Topic 2 Examples Solutions: Two-Sample T-tools

1. Does alcohol affect males and females differently? A study involving males and females with similar physical characteristics was conducted. In a controlled setting, each individual was asked to consume 4 ounces of alcohol. One hour after consumption each participant took a Breathalyzer to measure his or her blood alcohol level. The following results were obtained:

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
BAC	male	20	.086563	.0134962	.0030178
	female	20	.098639	.0139670	.0031231

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
BAC	Equal variances assumed	.071	.792	-2.781	38	.008	-.0120762	.0043430	-.0208680	-.0032843
	Equal variances not assumed			-2.781	37.955	.008	-.0120762	.0043430	-.0208684	-.0032840

a) Is there a real difference in blood alcohol levels between males and females? Carry out an appropriate test to answer this question.

Y_1 – *BAC for males,* Y_2 – *BAC for females*

We are interested in the parameter, $\mu_1 - \mu_2$.

$H_0 : \mu_1 - \mu_2 = 0$ (*no difference*)

$H_a : \mu_1 - \mu_2 \neq 0$ (*some difference*)

From random samples of $n_1 = 20$ *and* $n_2 = 20$, *from populations 1 and 2 respectively, it was observed that* $\bar{y}_1 = 0.086563, \bar{y}_2 = 0.098639,$ $s_1 = 0.0134962,$ *and* $s_2 = 0.0139670.$

$$s_p = \sqrt{\frac{19(0.0134962)^2 + 14(0.0139670)^2}{38}} = 0.013734$$

$$t_0^* = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.086563 - 0.098639}{0.013734\sqrt{\frac{1}{20} + \frac{1}{20}}} = \frac{-0.012076}{0.004343} = -2.781.$$

$$p-value = 2P(t_{38} > 2.781) \approx 2P(t_{30} > 2.781) \in 2(0.0025, 0.005) = (0.005, 0.01).$$

With a p-value $\in (0.005, 0.01)$, *there is fairly strong evidence to indicate that average BAC is different for males and females.*

b) Calculate a 95% confidence interval for the difference in blood alcohol levels between males

$$Estimate = \bar{y}_1 - \bar{y}_2 = -0.0120762, S.E.(\bar{Y}_1 - \bar{Y}_2) = 0.004343.$$

$$For\ 95\%\ confidence,\ C.V.\ =\ t^*_{n_1+n_2-2,\ 1\text{-}\alpha/2} = t^*_{33,0.975} \approx t^*_{30,0.975} = 2.042.$$

$$\therefore A\ 95\%\ C.I.\ for\ \mu_1 - \mu_2\ is\ then:$$

$$\Rightarrow -0.0120762 \pm 2.042(0.004343) \qquad \Rightarrow (-0.02094, -0.00321).$$

It is estimated with 95% confidence that the true difference in the mean BAC for males vs. females 1 hour after consuming 4 ounces of alcohol is between -0.02094 and -0.00321.

2. **Cloud Seeding to increase Rainfall:** To test the hypothesis that injection of silver iodide into cumulus clouds can lead to an increase in rainfall, a randomized experiment was conducted in Florida between 1968 and 1972. On each of 52 days that were deemed suitable for cloud seeding, a plane went on a seeding run. A random mechanism was used to determine if the seeding mechanism was to be loaded or not. To avoid bias, the pilot was unaware if the seeding mechanism was loaded or not. The precipitation (measured as total rain volume (acre-ft) falling from the cloud) as measured by radar, was observed following the seeding run. The results for 26 seeded days and 26 unseeded days (control) were:

Rainfall (Original Data)	
Unseeded	**Seeded**
1202.60	2745.60
830.10	1697.80
372.40	1656.00
345.50	978.00
321.20	703.40
244.30	489.10
163.00	430.00
147.80	334.10
95.00	302.80
87.00	274.70
81.20	274.70
68.50	255.00
47.30	242.50
41.10	200.70
36.60	198.60
29.00	129.60
28.60	119.00
26.30	118.30
26.10	115.30
24.40	92.40
21.70	40.60
17.30	32.70
11.50	31.40
4.90	17.50
4.90	7.70
1.00	4.10

➔

Natural Log of Rainfall (Transformed Data)	
Unseeded	**Seeded**
ln(1202.6) = 7.09	7.92
6.72	7.44
5.92	7.41
5.84	6.89
5.77	6.56
5.50	6.19
5.09	6.06
5.00	5.81
4.55	5.71
4.47	5.62
4.40	5.62
4.23	5.54
3.86	5.49
3.72	5.30
3.60	5.29
3.37	4.86
3.35	4.78
3.27	4.77
3.26	4.75
3.19	4.53
3.08	3.70
2.85	3.49
2.44	3.45
1.59	2.86
1.59	2.04
0.00	1.41

(Data from J. Simpson, A. Olsen, and J. Eden, "A Bayesian Analysis of a Multiplicative Treatment Effect in Whether Modification," *Technometrics* 17 (1975): 161-66.)

a) Carry out a preliminary analysis to check the assumptions.
b) Does cloud seeding increase rainfall? Carry out a test.
c) Calculate a 95% confidence interval for the effect of cloud seeding.

What types of inferences can be made?

Since randomization was used to determine which days were seeded and which were not, it is safe to conclude that the seeding caused the difference (increase) in rainfall.

Population inferences can only be made if it is assumed that these selected days were representative of the set of all days. This seems like a reasonable assumption.

a) Carry out a preliminary analysis to check the assumptions.

Model Diagnostics: Are the T-tools appropriate?

If you refer to the graphs in the question file, it is evident from box-plots on the original scale that there is a violation of the equal variability assumption. In addition, from the histograms and normal probability plots there is a clear violation of normality. The T-tools are not appropriate on the original scale. After a natural log transformation on rainfall, there do not appear to be violations in the model assumptions for the T-tools. Thus, I will continue with the T-tools on the log-transformed data. The results will be interpreted on the original scale at the very end.

b) Does cloud seeding increase rainfall? Carry out a test.

$Define: Y_1 -$ rainfall for unseeded days, $Y_2 -$ rainfall for seeded days

$Z_1 = \ln(Y_1) - natural$ logged rainfall for unseeded days,

$Z_2 = \ln(Y_2) - natural$ logged rainfall for seeded days

$\mu_{Z_1} -$ average of the natural logged rainfall for unseeded days

$\mu_{Z_2} -$ average of the natural logged rainfall for seeded days

$Parameter: \mu_{Z_1} - \mu_{Z_2}$

$H_0: \mu_{Z_1} - \mu_{Z_2} = 0$

$H_0: \mu_{Z_1} - \mu_{Z_2} < 0$

From SPSS,

$$t_0^* = \frac{\bar{z}_1 - \bar{z}_2}{S.E.(\bar{Z}_1 - \bar{Z}_2)} = \frac{-1.1438}{0.44953} = -2.544$$

$$p - value = P(t_{50} < -2.544) \approx \frac{0.014}{2} = 0.007$$

(From t-table: $p - value = P(t_{50} < -2.544) \in (0.005, 0.01)$)

With a p-value ≈ 0.007, there is very convincing evidence that the average logged rainfall for seeded days is higher than the average logged rainfall for unseeded days. This also implies convincing evidence that $Median(Y_1) / Median(Y_2) < 1.$

c) Calculate a 95% confidence interval for the effect of cloud seeding.

A 95% $C.I.$ for $\mu_{Z_1} - \mu_{Z_2}$ is:

$Estimate \pm C.V. \times S.E.(estimate)$

$\Rightarrow \bar{z}_1 - \bar{z}_2 \pm t^*_{50,0.975} \times S.E.(\bar{Z}_1 - \bar{Z}_2)$

$\Rightarrow -1.1438 \pm 2.009(0.44953)$

$\Rightarrow -1.1438 \pm 0.90311$

$\Rightarrow (-2.047, -0.241)$

It is estimated with 95% confidence that the average logged rainfall for unseeded days is between 0.241 to 2.047 less than the average logged rainfall for seeded days.

On the original scale: A 95% $C.I.$ for $Median(Y_1) / Median(Y_2)$ is:

$\Rightarrow (e^{-2.047}, e^{-0.241}) \Rightarrow (0.13, 0.79)$

It is estimated with 95% confidence that the median rainfall for unseeded days is between 0.13 to 0.79 times ("13% to 79% of" or "21% to 87% less than") the median rainfall for seeded days.

Or, a 95% $C.I.$ for $Median(Y_2) / Median(Y_1)$ is:

$\Rightarrow (e^{0.241}, e^{2.047}) \Rightarrow (1.27, 7.75)$

It is estimated with 95% confidence that the median rainfall for seeded days is between 1.27 to 7.75 times ("127% to 775% of" or "27% to 675% more than) the median rainfall for unseeded days.