we cannot know the difference between the categories. I made it start from 0 because there is a “Never” category.

**Part 2:** I scored the items in Part 2 using a 0 to 3 scale that I considered as an ordinal score because we can’t know exactly what is the distance between the different answers (for example knowing something about a topic can have a large range within the same category) and also we cannot know the difference between the categories. I made it start from 0 because there is a “Never heard about it” category.

**Part 3:** I coded gender as dummy (0 or 1) because, although the option “other” is included, there is no one who answered it. Thus, I considered it as dichotomous. Favorite color was recoded using nominal scores from 1 to 4, and “Friendlier country” with nominal scores from 1 to 5. These variables are all coded as nominal, because there is no distance or natural order between the possible values.

b)

For Parts 1 and 2 the summary statistics used are the median and the mode, because there is no real meaning in using the arithmetic mean to describe distribution of ordinal variables.

For Part 3, since all the variables are nominal in nature, the summary statistic used is the mode.

Part 1 Question 2: In everyday life, how often do you usually use the Internet to better understand issues related to, for example, your health or illnesses, financial matters or environmental issues?

Median = 3 (At least once a week but not every day)

Mode = 4 (Every day, frequency of 6)

Comment: We can see that the median and the mode don’t coincide, which suggests that the distribution of the item is skewed.

Part 2 Question 5: How informed are you about the following environmental issues? Air pollution.

Median = 2 (I know something about this and could explain the general issue)

Mode = 2 (I know something about this and could explain the general issue, frequency of 13)

Comment: We can see that the median and mode coincide. However, we can also see that the mode holds a frequency of 13 out of 17, which suggests a highly skewed distribution. Moreover, this is so highly skewed that it might make sense to code it a dichotomous variable.

c)

I would use the mean of the total scores

Really sorry, I did not have time to finish the last two parts of Exercise 3. If it is possible I can send them next Monday.

# Annex: R-studio code

library(dplyr)

#Recode values rules#

levelkeyP1 <- c("Never" = 0, "Less than once a month" = 1, "Less than once a week but at least once a month" = 2, "At least once a week but not every day"= 3, "Every day" = 4)

levelkeyP2 <- c("I have never heard of this" = 0, "I have heard about this but I would not be able to explain what it is really about" = 1, "I know something about this and could explain the general issue"= 2, "I am familiar with this and I would be able to explain this well" = 3)

levelkeyP3X1 <- c("Female" = 0, "Male" = 1, "Other" = 2)

levelkeyP3X2 <- c("Blue" = 1, "Green" = 2, "Red" = 3, "Yellow" = 4)

levelkeyP3X3 <- c("Denmark" = 1, "Finland"= 2, "Iceland" = 3, "Norway" = 4, "Sweden" = 5)

#Extracting variables and recoding

P1X1 <- QT1$X1..Use.email.

P1X1.recode <- recode(P1X1, !!!levelkeyP1)

P1X2 <- QT1$X2..Use.the.Internet.to.better.understand.issues.related.to..for.example..your.health.or.illnesses..financial.matters.or.environmental.issues.

P1X2.recode <- recode(P1X2, !!!levelkeyP1)

P1X3 <- QT1$X3..Conduct.transactions.on.the.Internet..for.example.buying.or.selling.products.or.services..or.banking.

P1X3.recode <- recode(P1X3, !!!levelkeyP1)

P1X4 <- QT1$X4..Use.spreadsheet.software..for.example.Excel.

P1X4.recode <- recode(P1X4, !!!levelkeyP1)

P1X5 <- QT1$X5..Use.a.word.processor..for.example.Word.

P1X5.recode <- recode(P1X5, !!!levelkeyP1)

P1X6 <- QT1$X6..Use.a.programming.language.to.program.or.write.computer.code.

P1X6.recode <- recode(P1X6, !!!levelkeyP1)

P1X7 <- QT1$X7...Participate.in.real.time.discussions.on.the.Internet..for.example.online.conferences.or.chat.groups.

P1X7.recode <- recode(P1X7, !!!levelkeyP1)

P2X1 <- QT1$X1..The.increases.of.greenhouse.gases.in.the.atmosphere.

P2X1.recode <- recode(P2X1, !!!levelkeyP2)

P2X2 <- QT1$X2..The.use.of.genetically.modified.organisms.

P2X2.recode <- recode(P2X2, !!!levelkeyP2)

P2X3 <- QT1$X3..Nuclear.waste.

P2X3.recode <- recode(P2X3, !!!levelkeyP2)

P2X4 <- QT1$X4..The.consequences.of.clearing.forests.for.other.land.use.

P2X4.recode <- recode(P2X4, !!!levelkeyP2)

P2X5 <- QT1$X5...Air.pollution.

P2X5.recode <- recode(P2X5, !!!levelkeyP2)

P2X6 <- QT1$X6..Extinction.of.plants.and.animals.

P2X6.recode <- recode(P2X6, !!!levelkeyP2)

P2X7 <- QT1$X7..Water.shortage.

P2X7.recode <- recode(P2X7, !!!levelkeyP2)

P3X1 <- QT1$What.is.your.gender.

P3X1.recode <- recode(P3X1, !!!levelkeyP3X1)

P3X2 <- QT1$Out.of.the.following..which.is.your.favorite.colour.

P3X2.recode <- recode(P3X2, !!!levelkeyP3X2)

P3X3 <- QT1$Out.of.the.following..which.country.has.the.friendliest.people.

P3X3.recode <- recode(P3X3, !!!levelkeyP3X3)

#frequency tables

P1T1 <- table(P1X1)

P1T2 <- table(P1X2)

P1T3 <- table(P1X3)

P1T4 <- table(P1X4)

P1T5 <- table(P1X5)

P1T6 <- table(P1X6)

P1T7 <- table(P1X7)

P2T1 <- table(P2X1)

P2T2 <- table(P2X2)

P2T3 <- table(P2X3)

P2T4 <- table(P2X4)

P2T5 <- table(P2X5)

P2T6 <- table(P2X6)

P2T7 <- table(P2X7)

P3T1 <- table(P3X1)

P3T2 <- table(P3X2)

P3T3 <- table(P3X3)

print(P2T5) #mode

print(P1T2) #mode

#Summary statistics (median)

P1S2X2 <- median(P1X2.recode)

P2S2X5 <- median(P2X5.recode)

print(P2X5.recode)