

**Ethics Pledge**

**Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.**

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Feb 11th 2019

Please note that assignments in this class may be submitted to www.turnitin.com, a web- based anti-plagiarism system, for an evaluation of their originality.

1. **In an NYT/CBS poll, 56% of 2,000 randomly selected voters in New York City said they would vote for the incumbent in a certain two-person race.**

**(a)  Calculate a 95% confidence interval for the population proportion π. (Check whether the necessary ‘assumption’ is met.)**

**Answer:**

The sample is big enough, it should be normal distribution.

p = 0.56, n = 2000, α = 0.05

n \* p = 1120 > 5 and n \* (1 – p) = 880 > 5

Therefore, we can calculate the confidence interval by using this formula:

= 0.56 ± 1.96 \* 0.011 = 0.56 ± 0.02156

The confidence interval for the population proportion 0.56 are from 0.5384 to 0.5816.

**(b)  Carefully interpret the meaning of the confidence interval obtained in (a).**

**Answer:**

We are 95% confident that the true percentage of voters will vote for the incumbent in a certain two-person race is between 53.84% and 58.16%.

Although the interval from 0.5384 to 0.5816 may or may not contain the true proportion, 95% of intervals formed from samples of size 2000 in this manner will contain the true proportion.

**(c)  What is the margin of error?**

**Answer:**

Margin of error: = 1.96 \* 0.011 = 0.02156

**(d)  Assume we had no prior knowledge about the true proportion π. We want to construct a 95% confidence interval for π with margin of error 2%. How large a sample is needed? How does the sample size change if we want to be 99% confident?**

**Answer:**

α = 0.05, e = 0.02

Since we don’t have π value, conservatively assume π is 0.5.

Transform equation to

Therefore, = 2401

When confidence interval is 95%, the sample size should be 2401.

If the confidence interval changes to 99%, α = 2.58

Therefore, = 4161

When confidence interval is 99%, the sample size should be 4161.

1. **Many companies are experimenting with “flex-time,” allowing employees to choose their schedules within broad limits set by management. Among other things, flex-time is supposed to reduce absenteeism. One firm knows that in the past few years, employees have averaged 6.3 days off from work (apart from vacations). This year, the firm introduces flex-time. Management chooses a simple random sample of 100 employees to follow in detail, and at the end of the year, these employees average 5.5 days off from work, and the sample standard deviation (SD) is 2.9 days.** 
   1. **(a)  Did absenteeism really go down, or is this just chance variation? Formulate the null and alternative hypotheses and carry out the testing.**

**Answer:**

H0 : μ >= 6.3, H1 : μ < 6.3

n = 100, σ = 2.9, = 5.5, and suppose α = 0.05 (95% confidence interval)

Since σ is known, we use Z-test.

= -2.76

P = 0.0029

Since p-value = 0.0029 < α = 0.05, reject H0.

Therefore, when the confidence interval is 95%, there is sufficient evidence to conclude the absenteeism really went down.

* 1. **(b)  Repeat the above for a sample average of 5.9 days and an SD of 2.9 days.**

**Answer:**

H0 : μ >= 6.3, H1 : μ < 6.3

n = 100, σ = 2.9, = 5.9, and suppose α = 0.05 (95% confidence interval)

Since σ is known, we use Z-test.

= -1.38

P = 0.0838

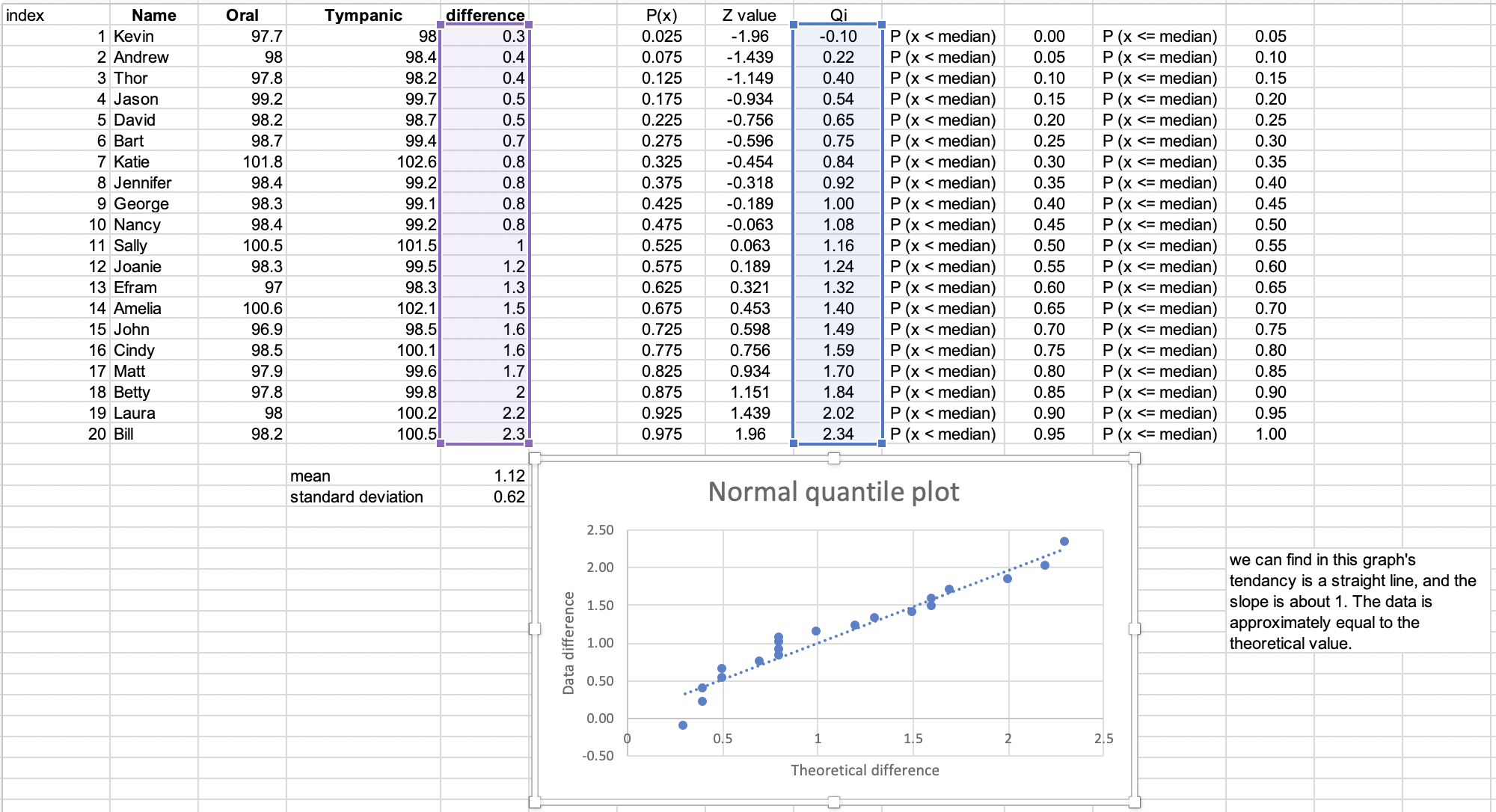
Since p-value = 0.0838 > α = 0.05, do not reject H0.

Therefore, when the confidence interval is 95%, we cannot prove the absenteeism really went down, this should be just chance variation.

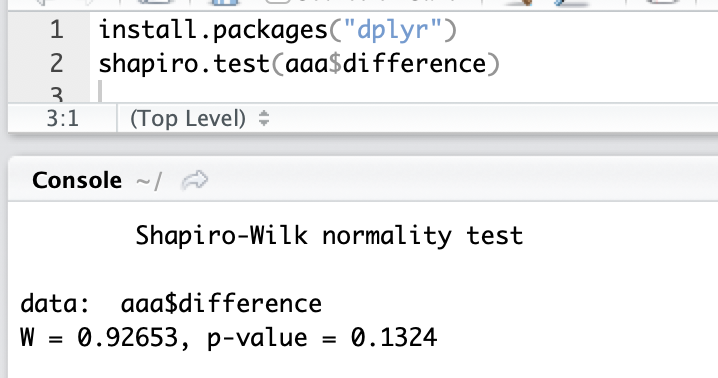
1. **Open up the data file Therm.dat (or Therm.xls) uploaded in Canvas. Carry out the hypothesis testing to check whether the temperature measurement difference > 0 at significance level 0.01. (Don’t forget to check whether an ‘underlying assumption’ holds or not, e.g., normal probability plot and Goodness- of-Fit Test for normality checking.)**

**Answer:**

Since the data is small, firstly, we use normal probability plot to see if it’s normal distribution or not.



And for Shapiro-Wilk Goodness-of-Fit Test:



When n = 20, and significant level 0.01, the W value we calculate is bigger than the critical threshold, so the assumption of a normal distribution will NOT be rejected.

H0 : μ <= 0, H1 : μ > 0

n = 20, S = 0.62, = 1.12, α = 0.01 (99% confidence interval)

Since the data is small, we use t-test.

= 8.081

Get the p value from t score calculator which from following website: (<https://www.socscistatistics.com/pvalues/tdistribution.aspx>)

The p value is < 0.00001 when d.f = 19 and t score = 8.08

Since p-value < α = 0.01, we can reject H0.

Therefore, when the confidence interval is 99%, we can prove the temperature measurement difference is greater than 0.

1. **(True, False) To make a t-test with 5 measurements, use Student’s t-distribution with 5 degrees of freedom.**

**Technical notes. The notion of “degrees of freedom” actually has to do with the notion of dimensions in linear algebra context (e.g., projection, subspace). Here’s the idea behind the phrase. The standard error (SE) for the average depends on the standard deviation (SD) of the measurements, and that in turn depends on the deviations (xi − x ̄, i = 1, . . . , n) from the average. But the sum of the deviations has to be 0 (by the definition of mean x ̄ = (x1 + · · · + xn)/n), so they cannot all vary freely. The constraint that the sum equals 0 eliminates one degree of freedom. For example, with 5 measurements, the sum of the 5 deviations is 0. If you know 4 of them, you can compute the 5th–so there are only 4 degrees of freedom.**

**Answer:**

False, for a t-test with 5 measurements, the degrees of freedom should only be n – 1 = 4.