Chapter 9: Comparing two means

Labcoat Leni's Real Research

You don't have to be mad here, but it helps

Problem

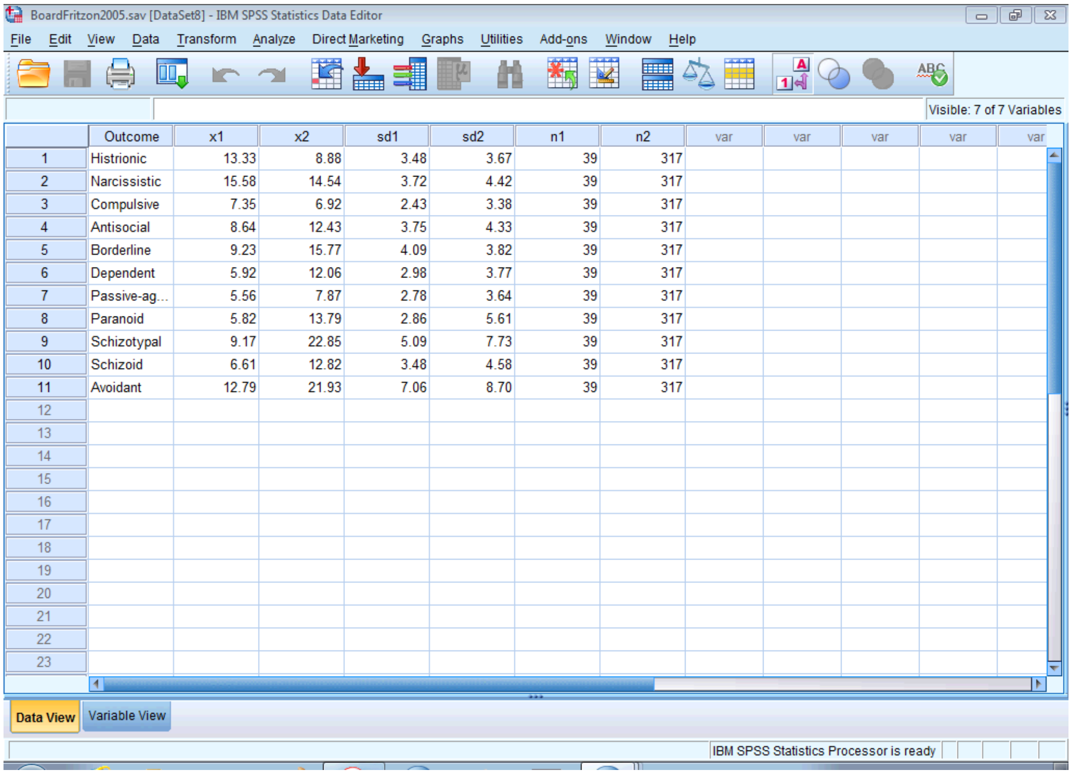
Board, B. J., & Fritzon, K. (2005). *Psychology, Crime & Law*, *11*(1), 17–32.

In the UK you often see the 'humorous' slogan 'You don't have to be mad to work here, but it helps' stuck up in work places. Well, Board and Fritzon (2005) took this a step further by measuring whether 39 senior business managers and chief executives from leading UK companies were mad (well, had personality disorders, PDs). They gave them the Minnesota Multiphasic Personality Inventory Scales for DSM III Personality Disorders (MMPI-PD), which is a well-validated measure of 11 personality disorders: histrionic, narcissistic, antisocial, borderline, dependent, compulsive, passive-aggressive, paranoid, schizotypal, schizoid and avoidant. They needed a comparison group, and what better one to choose than 317 legally classified psychopaths at Broadmoor Hospital (a famous high-security psychiatric hospital in the UK).

The authors report the means and SDs for these two groups in Table 2 of their paper. Using these values and the syntax file **Independent t from means.sps**, we can run *t*-tests on these means. Use the file **Board and Fritzon 2005.sav** and the syntax file to run *t*-tests to see whether managers score higher on personality disorder questionnaires than legally classified psychopaths. Report these results. What do you conclude?

Solution

The data look like this:



	Outcome	x1	x2	sd1	sd2	n1	n2
1	Histrionic	13.33	8.88	3.48	3.67	39	317
2	Narcissistic	15.58	14.54	3.72	4.42	39	317
3	Compulsive	7.35	6.92	2.43	3.38	39	317
4	Antisocial	8.64	12.43	3.75	4.33	39	317
5	Borderline	9.23	15.77	4.09	3.82	39	317
6	Dependent	5.92	12.06	2.98	3.77	39	317
7	Passive-ag...	5.56	7.87	2.78	3.64	39	317
8	Paranoid	5.82	13.79	2.86	5.61	39	317
9	Schizotypal	9.17	22.85	5.09	7.73	39	317
10	Schizoid	6.61	12.82	3.48	4.58	39	317
11	Avoidant	12.79	21.93	7.06	8.70	39	317

The columns represent the following:

Outcome: A string variable that tells us which personality disorder the numbers in each row relate to.

X1: Mean of the managers group.

X2: Mean of the psychopaths group.

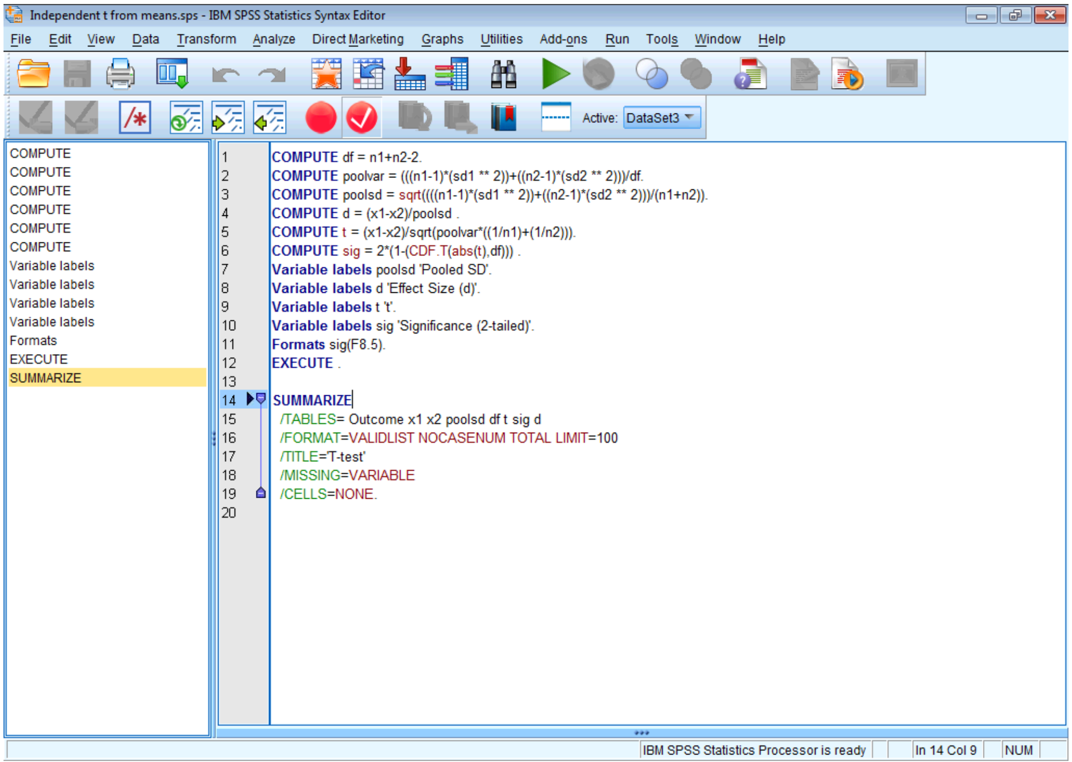
sd1: Standard deviation of the managers group.

sd2: Standard deviation of the psychopaths group.

n1: The number of managers tested.

n2: The number of psychopaths tested.

The syntax file looks like this:



We can run the syntax by selecting Run All . The output looks like this:

T-test[a]

	Outcome Measure	Mean of Managers	Mean of Psychopaths	Pooled SD	df	t	Significance (2-tailed)	Effect Size (d)
1	Histrionic	13.33	8.88	3.64	354.00	7.18	.00000	1.22
2	Narcissistic	15.58	14.54	4.34	354.00	1.41	.15977	.24
3	Compulsive	7.35	6.92	3.28	354.00	.77	.44185	.13
4	Antisocial	8.64	12.43	4.26	354.00	-5.23	.00000	-.89
5	Borderline	9.23	15.77	3.84	354.00	-10.01	.00000	-1.70
6	Dependent	5.92	12.06	3.68	354.00	-9.80	.00000	-1.67
7	Passive-aggressive	5.56	7.87	3.55	354.00	-3.83	.00015	-.65
8	Paranoid	5.82	13.79	5.37	354.00	-8.73	.00000	-1.48
9	Schizotypal	9.17	22.85	7.47	354.00	-10.76	.00000	-1.83
10	Schizoid	6.61	12.82	4.46	354.00	-8.18	.00000	-1.39
11	Avoidant	12.79	21.93	8.52	354.00	-6.31	.00000	-1.07

a. Limited to first 100 cases.

We can report that managers scored significantly higher than psychopaths on histrionic personality disorder, $t(354) = 7.18$, $p < .001$, $d = 1.22$. There were no significant differences between groups on narscissistic personality disorder, $t(354) = 1.41$, $p > .05$, $d = 0.24$, or compulsive personality disorder, $t(354) = 0.77$, $p > .05$, $d = 0.13$. On all other measures, psychopaths scored significantly higher than managers: antisocial personality disorder, $t(354)$

= −5.23, $p < .001$, $d = −0.89$; borderline personality disorder, $t(354) = −10.01$, $p < .001$, $d = −1.70$; dependent personality disorder, $t(354) = −9.80$, $p < .001$, $d = −1.67$; passive-aggressive personality disorder, $t(354) = −3.83$, $p < .001$, $d = −0.65$; paranoid personality disorder, $t(354) = −8.73$, $p < .001$, $d = −1.48$; schizotypal personality disorder, $t(354) = −10.76$, $p < .001$, $d = −1.83$; schizoid personality disorder, $t(354) = −8.18$, $p < .001$, $d = −1.39$; avoidant personality disorder, $t(354) = −6.31$, $p < .001$, $d = −1.07$.

The results show the presence of elements of PD in the senior business manager sample, especially those most associated with psychopathic PD. The senior business manager group showed significantly higher levels of traits associated with histrionic PD than psychopaths. They also did not significantly differ from psychopaths in narcissistic and compulsive PD traits. These findings could be an issue of power (effects were not detected but are present). The effect sizes *d* can help us out here, and these are quite small (0.24 and 0.13), which can give us confidence that there really isn't a difference between psychopaths and managers on these traits. Broad and Fritzon (2005) conclude that: 'At a descriptive level this translates to: superficial charm, insincerity, egocentricity, manipulativeness (histrionic), grandiosity, lack of empathy, exploitativeness, independence (narcissistic), perfectionism, excessive devotion to work, rigidity, stubbornness, and dictatorial tendencies (compulsive). Conversely, the senior business manager group is less likely to demonstrate physical aggression, consistent irresponsibility with work and finances, lack of remorse (antisocial), impulsivity, suicidal gestures, affective instability (borderline), mistrust (paranoid), and hostile defiance alternated with contrition (passive/aggressive.)' And these people are in charge of large companies. Hmm, suddenly a lot things make sense.

Bladder control

Problem

Tuk, M. A., et al. (2011). *Psychological Science, 22*(5), 627–633.

Visceral factors that require us to engage in self-control (such as a filling bladder) can affect our inhibtory abilities in unrelated domains. In a fascinating study by Tuk, Trampe, and Warlop (2011) participants were given five cups of water: one group was asked to drink them all, whereas another was asked to take a sip from each. This manipulation led one group to have full bladders and the other group relatively empty bladders (**Drink_Group**). Later on, these participants were given eight trials on which they had to choose between a small financial reward that they would receive soon (SS) or a large financial reward for which they would wait longer (LL). They counted how many trials participants choose the LL reward as an indicator of inhibitory control (**LL_Sum**). Do a *t*-test to see whether people with full bladders inhibited more than those without (**Tuk et al. (2011).sav**).

Solution

We will conduct an independent samples *t*-test on these data because there were different participants in each of the two groups (independent design). Looking at the means in the Group Statistics table below, we can see that on average more participants in the High Urgency group (*M* = 4.5) chose the large financial reward for which they would wait longer than participants in the Low Urgency group (*M* = 3.8). Looking at the Independent Samples Test table, we can see that this difference was significant, *p* = .03.

Group Statistics

	Drink Condition	N	Mean	Std. Deviation	Std. Error Mean
Number of LL Rewards chosen (out of 8)	High Urgency (Drink everything)	50	4.5000	1.59399	.22542
	Low Urgency (Take Sips from the Water)	52	3.8269	1.49143	.20682

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Number of LL Rewards chosen (out of 8)	Equal variances assumed	2.025	.158	2.203	100	.030	.67308	.30553	.06692	1.27923
	Equal variances not assumed			2.200	98.890	.030	.67308	.30593	.06604	1.28011

To calculate the effect size *r*, all we need is the value of *t* and the *df* from the Independent Samples Test table:

$$r = \sqrt{\frac{2.203^2}{2.203^2 + 100}} = \sqrt{\frac{4.853}{104.853}} = .215$$

If you think back to our benchmarks for effect sizes, this represents a small to medium effect (it is between .1 (small effect) and .3 (a medium effect)).

In this example the Independent Samples Test table tells us that the value of *t* was 2.20, that this was based on 100 degrees of freedom, and that it was significant at *p* = .03. We can also see the means for each group. We could write this as:

- ✓ On average, participants who had full bladders (*M* = 4.5, *SD* = 1.59) were more likely to choose the large financial reward for which they would wait longer than participants who had relatively empty bladders (*M* = 3.8, *SD* = 1.49), $t(100) = 2.20$, $p < .05$.

The beautiful people

Problem

Gelman, A., & Weakliem, D. (2009). *American Scientist, 97*, 310–316.

Apparently there are more beautiful women in the world than there are handsome men. Satoshi Kanazawa explains this finding in terms of good-looking parents being more likely to have a baby daughter as their first child than a baby son. Perhaps more controversially, he suggests that, from an evolutionary point of view, beauty is a more valuable trait for women than for men (Kanazawa, 2007). In a playful and very informative paper, Andrew Gelman and David Weakliem discuss various statistical errors and misunderstandings, some of which have implications for Kanazawa's claims. The 'playful' part of the paper is that to illustrate their point they collected data on the 50 most beautiful celebrities (as listed by *People* magazine) of 1995–2000. They counted how many male and female children they had as of 2007. If Kanazawa is correct, these beautiful people whould have produced more girls than boys. Do a *t*-test to find out whether they did. The data are in **Gelman & Weakliem (2009).sav**.

Solution

We need to run a paired samples *t*-test on these data because the researchers recorded the number of daughters and sons for each participant (repeated-measures design). Looking at **Error! Reference source not found.**the output below, we can see that there was a non-significant difference between the number of sons and daughters produced by the 'beautiful' celebrities.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1	Number of Sons - Number of Daughters	.059	1.166	.073	-.085	.203	.807	253	.420

We are going to calculate Cohen's *d* as our effect size. To do this we first need to get some descriptive statistics for these data – the means and standard deviations:

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Number of Sons	254	0	4	.68	.057	.901
Number of Daughters	254	0	7	.62	.057	.902
Valid N (listwise)	254					

We can now compute Cohen's *d* using the two means (.68 and .62) and the standard deviation of the control group (it doesn't matter which one you choose here, but I have chosen to use the sons):

$$d = \frac{\bar{X}_{\text{Daughters}} - \bar{X}_{\text{Sons}}}{s_{\text{Sons}}} = \frac{0.62 - 0.68}{0.901} = -0.07$$

This means that there is –0.07 of a standard deviation difference between the number of sons and daughters produced by the celebrities, which is a very small effect.

In this example the SPSS output tells us that the value of *t* was 0.81, that this was based on 253 degrees of freedom, and that it was non-significant, *p* = .420. We also calculated the means for each group. We could write this as follows:

- ✓ There was no significant difference between the number of daughters (*M* = 0.62, *SE* = 0.06) produced by the 'beautiful' celebrities and the number of sons (*M* = 0.68, *SE* = 0.06), $t(253) = 0.81$, $p > .05$, $d = -0.07$.