Evaluation Methods and Statistics revision#2

Chris Baber

Aims of this Module

- This module provides an introduction to the use of empirical, scientific methods, including experimental design and statistics.

- This module is targeted at computer scientists with an interest in:
  - Developing systems that support human activity (Human-Computer Interaction)
  - Building computational models of human behaviour
  - Understanding human behaviour as an inspiration for Robotics, Machine Learning and Artificial Intelligence
  - **Designing and analysing experiments to evaluate system performance**

Outcomes of this Module

- On successful completion of this module, you will be expected to be able to:
  - identify and apply research methodologies for investigating human behaviour;
  - recognise the appropriateness of statistical techniques in data analysis;
  - conduct and report statistical tests;
  - interpret and critique research findings that are supported by statistical tests;
  - demonstrate understanding of experimental design, including sampling, participant selection, task design and research ethics.

New Exam Requirements

- You will full details from the School.

- What do I expect…
  - Not all students will have access to R (or Statistics packages), so any calculations should be possible to perform by hand
  - If you are able to use Statistics packages for calculations, I expect enough information in your answer to demonstrate how this answer was produced (don’t just write a single number and expect full marks…)

Doing Experiments

- Ethics
  - What are the basic principles of the Declaration of Helsinki
  - How can you ensure that participants' identity is protected and that they will not suffer from participating in the experiment?
  - What are the basic principles of the (UK) Data Protection Act?
- Ecological Validity
  - How 'true to life' is the activity that you are asking participants to do?
  - How 'true to life' is the environment in which these tasks are performed?
- Experimental Design
  - What is the Hypothesis to be tested?
  - What are the Dependent and Independent variables?
  - For the Independent variables, which is the 'control' and the 'experimental' condition?
  - How can you manage confounding variables in the experiment?

Experimental Design template

Hypothesis: Reaction time to congruent words will be faster than reaction time to incongruent words	
Independent Variable: Congruent Words (colour of ink = name of word), Incongruent Words (colour of ink ≠ name of word)	
Control Condition: Congruent Words	**Experimental Condition:** Incongruent Words
Dependent Variable(s): Reaction Time	
Task: participants will be asked to read, as quickly as possible, single words on a display. The words will be the names of colours and will be presented either in the same colour as the word's name or in a different colour	
Confounding Variables: performance could be affected by ability to perceive colour ('colour-blindedness') and knowledge of the names of colour ('language skills')	

Hypothesis Testing

- Type I error
  - We could accept the Alternative hypothesis when it is false (false positive).
  - Many statistics tests are designed to minimise this error.
  - Type I errors define the significance level (α) that the experimenter will accept (conventionally 5%)

- Type II error
  - We could accept the Null hypothesis (fail to reject it) when it is false (false negative).
  - The probability of a Type II error is defined as β
  - The probability of correctly rejecting a false null hypothesis is defined as 1-β, which called Power.

Statistics

- To apply a Parametric statistical test, we need to show that the data follow a Normal distribution
  - Applying shapiro-wilk tests (next slides)
- If the data are measured on at least an interval scale and are normally distributed, then you can use the t-statistic to compare means between two groups (for more groups, you need ANOVA; if the data are not normally distributed then you apply non-parametric tests)

Shapiro-Wilk

- This is to test whether the set of data in an experiment are drawn from a normal distribution.
- The formula is:

$$\boxed{W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

  - You will recognise the denominator as the formula for the Sum of Squares of the data; the numerator defines the expected mean, variance, and covariance of a sample size (n)from a normal distributed sample.
  - While the sum of squares is easy to calculate, we use look-up tables for the expected values.
  - I will provide a step-by-step tutorial on how to do this…

Shapiro-Wilk tutorial

- Assume you collected these data:

Condition A	Condition B
65	74
61	35
63	72
86	68
70	45
55	58
M = 66.7	M = 59
sd = 10.7	sd = 15.8

Shapiro-Wilk tutorial

- Step 1:
  - Combine ALL of the data for the experiment into a single table
  - Order the data in terms of size (smallest to largest)

All Experiment Data
35
45
55
58
61
63
65
68
70
72
74
86

Shapiro-Wilk tutorial

- Step 2:
  - Calculate sum of squares
    - For each value, calculate difference between that value and the mean of the sample
    - Square this difference
    - Sum the squares

Data	X-M	(x-m)2
35	-27.67	765.4444
45	-17.67	312.1111
55	-7.67	58.77778
58	-4.67	21.77778
61	-1.67	2.777778
63	0.33	0.111111
65	2.33	5.444444
68	5.33	28.44444
70	7.33	53.77778
72	9.33	87.11111
74	11.33	128.4444
86	23.33	544.4444
M= 62.67		SS= 2008.667

Shapiro-Wilk tutorial

- Step 3:
  - Estimate the expected values for the data *if* they were drawn from a normal distribution
  - Define sample size, N. In this case, N = 12.
  - Use look-up table to define coefficients for a when N = 12.

n =	2	3	4	5	6	7	8	9	10	11	12	13	14
a1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251
a2			0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318
a3					0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460
a4							0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802
a5									0.0399	0.0695	0.0922	0.1099	0.1240
a6											0.0303	0.0539	0.0727
a7													0.0240

Shapiro-Wilk tutorial

- Step 4:
  - Apply these coefficients to your data.
    - Divide the data into pairs to correspond with the number of coefficients.
    - In this case, there are 6 values for a (from the table).
    - We divide our data into 6 pairs, subtracting the largest from the smallest values.

	xl- xs	a	a*(xl-xs)
x12-x1	51	0.5475	27.9225
x11-x2	29	0.3325	9.6425
x10-x3	17	0.2347	3.9899
x9-x4	12	0.1586	1.9032
x8-x5	7	0.0922	0.6454
x7-x6	2	0.0303	0.0606
			b = 44.1641
			b^2 = 1950.468

Shapiro-Wilk tutorial

- Step 5:
  - Calculate W

  W = 1950.468 / 2008.667

  = 0.971026

$$\boxed{W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

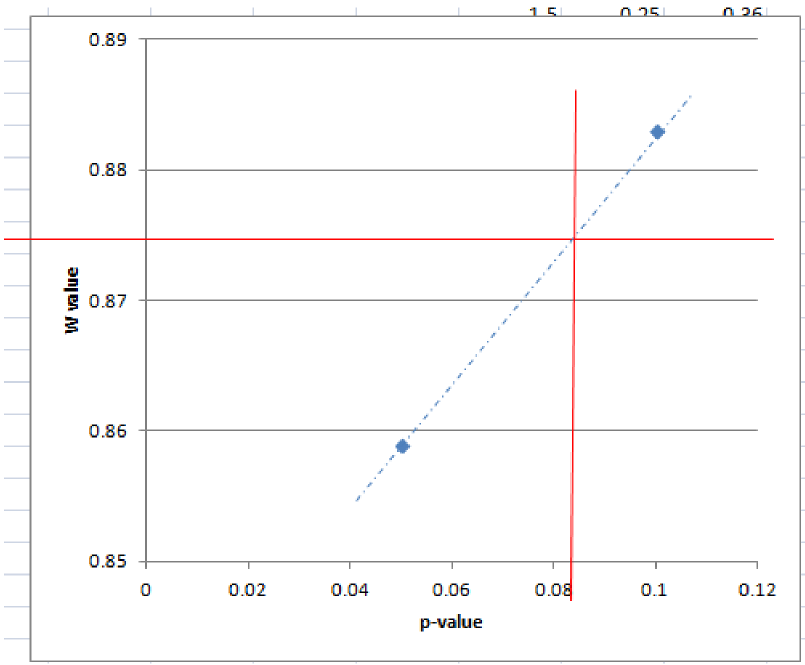
Shapiro-Wilk tutorial

- Step 7:
  - Determine significance of this result by using look-up tables
  - Our calculated value (0.971026) lies between 0.974 and 0.943, i.e., between p=0.9 and p=0.5.
  - As the smallest p-value is > 0.05,we accept the null hypothesis and conclude that the data <u>are</u> normally distributed

n \ P	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987

Shapiro-Wilk tutorial

- What if the value of W was 0.8752?
  - The corresponding p-values (for N = 12) are 0.05 and 0.1.
  - The table values are 0.859 and 0.883
  - By linear interpolation, we can estimate a p-value that is around 0.08. As this is >0.05, the data are normally distributed



T-statistic

- Should you use Independent or Repeated measures test?
- How do you define the significance level of the result?
- How do you use Cohen’s d to calculate effect size?

Tables for a (Shapiro-Wilk)

n =	2	3	4	5	6	7	8	9	10	11	12	13	14
a1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251
a2			0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318
a3					0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460
a4							0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802
a5									0.0399	0.0695	0.0922	0.1099	0.1240
a6											0.0303	0.0539	0.0727
a7													0.0240

n =	15	16	17	18	19	20	21	22	23	24	25	26
a1	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407
a2	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043
a3	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533
a4	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151
a5	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836
a6	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563
a7	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316
a8		0.0196	0.0359	0.0496	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089
a9				0.0163	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876
a10						0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672
a11								0.0122	0.0228	0.0321	0.0403	0.0476
a12									0.0000	0.0107	0.0200	0.0284
a13											0.0000	0.0094

More Tables for a (Shapiro-Wilk)

n =	27	28	29	30	31	32	33	34	35	36	37	38
a 1	0.4366	0.4328	0.4291	0.4254	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015
a 2	0.3018	0.2992	0.2968	0.2944	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774
a 3	0.2522	0.2510	0.2499	0.2487	0.2475	0.2463	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391
a 4	0.2152	0.2151	0.2150	0.2148	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110
a 5	0.1848	0.1857	0.1864	0.1870	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881
a 6	0.1584	0.1601	0.1616	0.1630	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686
a 7	0.1346	0.1372	0.1395	0.1415	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513
a 8	0.1128	0.1162	0.1192	0.1219	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356
a 9	0.0923	0.0965	0.1002	0.1036	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211
a 10	0.0728	0.0778	0.0822	0.0862	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075
a 11	0.0540	0.0598	0.0650	0.0697	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947
a 12	0.0358	0.0424	0.0483	0.0537	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824
a 13	0.0178	0.0253	0.0320	0.0381	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706
a 14	0.0000	0.0084	0.0159	0.0227	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592
a 15			0.0000	0.0076	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481
a 16					0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372
a 17							0.0000	0.0062	0.0119	0.0172	0.0220	0.0264
a 18									0.0000	0.0057	0.0110	0.0158
a 19											0.0000	0.0053

More Tables for a (Shapiro-Wilk)

n =	39	40	41	42	43	44	45	46	47	48	49	50
a1	0.3989	0.3964	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
a2	0.2755	0.2737	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
a3	0.2380	0.2368	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
a4	0.2104	0.2098	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
a5	0.1880	0.1878	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
a6	0.1689	0.1691	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
a7	0.1520	0.1526	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
a8	0.1366	0.1376	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
a9	0.1225	0.1237	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
a10	0.1092	0.1108	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
a11	0.0967	0.0986	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
a12	0.0848	0.0870	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
a13	0.0733	0.0759	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.9190	0.0932
a14	0.0622	0.0651	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
a15	0.0515	0.0546	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
a16	0.0409	0.0444	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
a17	0.0305	0.0343	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
a18	0.0203	0.0244	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
a19	0.0101	0.0146	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
a20	0.0000	0.0049	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
a21			0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
a22					0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
a23							0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
a24									0.0000	0.0037	0.0071	0.0104
a25											0.0000	0.0035

P-values (Shapiro-Wilk)

n \ p	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989

	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.990
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.939	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991