MATU9D2 : PRACTICAL STATISTICS
WEEKLY ASSIGNMENT 6 : SOLUTIONS

ALL TESTS WILL BE PERFORMED AT 5% SIGNIFICANCE LEVEL IN THE ASSIGNMENT

1. (i) There is one sample of quantitative data, estimating the standard deviation from the sample and the first question is about the mean so the appropriate test is a One Sample t test. This test assumes the data is a representative random sample from the population and is Normally distributed so this must be validated.

Subjectively, it looks as if the mean is slightly less than 105 (104.13 but with a relatively large standard deviation of 9.4) but not by much so unlikely to be statistically significant. The histogram (**Figure 1**) does not look very symmetric but with a small sample it is difficult to get a sensible scale. The Normal Probability Plot (**Figure 2**) looks fairly linear so we can assume the data is Normally distributed.

$$H_0: \mu = 105 \quad where\ \mu\ is\ the\ population\ mean\ radon\ level$$

$$H_1: \mu \neq 105$$

Table 1 shows that the Observed Test Statistics is -0.32 and p=0.755. Since $p>0.05$ we cannot reject H_o in favour of H_1 at 5% level i.e. conclude insufficient evidence that the mean radon level differs significantly from 105 picocuries per litre.

(ii) **Figure 1** and **Table 1** show that the 95% confidence interval for the population mean radon level is from 98.16 picocuries/litre and 110.10 picocuries/litre. i.e. 95% confident that true population mean radon level lies within this range. In particular, 105 is within this interval so we cannot reject H_o in favour of H_1 at 5% level i.e. conclude insufficient evidence that the mean radon level differs significantly from 105 picocuries per litre.

Figure 1

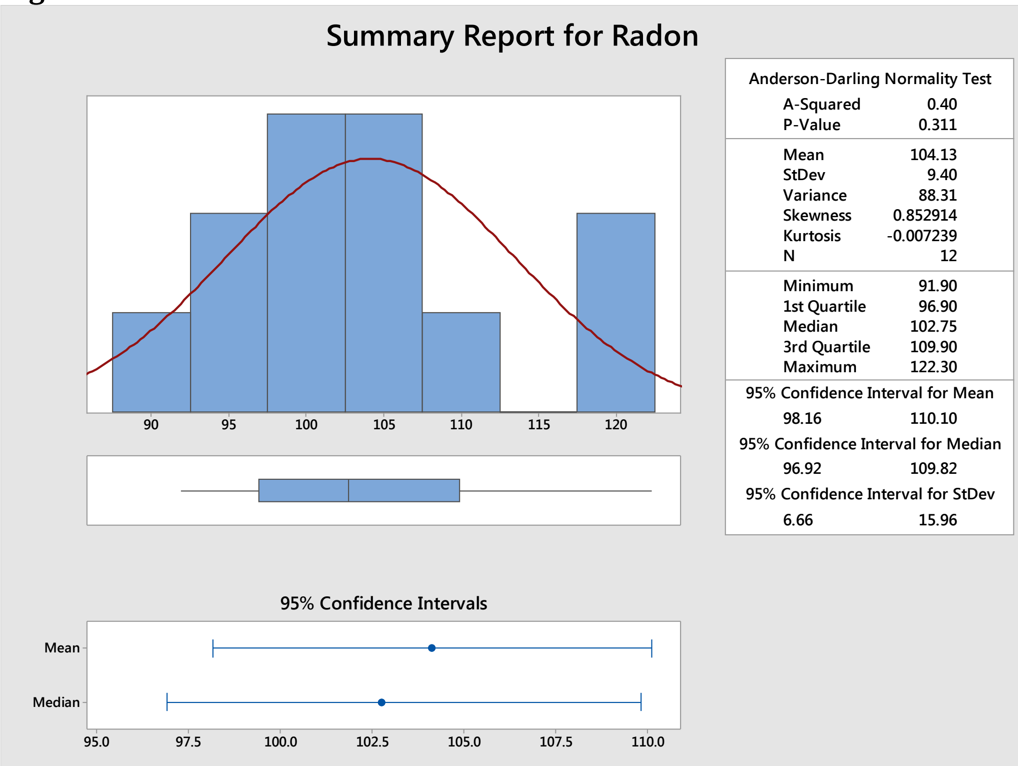

Figure 2.

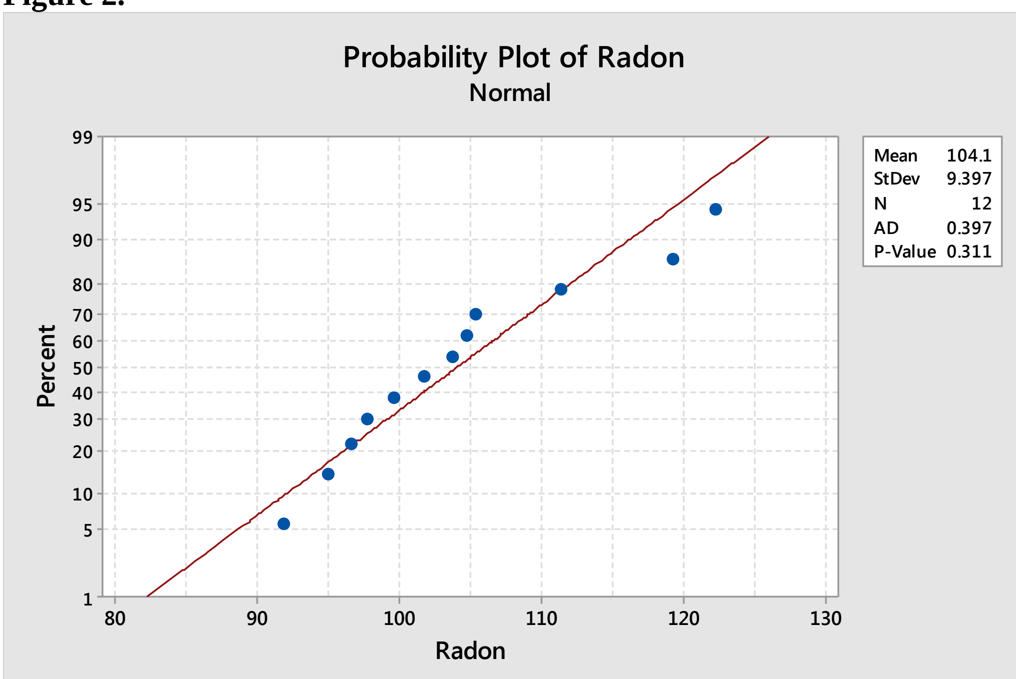

Table 1 : One-Sample T: Radon

Test of μ = 105 vs ≠ 105

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Radon	12	104.13	9.40	2.71	**(98.16, 110.10)**	**-0.32**	**0.755**

2. This question involves (i) Two Independent Samples; (ii) Quantitative Data and (iii) the question is about the difference in support (i.e. difference in the mean levels in villages and towns).

The appropriate formal technique would be an Unpaired t-test. There are two versions : one assumes equal variance and one does not so we must check whether the variances can be assumed to be equal (F test would be appropriate). We will do the F test in Practical 7 – so for this assignment use assumes equal variance.

Both the unpaired t-tests and the F test assume that both sets of data are Normally distributed and so this must be validated and they also assume that both sets are representative independent random samples from the two populations.

Informally, it looks as if neither set of data is very symmetric (**Figures 3 and 4**) but it does look as if there is a possible difference in the mean level of support (Villages 21.72% $\pm$ 12.49% and Towns 30.91% $\pm$ 9.2%). The boxplot in **Figure 5** confirms this subjective impression with the Towns having a greater but less variable level of support.

For this week, the appropriate Unpaired t-test assumes equal variance and is two-tailed since the original question asked about a difference. (Next time we will check – using an F test!!)

$H_0 \quad : \quad \mu_1 = \mu_2$ where μ_1, μ_2 are the population mean level of support in villages and towns respectively.

$H_1 \quad : \quad \mu_1 \neq \mu_2$

<u>Table 3</u> shows the results of this test. Pooled variance is 11.38%; Observed Test Statistic t=-2.11, df =27 and p=0.44. In conclusion, $p<0.05$ so we can reject H_o in favour of H_1 at 5% level i.e sufficient evidence that the mean level of support in villages and towns are significantly different. We can, therefore, conclude that there is a statistically significant difference and that the level of support is towns is between 0.25% and 18.13% greater than the level of support in villages on average with 95% confidence.

Figure 3

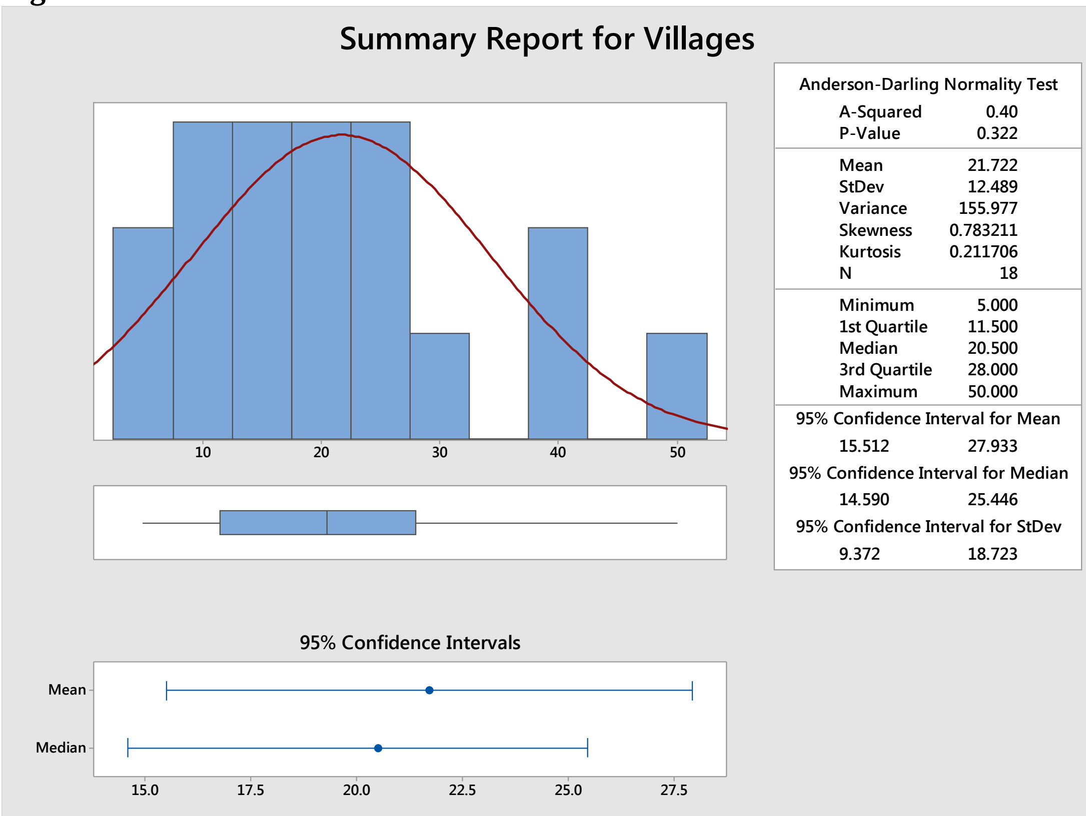

Figure 4

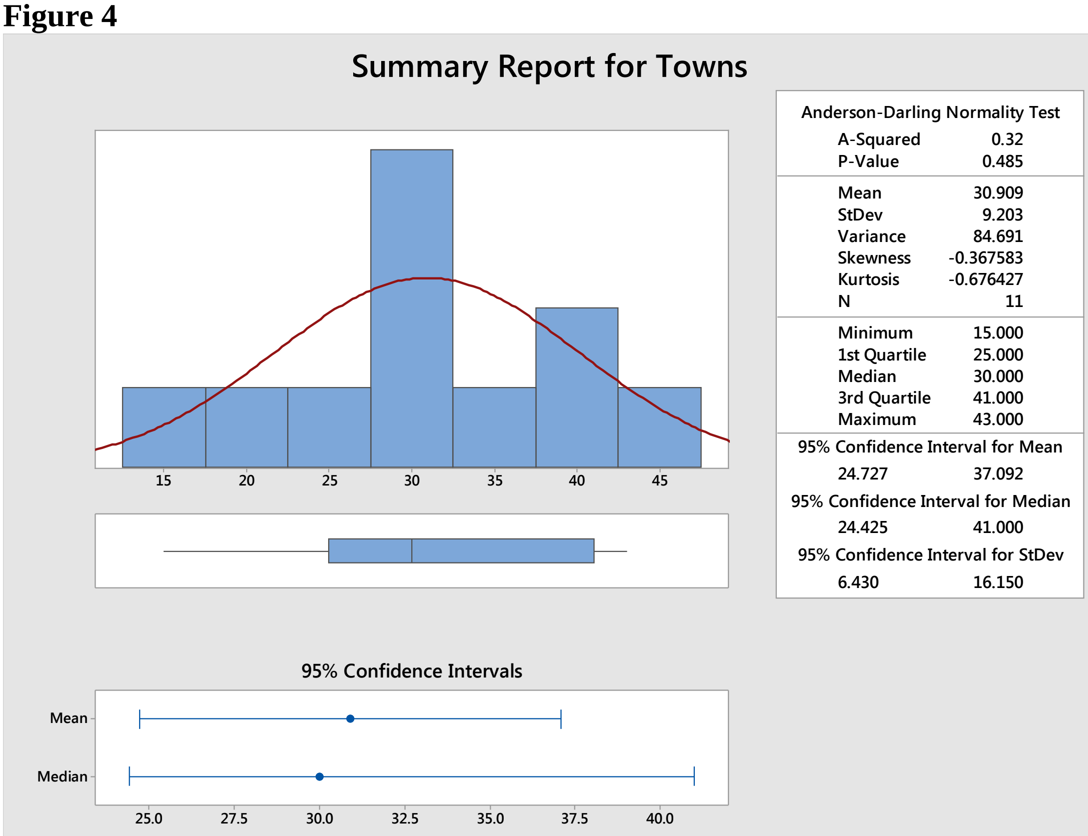

Figure 5.

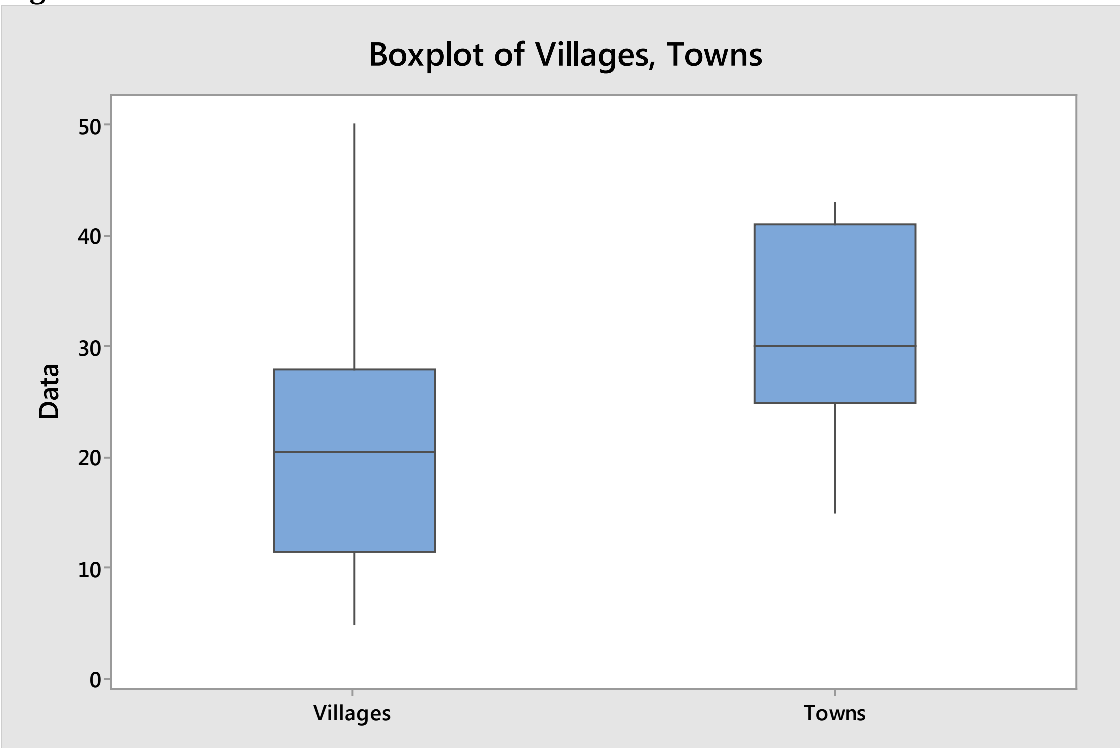

Figure 6.

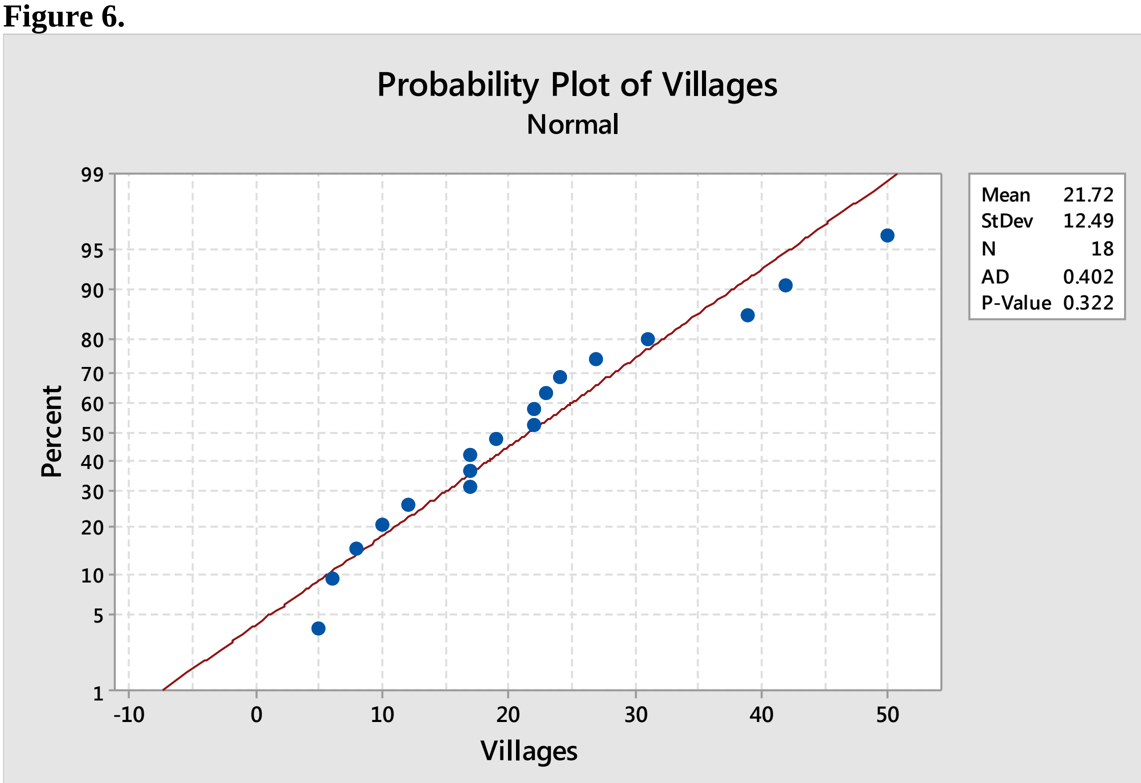

Figure 7.

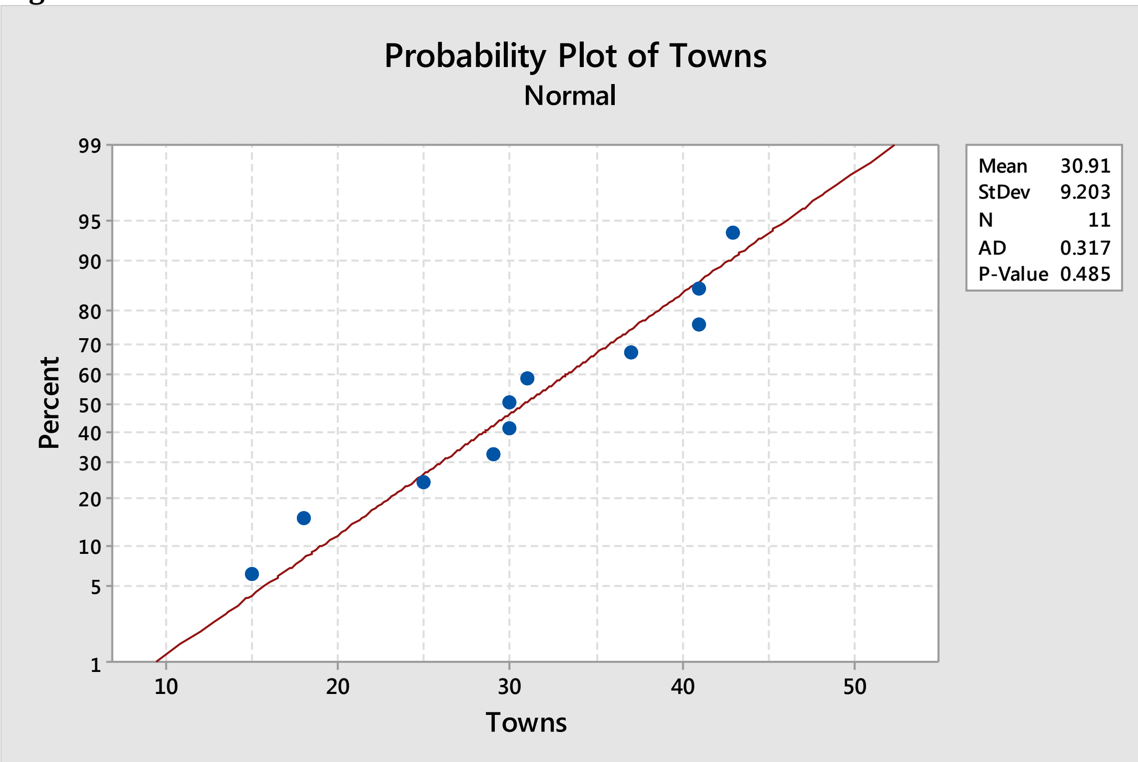

Table 3. Two-Sample T-Test and CI: Villages, Towns

```
Two-sample T for Villages vs Towns

           N   Mean  StDev  SE Mean
Villages  18   21.7   12.5      2.9
Towns     11  30.91   9.20      2.8


Difference = μ (Villages) - μ (Towns)
Estimate for difference:  -9.19
95% CI for difference:  (-18.13, -0.25)
T-Test of difference = 0 (vs ≠): T-Value = -2.11  P-Value = 0.044  DF = 27
Both use Pooled StDev = 11.3831
```

3. This question involves (i) Two Paired Samples i.e. the same person observed before and after training; (ii) Quantitative Data i.e. score on a test and (iii) the question is about an improvement in the scores. This suggests that the appropriate test would be a one tailed Paired t test. This assumes that the differences in the test scores is Normally distributed and it is one tailed because the question asks about an improvement.

Subjectively, **<u>Figure 8</u>** shows that the Minimum Improvement is 0 (After-Before) and the mean improvement is 28.5 with a standard deviation of 15.64 and so there seems to be an improvement especially as nobody scores less after the training. The histogram does not look very Normally distributed or even symmetric but **Figure 9** is linear so we can assume that the differences are Normally distributed so we can perform a one tailed Paired t test.

$H_0 : \mu_d = 0$ where μ_d is the population mean difference in Scores (After-Before)

$H_1 : \mu_d > 0$

<u>Table 4</u> shows the results for the one tailed Paired t test with the Observed Test Statistic t = 5.76 and the p =0.000 i.e. $p<0.001$. So we can reject H_o in favour of H_1 at 5% level i.e. sufficient evidence that the mean improvement in the verbal test significantly different from zero. We can, therefore, conclude that there is a statistically significant improvement and that the mean improvement is at least 19.43 on average with 95% confidence.

Figure 8.

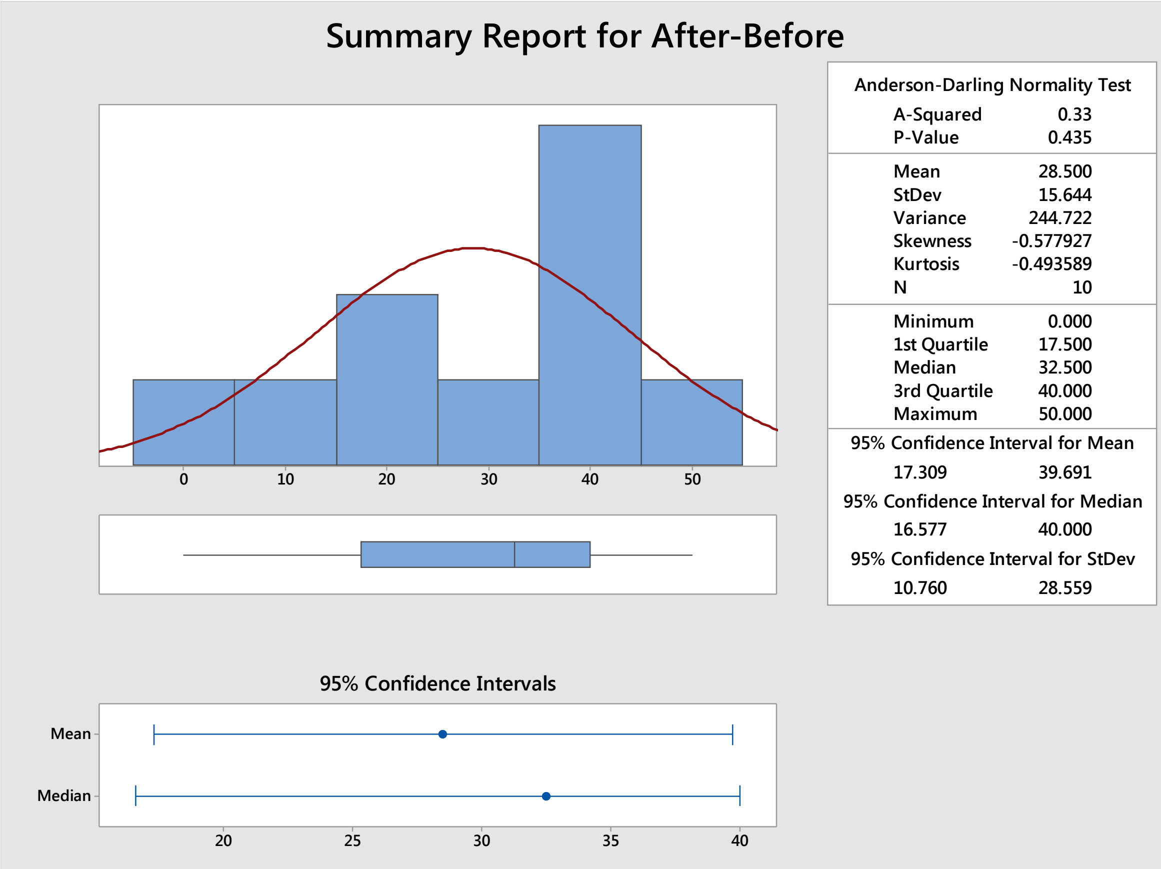

Figure 9.

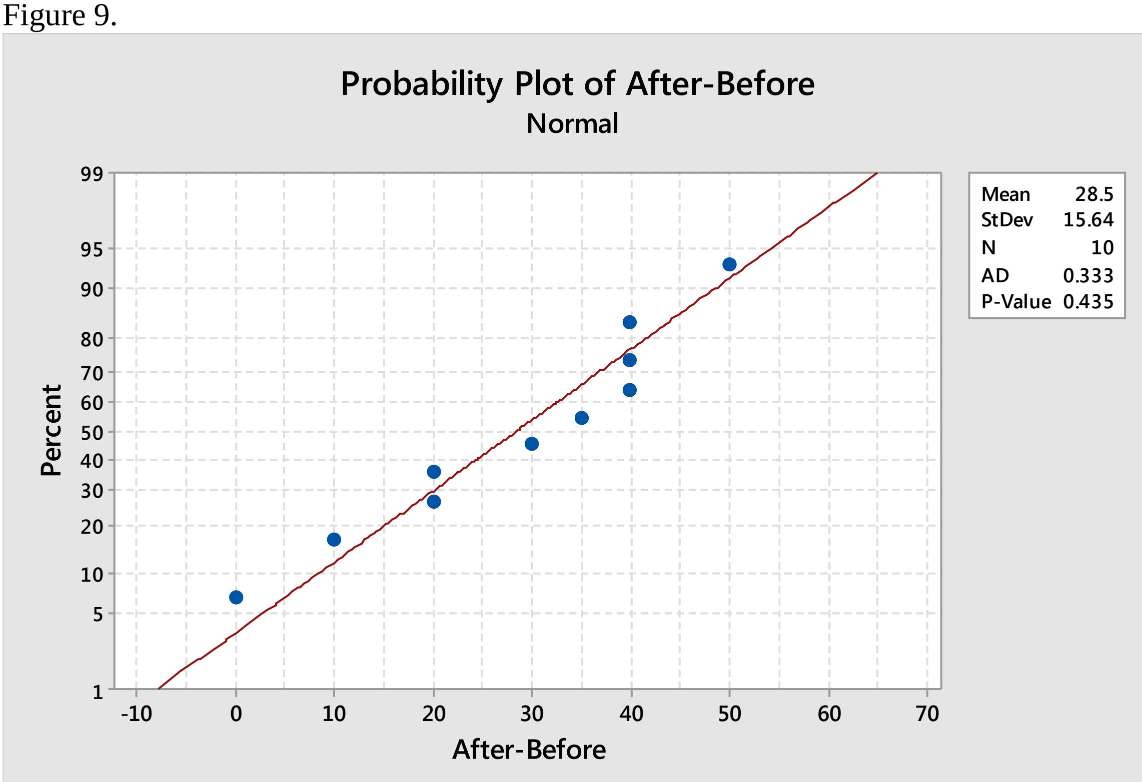

Table 4. Paired T-Test and CI: After, Before

Paired T for After - Before

	N	Mean	StDev	SE Mean
After	10	525.5	87.8	27.8
Before	10	497.0	85.1	26.9
Difference	10	28.50	15.64	4.95

95% lower bound for mean difference: 19.43
T-Test of mean difference = 0 (vs > 0): T-Value = 5.76 P-Value = 0.000