Dear all, this is a close book exam. Please turn in by 9:50am. Calculator (with no internet connection) is allowed. Please scan your exam in one PDF file and submit to the class Brightspace site. Academic dishonesty will result in severe consequence.

1. To study the effectiveness of wall insulation in saving energy for home heating, the energy consumption (in MWh) for 5 houses in Bristol, England, was recorded for two winters; the first winter was before insulation and the second winter was after insulation:

House	1	2	3	4	5
Before	12.1	10.6	13.4	13.8	15.5
After	12.0	11.0	14.1	11.2	15.3

   (a) Please provide a 95% confidence interval for the difference between the mean energy consumption before and after the wall insulation is installed. What assumptions are necessary for your inference?
   (b) Can you conclude that there is a difference in mean energy consumption before and after the wall insulation is installed at the significance level 0.05? What assumption(s) is (are) necessary for your inference?

Solution: This is inference on two population means, paired samples – which will reduce to the inference of one population mean based on the paired differences. Given that the sample size is small (n < 30), we will need the normal distribution assumption to use the exact t distribution for the pivotal quantity (and test statistic) in this exam situation.

House	1	2	3	4	5
Before	12.1	10.6	13.4	13.8	15.5
After	12.0	11.0	14.1	11.2	15.3
d	0.1	-0.4	**- 0.7**	2.6	0.2

(a). $\bar{d} = 0.36$, $s_d = 1.30$

The 95% C.I for μ_d is:

$$\bar{d} \pm t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}} = 0.36 \pm 2.776\frac{1.30}{\sqrt{5}} = [-1.25, 1.97]$$

(b). $H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$

(1) $t_0 = \dfrac{\bar{d} - 0}{s_d/\sqrt{n}} = \dfrac{0.36 - 0}{1.30/\sqrt{5}} \approx 0.619$

$t_{n-1,\alpha/2} = t_{4,0.025} = 2.776$

Since $|t_0| \approx 0.619$ is smaller than $t_{4,0.025} = 2.776$, we cannot reject H_0.

Assumptions for (a) and (b): the paired differences follow a normal distribution.

2. People at high risk of sudden cardiac death can be identified using the change in a signal averaged electrocardiogram before and after prescribed activities. The current method is about 80% accurate. The method was modified, hoping to improve its accuracy. The new method is tested on 50 people and gave correct results on 46 patients.

(a) Is this convincing evidence that the new method is more accurate? Please test at α =.05.
(b) Please construct a 95% confidence interval for the accuracy of the new method.
(c) (extra credit) How large a random sample should we select to guarantee the length of the 95% confidence interval to be no more than 0.02? Please consider the two cases: (i) we have an estimate of the accuracy based on the sample of 50 people shown in the problem; (ii) we do not have an estimate of the accuracy to be estimated.

Solution: This is inference on one population proportion. Here our sample size is n = 50, and we have 46 'successes' and 4 'failures'. In fact, we can only use the large sample z-pivotal quantity or test statistic based on the Central Limit Theorem if we use the threshold of at least 3 'successes' and at least 3 'failures' for a large sample test or confidence interval --- and not the more conservative threshold of at least 5 'successes' and 'failures.

(a) $H_0: p = 0.8$ versus $H_0: p > 0.8$

Using $n = 50, \hat{p} = \frac{46}{50} = 0.92$, the test statistic is:

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.92 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{50}}} = 2.121$$

Since $z_0 = 2.121 > z_{0.05} = 1.645$, we can reject the null hypothesis in favor of the alternative hypothesis, that is, this is convincing evidence that the new method is more accurate than the current method based on the given sample, at the significance level of $\alpha = 0.05$.

(b) The 95% C.I for p is:

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.92 \pm 1.96\sqrt{\frac{0.92(1-0.92)}{50}} = [0.845, 0.995]$$

(c) $\mathrm{L} = 0.02$

(i) if $\hat{p}$ is known,

$$\mathrm{n} = \frac{4\left(\mathrm{Z}_{\alpha/2}\right)^2}{\mathrm{L}^2}\hat{\mathrm{p}}(1-\hat{\mathrm{p}}) = 4\left(\frac{1.96}{0.02}\right)^2 * 0.92 * 0.08 \approx 2828$$

(ii) if $\hat{p}$ is unknown, we shall plug in $\hat{p}$ = ½ for the maximum possible sample size:

$$\mathrm{n} = \frac{4\left(\mathrm{Z}_{\alpha/2}\right)^2}{\mathrm{L}^2}\hat{\mathrm{p}}(1-\hat{\mathrm{p}}) \le \frac{4\left(\mathrm{Z}_{\frac{\alpha}{2}}\right)^2}{\mathrm{L}^2} * \frac{1}{4} = \left(\frac{1.96}{0.02}\right)^2 = \mathbf{9604}$$

3. In response to student complaints and financial considerations, a high school decides to close its kitchen and contract a food service, the "Tigress Express", to provide school lunches. The previous year, when food was prepared in the high school kitchen, about **60%** of the students purchased lunch on a daily basis. The daily proportions of students using the food service from **10** randomly selected days during the fourth month of the contract are (in %): **68, 61, 65, 74, 68, 66, 70, 63, 74, 65**.

(a) According to the data, can you conclude, at the significance level of **0.05**, that compared to the previous year, there is an increase in the average daily proportion of students purchasing lunches provided by the food service, the "Tigress Express"?

(b) Please construct a 95% confidence interval for the average daily proportion of students purchasing lunches provided by the "Tigress Express".

Solution: This is inference on one population mean. Given that the sample size is small (n < 30), we will need the normal distribution assumption to use the exact t distribution for the pivotal quantity (and test statistic) in this exam situation.

(a) $\begin{cases} H_0: \mu = 60 \\ H_1: \mu > 60 \end{cases}$

$$n = 10, \bar{X} = 67.4, s^2 = 18.71, s = 4.33.$$
$$t_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = 5.40 > t_{9,0.05} = 1.833,$$

so we can conclude that there is an increase in the average daily proportion of students purchasing lunches provided by the food service, the "Tigress Express" at the significance level of 0.05.

(b) $n = 10, \bar{X} = 67.4, s^2 = 18.71, s = 4.33, t_{9,0.025} = 2.262$

The 95% C.I for μ is:

$$\bar{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = 67.4 \pm 2.262\frac{4.33}{\sqrt{10}} = [64.30, 70.50]$$