Assignment 1 solutions

9/10/2024

EPSY 6637 Assignment #1 Fall 2024

This assignment is due on Monday September 9th.

1) *In the week 2 workspace there is a dataset called astro. It contains correct and incorrect responses from a 27 question astronomy exam.*

a) *Calculate the alpha for the test. Explain what it signifies.*

```
library(psych)
alpha(astro)

## Warning in alpha(astro): Some items were negatively correlated with the fi
rst principal component and probably
## should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option
```

The alpha is estimated to be 0.83. This provides an estimate of the reliability of the instrument based on the internal consistency of the items. It signifies that 83% of the observed variance is so-called true score variance, and so 17% is error variance.

b) *Calculate the standard error of measurement.*

If 83% of the observed variance is true score variance, then 17% is error variance. We take the observed variance, and multiply it by 0.17. Then we take the square root to get the standard error of measurement.

```
observedvariance <- var(apply(astro,1,sum))

sqrt(0.17*observedvariance)

## [1] 2.180834
```

The standard error of measurement is 2.2.

c) *Give a 95% confidence interval for a person who scored 22 out of 27.*

We create the 95% CI by extending an interval +/- 2SE from the observed score. Therefore the interval is: 22 +/- 2*2.2 or [17.6, 26.4]. The person scored 22, but we might allow a wider range as to where the "true score" actually is. (The apple doesn't fall far from the tree)

d) *Study the list of item-total correlations. Are there any problematic items on the test?*

```
alpha(astro)$item.stats

## Warning in alpha(astro): Some items were negatively correlated with the fi
rst principal component and probably
```

```
## should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option

## Some items ( V6 ) were negatively correlated with the first principal comp
onent and
## probably should be reversed.
## To do this, run the function again with the 'check.keys=TRUE' option

##       n      raw.r      std.r      r.cor     r.drop       mean        sd
## V1  130  0.2828513  0.2860600  0.2455115  0.2098468 0.20769231 0.4072246
## V2  130  0.4717641  0.4633016  0.4330066  0.3944625 0.40769231 0.4933064
## V3  130  0.4386927  0.4408410  0.4137174  0.3628659 0.67692308 0.4694609
## V4  130  0.4688009  0.4733079  0.4471036  0.4021033 0.76153846 0.4277913
## V5  130  0.5026918  0.5044072  0.4854192  0.4294637 0.63076923 0.4844634
## V6  130 -0.2239135 -0.1955811 -0.2818942 -0.2731462 0.08461538 0.2793851
## V7  130  0.4208273  0.4158259  0.3815873  0.3381424 0.49230769 0.5018748
## V8  130  0.5067195  0.5139459  0.4954782  0.4414294 0.74615385 0.4368942
## V9  130  0.2963681  0.2874549  0.2409701  0.2064065 0.50000000 0.5019342
## V10 130  0.5878349  0.5986943  0.6025052  0.5303494 0.75384615 0.4324357
## V11 130  0.3898043  0.4045326  0.3625571  0.3226232 0.80000000 0.4015474
## V12 130  0.3356269  0.3326037  0.2861191  0.2486873 0.42307692 0.4959586
## V13 130  0.5593180  0.5490830  0.5372756  0.4908176 0.61538462 0.4883863
## V14 130  0.4724400  0.4667770  0.4363426  0.3966628 0.63076923 0.4844634
## V15 130  0.2367570  0.2229026  0.1709108  0.1446595 0.46923077 0.5009829
## V16 130  0.4551646  0.4831546  0.4652354  0.4136844 0.92307692 0.2675002
## V17 130  0.4574563  0.4488878  0.4158952  0.3795335 0.60769231 0.4901535
## V18 130  0.5830032  0.5780879  0.5774377  0.5206061 0.69230769 0.4633239
## V19 130  0.5068603  0.5081421  0.4865497  0.4422873 0.24615385 0.4324357
## V20 130  0.5380953  0.5337247  0.5176156  0.4691723 0.65384615 0.4775834
## V21 130  0.4830450  0.4919356  0.4637550  0.4198292 0.78461538 0.4126792
## V22 130  0.5449818  0.5391239  0.5242474  0.4738909 0.57692308 0.4959586
## V23 130  0.5919166  0.5762939  0.5688080  0.5251891 0.54615385 0.4997913
## V24 130  0.3200393  0.3203191  0.2717696  0.2423233 0.26153846 0.4411726
## V25 130  0.5809020  0.5767094  0.5684971  0.5133602 0.56923077 0.4970995
## V26 130  0.3889486  0.3892773  0.3566644  0.3155440 0.25384615 0.4368942
## V27 130  0.2564966  0.2609189  0.2128184  0.1733978 0.29230769 0.4565824
```

As the warnings at the start and end of the alpha command have indicated, there is an item negatively correlated with the scale. Item 6 has a negative discrimination of -0.22. This is in contrast to the relatively strong positive correlations for all the other items. It suggests that the item may have been incorrectly scored, in particular that the wrong answer may have been scored as correct. The magnitude of the negative correlation here is large enough to be of concern. The negative discrimination would translate in IRT terms into an ICC that ran **down** from 1 to 0 as ability increases.

If this item can't be fixed (that is if it doesn't turn out that the answer key was wrong, or it can't be improved on future tests) it should be dropped.

e) *Make a correlation matrix of the items 17, 18 and 19. An IRT analysis of these data would assume conditional independence of the items. Does the correlation matrix of these three items contradict the conditional independence assumption? Explain.*

```
round(cor(astro[,c(17,18,19)]),2)
```

```
##      V17  V18  V19
## V17 1.00 0.22 0.24
## V18 0.22 1.00 0.19
## V19 0.24 0.19 1.00
```

First of all, we *expect* to see correlations in this analysis. The assumption underlying the overall test is that the items all measure the same construct, and so it follows that the scores should be positively correlated. The conditional independence assumption states that if we have a sample with the *same* ability - that is, ability is "controlled for" or "conditioned on" - then the items are no longer correlated. But in the full sample we should expect positive correlations. In class we will look at the different results when we correlated items in the general sample, and when we did it by narrowing the sample to people with very similar total scores.

The correlations in this output are relatively modest, but they are positive. They are consistent with a set of binary items measuring the same construct.

f) *Make the tetrachoric correlation matrix for three items (17,18,19).*

```
round(tetrachoric(astro[,c(17,18,19)])$rho,2)
```

```
##      V17  V18  V19
## V17 1.00 0.35 0.43
## V18 0.35 1.00 0.36
## V19 0.43 0.36 1.00
```

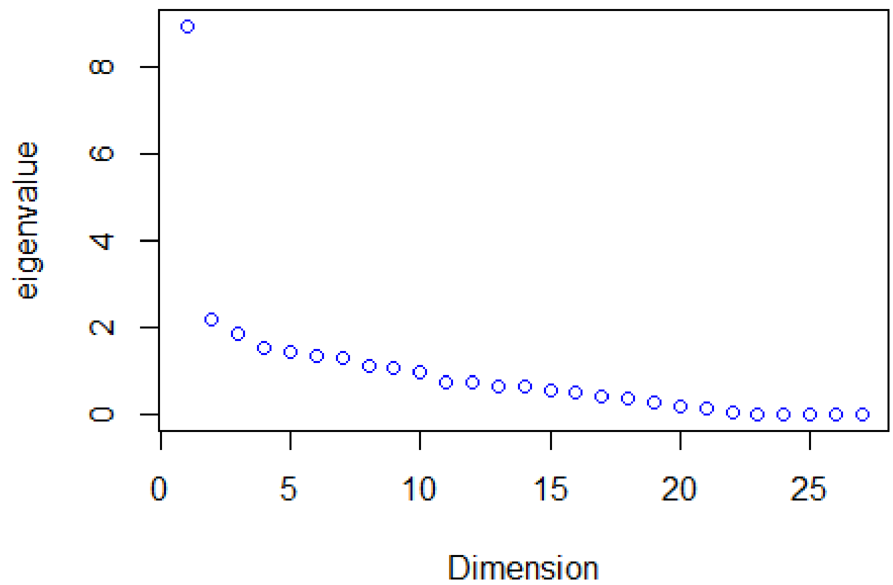
The tetrachoric correlation is calculated based on the idea that the 0,1 correct/incorrect observation is a consequence of an underlying continuous variable with a threshold above which we see 1, and below which we see 0. The tetrachoric correlation estimates what the correlation would be of the underlying continuous variables. As such, it attempts to fix some of the limitations of correlations based on binary data with unbalanced margins. Note that the estimated correlations are higher than in part (e).

g) *Make a plot of the eigenvalues. Is the plot consistent with the assumption of unidimensionality that would be part of an IRT analysis?*

The plot shows a single dominant eigenvector. The second dimension looks like it is close to the scree. The graph is generally consistent with a unidimensional construct.

```
plot(c(1:27),eigen(tetrachoric(astro)$rho)$values, ylab="eigenvalue",xlab="Di
mension",col="blue")
```

```
## Warning in cor.smooth(mat): Matrix was not positive definite, smoothing wa
s
## done
```



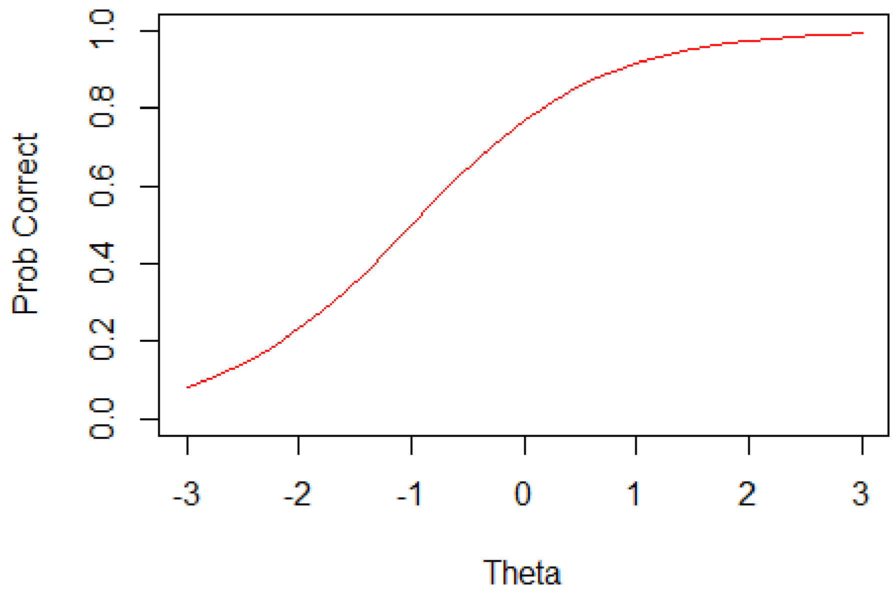
2) *The following questions are about the ICC curves for common IRT Models.*

a) *Given the 3PL model, calculate probability of a correct response if θ = -1.5, and b = -1, and a = 1.2.*

Here is the ICC with the requested parameters.

```
b<- -1
a<-1.2
thetas<-seq(-3,3,0.2)

plot(thetas, 1/(1 + exp(-a*(thetas - b))),col="red", type="l",ylab="Prob Corr
ect",xlab="Theta",ylim=c(0,1))
```



We can see that for = -1.5, the probablity is close to 0.3. We can calculate exactly by substituting into the equation:

```
1/(1 + exp(-a*(-1.5 - b)))

## [1] 0.3543437
```

b) *Plot the ICC for an item with b = 1.0, a = 1.5 Also plot another curve for an item that is easier but has a steeper discrimination.*

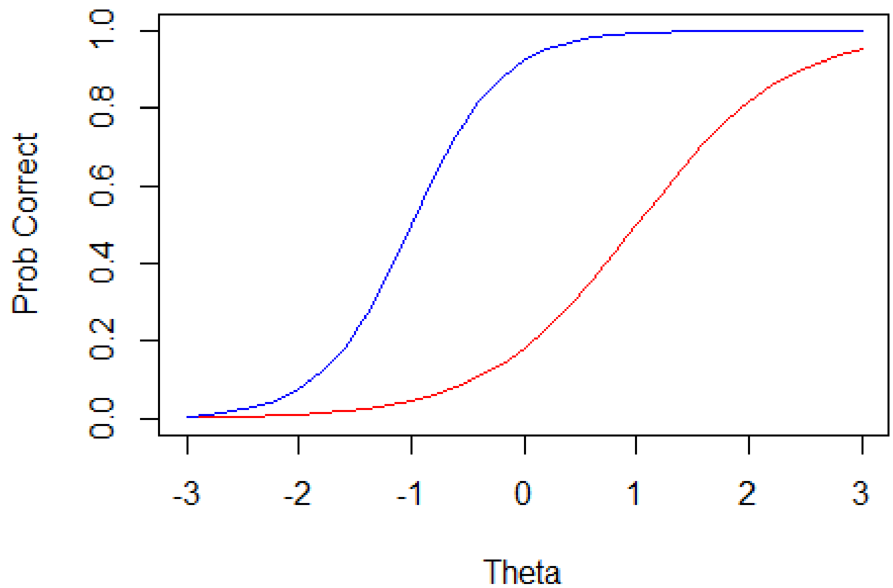
In red we have the requested item. An easier item would have lower b, and a steeper discrimination means a > 1.5. Let's go with b = -1 and a = 2.5, and make it blue.

```
b<- 1
a<-1.5
c <- 0
thetas<-seq(-3,3,0.2)

plot(thetas, c + (1-c)*1/(1 + exp(-a*(thetas - b))),col="red", type="l",ylab=
"Prob Correct",xlab="Theta",ylim=c(0,1))

b<- -1
a<-2.5
c <- 0

points(thetas, c + (1-c)*1/(1 + exp(-a*(thetas - b))),col="blue", type="l")
```



c) *Suppose a test had 20 items, and the b parameters were greater than 0 for all the items? What could you say about the properties of the test?*

It is certainly possible to have a test with 20 items with b values greater than zero. This simply means that it would be a difficult test, composed entirely of items that most people do not answer correctly. Therefore the item percent corrects will be very low. Because the ICC curves will show low probability of a correct response (at guessing even) for much of the sample, we can predict that this test will give very little *information* about the abilities of people in the lower half of the ability distribution. On the other hand, it will be informative - maybe even highly so - as a measure of the ability of above average participants.