Probability and Statistics

Dr. Faisal Bukhari

Punjab University College of Information Technology (PUCIT)

References

Readings for these lecture notes:

- Probability & Statistics for Engineers & Scientists, Ninth edition, Ronald E. Walpole, Raymond H. Myer

- www.unm.edu/~marley/statppt/fall06/002/day12.ppt

These notes contain material from the above resources.

Independent and Dependent Samples.

- Two samples are **independent** if the sample values selected from **one population** are **not related to or somehow paired or matched** with the sample values selected from the other population.

- Two samples are **dependent** (or consist of **matched pairs**) if the members of one sample can be used to determine the members of the other sample. [Samples consisting of **matched pairs** (such as husband wife data) are **dependent**.

- In addition to **matched pairs of sample data**, **dependence** could also occur with samples related **through associations** such as **family members**.]

Confidence Interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observations

If $\bar{d}$ and s_d are the **mean** and **standard deviation**, respectively, of the normally distributed differences of n **random pairs of measurements**, a 100(1 − α)% confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{(\alpha/2, n-1)} \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{(\alpha/2,\, n-1)} \frac{s_d}{\sqrt{n}}$$

Where,

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n-1}} \text{ OR } s_d = \sqrt{\frac{1}{n(n-1)}\left\{n\sum_{i=1}^{n} d^2{}_i - \left(\sum_{i=1}^{n} d^2{}_i\right)^2\right\}}$$

$$s_d^2 = \frac{\sum(d - \bar{d})^2}{n-1} \text{ OR } s_d^2 = \frac{1}{n(n-1)}\left\{n\sum_{i=1}^{n} d^2{}_i - \left(\sum_{i=1}^{n} d^2{}_i\right)^2\right\}$$

$$d_i = x_{1i} - x_{2j} \text{ OR } d_i = x_{2i} - x_{1j},\ \bar{d} = \frac{\sum_{i=1}^{n} di}{n}$$

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

Testing Hypothesis about Paired Observation

a) H_o: $\mu_d = 0$

H_1:$\mu_d < 0$ (One tailed test)

b) H_o: $\mu_d = 0$

H_1:$\mu_d > 0$ (One tailed test)

c) H_o: $\mu_d = 0$

H_1:$\mu_d \neq 0$ (Two tailed test)

Test statistic:

$$t_{cal} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}},$$

Where $d_i = x_{1i} - x_{2j}$ OR $d_i = x_{2i} - x_{1j}$

$$\bar{d} = \frac{\sum_{i=1}^{n} di}{n}$$

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n-1}} \text{ OR}$$

$$s_d = \sqrt{\frac{1}{n(n-1)}\{n\sum_{i=1}^{n} d^2{}_i - (\sum_{i=1}^{n} d_i)^2\}}$$

The *t* Test for Dependent Samples: An Example

Eight individuals indicated their attitudes toward socialized medicine before and after listening to a pro–socialized medicine lecture. Attitudes were assessed on a scale from 1 to 7, with higher scores indicating more positive attitudes. The attitudes before and after listening to the lecture were as indicated in the second and third columns of the table. Test for a relationship between the time of assessment and attitudes toward socialized medicine using a correlated groups *t* test.

Individual	Before speech	After speech
1	3	6
2	4	6
3	3	3
4	5	7
5	2	4
6	5	6
7	3	7
8	4	6

Solution

$\mu_D = 0$ (Population mean)

$n = 8$ (Sample size)

$\alpha = 0.05$ (Level of significance)

$\bar{d} = ?$

$s_d = ?$

1. We state our hypothesis as:

H_o: $\mu_d = 0$

H_1: $\mu_d \neq 0$ (Two tailed test)

2. The level of significance is set $\alpha = 0.05$

3. Test statistic to be used is

$$t_{cal} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

4. Calculations:

Before speech	After speech	$d_i = x_{1i} - x_{2j}$	$d^2{}_i$
3	6	-3	9
4	6	-2	4
3	3	0	0
5	7	-2	4
2	4	-2	4
5	6	-1	1
3	7	-4	16
4	6	-2	4
Sum		$\sum_{i=1}^{n} d_i = -16$	$\sum_{i=1}^{n} d^2{}_i = 42$

$$\bar{d} = \frac{\sum_{i=1}^{n} \mathrm{di}}{n} = -16/8 = \mathbf{-2}$$

$$s_d = \sqrt{\frac{1}{n(n-1)}\{n\sum_{i=1}^{n} d^2{}_i - (\sum_{i=1}^{n} d_i)^2\}}$$

$$s_d = \sqrt{\frac{1}{8(8-1)}\{8(42) - (-16)^2\}} = \sqrt{\frac{80}{8(8-1)}} = 1.1952$$

$$\mathbf{t_{cal}} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = t_{cal} = \frac{-2 - 0}{\frac{1.1952}{\sqrt{8}}} = \frac{-2}{0.4226}$$

$t_{cal} = -\ 4.7326$

$|\mathbf{t_{cal}}| = 4.7326$

5. **Critical region:**

$|\mathbf{t_{cal}}| > \mathbf{t_{tab}}$**, where** $\mathbf{t_{tab} = t_{(\alpha/2,\ n-1)}}$

Where $t_{tab} = t_{(\alpha/2,\ n-1)} = t_{(0.0250,\ 7)} = 2.365$

6. **Conclusion:** Since calculated value of t_{cal} is greater than t_{tab}, so we reject H_O

Interpret your results.

After the **pro-socialized medicine lecture**, individuals' attitudes toward **socialized medicine** were significantly more positive than before the lecture.

Table A.4 Critical Values of the t-Distribution

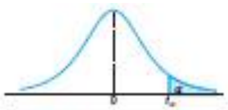

Table A.4 Critical Values of the t-Distribution

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Table A.4 (continued) Critical Values of the t-Distribution

	α						
v	0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290