HYPOTHESIS TESTING

Statistical Inference

Type of errors

An error in the analysis!

More work is needed.

	H_0 is True	H_0 is False
Reject H_0	Type I error (α) (False Positive)	Correct inference (True Positive)
Fail to reject H_0	Correct inference (True Negative)	Type II error (β) (False Negative)

$$P(H_1|H_0) = \propto$$
$$P(H_o|H_1) = \beta$$

$$P(H_1|H_1) = 1 - \beta$$

Type of errors

- Type I error: "rejecting the null hypothesis when it is true".
- Type II error: "accepting the null hypothesis when it is false".
- Type III error: “solving the wrong problem [representation]“.
- Type IV error: "the incorrect interpretation of a correctly rejected hypothesis”.

Joking?

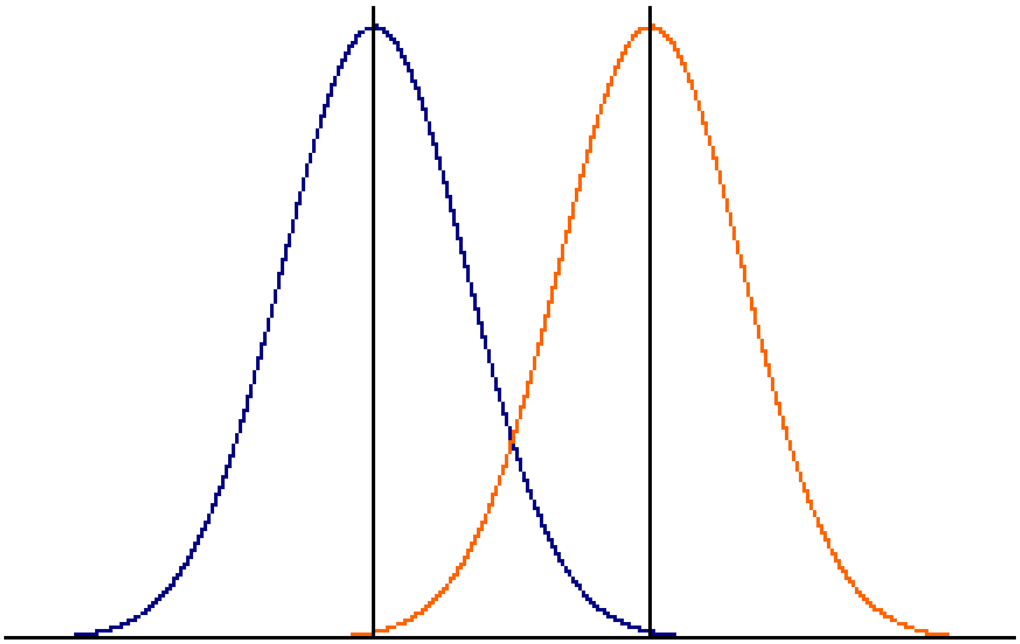
Dirty Rotten Strategies: How We Trick Ourselves and Others into Solving the Wrong Problems Precisely (High Reliability and Crisis Management), Ian I. Mitroff, Abraham Silvers. ISBN-13: 978-0804759960

Type of errors

- Type 1 errors are more important than Type 2 errors → evidence.
  - *p*-values are only correlated with evidence.
  - Evidence in science is necessarily *relative*. When data is more likely assuming one model is true (e.g., a null model) compared to another model (e.g., the alternative model), we can say the model provides **evidence** for the null compared to the alternative hypothesis. *P*-values only give you the probability of the data under one model – what you need for evidence is the relative likelihood of two models.

Equal variances test

Comparison of two configurations with equal variances.



Central limit theorem

- In probability theory, the central limit theorem (CLT) states that, given certain conditions, the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed.

Comparison of two configurations with equal variances.

- We define the hypothesis test:
  - H0: μA = μB
  - H1: μA > μB
- Thanks the central limit theorem we obtain that:

$$\overline{y}_A \approx N(\mu_A, \frac{\sigma_A}{\sqrt{n_A}})$$

$$\overline{y}_B \approx N(\mu_B, \frac{\sigma_B}{\sqrt{n_B}})$$

Comparison of two configurations with equal variances.

- We can deduce that:

$$\bar{y}_A - \bar{y}_B \approx N(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}})$$

$$\frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \approx N(0,1)$$

Comparison of two configurations with equal variances.

- We define the test, and calculate s, the common sample variance:

$$\frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_A)}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \approx t_n$$

- Where $n=n_A+n_B-2$

Comparison of two configurations with equal variances.

- The test is defined as is shown:

$$\frac{\bar{y}_A - \bar{y}_B}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{1-\alpha,n}$$

- We reject H_0 if this is true.

Example

Equal variances

Example

Observation	Values for pop A	Values for pop B
1	24.3	24.4
2	25.6	21.5
3	26.7	25.1
4	22.7	22.8
5	24.8	25.2
6	23.8	23.5
7	25.9	22.2
8	26.4	23.5
9	25.8	23.3
10	25.4	24.7

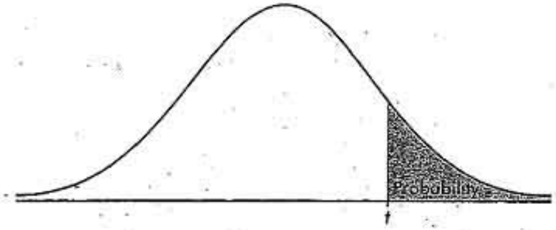


TABLE B: *t*-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Example

- Mean of the sample.
  - A=25.14; B=23.62

  - H0: $\mu A = \mu B$
  - H1: $\mu A > \mu B$

Example

- Equal variances:

$$\frac{\bar{y}_A - \bar{y}_B}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{1-\alpha,n}$$

Example

- The standard deviation is:
  - A=1.242; B=1.237

$$\frac{25.14-23.62}{1.24\sqrt{\frac{1}{10}+\frac{1}{10}}}=2.74>t_{0.05,18}=1.734$$

  - Reject H_0

No equal variances

Two configurations comparison

- If we cannot assume equal variances.

$$t' = \frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Two configurations comparison.

- If $nA = nB = n$, the signification level is determined using as a reference distribution a **t of Student** with n-1 degrees of freedom.
- If $nA \neq nB$, with the value calculated of t' we can find different signification values pA and pB in the student distributions, with nA-1 and nB-1 degrees of freedom respectively.

Two configurations comparison

- The signification level of the test:

$$\alpha = \frac{\omega_A p_A + \omega_B p_B}{\omega_A + \omega_B}$$

- with:

$$\omega_A = \frac{S_A^2}{n_A} \qquad \omega_B = \frac{S_B^2}{n_B}$$

Equal variance test

- Hypothesis test:
  - H0: $\sigma_A^2 = \sigma_B^2$
  - H1: $\sigma_A^2 \neq \sigma_B^2$

$$F = \frac{variance\ between\ treatments}{variance\ within\ treatments}$$

$$\frac{S_A^2}{S_B^2} \approx F_{n,m}$$

- F of Snedecor
  - $n = n_A - 1$
  - $m = n_B - 1$.

Example

- $S_A{}^2 = 1.54$
- $S_B{}^2 = 2.18$

$$\frac{S_B^2}{S_A^2} = \frac{2.18}{1.54} = 1.42 < F_{0.05,9,9} = 3.18$$

- Accept H_0

Example

- $S_A{}^2 = 1.54$
- $S_B{}^2 = 16.3$

$$\frac{S_B^2}{S_A^2} = \frac{16.3}{1.54} = 10.58 > F_{0.05,9,9} = 3.18$$

- Discard H_0

To know more

- Part III: Statistics and Chapter 9.4: Statistical Inference I of **Probability and Statistics for Computer Scientists** (2014 Ed.)