Effect Sizes and P-Value

Qu Jin

October 9, 2013

Outline

- Review of Some Statistics Concepts
  - Hypothesis Testing
  - P-Value
- Effect Size
- Example Results from YES data

Review of Statistics Concepts

Hypothesis Testing

Hypothesis testing is a method to judge **hypotheses** using **test statistics** and their **distribution**

- H_0: null hypothesis vs. H_1: alternative hypothesis
  - Example 1: H_0: $\mu = 0$ vs. H_1: $\mu \neq 0$
  - Example 2: H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$
- Test statistics

  Measures the amount of deviation of estimates from H_0

  Example 1:

$$T = \frac{\bar{Y} - 0}{S/\sqrt{n}} \quad \sim^{H_0} \quad t(n-1)$$

$\bar{Y}$ is the sample mean

S is the sample standard deviation

n is the sample size

df = n - 1

Review of Statistics Concepts

Hypothesis Testing

Example 2:

$$T = \frac{(\bar{Y}_2 - \bar{Y}_1) - 0}{S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \sim^{H_0} \quad t(n_1 + n_2 - 2) \qquad S_{pool}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$\bar{y}_1$ and $\bar{y}_2$ are the sample means

S_1 and S_2 are the sample standard deviations

n_1 and n_2 are the sample sizes

$df = n_1 + n_2 - 2$

- Decision Rules

  Given significance level α (usually 0.05), there are two approaches

  - Compare the observed test statistic with critical value
  - Compare the p-value of observed test statistic with α

    Reject H_0 if p-value ≤ α

Review of Statistics Concepts

Hypothesis Testing

- In Example 1

If we collected 10 samples with mean of 1.01 and SD^2 of 3.49

y =1.01, S^2 =3.49, n=10

$$T1 = \frac{y - 0}{S/\sqrt{n}} = \frac{1.01 - 0}{\sqrt{3.49/10}} = 1.71$$

- In Example 2

If we collected 10 samples from each of the two groups

$y1$ =1.01, $S1^2$ =3.49, n_1=10; $y2$ =1.01, $S2^2$ =3.49, n_2=10

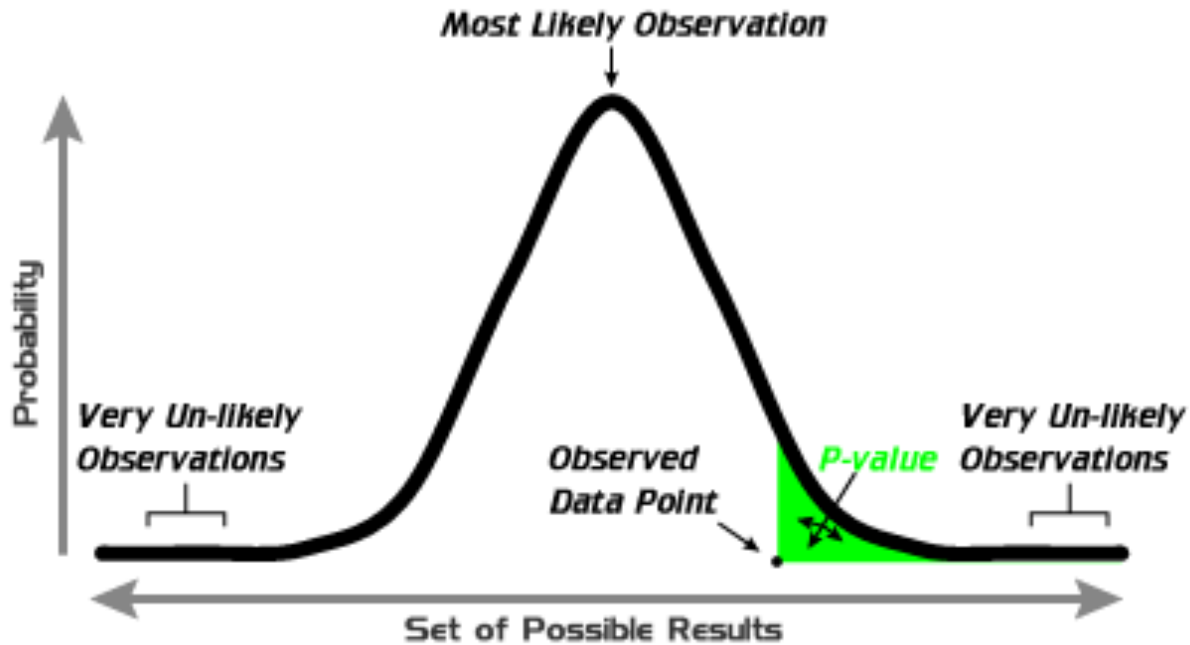
$$S_{pool}^2 = \frac{(n1-1)S1^2 + (n2-1)S2^2}{n1+n2-2} = 1.4$$

$$T2 = \frac{(y2 - y1) - 0}{Spool\sqrt{1/n1 + 1/n2}} = 2.22$$

Review of Statistics Concepts
P-Value

- Definition "the tale"

  P-value is the probability that test statistics takes on a value that is at least as extreme as the observed value of the statistic when H_0 is true



A p-value (shaded green area) is the probability of an observed (or more extreme) result arising by chance

Review of Statistics Concepts

P-Value

For example, $T_1 = 1.71$ df = 9
From the table, we can get
that the $p = 0.12 > 0.05$
Fail to reject H_0

For example, $T_2 = 2.22$ df = 18
From the table, we can get
that the $p = 0.04 < 0.05$
Reject H_0

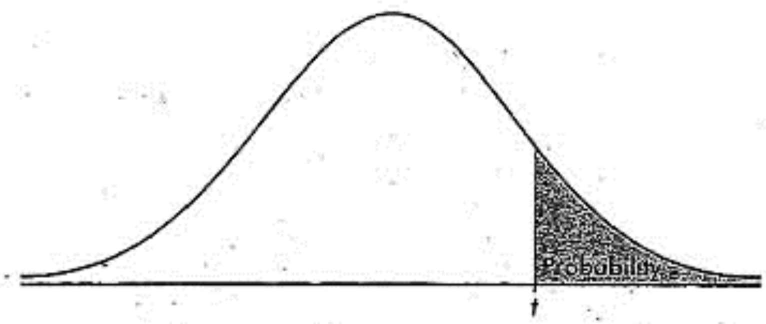


TABLE B: t-DISTRIBUTION CRITICAL VALUES

	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

Review of Statistics Concepts

P-Value

$$T = \frac{\bar{Y} - 0}{S/\sqrt{n}} \overset{H_0}{\sim} t(n-1)$$

n↑ T↑ p↓

TABLE B: *t*-DISTRIBUTION CRITICAL VALUES

df	Tail probability *p*											
	.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*											

Review of Statistics Concepts

P-Value

- P-value is dependent on sample size
- When sample size increases, p-value gets smaller and smaller
  - In Example 2, if $n_1 = n_2 = 100$, $p = .000$
  - When sample size is really large (> 1000), very small difference in sample means can result very small p-value
  - P-value can tell if the difference is **significant** or not
  - However, p-value cannot tell how **large** the difference is

Effect Sizes

- A measure of the strength of a relationship or effect
- Is not dependent on sample size
- Commonly used effect sizes
  - Cohen's d: standardized difference in sample means
    Can represent pair-wise difference

$$d = \frac{\bar{Y}_2 - \bar{Y}_1}{S_{pool}} \qquad S^2_{pool} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

  - Eta-square: measures the degree of association between the independent variable and dependent variable
    Cannot represent pair-wise difference

$$\eta^2 = \frac{SS(\text{Between groups})}{SS(\text{Total})}$$

SS(Between groups) is the sum of squares for the independent variable

SS(Total) is the total sum of squares

Example Results from YES Data

Scale	η^2	Engineering Undergrad		Business Undergrad		STM Undergrad		Other Undergrad	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Entrepreneurial Intent	**.075**	2.93	1.06	3.54	1.01	2.54	1.09	2.87	1.14
Sample Size		518		471		668		1230	

$$d = \frac{\bar{Y}_2 - \bar{Y}_1}{S_{pool}}$$

$$S^2_{pool} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$\eta^2 = \frac{SS(\text{Between groups})}{SS(\text{Total})}$$

Cohen's d			P-Value		
Engr-Bus	Engr-STM	Engr-Other	Engr-Bus	Engr-STM	Engr-Other
.58	**.36**	.06	< .0001	< .0001	.305

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Entrepreneurial Intent	Between Groups	276.456	3	92.152	77.204	.000
	Within Groups	3431.635	2875	1.194		
	Total	3708.091	2878			

Cohen's d Guidelines: (Cohen, 1988)

- **small d = 0.2 - 0.5**
- **medium d = 0.5 - 0.8**
- **large d ≥ 0.8**

Eta-squared (η^2) Guidelines: (Cohen, 1988)

- **small 0.01 - 0.06**
- **medium 0.06 - 0.14**
- **large > 0.14**

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciencies*: Routledge.

Backup

Review of Statistics Concepts

Type I and Type II Errors

- Type I error: when H_0 is true, reject H_0

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

- Type II error: when H_0 is false, not reject H_0

$$\beta = P(\text{type II error}) = P(\text{not reject } H_0 \mid H_0 \text{ is false})$$

Decision \ **Actual**	Null Hypothesis True	Null Hypothesis False
Reject Null Hypothesis	Type I Error	Correct
Fail to Reject Null Hypothesis	Correct	Type II Error

Review of Statistics Concepts
Power

The probability of correctly rejecting a false H_0

$$\text{Power} = 1 - \beta = P(\text{reject } H_0 \mid H_0 \text{ is false})$$

$$d = \frac{|\mu_2 - \mu_1|}{2\sigma}$$

Power is a function of the d and the sample size n

- Power increases if n increases
- Power increases if d increases
- Power calculator for two-sample t-test: http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

- Desirable power is 0.8; typical power in social science is 0.6