Tests Before Factor Analysis

Luis Carlos Castillo

University of Urbino and Universty of Bremen

2024-06-26

Data Preparation for Factor Analysis

This section will present the different statistics and test before constructing the composite index. It includes the following:

- **ANOVA (Analysis of Variance) and Chi-Squared tests** are fundamental statistical methods used to assess differences in data. ANOVA evaluates whether there are significant differences in the mean scores of continuous variables across different groups, such as different years. Chi-Squared tests, on the other hand, determine if there are significant differences in the distribution of categorical variables across groups.

  In this study, these tests were applied to evaluate the significant differences in the means and distributions of variables across six years. This evaluation is crucial in the preliminary stages of factor analysis, as it ensures the suitability and consistency of the data. Significant differences highlighted by these tests can indicate variability that might impact the factor analysis. Understanding the data distribution and ensuring data quality through these tests helps validate the assumptions required for factor analysis. Thus, conducting ANOVA and Chi-Squared tests provides a comprehensive understanding of the data, ensuring it is appropriately structured and reliable for subsequent factor analysis.

Interpreting the results

Interpreting the results of hypothesis testing involves understanding the null (H0) and alternative (H1) hypotheses. The null hypothesis states there is no significant difference between groups, while the alternative suggests a significant difference exists. We calculate a p-value to test these hypotheses. If the p-value is less than 0.05, we reject the null hypothesis, indicating a significant difference. If it's greater, we fail to reject the null hypothesis, suggesting any differences are likely due to chance.

In this study, rejecting the null hypothesis in ANOVA tests indicates significant differences in mean scores of continuous variables across years. For Chi-Squared tests, it suggests significant differences in the distribution of categorical variables over time. These results help validate the data's variability, ensuring its suitability for factor analysis.

ANOVA and Chi-Squaredd for ACCESS Over the Six Yeaars

Variable	Test	Test_Statistic	P_Value	Decision
A2_A2	ANOVA	498.4083	0	Reject Null Hypothesis
A2_C2	ANOVA	458.8307	0	Reject Null Hypothesis
A1_B2b	Chi-Squared Test	1266.0190	0	Reject Null Hypothesis
A2_C5a	Chi-Squared Test	976.1822	0	Reject Null Hypothesis
A2_C4_low	Chi-Squared Test	8350.1290	0	Reject Null Hypothesis
A2_C4_high	Chi-Squared Test	8350.1290	0	Reject Null Hypothesis
size_rev_small	Chi-Squared Test	390.4181	0	Reject Null Hypothesis
size_rev_medium	Chi-Squared Test	138.5187	0	Reject Null Hypothesis
size_rev_large	Chi-Squared Test	186.4343	0	Reject Null Hypothesis

Chi-Squaredd for SKILLS Over the Six Yeaars

Variable	Test	Test_Statistic	P_Value	Decision
S_B1	Chi-Squared Test	184.66335	0	Reject Null Hypothesis
S_B2a	Chi-Squared Test	992.49448	0	Reject Null Hypothesis
S_B5a1	Chi-Squared Test	59.53903	0	Reject Null Hypothesis
S_B5b1	Chi-Squared Test	72.89928	0	Reject Null Hypothesis
S_B5c1	Chi-Squared Test	53.64918	0	Reject Null Hypothesis
S_B5d1	Chi-Squared Test	78.98737	0	Reject Null Hypothesis
S_B5e1	Chi-Squared Test	40.25756	0	Reject Null Hypothesis
S_B5f1	Chi-Squared Test	77.46799	0	Reject Null Hypothesis
S_B5g1	Chi-Squared Test	190.34061	0	Reject Null Hypothesis
size_rev_small	Chi-Squared Test	390.41808	0	Reject Null Hypothesis
size_rev_medium	Chi-Squared Test	138.51869	0	Reject Null Hypothesis
size_rev_large	Chi-Squared Test	186.43426	0	Reject Null Hypothesis

Chi-Squaredd for USAGE Over the Six Yeaars

Variable	Test	Test_Statistic	P_Value	Decision
UMK_C7	Chi-Squared Test	698.4772	0	Reject Null Hypothesis
UC_C8a	Chi-Squared Test	277.0003	0	Reject Null Hypothesis
UMK_C8c	Chi-Squared Test	154.7722	0	Reject Null Hypothesis
UM_C8g	Chi-Squared Test	237.9436	0	Reject Null Hypothesis
UMK_C8h	Chi-Squared Test	3141.1177	0	Reject Null Hypothesis
UM_E1	Chi-Squared Test	125.3922	0	Reject Null Hypothesis
UM_E2b	Chi-Squared Test	1202.3386	0	Reject Null Hypothesis
UM_E2a	Chi-Squared Test	995.8050	0	Reject Null Hypothesis
UC_J7	Chi-Squared Test	18160.3212	0	Reject Null Hypothesis
UMK_C10a	Chi-Squared Test	2553.6938	0	Reject Null Hypothesis
UMK_C10c	Chi-Squared Test	1433.1330	0	Reject Null Hypothesis
size_rev_small	Chi-Squared Test	390.4181	0	Reject Null Hypothesis
size_rev_medium	Chi-Squared Test	138.5187	0	Reject Null Hypothesis
size_rev_large	Chi-Squared Test	186.4343	0	Reject Null Hypothesis

- **Correlation of Variables:** Understanding correlations between variables is foundational in factor analysis. This assessment helps identify whether groups of variables share common patterns or underlying factors, which justifies their reduction into fewer dimensions.
- **Internal Consistency and Cronbach's Alpha:**

  Internal Consistency refers to the reliability of items within a scale or test to consistently measure the same underlying construct. It is a crucial aspect of scale reliability and validity. Cronbach's Alpha is a specific statistic used to measure this internal consistency. It quantifies how well a set of items measure an underlying (unobserved) construct.

  - Thresholds for Cronbach's Alpha:
    - $\alpha \geq 0.90$: Excellent
    - $0.80 \leq \alpha < 0.90$: Good
    - $0.70 \leq \alpha < 0.80$: Acceptable

  A high Cronbach's Alpha indicates that the items are highly correlated and thus reliably measure the same construct. This helps to ensure that the dimensions extracted from factor analysis are reliable.

Data Preparation for Factor Analysis II

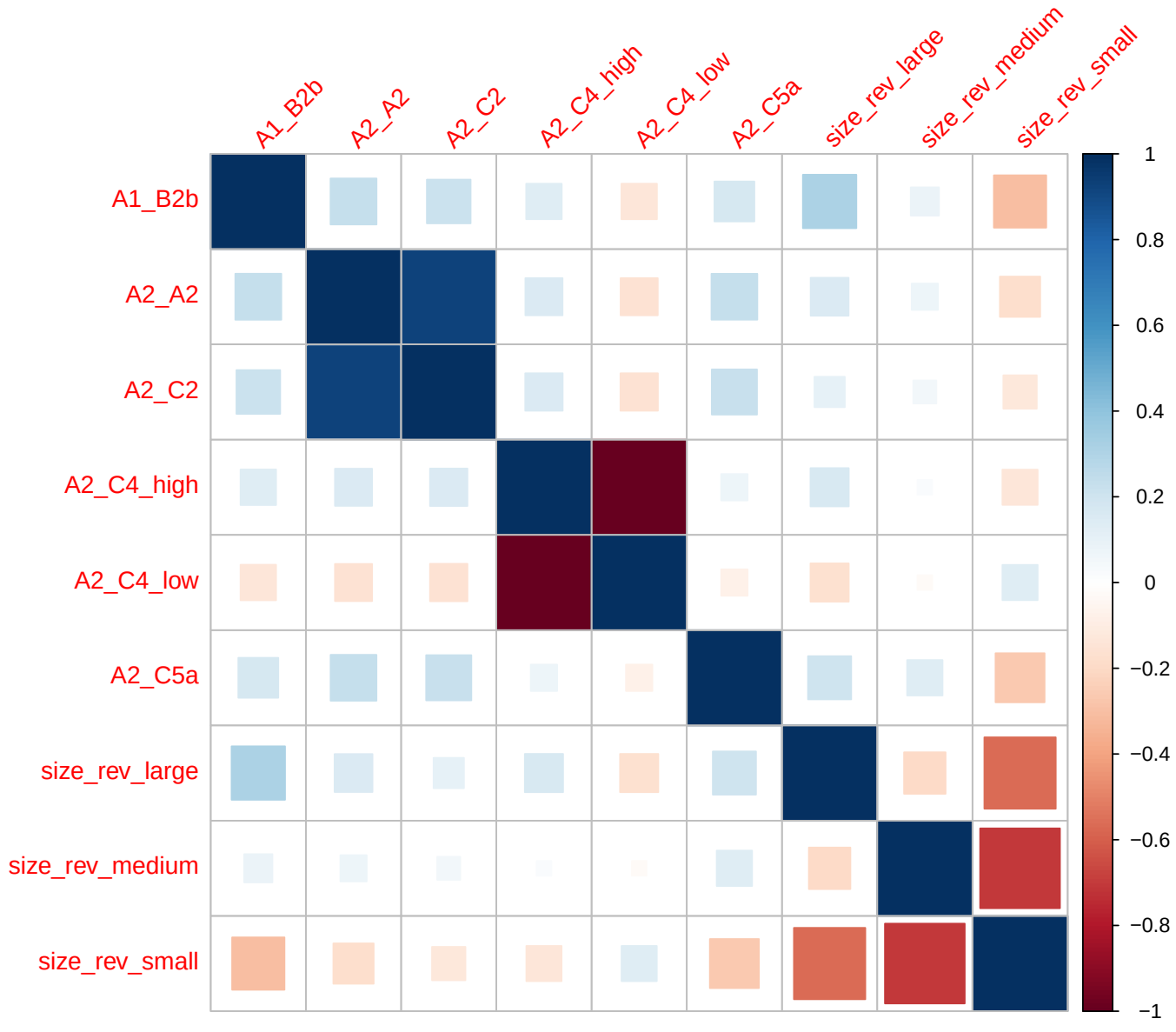
- **Bartlett's Test of Sphericity:** This test checks whether the correlation matrix is an identity matrix, which would indicate that the variables are unrelated and unsuitable for structure detection. A significant result from Bartlett's test supports the factorability of the correlation matrix.
  - **Null Hypothesis (H0)**: The correlation matrix is an identity matrix, indicating that the variables are unrelated and unsuitable for structure detection.
  - **Alternative Hypothesis (H1)**: The correlation matrix is not an identity matrix, indicating significant correlations among the variables, making them suitable for structure detection.

Interpretation of the p-value:

- $\mathbf{p} < 0.05$: Reject the null hypothesis. There are significant correlations among the variables, suggesting that the correlation matrix is factorable. This means that the data is suitable for factor analysis.
- $\mathbf{p} \geq 0.05$: Fail to reject the null hypothesis. There is not enough evidence to suggest that the correlation matrix is different from an identity matrix, indicating that the variables may not be suitable for structure detection.

1. 2014 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2014

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.9628538	0.962927	0.9285045	0.9285045	25.97377

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.6673214	-1.751375	-1.422199	-0.1000314	-0.6365453

- Combining continuous and binary variables

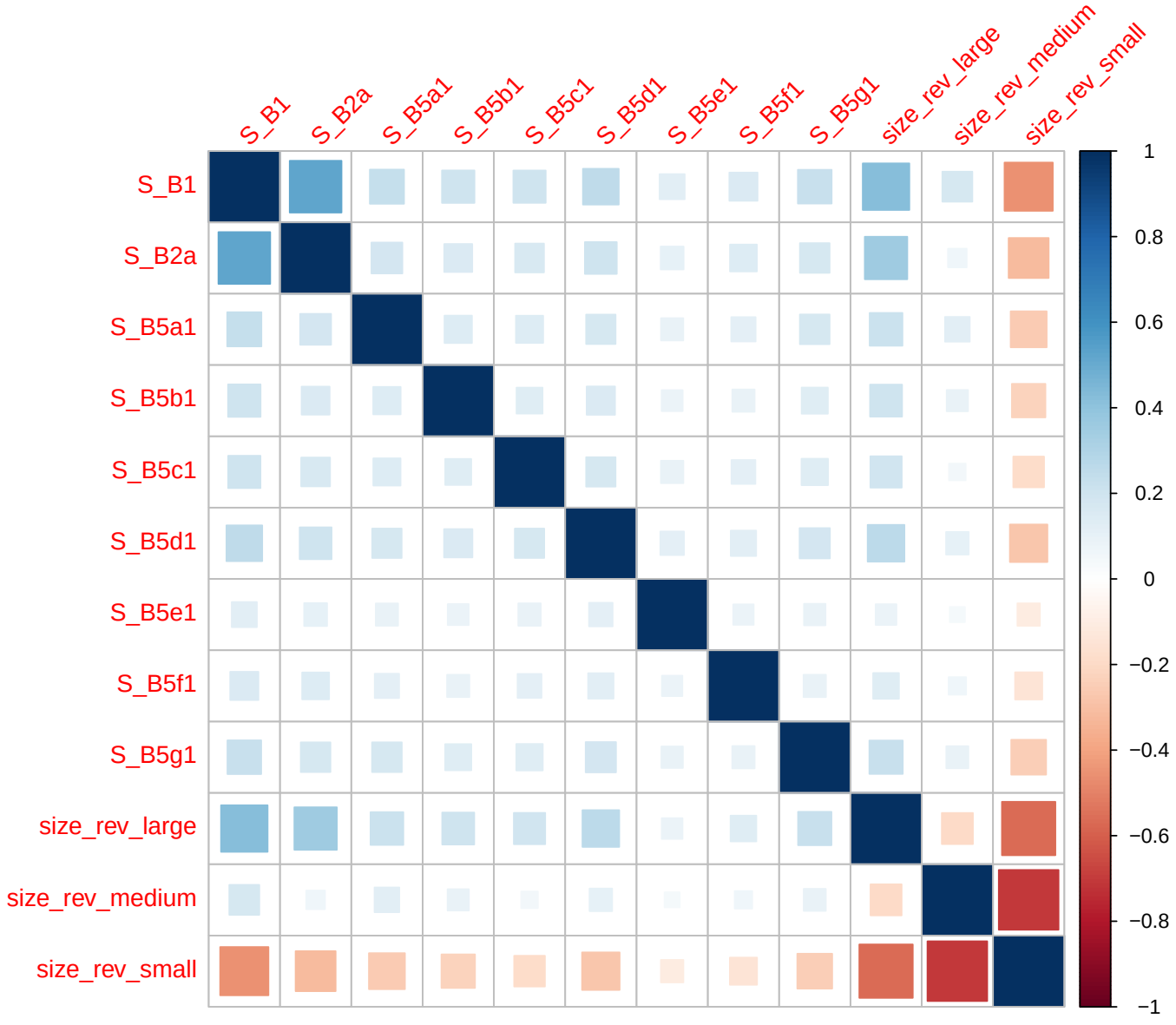
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7099288	-0.0450952	0.1920318	-0.0048175	-0.0431493

1.3. Bartlett's Test of Sphericity Access 2014

Statistic	Degrees_of_Freedom	p_Value
1245763	36	0

2. 2014 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2014

- For binary varibles

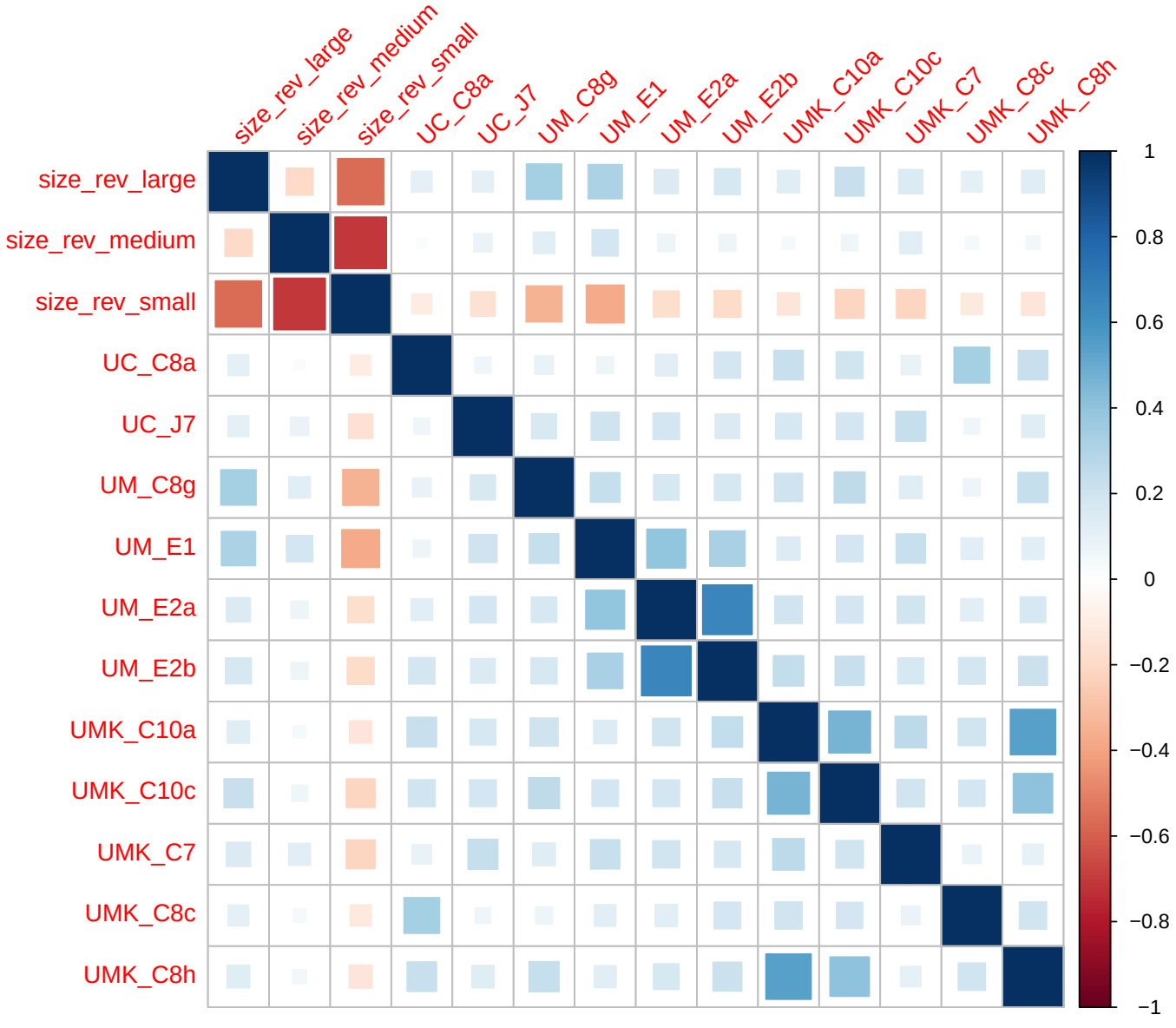
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7263091	0.4959274	0.5152401	0.0757743	0.9838411

2.3. Bartlett's Test of Sphericity Skills 2014

Statistic	Degrees_of_Freedom	p_Value
558135.9	66	0

3. 2014 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2014

- For binary varibles

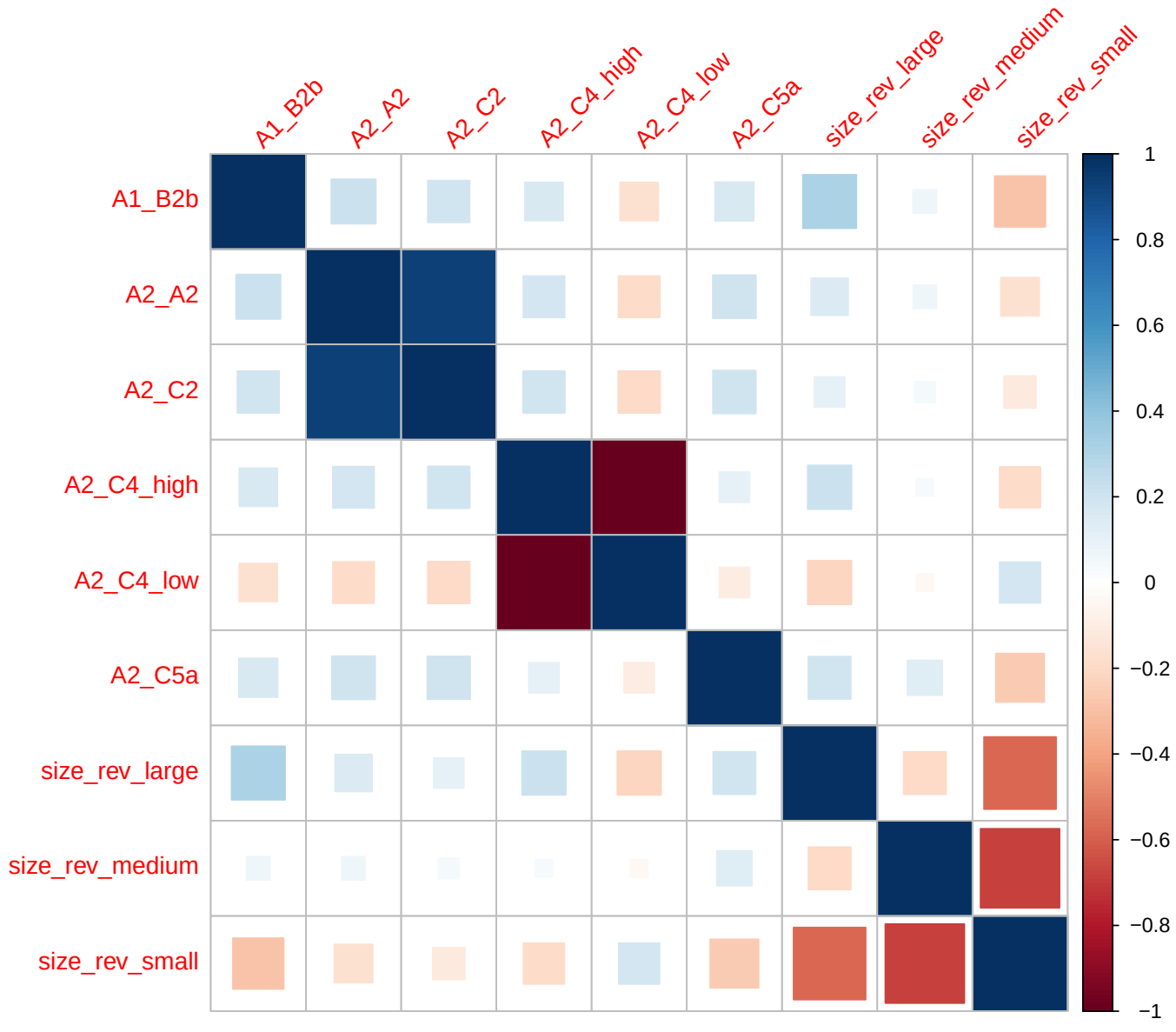
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7668949	0.6481703	0.6880681	0.1162891	1.842284

3.3. Bartlett's Test of Sphericity Usage 2014

Statistic	Degrees_of_Freedom	p_Value
580456.4	91	0

1. 2015 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2015

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.966592	0.9666152	0.9353874	0.9353874	28.95372

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.6830248	-1.811399	-1.476293	-0.1013745	-0.6443052

- Combining continuous and binary variables

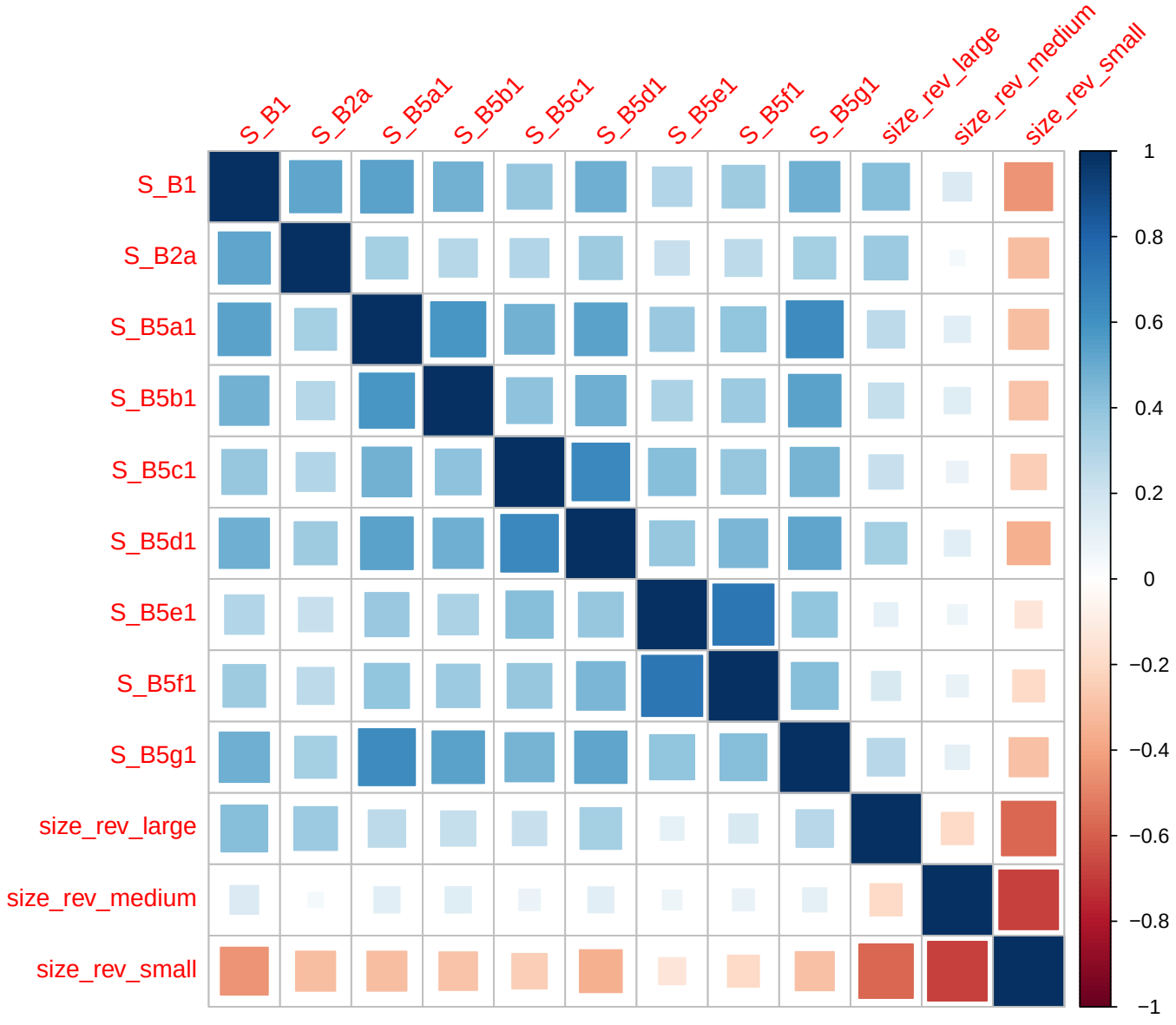
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7194492	-0.0772024	0.1701028	-0.0080272	-0.0716693

1.3. Bartlett's Test of Sphericity Access 2015

Statistic	Degrees_of_Freedom	p_Value
1308803	36	0

2. 2015 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2015

- For binary varibles

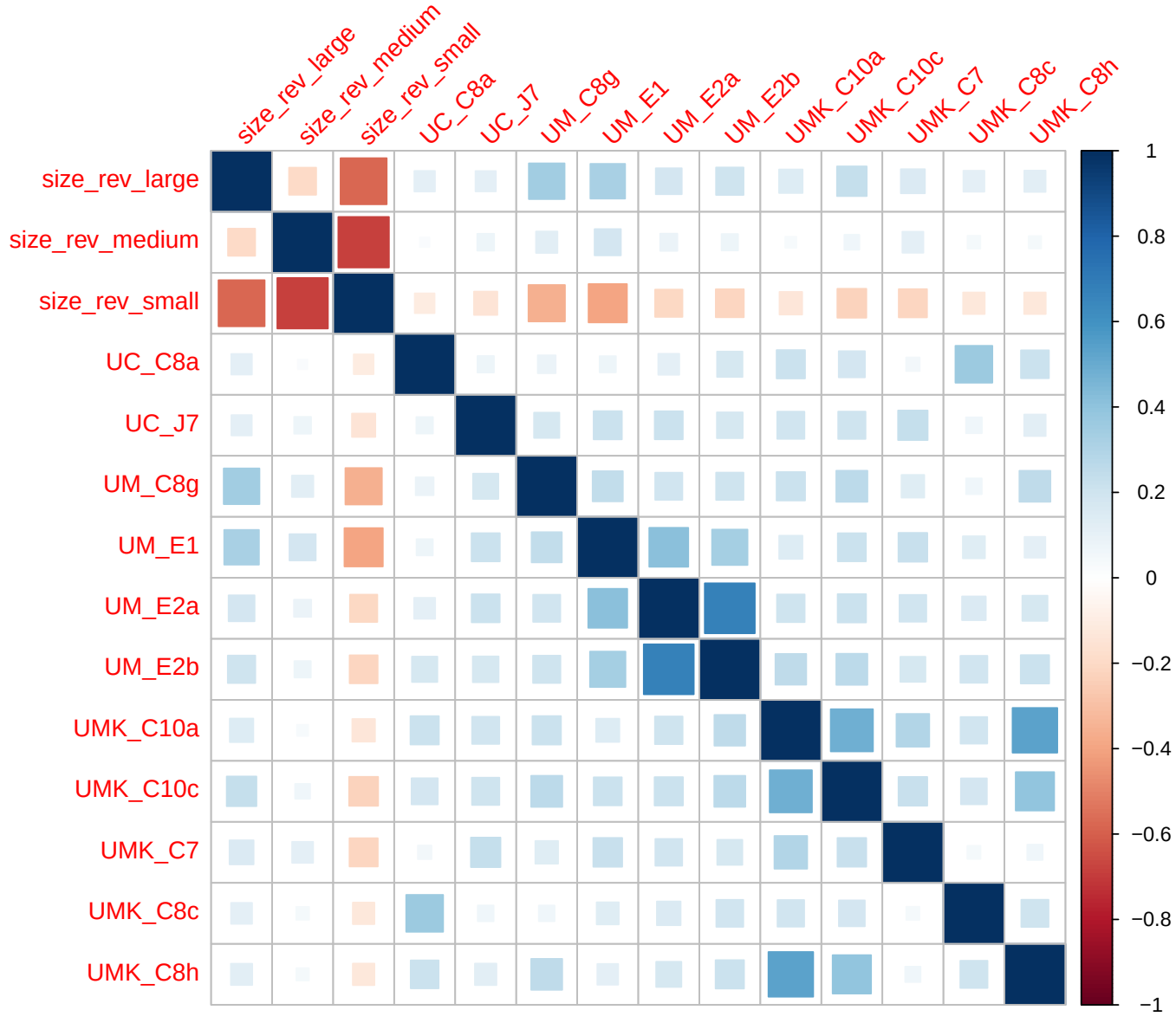
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.8644353	0.7775615	0.8189146	0.2255878	3.495624

2.3. Bartlett's Test of Sphericity Skills 2015

Statistic	Degrees_of_Freedom	p_Value
634071.9	66	0

3. 2015 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2015

- For binary varibles

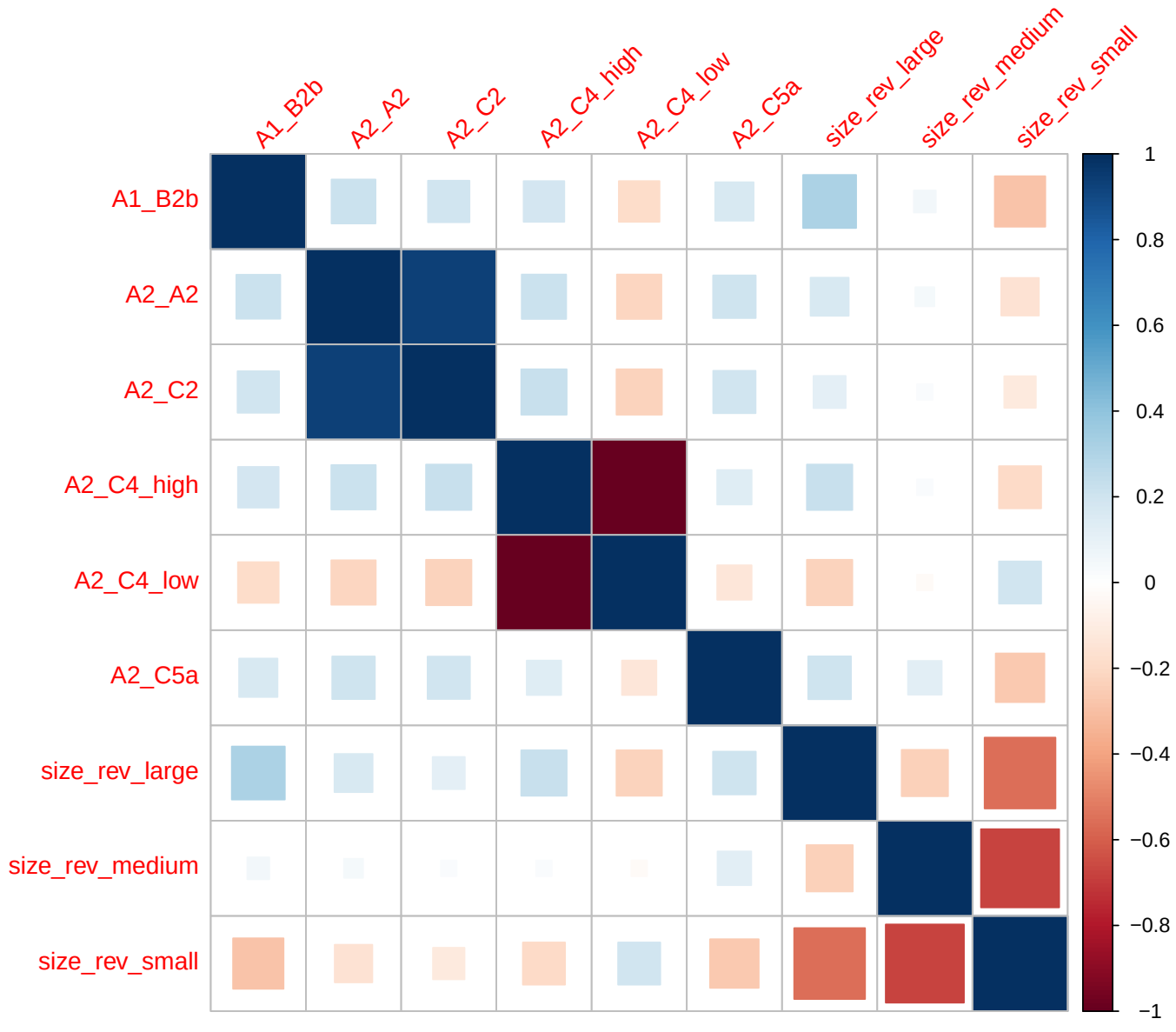
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7742574	0.6558194	0.6995897	0.1197986	1.905451

3.3. Bartlett's Test of Sphericity Usage 2015

Statistic	Degrees_of_Freedom	p_Value
600664.8	91	0

1. 2016 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2016

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.9654554	0.9654639	0.9332336	0.9332336	27.95519

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.6895462	-1.871075	-1.52412	-0.1026571	-0.6516984

- Combining continuous and binary variables

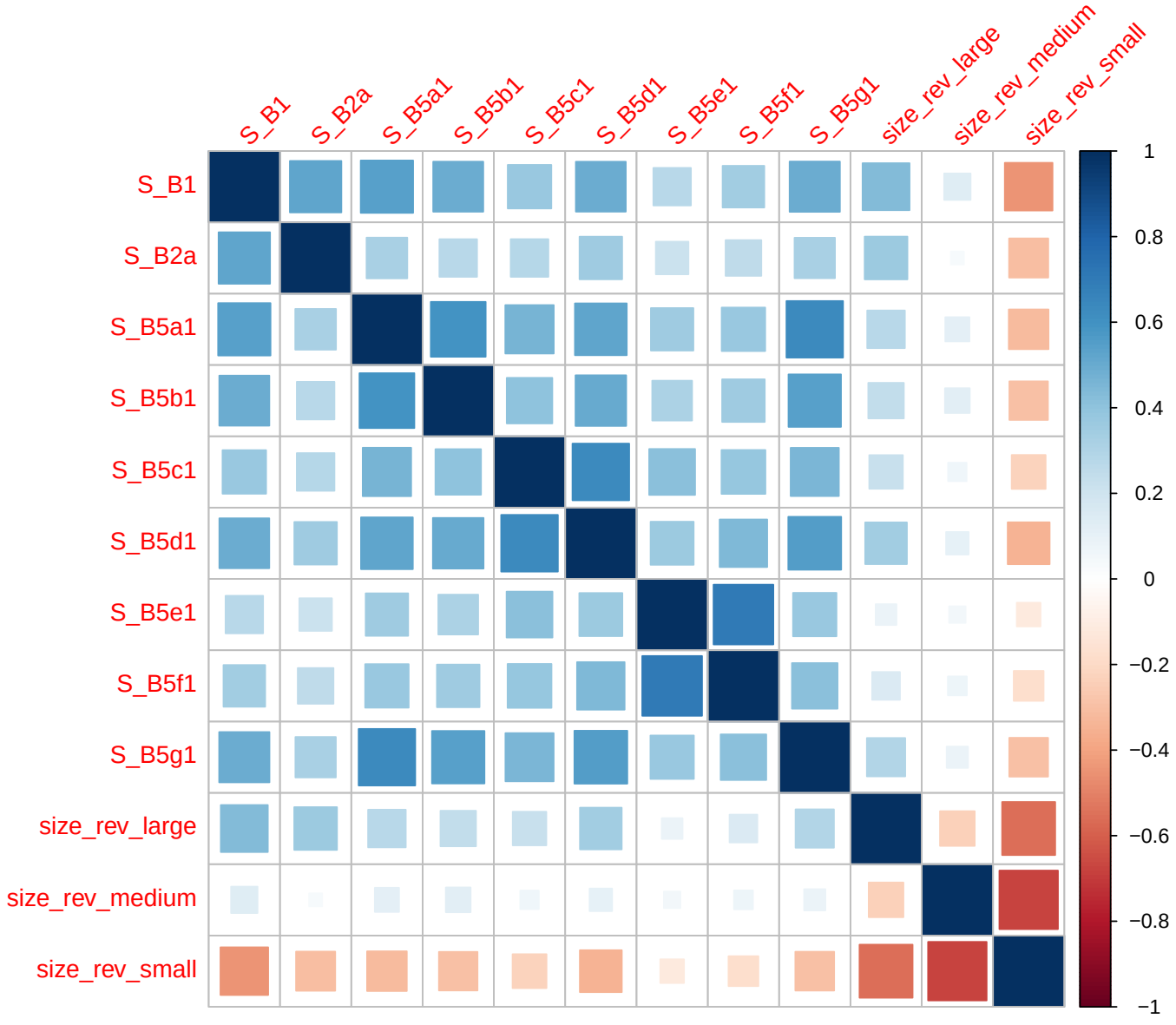
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7246086	-0.0923606	0.1585757	-0.0094837	-0.0845513

1.3. Bartlett's Test of Sphericity Access 2016

Statistic	Degrees_of_Freedom	p_Value
1309465	36	0

2. 2016 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2016

- For binary varibles

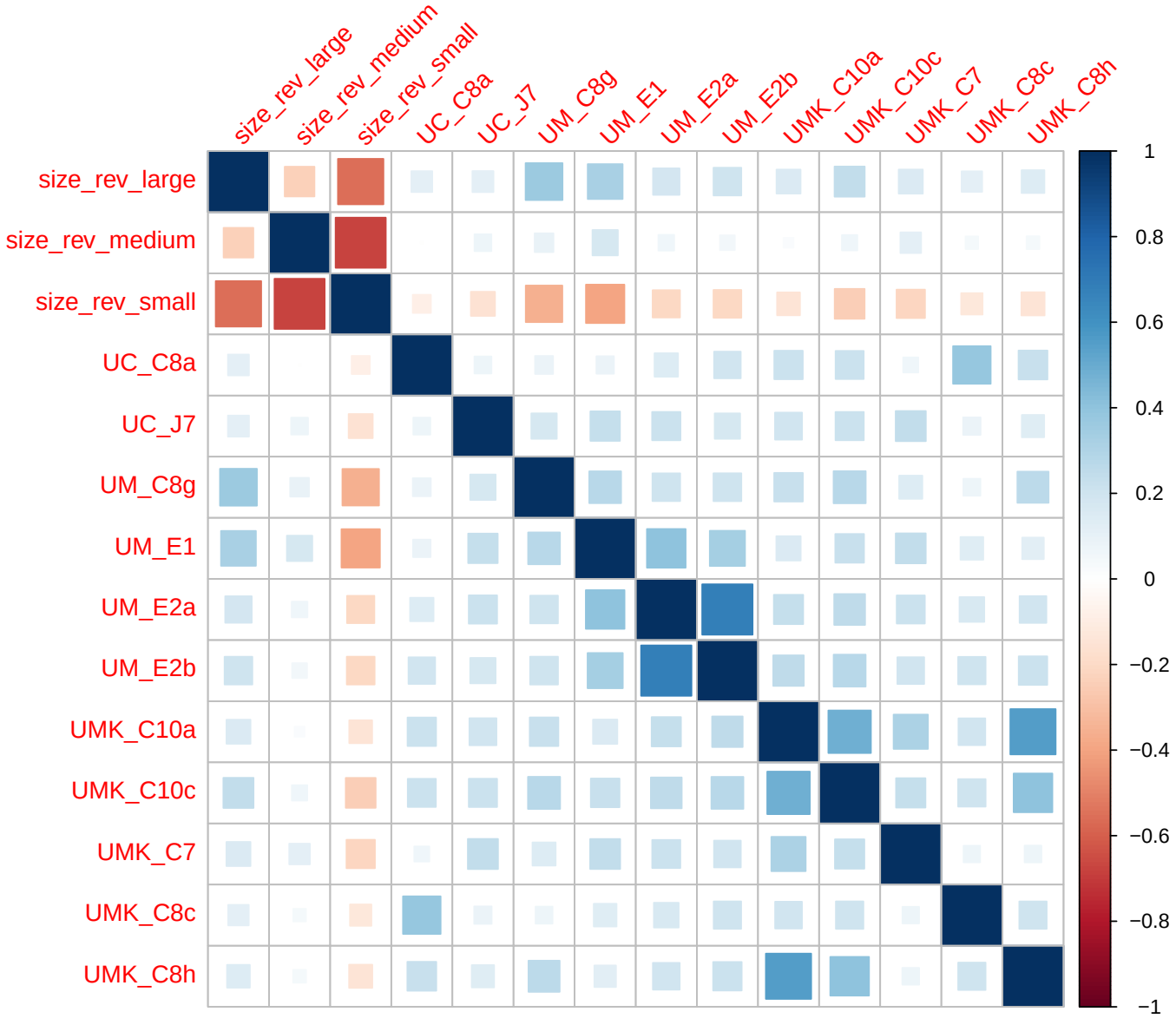
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.8606735	0.7735811	0.8161382	0.2216179	3.416593

2.3. Bartlett's Test of Sphericity Skills 2016

Statistic	Degrees_of_Freedom	p_Value
649206.1	66	0

3. 2016 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2016

- For binary varibles

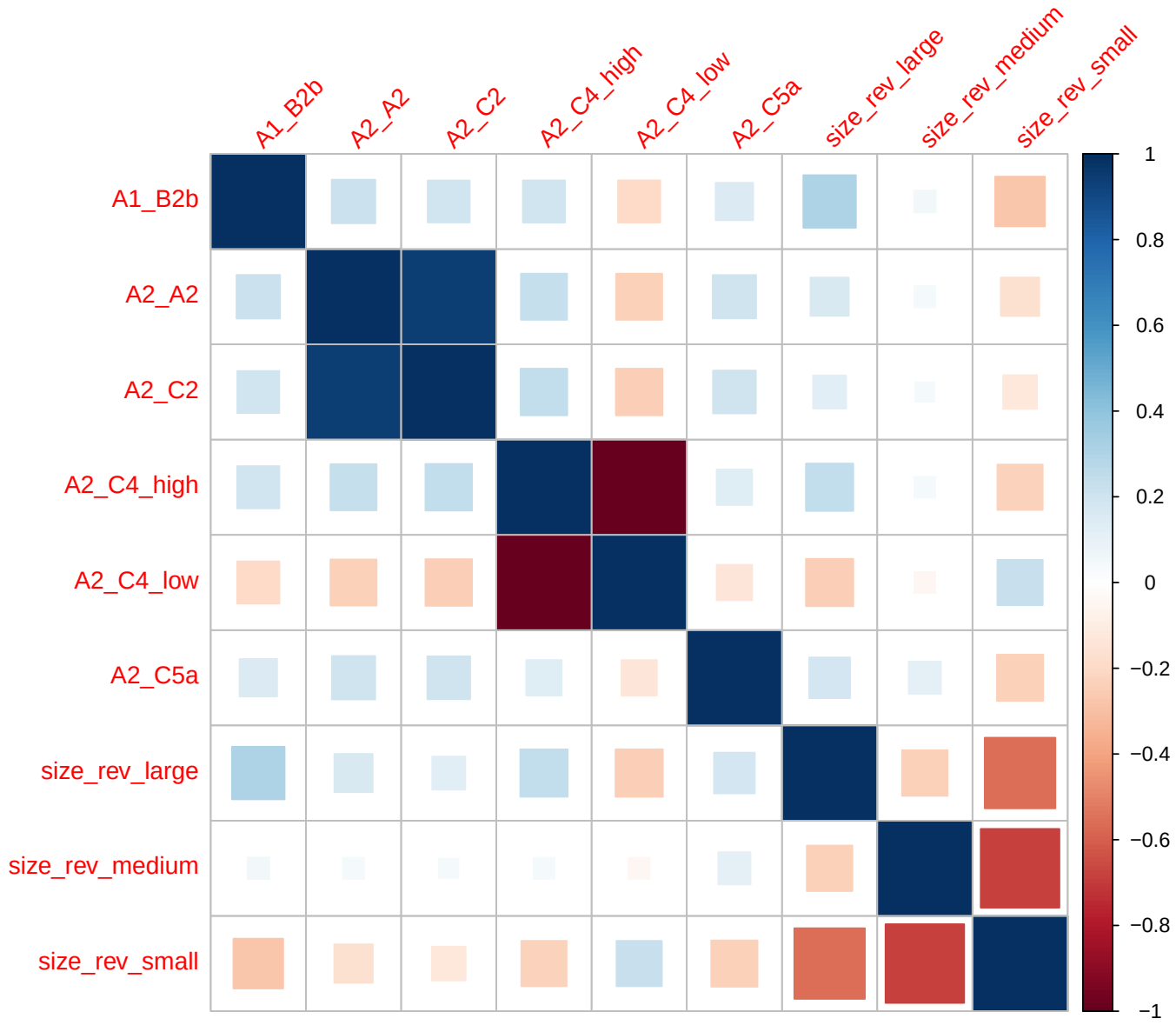
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7812859	0.6678692	0.7113806	0.1255936	2.010862

3.3. Bartlett's Test of Sphericity Usage 2016

Statistic	Degrees_of_Freedom	p_Value
619288.9	91	0

1. 2017 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2017

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.9694792	0.9694812	0.94077	0.94077	31.7667

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7007189	-1.888351	-1.543957	-0.1030191	-0.6537817

- Combining continuous and binary variables

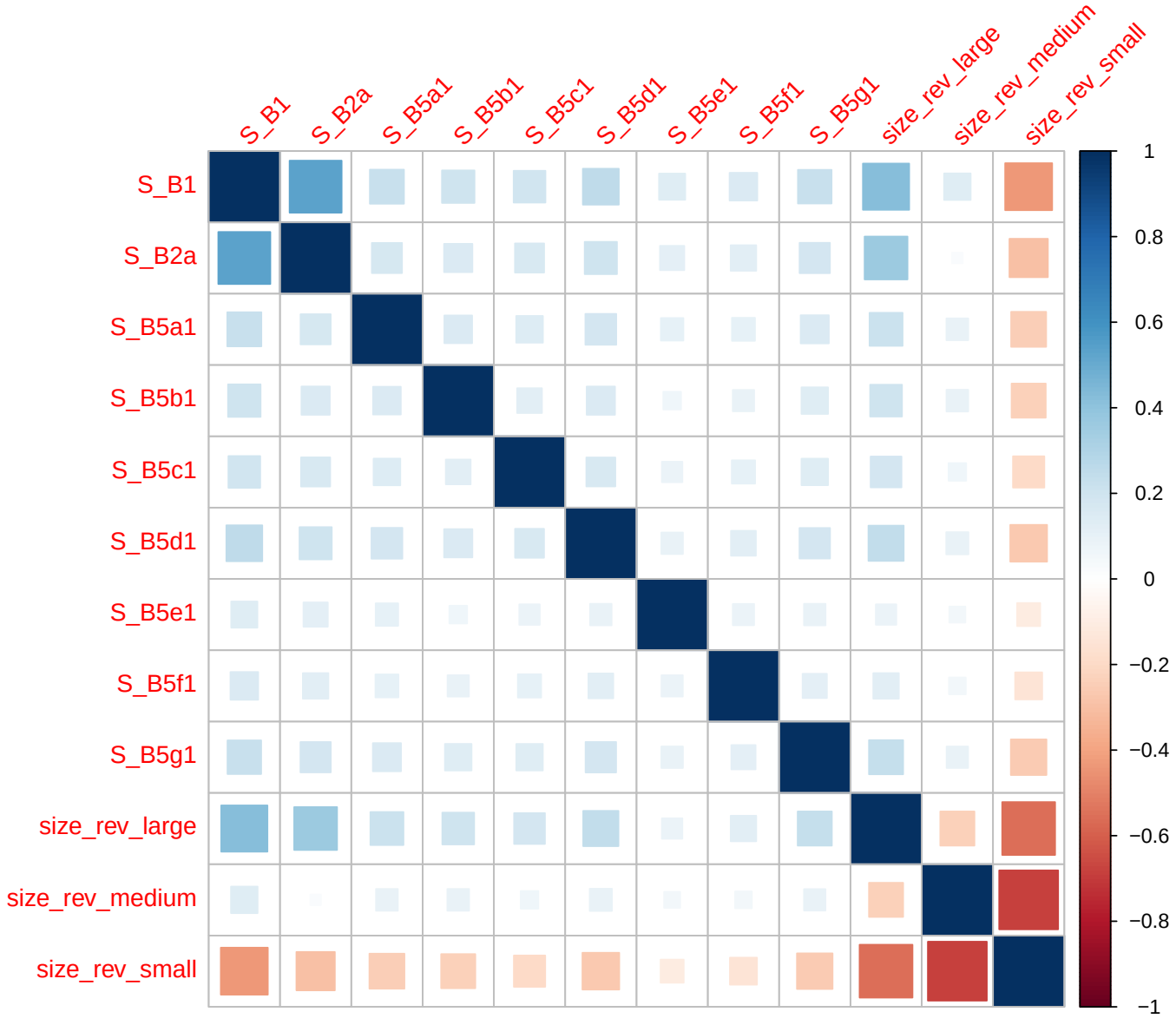
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7354846	-0.0861889	0.1641511	-0.0088951	-0.0793498

1.3. Bartlett's Test of Sphericity Access 2017

Statistic	Degrees_of_Freedom	p_Value
Inf	36	0

2. 2017 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2017

- For binary varibles

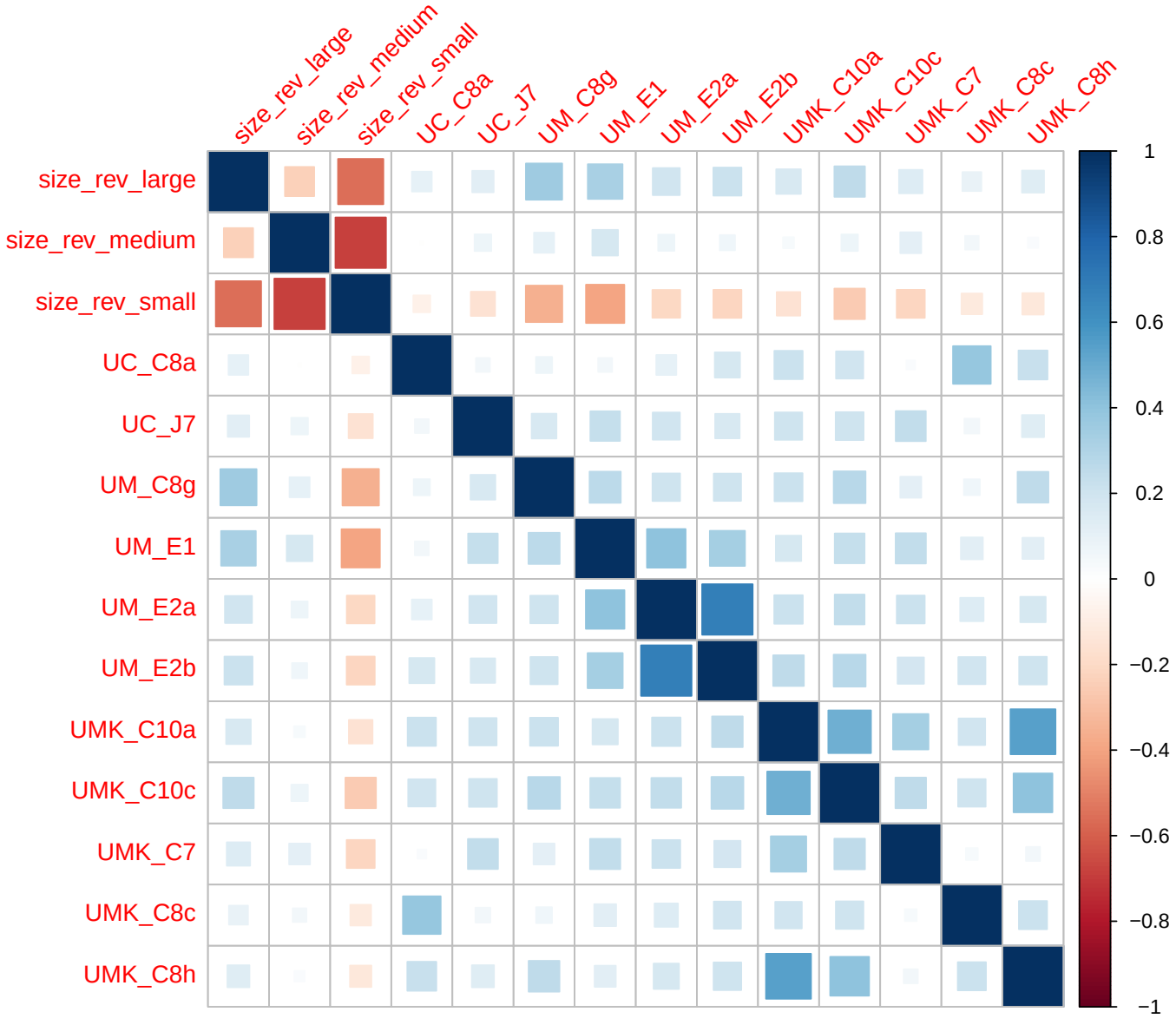
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7224824	0.4873174	0.5079796	0.0733966	0.9505245

2.3. Bartlett's Test of Sphericity Skills 2017

Statistic	Degrees_of_Freedom	p_Value
641579	66	0

3. 2017 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2017

- For binary varibles

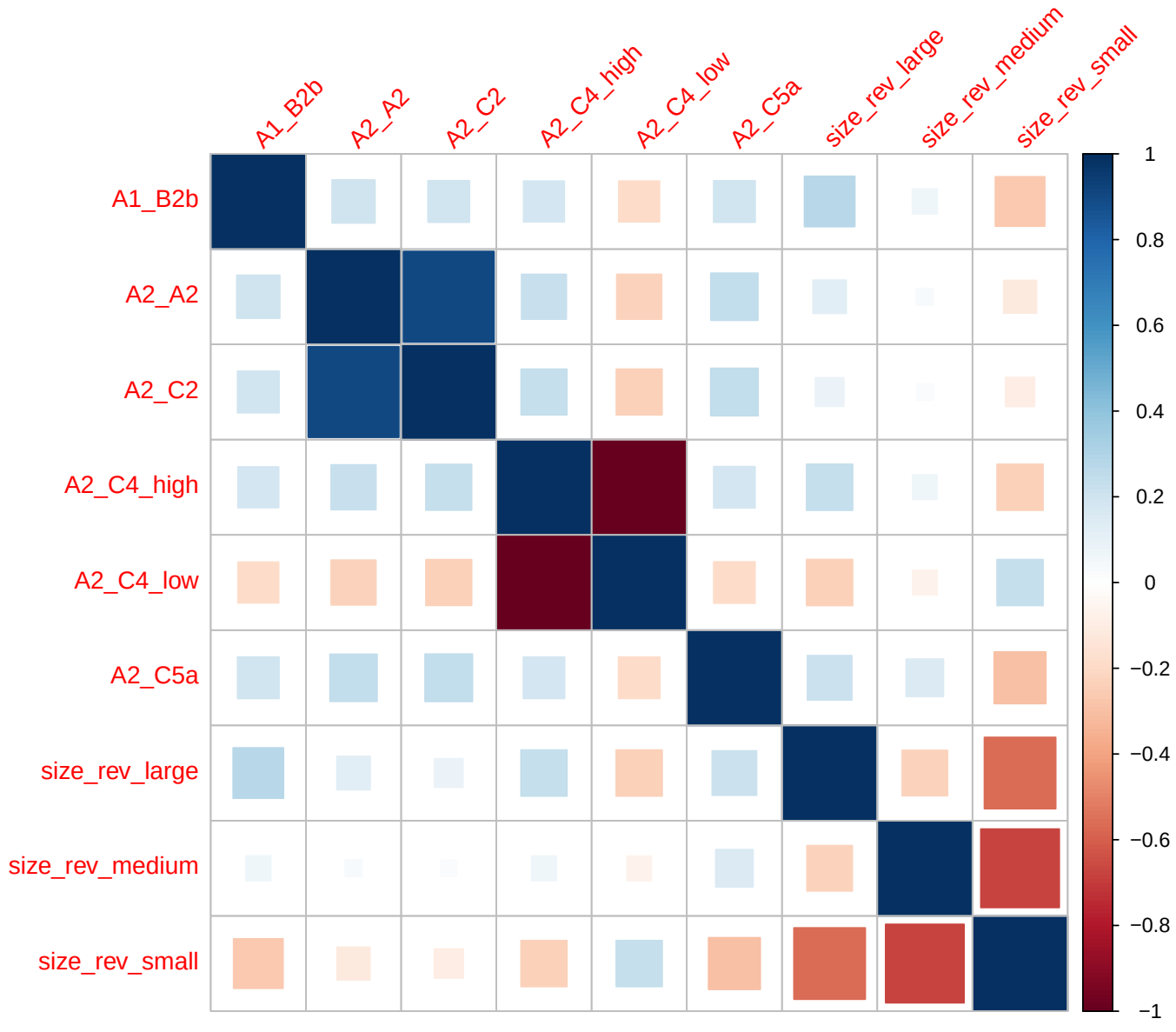
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7765604	0.6591832	0.706429	0.1213827	1.934127

3.3. Bartlett's Test of Sphericity Usage 2017

Statistic	Degrees_of_Freedom	p_Value
674139	91	0

1. 2018 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2018

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.9509248	0.9509287	0.9064482	0.9064482	19.37853

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7130943	-1.794827	-1.455185	-0.1010091	-0.6421961

- Combining continuous and binary variables

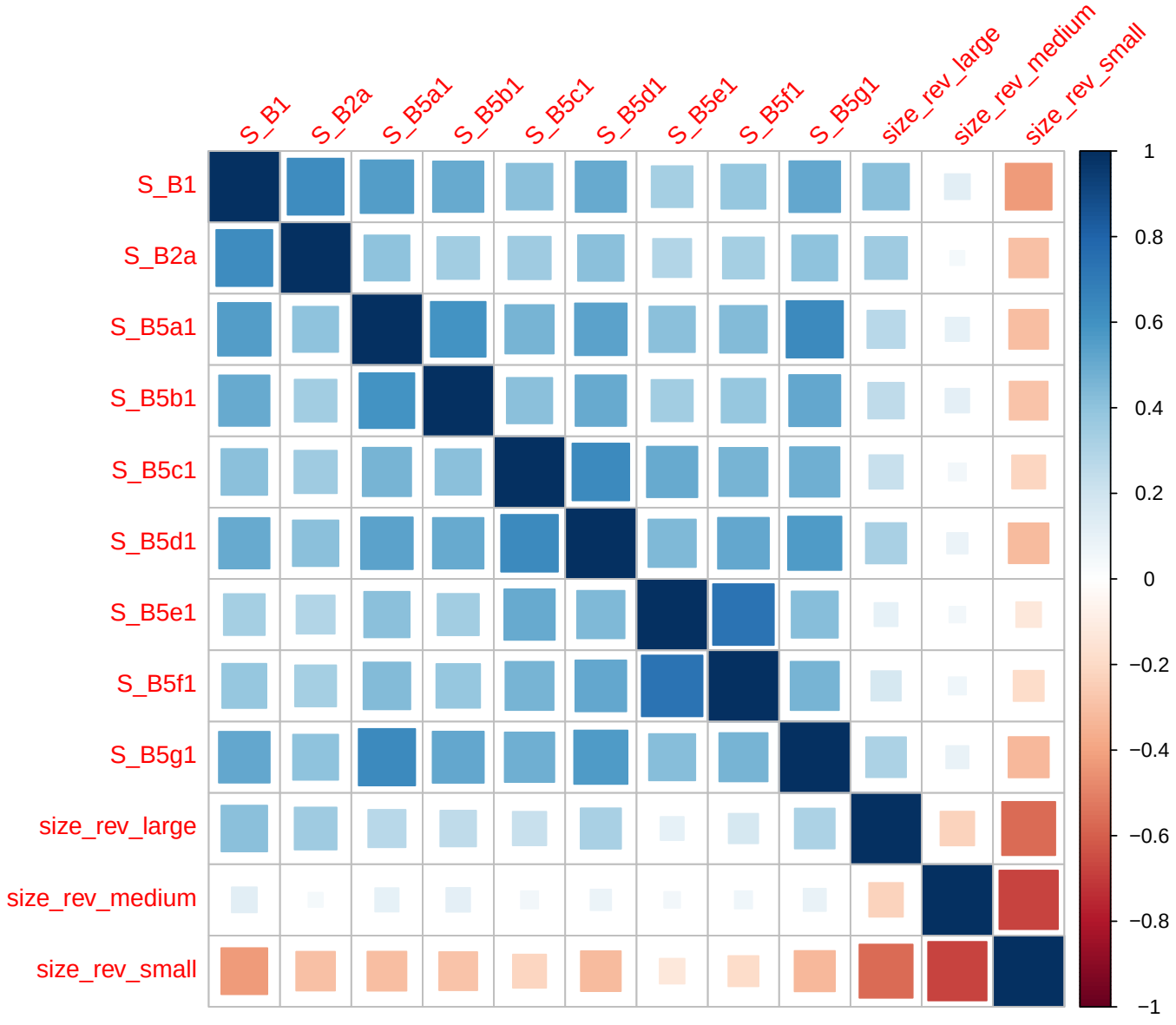
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7381329	-0.0709085	0.1649595	-0.0074116	-0.0662134

1.3. Bartlett's Test of Sphericity Access 2018

Statistic	Degrees_of_Freedom	p_Value
1449356	36	0

2. 2018 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2018

- For binary varibles

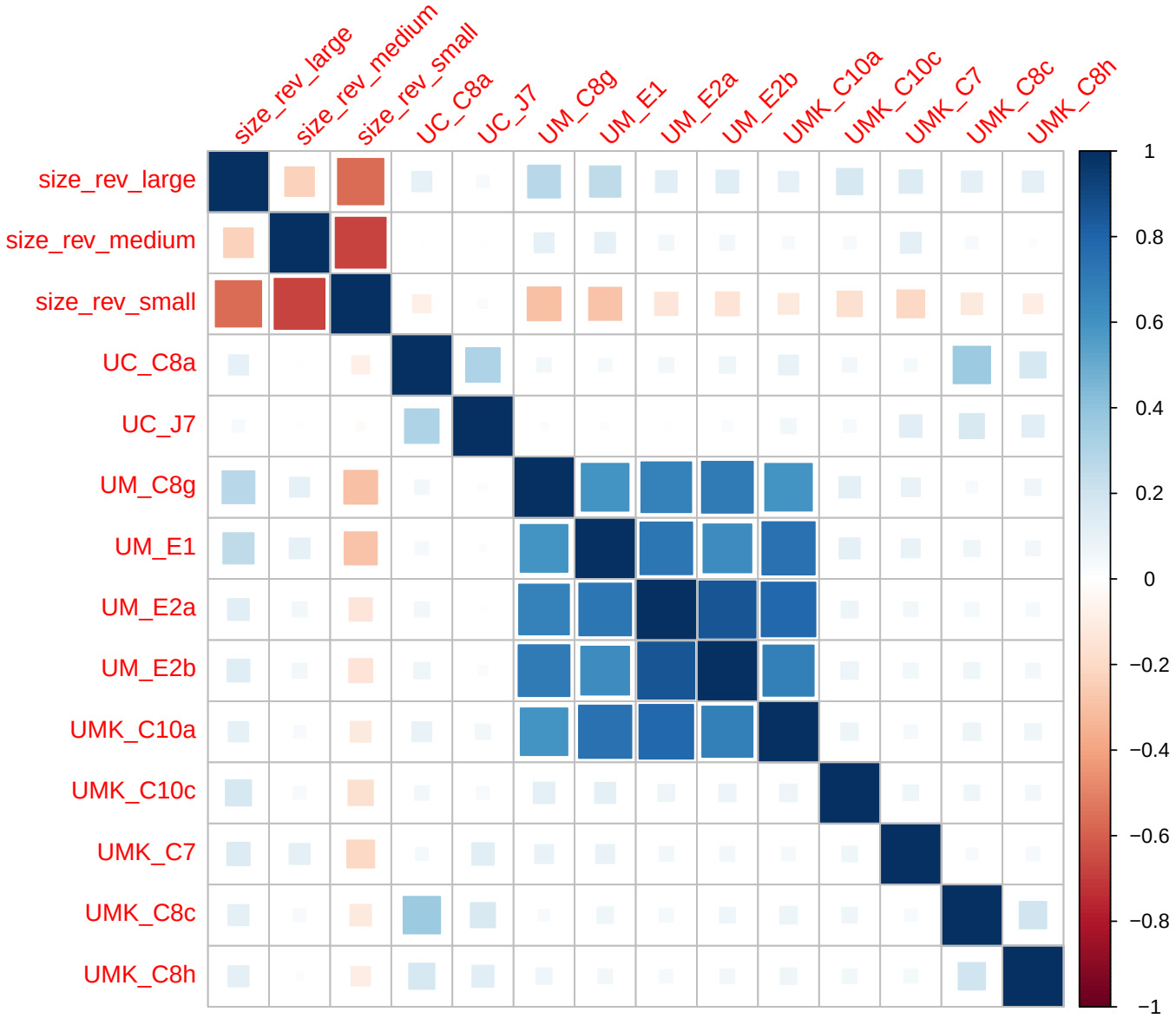
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.8705868	0.7949196	0.8348444	0.2441486	3.876136

2.3. Bartlett's Test of Sphericity Skills 2018

Statistic	Degrees_of_Freedom	p_Value
718070.5	66	0

3. 2018 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2018

- For binary varibles

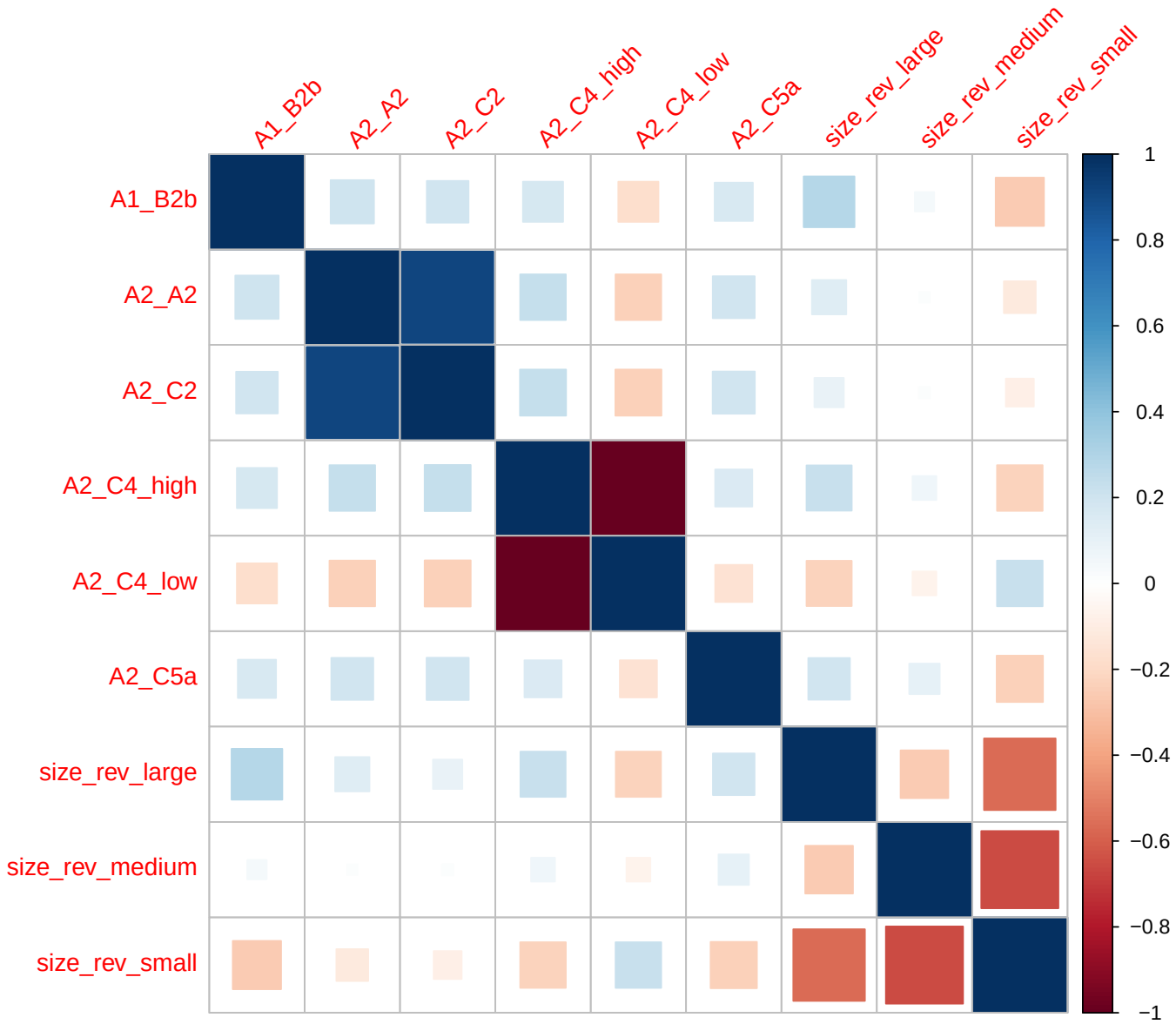
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7527269	0.6235321	0.7045998	0.1057895	1.656269

3.3. Bartlett's Test of Sphericity Usage 2018

Statistic	Degrees_of_Freedom	p_Value
716629.6	91	0

1. 2019 - Access - Correlation, Cronbach, and Bartlett's Ttests

1.1. Correlations



1.2. Cronbach's Alpha Access 2019

- For continuous varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.9572764	0.9572806	0.9180616	0.9180616	22.40857

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.6965247	-1.93966	-1.590836	-0.1040704	-0.6598246

- Combining continuous and binary variables

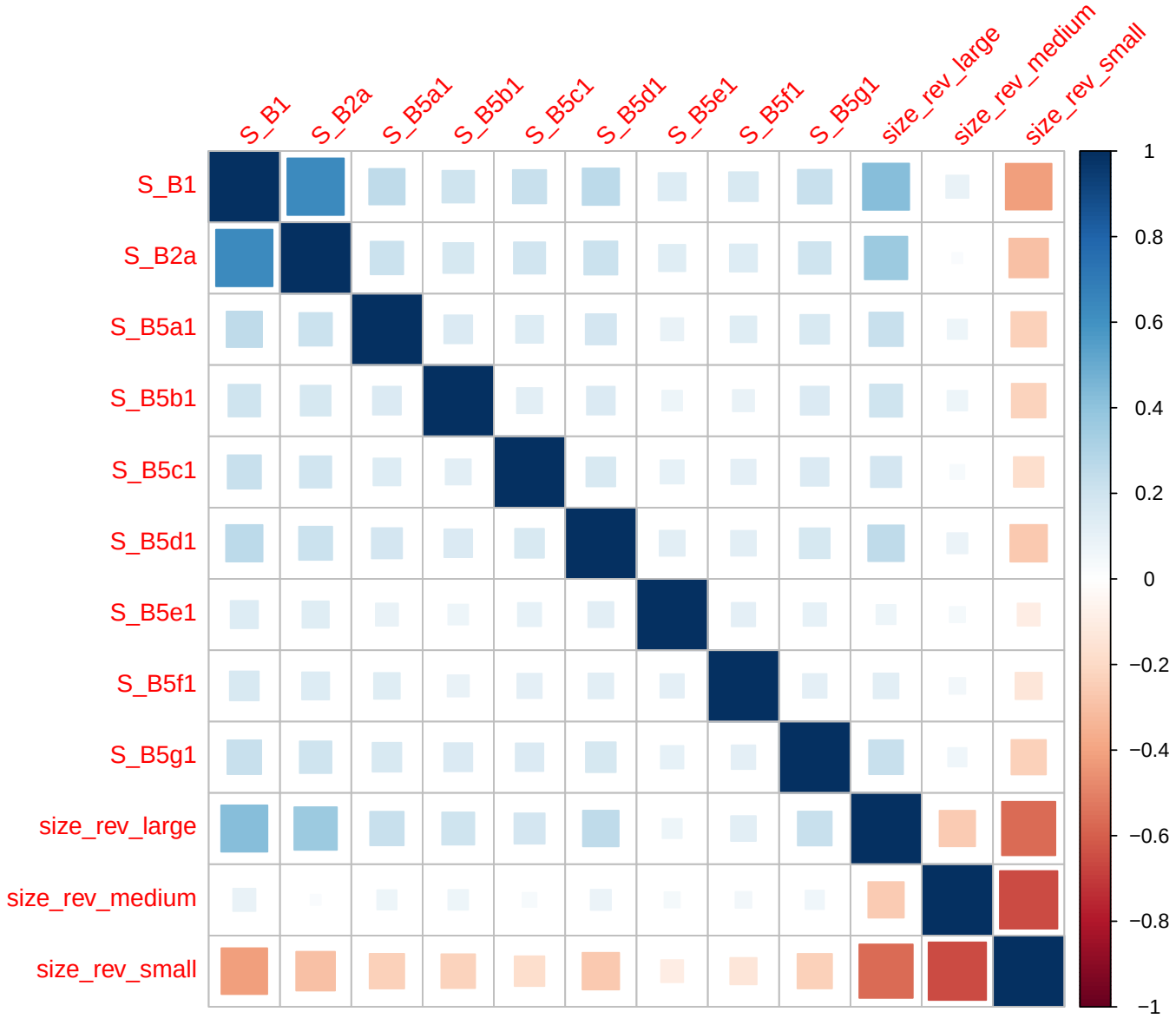
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7249646	-0.1168043	0.1315451	-0.0117575	-0.1045879

1.3. Bartlett's Test of Sphericity Access 2019

Statistic	Degrees_of_Freedom	p_Value
558278.5	28	0

2. 2019 - Skills - Correlation, Cronbach, and Bartlett's Ttests

2.1. Correlations



2.2. Cronbach's Alpha Skills 2019

- For binary varibles

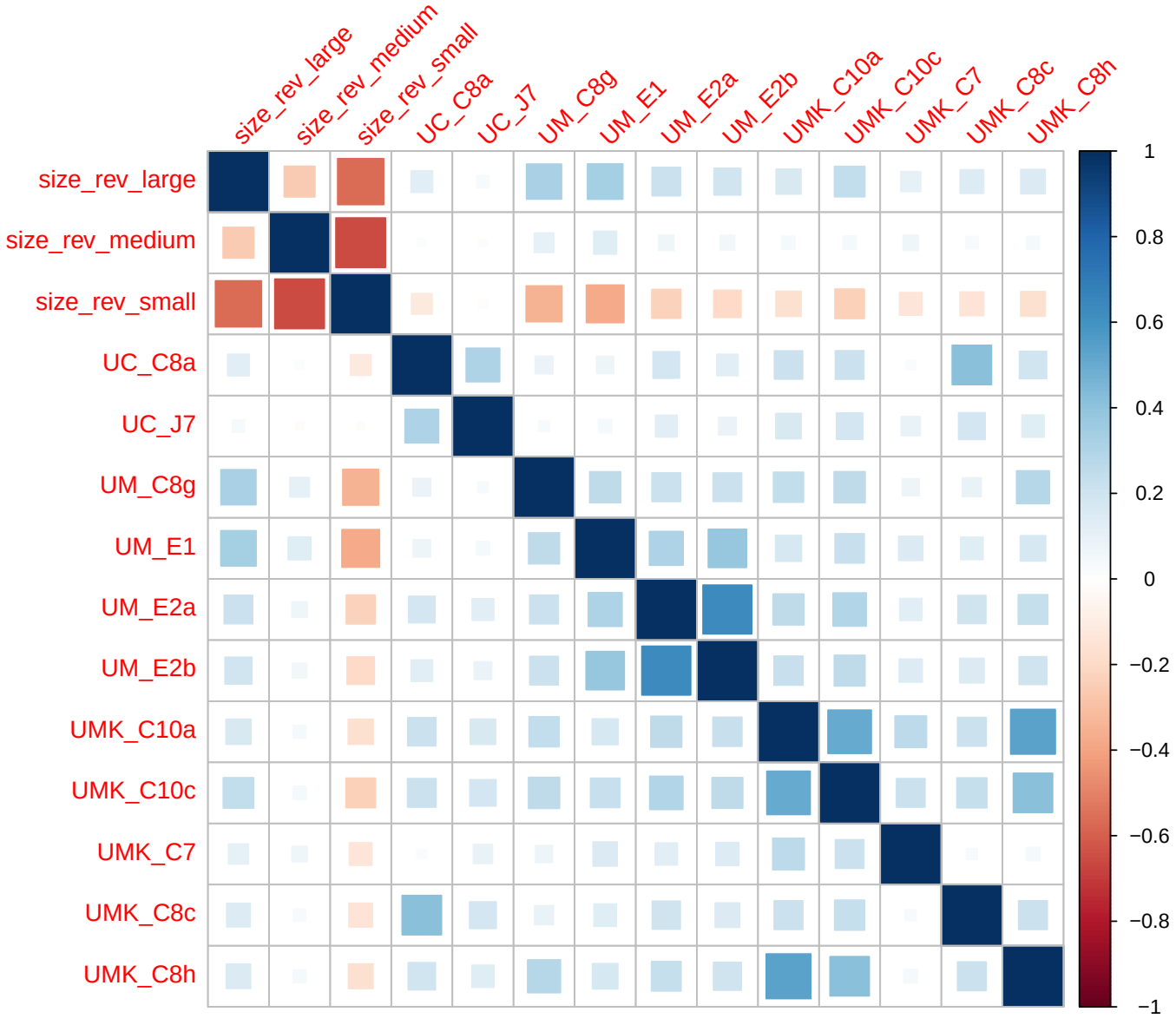
raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7295966	0.5095596	0.5360509	0.0796829	1.038984

2.3. Bartlett's Test of Sphericity Skills 2019

Statistic	Degrees_of_Freedom	p_Value
542837.8	66	0

3. 2019 - Usage - Correlation, Cronbach, and Bartlett's Ttests

3.1. Correlations



3.2. Cronbach's Alpha Usage 2019

- For binary varibles

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.7738189	0.6481937	0.6926948	0.1162996	1.842474

3.3. Bartlett's Test of Sphericity Usage 2019

Statistic	Degrees_of_Freedom	p_Value
564007.8	91	0