RLADIES : HYPOTHESIS TESTING

JUNE 2018

WHAT IS A HYPOTHESIS?

A **hypothesis** is a proposed explanation for a phenomenon. For a hypothesis to be a scientific hypothesis, the scientific method requires that one can test it.*

** Wikipediac*

A NULL HYPOTHESIS

Null hypothesis is a general statement or default position that there is no relationship between two measured phenomena, or no association among groups.*

* *Wikipediac*

WHAT WE CAN WE DO WITH A NULL HYPOTHESIS

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct

ERROR TYPE I REDUCTION AS THE GOAL

Ideally, we should decrease both types of errors but since the Error Type I is more serious, we try to reduce it by choosing a significance level (Alpha) and reduce Error Type II by increasing our sample size.

WHAT IS A P-VALUE ?

p-value Is the probability for a given statistical model that, when the null hypothesis is true, the statistical summary would be the same as or of greater magnitude than the actual observed results.*

** Wikipediac*

SAMPLING DISTRIBUTION AND CLT

Go to R !

WHAT IS STATISTICAL SIGNIFICANCE?

Statistical significance refers to the likelihood or probability that a statistic derived from a sample represent some genuine phenomenon in the population from which the sample was selected.*

** Statistics in Plain English*

STATISTICAL SIGNIFICANCE : A DEEPER LOOK

- Statistical significance Vs. Practical Significance
- Effect Size
- Confidence Interval

A GENERAL METHOD TO CALCULATE P-VALUE

P(Getting something by chance and not any specific reason)

SOME OF THE SCENARIOS WE FACE

- Is a sample coming from a specific population?
- Are two seemingly different samples belong to the same population?
- Does the same sample (or two samples related together) in different situations have different statistics?

Z-SCORE AND Z-TEST

$$z = \frac{X - \bar{X}}{s}$$

Standard Normal Probabilities

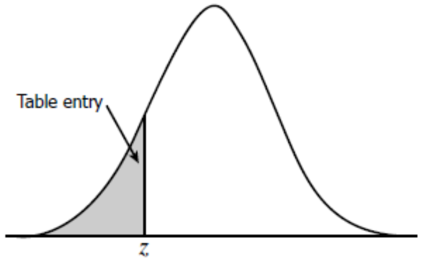


Table entry for z is the area under the standard normal curve to the left of z.

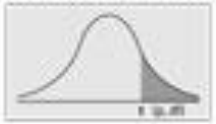
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143

T-STATISTICS AND T-TEST

$$t = \frac{|\bar{x} - \mu|}{s / \sqrt{n}}$$

The *t*-Table

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	—	—	80%	90%	95%	98%	99%	99.9%

INDEPENDENT T-TEST

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

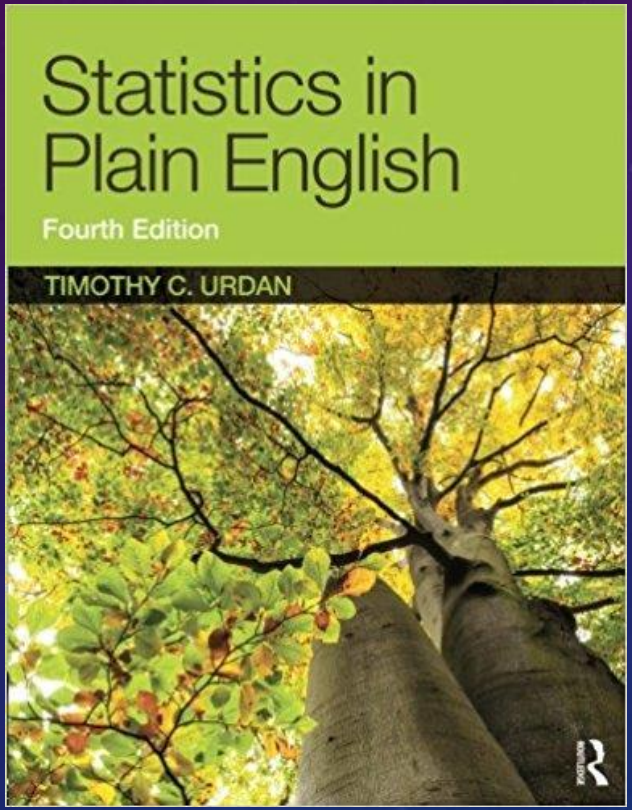
PAIRED T-TEST

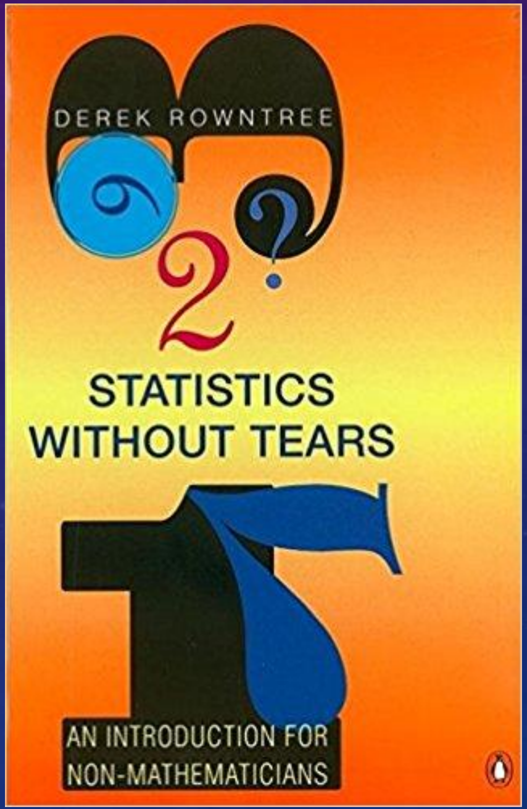
$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\left[(n_1-1)s_1^2+(n_2-1)s_2^2\right] \ (n_1+n_2)}{(n_1+n_2-2)(n_1n_2)}}}$$

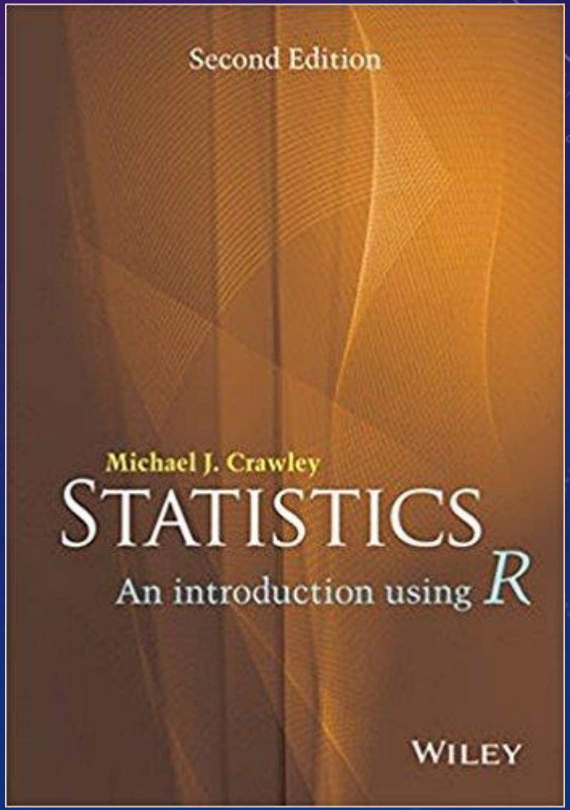
ANOVA

$$Vr = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n} x^2 - \frac{\sum_{i=1}^{K}\left(\sum_{j=1}^{n} x\right)^2}{n} - \frac{Sg}{k} + \frac{\left(\sum_{i=1}^{K}\sum_{j=1}^{n} x\right)^2}{N}}{(N-1)-(k-1)-(n-1)}$$

BOOKS TO READ







navid.nobani@gmail.com
www.linkedin.com/in/navidnobani/