Mert Dogan
Z1910844

Assignment 2
09/25/2023

1. A stats teacher wants to determine the effectiveness of her statistics lesson. She gives a simple skills test to nine students before the start of class (a pre-test) and another test to the same students at the end of class (a post-test). Conduct the appropriate test to determine her effectiveness.

Student	Pre-test	Post-test
1	78	80
2	67	69
3	56	70
4	78	79
5	96	96
6	82	84
7	84	88
8	90	92
9	87	92

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$

d values: {2,2,14,1,0,2,4,2,5}

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

d = Post-test – Pre-test
d-bar = 3.5556; s_d = 4.1866;
t = 2.5478, p-value = 0.0343
$t_{critical}$ = $\pm$ 2.306 at 8df

Reject NULL. There is enough evidence to conclude at .05 level that her stats class was effective. We can also conclude that, t-value being positive, it was effective in a positive way. (I didn't presume positive effect by doing right tailed t test and tested if the class was effective or not.)

2. Using the Facebook dataset, conduct the appropriate test to determine if males and females (0 = female, 1 = male in sas dataset) differ in terms of

a. number of friends $H_o: \mu_{male} = \mu_{female}$ $H_1: \mu_{male} \neq \mu_{female}$

b. posts $H_o: \mu_{male} = \mu_{female}$ $H_1: \mu_{male} \neq \mu_{female}$

c. number of likes $H_o: \mu_{male} = \mu_{female}$ $H_1: \mu_{male} \neq \mu_{female}$

Variable: Friends (Friends)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		335	775.0	680.8	37.1952	24.0000	4974.0
1		373	632.0	510.4	26.4271	3.0000	4999.0
Diff (1-2)	Pooled		143.0	597.1	44.9452		
Diff (1-2)	Satterthwaite		143.0		45.6276		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		775.0	701.8	848.1	680.8	632.8	736.6
1		632.0	580.0	684.0	510.4	476.2	549.9
Diff (1-2)	Pooled	143.0	54.7294	231.2	597.1	567.5	630.0
Diff (1-2)	Satterthwaite	143.0	53.3669	232.6			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	706	3.18	0.0015
Satterthwaite	Unequal	615.5	3.13	0.0018

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	334	372	1.78	<.0001

Variable: Posts (Posts)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		335	36.5075	79.7611	4.3578	0	746.0
1		373	39.0483	99.8448	5.1698	0	794.0
Diff (1-2)	Pooled		-2.5408	90.8982	6.8422		
Diff (1-2)	Satterthwaite		-2.5408		6.7614		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		36.5075	27.9352	45.0797	79.7611	74.1441	86.3061
1		39.0483	28.8826	49.2139	99.8448	93.1570	107.6
Diff (1-2)	Pooled	-2.5408	-15.9743	10.8927	90.8982	86.3946	95.9009
Diff (1-2)	Satterthwaite	-2.5408	-15.8160	10.7344			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	706	-0.37	0.7105
Satterthwaite	Unequal	696.7	-0.38	0.7072

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	372	334	1.57	<.0001

Variable: Likes (Likes)

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		335	135.9	271.3	14.8222	0	2619.0
1		373	151.2	252.2	13.0564	0	1857.0
Diff (1-2)	Pooled		-15.3015	261.4	19.6753		
Diff (1-2)	Satterthwaite		-15.3015		19.7526		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		135.9	106.7	165.0	271.3	252.2	293.6
1		151.2	125.5	176.8	252.2	235.3	271.7
Diff (1-2)	Pooled	-15.3015	-53.9305	23.3276	261.4	248.4	275.8
Diff (1-2)	Satterthwaite	-15.3015	-54.0845	23.4816			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	706	-0.78	0.4370
Satterthwaite	Unequal	683.78	-0.77	0.4388

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	334	372	1.16	0.1695

a) Since the p-value at Equality of Variances is less than alpha (.05), we can conclude that the two variances are unequal. Therefore, we need to check the Satterthwaite test. The P-value is calculated as 0.0018 < a (0.05), concluding reject NULL. There is a significant difference in the number of friends between males and females.

b) Since the p-value at Equality of Variances is less than alpha (.05), we can conclude that the two variances are unequal. Therefore, we need to check the Satterthwaite test. The P-value is calculated as 0.7072 > a (0.05), concluding fail to reject NULL. There is not a significant difference in posts between males and females.

c) Since the p-value at Equality of Variances is greater than alpha (.05), we can conclude that the two variances are equal. Therefore, we can check the Pooled test. The P-value is calculated as 0.4370 > a (0.05), concluding fail to reject NULL. There is not a significant difference in the number of likes between males and females.

3. Records of 40 used passenger cars and 40 used pickup trucks (none used commercially) were randomly selected to investigate whether there was any difference in the meantime in years that they were kept by the original owner before being sold. For cars, the mean was 5.3 years with standard deviation 2.2 years. For pickup trucks, the mean was 7.1 years with standard deviation 3.0 years. Test the hypothesis that there is a difference in the means against the null hypothesis that there is no difference. Use the 1% level of significance.

H_0: $\mu_{truck} = \mu_{car}$

H_1: $\mu_{truck} \neq \mu_{car}$

$$t = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \qquad t = \frac{(7.1-5.3)}{\sqrt{\frac{3.0^2}{40} + \frac{2.2^2}{40}}} = 3.0601$$

df = minimum of (n_1-1) or (n_2-1) So df = 39

t-critical = $\pm$ 2.7079 p-value (0.004) < a (0.01)

t-calculated = 3.0601

Reject H_o – we can conclude at .1 level that there is a significant difference in the meantime that truck owners and car owners keep their vehicles before they sell them.

4. Using the cars dataset, conduct the appropriate test to determine if there is a significant difference between the miles per gallon that we can expect between city driving and highway driving conditions (MPG_city and MPG_highway)

H_o : $\mu_d = 0$ H_1 : $\mu_d \neq 0$

Difference: MPG_City - MPG_Highway

N	Mean	Std Dev	Std Err	Minimum	Maximum
428	-6.7827	1.9495	0.0942	-11.0000	8.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-6.7827	-6.9679	-6.5975	1.9495	1.8270	2.0896

DF	t Value	Pr > |t|
427	-71.98	<.0001

With a t Value of -71.98 and a very small p-value of less than .0001, which is less than our nominal alpha of .05, we reject the NULL and conclude that there is a significant difference between the miles per gallon used between city and highway driving.