CONTINUOS MEDICAL EDUCATION
pISSN09763325 | eISSN22296816
Open Access Article (CC BY-SA)
www.njcmindia.com
DOI: 10.55489/njcm.130820221149

An R Function for Cronbach's Alpha Analysis: A Case-Based Approach

Himani Kotian[1], Aiswarya Liz Varghese[2], Rohith Motappa[3]

[1]Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India
[2]Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India
[3]Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India

ABSTRACT

Cronbach's alpha is a very commonly used method in biomedical research. Cronbach's alpha Indicates the extent to which the items in your questionnaire are related to each other, a useful coefficient for assessing the internal consistency of the items. Although this method is commonly used in medical research, the statistical software packages do not have the direct menu-driven operation for Cronbach's alpha. Hence this paper intends to provide an R function (Cronbach. Alpha) for Cronbach's alpha analysis.

Keywords: Internal consistency, Reliability, Cronbach's alpha

INTRODUCTION

Many quantities in medicine, such as anxiety, stress, or degree of handicap, are not possible to measure explicitly. Instead, we ask a series of questions and combine the answers into a single numerical value. Often this is done by simply adding a score from each answer. When items are used to form a scale, they need to have internal consistency. The items should all measure the same concept, so they should be correlated with one another. Cronbach's alpha, *is* a correlation measure, if a scale consists of several items that are *identical*, then it indicates a questionnaire is very poorly formulated. So, the general idea that correlated items are the best ones has its flaws and we need to remember that.

CRONBACH'S ALPHA

Cronbach's alpha is a measure used to assess the reliability, or internal consistency, of a set of scale or test items. In other words, the reliability of any given measurement refers to the extent to which it is a consistent measure of a concept, and Cronbach's alpha is one way of measuring the strength of that consistency.

Cronbach's alpha is computed by correlating the score for each scale item with the total score for each observation (usually individual survey respondents or test takers), and then comparing that to the variance for all individual item scores. The formula is

$$\alpha = \frac{K}{k-1}\left(1 - \frac{\sum S_i^2}{S_T^2}\right)$$

where k is the number of items, s_i^2 is the variance of the i^{th} item and s_T^2 is the variance of the total score formed by summing all the items. If the items are not simply added to make the score, but first multiplied by weighting coefficients, we multiply the item by its coefficient before calculating the variance s_i^2. Clearly, we must have at least two items-that is k >1, or α will be undefined.

HOW CRONBACH'S ALPHA CAN BE COMPUTED USING R FUNCTION

Cronbach alpha can be calculated manually, we even have different software like SPSS, Stata and SAS. This paper intends to show the computation of Cronbach's alpha using R function. There are many ways of calculating Cronbach's alpha in R using a va-

How to cite this article: Kotian H, Varghese AL, Rohith M. An R Function for Cronbach's Alpha Analysis: A Case-Based Approach. Natl J Community Med 2022;13(8):571-575. DOI: 10.55489/njcm.130820221149

Financial Support: None declared | **Conflict of Interest:** None declared | **Date of Submission**: 28-03-2022 | **Date of Acceptance**: 24-06-2022 | **Date of Publication**: 31-08-2022

Correspondence: Rohith M (Email Id: drrohith88@gmail.com)

riety of different packages. **Case scenario -1** To demonstrate the calculation involved in Cronbach alpha, by framing hypothetical data frame and by reading excel sheet both the scenarios are explained in this paper. The ltm package is used to create hypothetical data frame. The data frame consists of 3 items with 10 responses each, the syntax below can be modified accordingly involving more items and responses as per questionnaire is concerned, also 95% confidence interval can also be generated using second part of syntax. **Case scenario -2 To** estimate internal consistency using Cronbach's alpha by reading excel data is done using following syntax, raw data is taken from a study development and validation of awareness, knowledge, attitude and skill questionnaire for tele rehabilitation in stuttering among speech language pathologists. There are 14 questions in the questionnaire. **Case scenario -3** To establish internal consistency, raw data was taken from a study done to assess the factors influencing daily regimen of anti-tubercular drugs in Mangalore city, A semi-structured, pretested questionnaire was developed and administered to individuals who are on treatment, with modifications relevant to local conditions. The questionnaire consisted of 5 parts i) Socio-demographic variables ii) Knowledge about TB drugs iii) Treatment phase/ Category iv) HIV status and anti-retroviral therapy adherence. v) Factors influencing the compliance with the treatment. vi) Health facility details and other features.

Case scenario1

```
> #ltm library function is used to calculate Cronbach alpha
> library(ltm)
> #enter survey responses as a data frame
> data <- data.frame(Q1=c(1, 2, 2, 3, 2, 2, 3, 3, 2, 3), Q2=c(1, 1, 1, 2, 3, 3, 2, 3, 3, 3), Q3=c(1, 1, 2, 1, 2, 3, 3, 3, 2, 3))
> #calculate Cronbach's Alpha
> cronbach.alpha(data)
```

```
Cronbach's alpha for the 'data' data-set
Items: 3
Sample units: 10
alpha: 0.773
```

```
> #we can also specify CI=True to return a 95% confidence interval for Cronbach's Alpha
> cronbach.alpha(data, CI=TRUE)
```

```
Cronbach's alpha for the 'data' data-set
Items: 3
Sample units: 10
alpha: 0.773

Bootstrap 95% CI based on 1000 samples
2.5% 97.5%
0.174 0.930
```

Case scenario 2

```
> #psych library function is used to calculate cronbach's alpha
> library("psych")
> #enter the file directory path where csv file is placed in the system
> filePath <- "C:/Users/MAHE/Documents/Stuttering-SLP.csv"
> #reading the csv file
> survdata<-read.csv(filePath)
> #Estimate cronbach's alpha
> result <- alpha(survdata[,c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9","Q10","Q11","Q12","Q13","Q14")],
check.keys=TRUE)
> #Printing the results of cronbach's alpha
> print(result)

Reliability analysis
Call: alpha(x = survdata[, c("Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8", "Q9", "Q10", "Q11", "Q12", "Q13", "Q14")],
check.keys = TRUE)
```

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.87	0.89	0.85	0.36	7.8	0.019	1.9	0.37	0.2

	95% confidence boundaries		
	lower	alpha	upper
Feldt	0.83	0.87	0.90
Duhachek	0.83	0.87	0.91

Reliability if an item is dropped:

raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Q1	0.86	0.87	0.85	0.34	6.8	0.020	0.11	0.2
Q2-	0.85	0.88	0.84	0.35	7.1	0.021	0.11	0.2
Q3-	0.86	0.88	0.85	0.36	7.5	0.021	0.12	0.2
Q4	0.86	0.87	0.85	0.34	6.8	0.020	0.11	0.2
Q5-	0.85	0.88	0.84	0.35	7.1	0.021	0.11	0.2
Q6-	0.86	0.88	0.85	0.36	7.5	0.021	0.12	0.2
Q7	0.86	0.87	0.85	0.34	6.8	0.020	0.11	0.2
Q8-	0.85	0.88	0.84	0.35	7.1	0.021	0.11	0.2
Q9	0.86	0.87	0.85	0.34	6.8	0.020	0.11	0.2
Q10	0.86	0.87	0.82	0.34	6.8	0.020	0.11	0.2
Q11-	0.85	0.88	0.84	0.35	7.1	0.021	0.11	0.2
Q12-	0.86	0.88	0.85	0.36	7.5	0.021	0.12	0.2
Q13	0.87	0.89	0.84	0.39	8.3	0.019	0.12	0.2
Q14	0.87	0.89	0.84	0.38	8.0	0.019	0.12	0.2

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
Q1	108	0.61	0.75	0.71	0.56	1.1	0.40
Q2-	108	0.70	0.67	0.60	0.63	2.9	0.65
Q3-	108	0.70	0.57	0.49	0.60	2.3	0.89
Q4	108	0.61	0.75	0.71	0.56	1.1	0.40
Q5-	108	0.70	0.67	0.60	0.63	2.9	0.65
Q6-	108	0.70	0.57	0.49	0.60	2.3	0.89
Q7	108	0.61	0.75	0.71	0.56	1.1	0.40
Q8-	108	0.70	0.67	0.60	0.63	2.9	0.65
Q9	108	0.61	0.75	0.71	0.56	1.1	0.40
Q10	108	0.60	0.74	0.80	0.55	1.1	0.41
Q11-	108	0.70	0.67	0.60	0.63	2.9	0.65
Q12-	108	0.70	0.57	0.49	0.60	2.3	0.89
Q13	108	0.32	0.35	0.31	0.24	1.3	0.46
Q14	108	0.37	0.41	0.37	0.28	1.5	0.50

Non missing response frequency for each item

	1	2	3	4	5	miss
Q1	0.88	0.10	0.02	0.00	0.00	0
Q2	0.04	0.01	0.82	0.08	0.05	0
Q3	0.08	0.00	0.11	0.78	0.03	0
Q4	0.88	0.10	0.02	0.00	0.00	0
Q5	0.04	0.01	0.82	0.08	0.05	0
Q6	0.08	0.00	0.11	0.78	0.03	0
Q7	0.88	0.10	0.02	0.00	0.00	0
Q8	0.04	0.01	0.82	0.08	0.05	0
Q9	0.88	0.10	0.02	0.00	0.00	0
Q10	0.87	0.11	0.02	0.00	0.00	0
Q11	0.04	0.01	0.82	0.08	0.05	0
Q12	0.08	0.00	0.11	0.78	0.03	0
Q13	0.69	0.31	0.00	0.00	0.00	0
Q14	0.51	0.49	0.00	0.00	0.00	0

Case scenario-3

```
> #psych library function is used to calculate cronbach's alpha
> library("psych")
> #enter the file directory path where csv file is placed in the system
> filePath <- "C:/Users/MAHE/Documents/Appraisal.csv"
> #reading the csv file
> data <- read.csv(filePath)
> #Estimate cronbach's alpha
> result <- al-
pha(data[,c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9","Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19
","Q20")], check.keys=TRUE)
> #Printing the results of cronbach's alpha
> print(result)

Reliability analysis
Call: alpha(x = data[, c("Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8", "Q9", "Q10", "Q11", "Q12", "Q13", "Q14", "Q15", "Q16",
"Q17", "Q18", "Q19", "Q20")], check.keys = TRUE)
```

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.56	0.61	0.74	0.074	1.6	0.045	1.6	0.15	0.052

	95% lower	confidence alpha	boundaries upper
Feldt	0.46	0.56	0.64
Duhachek	0.47	0.56	0.65

Reliability if an item is dropped:

raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
Q1-	0.64	0.64	0.75	0.084	1.7	0.036	0.030	0.053
Q2	0.60	0.63	0.75	0.083	1.7	0.041	0.032	0.060
Q3	0.58	0.63	0.75	0.083	1.7	0.043	0.031	0.055
Q4	0.50	0.57	0.72	0.066	1.3	0.052	0.030	0.043
Q5	0.49	0.57	0.70	0.065	1.3	0.052	0.028	0.047
Q6	0.49	0.57	0.70	0.065	1.3	0.052	0.029	0.043
Q7	0.50	0.56	0.71	0.064	1.3	0.052	0.030	0.043
Q8	0.56	0.62	0.75	0.080	1.7	0.045	0.032	0.054
Q9	0.54	0.60	0.73	0.073	1.5	0.047	0.030	0.051
Q10	0.55	0.61	0.74	0.077	1.6	0.046	0.031	0.053
Q11	0.53	0.59	0.72	0.070	1.4	0.048	0.031	0.047
Q12	0.52	0.59	0.73	0.071	1.5	0.049	0.032	0.052
Q13	0.54	0.60	0.72	0.072	1.5	0.047	0.029	0.043
Q14-	0.57	0.63	0.75	0.082	1.7	0.044	0.030	0.053
Q15	0.57	0.62	0.75	0.080	1.7	0.044	0.032	0.054
Q16-	0.55	0.61	0.73	0.075	1.5	0.046	0.032	0.053
Q17-	0.56	0.62	0.74	0.079	1.6	0.045	0.032	0.060
Q18	0.53	0.58	0.72	0.068	1.4	0.048	0.029	0.052
Q19-	0.54	0.58	0.71	0.067	1.4	0.047	0.031	0.047
Q20	0.53	0.58	0.71	0.069	1.4	0.048	0.030	0.051

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
Q1-	200	0.197	0.087	-0.00066	-0.094	2.0	0.85
Q2	200	0.164	0.123	0.00877	-0.054	2.3	0.64
Q3	200	0.041	0.107	0.01071	-0.085	1.2	0.37
Q4	200	0.578	0.534	0.52146	0.449	1.6	0.49
Q5	200	0.611	0.572	0.59571	0.493	1.4	0.48
Q6	200	0.591	0.563	0.58456	0.468	1.4	0.48
Q7	200	0.589	0.595	0.59361	0.478	1.3	0.44
Q8	200	0.100	0.182	0.06996	0.019	1.1	0.24
Q9	200	0.360	0.362	0.32119	0.251	1.9	0.35
Q10	200	0.240	0.264	0.20851	0.132	1.1	0.33
Q11	200	0.430	0.432	0.40042	0.290	1.3	0.46
Q12	200	0.446	0.406	0.34132	0.313	1.7	0.45
Q13	200	0.388	0.382	0.36179	0.235	1.4	0.48
Q14-	200	0.133	0.151	0.06316	0.011	2.8	0.36
Q15	200	0.178	0.179	0.06360	0.018	1.7	0.47
Q16-	200	0.294	0.310	0.24498	0.180	2.1	0.35
Q17-	200	0.212	0.228	0.14913	0.072	2.2	0.42
Q18	200	0.420	0.477	0.46834	0.304	1.2	0.39
Q19-	200	0.417	0.502	0.49795	0.349	2.1	0.23
Q20	200	0.407	0.474	0.47322	0.303	1.1	0.34

Non missing response frequency for each item

	1	2	3	miss
Q1	0.37	0.28	0.35	0
Q2	0.10	0.50	0.41	0
Q3	0.84	0.16	0.00	0
Q4	0.40	0.60	0.00	0
Q5	0.65	0.35	0.00	0
Q6	0.64	0.36	0.00	0
Q7	0.74	0.26	0.00	0
Q8	0.94	0.06	0.00	0
Q9	0.14	0.86	0.00	0
Q10	0.88	0.12	0.00	0
Q11	0.70	0.30	0.00	0
Q12	0.27	0.73	0.00	0

Q13	0.63	0.37	0.00	0
Q14	0.85	0.15	0.00	0
Q15	0.33	0.67	0.00	0
Q16	0.14	0.86	0.00	0
Q17	0.22	0.78	0.00	0
Q18	0.82	0.18	0.00	0
Q19	0.06	0.94	0.00	0
Q20	0.86	0.14	0.00	0

INTERPRETATION OF OUTPUT

Cronbach's alpha has direct interpretation, a score of .70 or greater is generally considered to be acceptable 0.90 or greater indicates high consistent, 0.80-0.89 is good consistent,0.70-0.79 is acceptable consistent, 0.65-0.69 is marginal consistent and <0.5 indicates unacceptable consistent. For scales which are used as research tools to compare groups, alpha values of 0.7 to 0.8 are regarded as satisfactory. For the clinical application, much higher values of alpha are needed. The minimum is 0.90 and 0.95 is desirable.

CONCLUSION

Cronbach's alpha Indicates the extent to which the items in your questionnaire are related to each other, it ranges from 0 to 1 and not robust against missing data. It measures only internal consistency of the scale and higher values are always preferred over lower ones. Alpha is zero indicates items are not measuring what they supposed to measure. In this paper, the Cronbach's alpha is illustrated in a simplified way to help the researchers. Also, the **Cronbach. Alpha** function provided in this paper will help to generate the alpha value.

REFERENCES

1. Bland JM, Casey ATH, Crockard HA, Stevens J,Moskovich R, Ransford AO, Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy Ann Rheum Dis
2. McKinley RK,Manku-Scott T, Hastings AM , French DP,Baker R.Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the United Kingdom;development of a patient questionnaire. BMJ 1997;314:193-8.
3. Bosma H.Marmot MG,Hemingway H,Nicholson AC, Brunner E,Stansfield SA.Low job control and risk of coronary heart disease in whitehall II(prospective cohort)study, BMJ 1997;314;558-65 .
4. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-333.