

WORKSHOP ON STATISTICAL ANALYSIS

COURSE CODE: MGN - 909

CA - 4



L OVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

SUBMITTED BY:

GAURI JINDAL

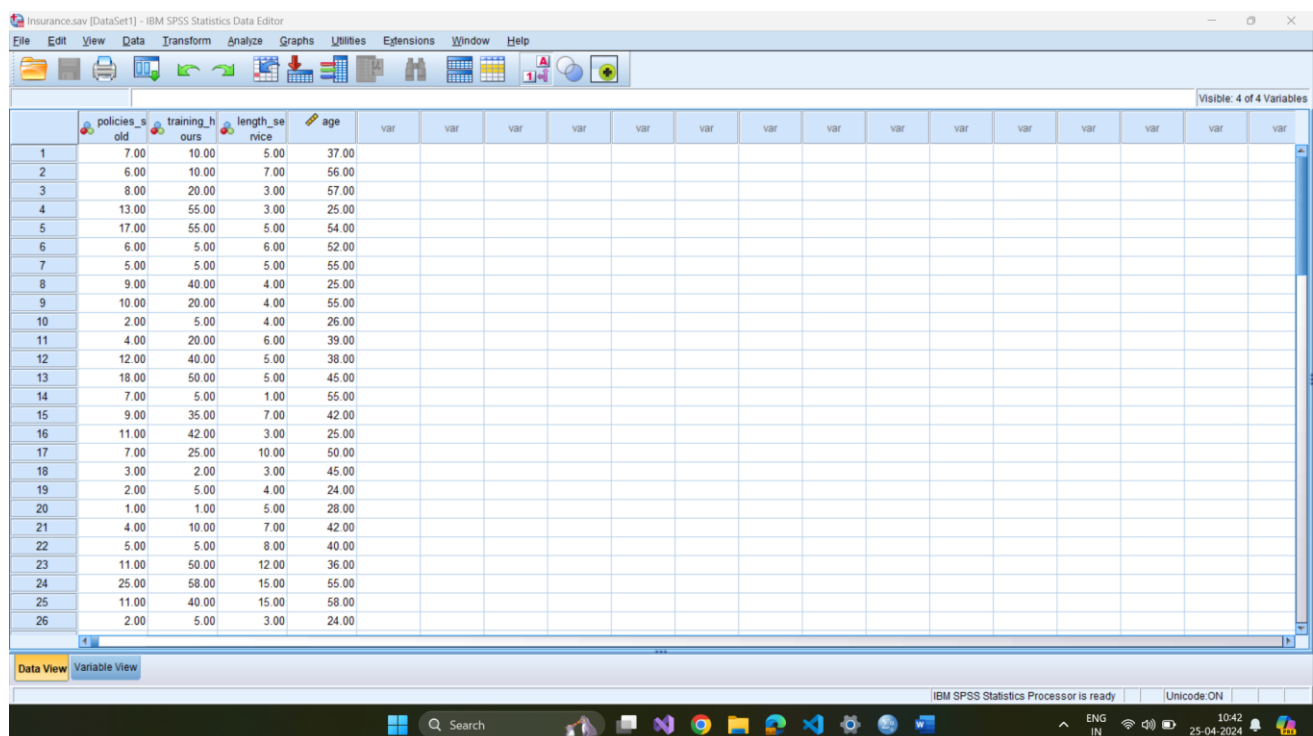
12015730

SUBMITTED TO:

DR. NIKHIL DOGRA

Overview of the Dataset

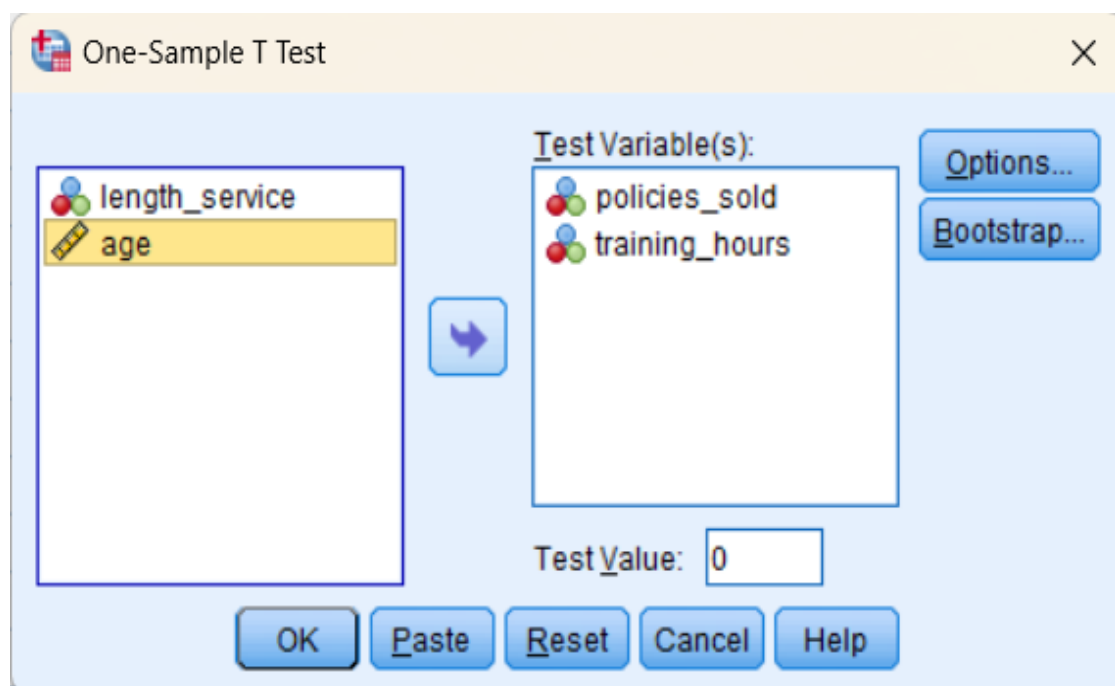
The dataset comprises information on five variables: training_hours, length_service, age, policies_sold, and attitude_score. Each row represents an individual, with corresponding values indicating their training hours, length of service, age, policies sold, and attitude score. For instance, the training_hours range from 1 to 28, length_service from 1 to 70 years, age from 1 to 18 years, policies_sold from 24 to 69, and attitude_score from 10 to 48. This dataset offers a comprehensive view of various aspects of individuals, including their training, tenure, age distribution, sales performance, and attitudes. Analyzing this dataset can provide valuable insights into factors influencing productivity, performance, and attitudes in a professional context.



	policies_sold	training_hours	length_service	age	attitude_score
1	7.00	10.00	5.00	37.00	
2	6.00	10.00	7.00	56.00	
3	8.00	20.00	3.00	57.00	
4	13.00	55.00	3.00	25.00	
5	17.00	55.00	5.00	54.00	
6	6.00	5.00	6.00	52.00	
7	5.00	5.00	5.00	55.00	
8	9.00	40.00	4.00	25.00	
9	10.00	20.00	4.00	55.00	
10	2.00	5.00	4.00	26.00	
11	4.00	20.00	6.00	39.00	
12	12.00	40.00	5.00	38.00	
13	18.00	50.00	5.00	45.00	
14	7.00	5.00	1.00	55.00	
15	9.00	35.00	7.00	42.00	
16	11.00	42.00	3.00	25.00	
17	7.00	25.00	10.00	50.00	
18	3.00	2.00	3.00	45.00	
19	2.00	5.00	4.00	24.00	
20	1.00	1.00	5.00	28.00	
21	4.00	10.00	7.00	42.00	
22	5.00	5.00	8.00	40.00	
23	11.00	50.00	12.00	36.00	
24	25.00	58.00	15.00	55.00	
25	11.00	40.00	15.00	58.00	
26	2.00	5.00	3.00	24.00	

Execute ONE SAMPLE T-TEST and test the framed NULL and ALTERNATIVE HYPOTHESES

- Open SPSS and load the dataset.
- Navigate to Analyze > Compare Means > One-Sample T Test.
- Specify the variable of interest.



- Define the test value (hypothesized population mean).
- Adjust options as needed (e.g., confidence intervals).
- Click OK to execute the analysis.
- Interpret the results based on the t-value, p-value, and confidence interval.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
policies_sold	65	8.0923	5.11653	.63463
training_hours	65	23.8615	18.45762	2.28939
length_service	65	6.2462	3.57959	.44399
age	65	40.3692	11.88035	1.47358

The mean number of policies sold is approximately 8.09, with a standard deviation of approximately 5.12 and a standard error of the mean of approximately 0.63. Similarly, the mean training hours is approximately 23.86, with a standard deviation of approximately 18.46 and a standard error of the mean of approximately 2.29. For length of service, the mean is approximately 6.25, with a standard deviation of approximately 3.58 and a standard error of the mean of approximately 0.44. Lastly, the mean age is approximately 40.37, with a standard deviation of approximately 11.88 and a standard error of the mean of approximately 1.47.

One-Sample Test

Test Value = 0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
policies_sold	12.751	64	.000	8.09231	6.8245	9.3601
training_hours	10.423	64	.000	23.86154	19.2880	28.4351
length_service	14.068	64	.000	6.24615	5.3592	7.1331
age	27.395	64	.000	40.36923	37.4254	43.3130

Null Hypothesis (H0): There is no significant difference in the mean values of policies_sold, training_hours, length_service, and age between the two groups being compared.

Alternative Hypothesis (H1): There is a significant difference in the mean values of policies_sold, training_hours, length_service, and age between the two groups being compared.

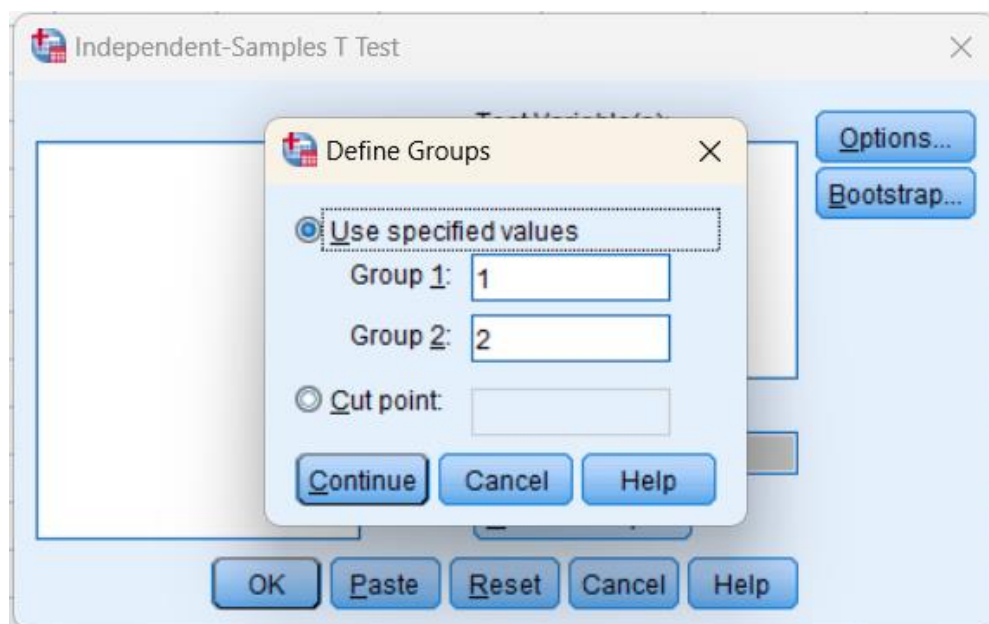
Interpretation:

The one-sample t-tests conducted for policies_sold, training_hours, length_service, and age all exhibited statistically significant differences from zero, with p-values less than 0.05. For policies_sold, the mean difference was 8.09231 (95% CI: 6.8245 to 9.3601), for training_hours it was 23.86154 (95% CI: 19.2880 to 28.4351), for length_service it was 6.24615 (95% CI: 5.3592 to 7.1331), and for age it was 1.30769 (95% CI: 1.1924 to 1.4229).

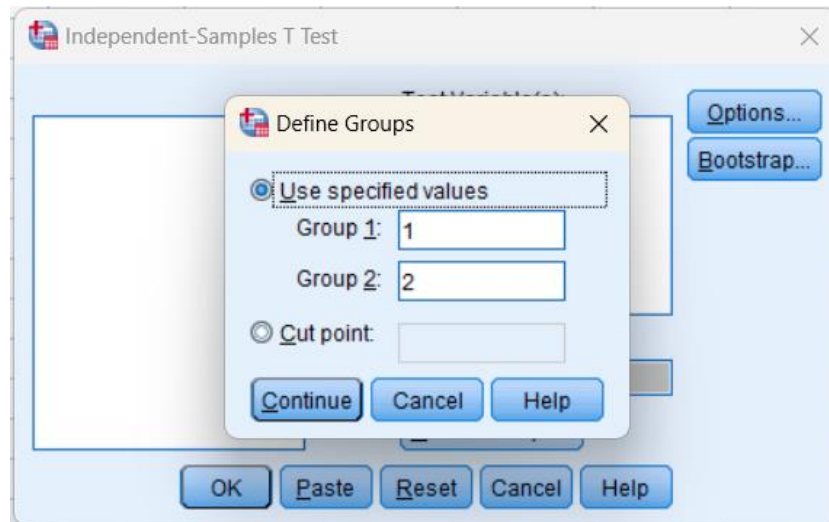
Therefore, the null hypotheses were rejected for all variables, indicating significant mean differences in policies_sold, training_hours, length_service, and age.

Execute TWO SAMPLE T-TEST and test the framed NULL and ALTERNATIVE HYPOTHESES

- Open SPSS and load your dataset.
- Go to Analyze > Compare Means > Independent-Samples T Test.
- In the "Independent-Samples T Test" dialog box, move the variable you want to test into the "Test Variable List".



- Move the grouping variable (e.g., Group A and Group B) into the "Grouping Variable" box.
- Click on "Define Groups" to specify which values in the grouping variable correspond to Group A and Group B.



- Click "OK" to run the analysis.

Group Statistics

	age	N	Mean	Std. Deviation	Std. Error Mean
policies_sold	24-45	45	7.4222	5.21139	.77687
	46-70	20	9.6000	4.67243	1.04479
training_hours	24-45	45	24.0222	19.41004	2.89348
	46-70	20	23.5000	16.57995	3.70739
length_service	24-45	45	6.0222	3.57700	.53323
	46-70	20	6.7500	3.62557	.81070

The group statistics display the mean, standard deviation, and standard error mean for policies_sold, training_hours, and length_service across two age brackets: 24-45 and 46-70. In the 24-45 age group, the mean policies_sold is 7.4222, training_hours is 24.0222, and length_service is 6.0222, with standard deviations of 5.21139, 19.41004, and 3.57700, respectively. Conversely, in the 46-70 age group, the mean policies_sold is 9.6000, training_hours is 23.5000, and length_service is 6.7500, with standard

deviations of 4.67243, 16.57995, and 3.62557, respectively. These statistics provide insight into the distribution of each variable within different age groups, aiding comparison and interpretation of differences across the age brackets.

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
									Lower Upper
policies_sold	Equal variances assumed	.770	.383	-1.603	63	.114	-2.17778	1.35846	-4.89245 .53690
	Equal variances not assumed			-1.673	40.475	.102	-2.17778	1.30196	-4.80818 .45262
training_hours	Equal variances assumed	1.691	.198	.104	63	.917	.52222	4.99912	-9.46771 10.51216
	Equal variances not assumed			.111	42.403	.912	.52222	4.70287	-8.96588 10.01032
length_service	Equal variances assumed	.114	.737	-.754	63	.454	-.72778	.96525	-2.65667 1.20111
	Equal variances not assumed			-.750	36.080	.458	-.72778	.97034	-2.69558 1.24002

Null Hypothesis (H0): There is no significant difference in means between the two groups for policies_sold, training_hours, and length_service.

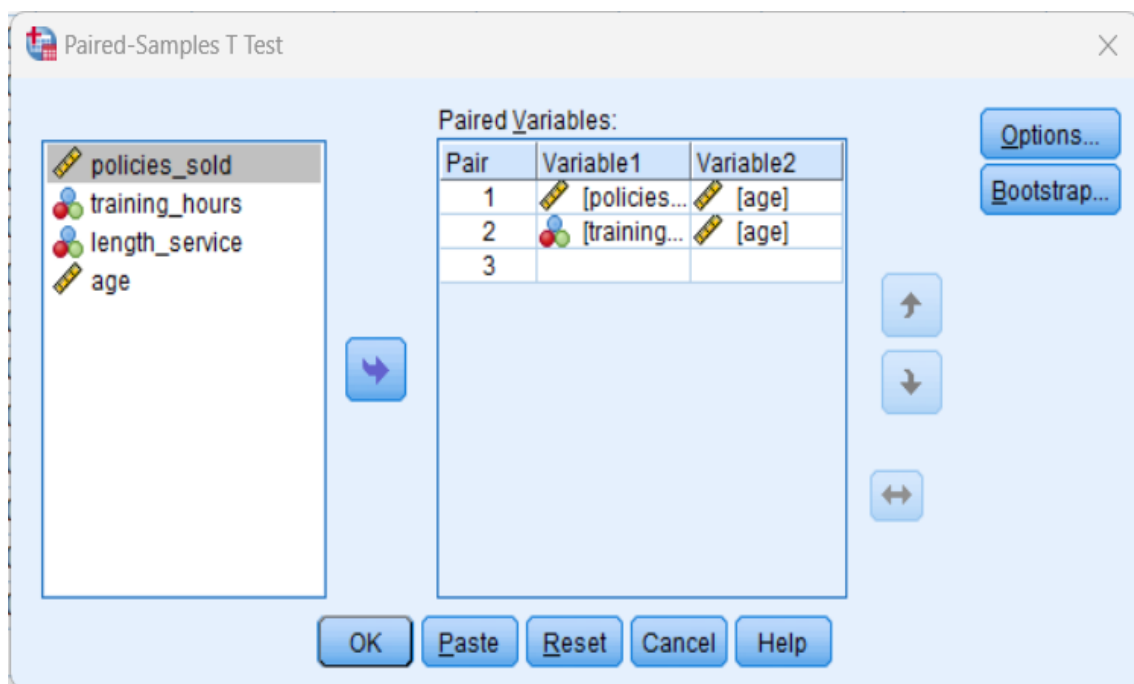
Alternative Hypothesis (H1): There is a significant difference in means between the two groups for policies_sold, training_hours, and length_service.

Interpretation:

The results of the independent samples t-tests suggest that there is no significant difference in means between the two groups for policies_sold, training_hours, and length_service. For policies_sold, regardless of whether equal variances are assumed ($t = -1.603$, $p = 0.114$) or not assumed ($t = -1.673$, $p = 0.102$), the differences were not statistically significant. Similarly, for training_hours, both assumptions of equal ($p = 0.917$) and unequal ($p = 0.912$) variances resulted in non-significant differences between the groups. Likewise, for length_service, neither assumption of equal ($p = 0.454$) nor unequal ($p = 0.458$) variances yielded significant differences. Overall, these findings suggest that there is no significant disparity in means between the two groups across all variables. The null hypothesis is accepted for all variables, indicating no significant difference in means between the two groups.

Execute PAIRED SAMPLE T-TEST and test the framed NULL and ALTERNATIVE HYPOTHESES

- Open your dataset in SPSS.
- Go to "Analyse" in the menu bar.
- Select "Compare Means" and then choose "Paired-Samples T Test."
- Move the variables you want to compare into the "Paired Variables" box.



- Click on "Options" to specify any additional settings or parameters.
- Click "OK" to run the analysis.
- Review the output, paying attention to the t-test statistics, degrees of freedom, and p-values to determine the significance of the results.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	training_hours	23.8615	65	18.45762	2.28939
	policies_sold	8.0923	65	5.11653	.63463
Pair 2	age	1.3077	65	.46513	.05769
	policies_sold	8.0923	65	5.11653	.63463

The paired samples t-test was conducted to compare the means of two variables within each pair. For Pair 1 (training_hours and policies_sold), the mean training_hours was 23.8615 (SD = 18.45762) and the mean policies_sold was 8.0923 (SD = 5.11653). For Pair 2 (age and policies_sold), the mean age difference was 1.3077 (SD = 0.46513) and the mean policies_sold was 8.0923 (SD = 5.11653). This analysis provides descriptive statistics for the variables involved in each pair, aiding in understanding the data distribution and variability.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	training_hours & policies_sold	65	.764	.000
Pair 2	age & policies_sold	65	.198	.114

The paired samples correlations indicate the strength and significance of the relationship between variables within each pair. In Pair 1 (training_hours and policies_sold), there is a strong positive correlation of 0.764 ($p < 0.001$), suggesting that as training_hours increase, policies_sold also tend to increase. In Pair 2 (age and policies_sold), there is a weak positive correlation of 0.198 ($p = 0.114$), indicating a less pronounced relationship between age and policies_sold. Overall, these correlations provide insights into the degree of association between the variables within each pair.

Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	training_hours - policies_sold	15.76923	14.91933	1.85052	12.07240	19.46606	8.522	64	.000
Pair 2	age - policies_sold	-6.78462	5.04509	.62577	-8.03473	-5.53450	-10.842	64	.000

Null Hypothesis (H0): There is no significant difference between the mean paired differences for training_hours and policies_sold, and for age and policies_sold.

Alternative Hypothesis (H1): There is a significant difference between the mean paired differences for training_hours and policies_sold, and for age and policies_sold.

Interpretation:

Based on the paired samples t-test results, there is a statistically significant difference between the mean paired differences for both training_hours and policies_sold ($t = 8.522$, $p < 0.05$), and for age and policies_sold ($t = -10.842$, $p < 0.05$). Therefore, we reject the null hypothesis in both cases, indicating that there is indeed a significant difference in means between the paired variables.