Factors affect p value

Presenter: Bui Minh Tri

Objective + Outline

- Objective

Factors affect p value

- Outline

1. **Overview**
2. **p of Independent-means t-test**

Overview

p value

- Ho: $\mu_1 = \mu_2$
- H1: $\mu_1 \neq \mu_2$

✓ **p ≤ .05 ➢ Reject Null Hypothesis** ➢ We have **enough evidence** to conclude that the difference between groups is statistically significant.

✓ **p > .05 ➢ Failed to reject Null Hypothesis** ➢ We **don’t have enough evidence** to conclude that the difference between groups is statistically significant.

p of
Independent-means t-test

Features - Comparing 2 groups

$p \le 0.05$:

1. Larger difference between 2 groups ($|\bar{x}_1 - \bar{x}_2|$ ↑)
2. Lower variability in each group (**s_1 s_2** ↓)
3. Increase sample size (**n_1 n_2** ↑)

Standard Deviation of sample

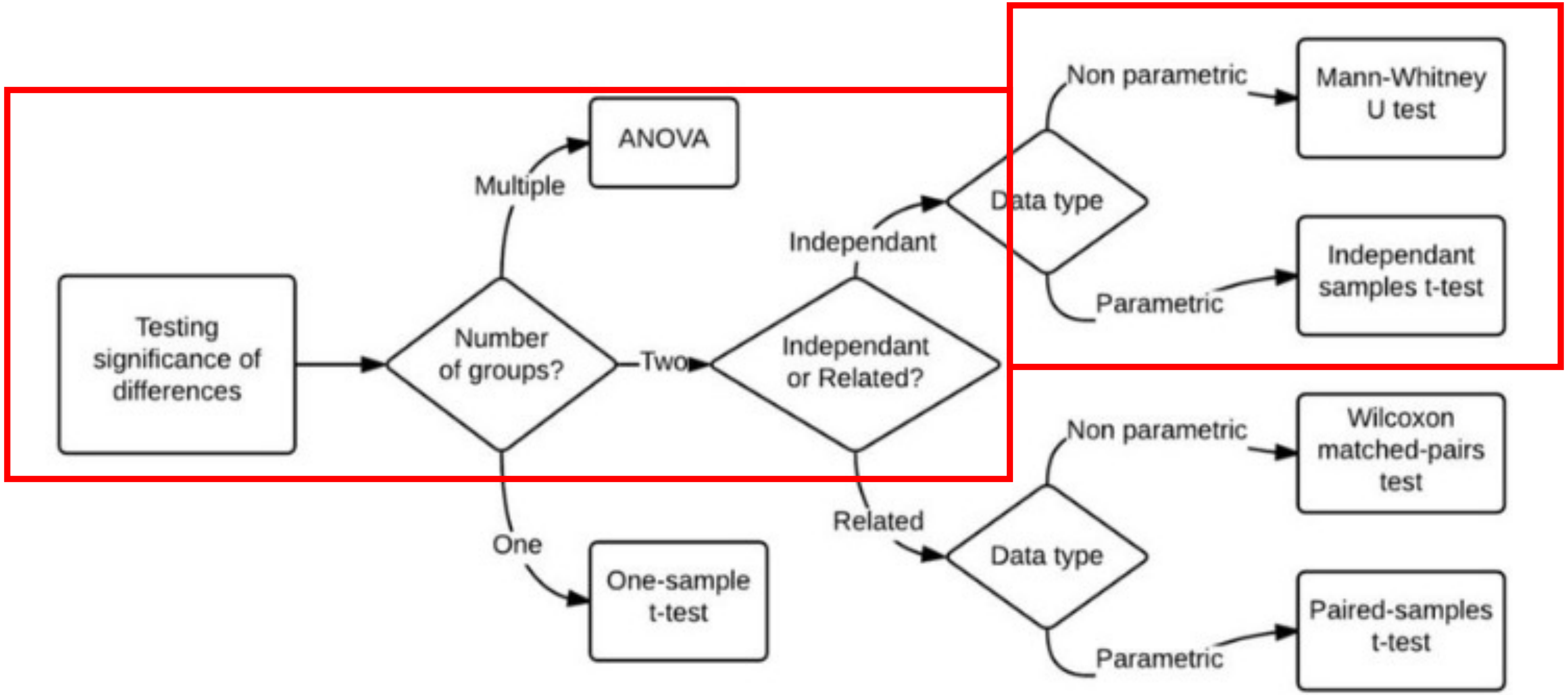
- Spread out of data on both side of mean:

  ✓ $$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}}$$

- Small s = more agreement in data.

Sample size = 10	9	5	0
	9	5	0
	9	5	0
	9	5	0
	9	5	0
	10	10	10
	10	10	10
	10	10	10
	10	10	10
	10	10	10
Mean	9.5	7.5	5
Standard Deviation	0.5	2.5	5

Decision tree for statistical analysis



Statistical analysis decision tree for testing significance of differences

Borghini YC. *An Assessment and Learning Analytics Engine for Games-based Learning* (Doctoral dissertation, University of the West of Scotland).

Assumptions of Independent-means t-test

a. **Independence:** Observations are independent of one another.

b. **Measurement scale:** numeric data **-** interval level or ratio level.

c. **Normality:** The variable is normally distributed in each population.

d. **Homogeneity of variance**: Equal variances between groups.

Calculate manually – t and df

	Equal Variance	Unequal Variance
t formula	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$ $\quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
df	$df = n_1 + n_2 - 2$	$df = \frac{(\frac{s_1^2}{n1} + \frac{s_2^2}{n2})^2}{\frac{(\frac{s_1^2}{n1})^2}{n1-1} + \frac{(\frac{s_2^2}{n2})^2}{n2-1}}$

$\bar{x}_1$; $\bar{x}_2$: Mean

s_1 ; s_2 : Standard deviation

n_1 ; n_2 : Sample size

Calculate manually – Critical value of t for two-tailed test

Example:

- t = 3.4
- df = 154
  - ➢ $p < 0.001$

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

```
```{r}
t.test(tuoilucchandoanlandau ~ Group_Change, data = DB_BB, var.equal = T, paired = F )
```

        Two Sample t-test

data: tuoilucchandoanlandau by Group_Change
t = 3.4077, df = 154, p-value = 0.0008361
alternative hypothesis: true difference in means between group BB and group DB is not equal to 0
95 percent confidence interval:
 2.392257 8.991695
sample estimates:
mean in group BB mean in group DB
        68.05263   62.36066
```

Calculate manually – Critical value of t for two-tailed test

Center for Molecular Biomedicine

- Similar df: t↑ ⇨ p↓
- df↑ and t↑ ⇨ p↓

- **Compare 2 groups**
  - **df↑ and t↑ ⇨p↓**

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	df	p
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑			↓
s_1 ; s_2 ↓			↓
n_1 ; n_2 ↑			↓

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	**df**	**p**
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑			
s_1 ; s_2			
n_1 ; n_2			

	p						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	df	p
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑		
s_1 ; s_2			
n_1 ; n_2			

	p						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	**df**	**p**
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑		↓
s_1 ; s_2			
n_1 ; n_2			

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965

Case 1: $|\bar{x}_1 - \bar{x}_2| \uparrow \Rightarrow p \downarrow$

Group1: 60 value, mean1 = 55, s1 = 10.48, [44,94]
Group2: 60 value, mean2 = 62, s2 = 9.68 , [42,83]

↓

Run t-test for group1, group2

↓

Group1: mean1 = mean1+1

↺ Target mean = 80

Case 1: $|\bar{x}_1 - \bar{x}_2|\uparrow \Rightarrow p\downarrow$

	Equal variance	
	t	p
$|\bar{x}_1 - \bar{x}_2|$ ↑	↑	↓
s_1 ; s_2 ↓		↓
n_1 ; n_2 ↑		↓

n1	n2	s1	s2	mean1	mean2	mean_diff	t	p
<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
60	60	8.77	8.49	56.15	61.7	-5.55000000	-3.522	0.0006098592841
60	60	8.94	8.49	57.03	61.7	-4.66666667	-2.932	0.0040434343426
60	60	9.07	8.49	57.95	61.7	-3.75000000	-2.339	0.0210325249112
60	60	9.15	8.49	58.90	61.7	-2.80000000	-1.738	0.0848199539226
60	60	9.23	8.49	59.85	61.7	-1.85000000	-1.143	0.2555185179111
60	60	9.32	8.49	60.80	61.7	-0.90000000	-0.553	0.5813449645668
60	60	9.39	8.49	61.77	61.7	0.06666667	0.041	0.9675171476247
60	60	9.45	8.49	62.73	61.7	1.03333333	0.630	0.5298412793136
60	60	9.49	8.49	63.72	61.7	2.01666667	1.227	0.2221699786141
60	60	9.52	8.49	64.70	61.7	3.00000000	1.822	0.0710151861847
60	60	9.56	8.49	65.68	61.7	3.98333333	2.414	0.0173313008506
60	60	9.56	8.49	66.68	61.7	4.98333333	3.020	0.0031030258097
60	60	9.56	8.49	67.68	61.7	5.98333333	3.626	0.0004269765376
60	60	9.56	8.49	68.68	61.7	6.98333333	4.231	0.0000461364306
60	60	9.56	8.49	69.68	61.7	7.98333333	4.837	0.0000040130585
60	60	9.56	8.49	70.68	61.7	8.98333333	5.443	0.0000002885123
60	60	9.56	8.49	71.68	61.7	9.98333333	6.049	0.0000000176081
60	60	9.52	8.49	72.67	61.7	10.96666667	6.660	0.0000000009135
60	60	9.48	8.49	73.65	61.7	11.95000000	7.273	0.0000000000418
60	60	9.45	8.49	74.63	61.7	12.93333333	7.888	0.0000000000017
60	60	9.42	8.49	75.62	61.7	13.91666667	8.505	0.0000000000001
60	60	9.35	8.49	76.58	61.7	14.88333333	9.130	0.0000000000000
60	60	9.26	8.49	77.53	61.7	15.83333333	9.766	0.0000000000000
60	60	9.17	8.49	78.48	61.7	16.78333333	10.405	0.0000000000000
60	60	9.03	8.49	79.40	61.7	17.70000000	11.065	0.0000000000000

- A larger mean difference between the two groups will increase the likelihood of finding a statistically significant result ($|\bar{x}_1 - \bar{x}_2|\uparrow \Rightarrow p\downarrow$)

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	df	p
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑	/	↓
s_1 ; s_2 ↓		/	
n_1 ; n_2			

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	df	p
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑		↓
s_1 ; s_2 ↓	↑		↓
n_1 ; n_2			

Case 2: s_1 s_2↓ ⇨ p↓

Group1: 60 value, mean1 = 64, s1 = 10.48, [44,94]
Group2: 60 value, mean2 = 62, s2 = 9.68 , [42,83]

↓

Run t-test for group1, group2

↓

Group1: s1 = s1 - 0.5
Group2: s2 = s2 - 0.5

Target s1 = 0.5

Case 2: s_1 s_2↓ ⇨ p↓

	Equal variance	
	t	p
$\lvert \bar{x}_1 - \bar{x}_2 \rvert$ ↑	↑	↓
s_1 ; s_2 ↓	↑	↓
n_1 ; n_2 ↑		↓

n1	n2	mean1	mean2	mean_diff	s1	s2	t	p
<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
60	60	64.70	61.70	3.000000	9.52	8.49	1.822	0.0710151861847
60	60	64.70	61.68	3.016667	9.16	8.02	1.919	0.0573939189178
60	60	64.58	61.80	2.783333	8.68	7.68	1.860	0.0653288441832
60	60	64.55	61.70	2.850000	8.14	7.24	2.028	0.0448448677426
60	60	64.52	61.72	2.800000	7.77	6.81	2.099	0.0379165334710
60	60	64.55	61.73	2.816667	7.25	6.41	2.254	0.0260235676614
60	60	64.48	61.82	2.666667	6.79	5.92	2.292	0.0236502683457
60	60	64.47	61.82	2.650000	6.40	5.44	2.445	0.0159802319858
60	60	64.47	61.83	2.633333	5.92	5.02	2.626	0.0097792683378
60	60	64.40	61.83	2.566667	5.41	4.58	2.806	0.0058761436492
60	60	64.42	61.78	2.633333	5.01	4.17	3.129	0.0022080829916
60	60	64.37	61.87	2.500000	4.54	3.73	3.297	0.0012897422665
60	60	64.30	61.87	2.433333	4.14	3.26	3.573	0.0005114605154
60	60	64.25	61.92	2.333333	3.64	2.77	3.949	0.0001338912332
60	60	64.18	61.87	2.316667	3.15	2.38	4.545	0.0000133995746
60	60	64.17	61.90	2.266667	2.73	1.94	5.243	0.0000007036262
60	60	64.07	61.98	2.083333	2.26	1.60	5.826	0.0000000502790
60	60	64.15	61.97	2.183333	1.79	1.07	8.094	0.0000000000006
60	60	64.08	61.98	2.100000	1.41	0.65	10.501	0.0000000000000
60	60	64.07	62.00	2.066667	0.95	0.00	16.775	0.0000000000000

- A smaller standard deviation results in less variability within each group, making it easier to detect differences between groups.
- Lower variability makes it more likely to be statistically significant (s_1 s_2 ↓ ⇨ p↓)

Components affect t and df

Equal Variance

t formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1)n_1 + (n_2 - 1)n_1} + \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1)n_2 + (n_2 - 1)n_2}}}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	**df**	**p**
$\lvert \bar{x}_1 - \bar{x}_2 \rvert$ ↑	↑		↓
s_1 ; s_2 ↓	↑		↓
n_1 ; n_2 ↑			

Components affect t and df

Equal Variance

t formula

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

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1)n_1 + (n_2 - 1)n_1} + \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1)n_2 + (n_2 - 1)n_2}}}$$

df

$$df = n_1 + n_2 - 2$$

	Equal variance		
	t	**df**	**p**
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑		↓
s_1 ; s_2 ↓	↑		↓
n_1 ; n_2 ↑	↑	↑	↓

Case 3: n_1 n_2↑ ⇨ p↓

Group1: 60 value, mean1 = 64, s1 = 10.48, [44,94]
Group2: 60 value, mean2 = 62, s2 = 9.68 , [42,83]

Run t-test for group1, group2

Group1: n1 = n1 + 1
Group2: n2 = n2 + 1
Mean different between group unchanged

Target n = 500

Case 3: n_1 n_2↑ ⇨ p↓

	Equal variance		
	t	df	p
$\|\bar{x}_1 - \bar{x}_2\|$ ↑	↑		↓
s_1 ; s_2 ↓	↑		↓
n_1 ; n_2 ↑	↑	↑	↓

n1 <int>	n2 <int>	mean1 <dbl>	mean2 <dbl>	mean_diff <dbl>	s1 <dbl>	s2 <dbl>	p <dbl>
60	60	64.70	61.70	3	9.52	8.49	0.0710151862
61	61	64.80	61.80	3	9.48	8.45	0.0675101485
62	62	65.00	62.00	3	9.53	8.53	0.0670661043
63	63	64.81	61.81	3	9.57	8.59	0.0664586187
64	64	64.98	61.98	3	9.60	8.64	0.0653550990
65	65	64.91	61.91	3	9.54	8.59	0.0618430284
66	66	65.14	62.14	3	9.65	8.72	0.0632156081
67	67	65.12	62.12	3	9.58	8.66	0.0593372603
68	68	65.51	62.51	3	10.05	9.19	0.0714981369
69	69	65.90	62.90	3	10.47	9.67	0.0825698606
70	70	66.10	63.10	3	10.53	9.74	0.0823870564
71	71	65.97	62.97	3	10.51	9.73	0.0797792034
72	72	65.82	62.82	3	10.52	9.75	0.0780049361
73	73	65.93	62.93	3	10.49	9.73	0.0752392865
74	74	65.81	62.81	3	10.47	9.72	0.0728135763
75	75	65.88	62.88	3	10.41	9.67	0.0695064392
76	76	66.11	63.11	3	10.53	9.80	0.0710500406
77	77	65.92	62.92	3	10.58	9.87	0.0708386177
78	78	65.90	62.90	3	10.51	9.81	0.0673062685
79	79	65.92	62.92	3	10.45	9.75	0.0639339559
80	80	66.19	63.19	3	10.65	9.97	0.0676915038
81	81	66.48	63.48	3	10.91	10.25	0.0731613410
82	82	66.74	63.74	3	11.10	10.46	0.0767513869
83	83	66.88	63.88	3	11.10	10.47	0.0751039762

Case 3: n_1 n_2↑ ⇨ p↓

n1	n2	mean1	mean2	mean_diff	s1	s2	p
<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
60	60	64.70	61.70	3	9.52	8.49	0.0710151862
61	61	64.80	61.80	3	9.48	8.45	0.0675101485
62	62	65.00	62.00	3	9.53	8.53	0.0670661043
63	63	64.81	61.81	3	9.57	8.59	0.0664586187
64	64	64.98	61.98	3	9.60	8.64	0.0653550990
65	65	64.91	61.91	3	9.54	8.59	0.0618430284
66	66	65.14	62.14	3	9.65	8.72	0.0632156081
67	67	65.12	62.12	3	9.58	8.66	0.0593372603
68	68	65.51	62.51	3	10.05	9.19	0.0714981369
69	69	65.90	62.90	3	10.47	9.67	0.0825698606
70	70	66.10	63.10	3	10.53	9.74	0.0823870564
71	71	65.97	62.97	3	10.51	9.73	0.0797792034
72	72	65.82	62.82	3	10.52	9.75	0.0780049361
73	73	65.93	62.93	3	10.49	9.73	0.0752392865
74	74	65.81	62.81	3	10.47	9.72	0.0728135763
75	75	65.88	62.88	3	10.41	9.67	0.0695064392
76	76	66.11	63.11	3	10.53	9.80	0.0710500406
77	77	65.92	62.92	3	10.58	9.87	0.0708386177
78	78	65.90	62.90	3	10.51	9.81	0.0673062685
79	79	65.92	62.92	3	10.45	9.75	0.0639339559
80	80	66.19	63.19	3	10.65	9.97	0.0676915038
81	81	66.48	63.48	3	10.91	10.25	0.0731613410
82	82	66.74	63.74	3	11.10	10.46	0.0767513869
83	83	66.88	63.88	3	11.10	10.47	0.0751039762

n1	n2	mean1	mean2	mean_diff	s1	s2	p
<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
122	122	66.47	63.47	3	12.18	11.80	0.0519266348
123	123	66.58	63.58	3	12.19	11.82	0.0512204312
124	124	66.42	63.42	3	12.27	11.90	0.0518040816
125	125	66.38	63.38	3	12.23	11.86	0.0500644098
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
126	126	66.52	63.52	3	12.28	11.92	0.0501655877
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
127	127	66.62	63.62	3	12.28	11.92	0.0493105332
128	128	66.69	63.69	3	12.26	11.90	0.0479632565
129	129	66.62	63.62	3	12.23	11.87	0.0466819043
130	130	66.73	63.73	3	12.25	11.89	0.0461819168
130	130	66.73	63.73	3	12.25	11.89	0.0461819168
130	130	66.73	63.73	3	12.25	11.89	0.0461819168

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)n_1 + (n_2-1)n_1} + \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)n_2 + (n_2-1)n_2}}}$$

Case 3: n_1 n_2↑ ⇨ p↓

n1 <int>	n2 <int>	mean1 <dbl>	mean2 <dbl>	mean_diff <dbl>	s1 <dbl>	s2 <dbl>	p <dbl>
486	486	68.49	65.49	3	10.72	10.61	0.0000129279
486	486	68.49	65.49	3	10.72	10.61	0.0000129279
487	487	68.46	65.46	3	10.73	10.62	0.0000128194
488	488	68.47	65.47	3	10.72	10.61	0.0000123516
489	489	68.46	65.46	3	10.71	10.60	0.0000118988
490	490	68.47	65.47	3	10.70	10.59	0.0000114468
491	491	68.50	65.50	3	10.71	10.61	0.0000115395
492	492	68.51	65.51	3	10.71	10.60	0.0000111890
493	493	68.51	65.51	3	10.70	10.59	0.0000107792
494	494	68.53	65.53	3	10.70	10.59	0.0000105531
494	494	68.53	65.53	3	10.70	10.59	0.0000105531
495	495	68.51	65.51	3	10.70	10.59	0.0000103329
496	496	68.52	65.52	3	10.69	10.59	0.0000099717
496	496	68.52	65.52	3	10.69	10.59	0.0000099717
496	496	68.52	65.52	3	10.69	10.59	0.0000099717
496	496	68.52	65.52	3	10.69	10.59	0.0000099717
497	497	68.52	65.52	3	10.68	10.58	0.0000095707
498	498	68.54	65.54	3	10.69	10.58	0.0000095402
499	499	68.53	65.53	3	10.69	10.58	0.0000093081

- Sample size increases, more likely to detect statistically significant differences (n_1 n_2↑ ⇨ p↓)

Components affect t and df

	Equal Variance	Unequal Variance
t formula	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$ $\quad s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
df	$df = n_1 + n_2 - 2$	$df = \dfrac{(\frac{s_1^2}{n1} + \frac{s_2^2}{n2})^2}{\frac{(\frac{s_1^2}{n1})^2}{n1-1} + \frac{(\frac{s_2^2}{n2})^2}{n2-1}}$

	Equal variance			Unequal variance		
	t	df	p	t	df	p
$\lvert\bar{x}_1 - \bar{x}_2\rvert$ ↑	↑	/	↓	↑	/	↓
s_1 ; s_2 ↓	↑	/	↓	↑	↑↓	↓
n_1 ; n_2 ↑	↑	↑	↓	↑	↑↓	↓

Conclusion p
Independent-means t-test

$p \leq 0.05$:

1. Larger difference between 2 groups ($|\bar{x}_1 - \bar{x}_2|$ |med1−med2| ↑)
2. Lower variability in each group (s_1 s_2 ↓)
3. Increase sample size (n_1 n_2 ↑)

References

1. Andrade C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. Indian journal of psychological medicine. 2019 May;41(3):210-5.
2. Baker M. Statisticians issue warning over misuse of P values. Nature 531 (March (7593)), 151.
3. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Annals of Ibadan postgraduate medicine. 2008;6(1):21-6.
4. Diez DM, Barr CD, Cetinkaya-Rundel M. OpenIntro statistics. Boston, MA, USA:: OpenIntro; 2012.
5. Ganesh S, Cave V. P-values, p-values everywhere!. New Zealand Veterinary Journal. 2018 Mar 4;66(2):55-6.