PARKING LOT CODING SHEET – DIFFERENCE BETWEEN THE MEANS

HYPOTHESIS: HIGHER INCOME PERSONS DRIVE MORE EXPENSIVE CARS

Independent variable: income
Measure: Student or faculty/staff
Categorical/nominal

Income → Car Value

Dependent variable: car value
Measure: Car value, 1-5
Continuous

STUDENT LOT

Car	Score	Mean	Diff.	Sq.
1	___	___	___	___
2	___	___	___	___
3	___	___	___	___
4	___	___	___	___
5	___	___	___	___
6	___	___	___	___
7	___	___	___	___
8	___	___	___	___
9	___	___	___	___
10	___	___	___	___
			Sum	____

Variance s^2 ___ (Sum / n-1)

FACULTY/STAFF LOT

Car	Score	Mean	Diff.	Sq.
1	4	3.3	.7	.49
2	5	3.3	1.7	2.89
3	2	3.3	1.3	1.69
4	5	3.3	1.7	2.89
5	2	3.3	1.3	1.69
6	1	3.3	2.3	5.29
7	4	3.3	.7	.49
8	4	3.3	.7	.49
9	5	3.3	1.7	2.89
10	1	3.3	2.3	5.29
			Sum	24.10

Variance s^2 2.68 (Sum / n-1)

CALCULATE THE *t* STATISTIC

1. Obtain the "pooled sample variance" $\mathbf{S_P^2}$ (simplified method – midpoint between the two sample **variances**)

$$S_p^2 = \frac{s^2_1 + s^2_2}{2}$$

Answer _________

2. Compute the standard error of the difference between means (how much we would expect means to differ by chance alone)

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Answer _________

3. Compute the t statistic (actual diff. betw. means / standard error of the difference between means)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

Answer _________ (sign is irrelevant – only important thing is the actual difference)

4. Degrees of freedom (stand-in for number of cases)

$df = (n_1 + n_2) - 2$ Answer ________

5. To confirm the WORKING hypothesis (i.e., that higher-income persons drive more expensive cars) the actual (obtained) difference between the means must be sufficiently large to overcome the NULL hypothesis of no significant difference.

Check the t-table. The NULL hypothesis of no significant difference between the means can be rejected if its probability of being true is not greater than five in one-hundred.

	Level of Significance for One-Tailed Test					
df	***.10***	***.05***	***.025***	***.01***	***.005***	***.0005***
	Level of Significance for Two-Tailed Test					
	.20	***.10***	***.05***	***.02***	***.01***	***.001***
1	3.078	6.314	12.706	31.821	63.657	636.62
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850

If our hypothesis predicted the direction of the effect (lower income → cheaper car) we would use the ONE-tail test. If it did not (simply, income → car value), we would use a TWO-tail test.

Here, our coefficient (again, sign is irrelevant) is significant for either test. Note that it's not quite large enough to go in the far right column, so it must be placed in the column to the left.

- If the (working) hypothesis was one tailed, there are less than five chances in one-thousand that the NULL hypothesis is correct.

- If the (working) hypothesis was two-tailed, there is less than one chance in one-hundred that the NULL hypothesis is correct.

Either way, the NULL hypothesis is rejected and the WORKING hypothesis is confirmed. It's like any test of significance. The LARGER the coefficient, the greater the probability that we can reject the NULL hypothesis and confirm the WORKING hypothesis.