Dear all, this homework is due before class on Tuesday, February 13, 2024. Please submit it to SBU Brightspace. Quiz 2 of similar content (please review the entire lecture notes 4 and 5), will be given the week of February 13 as well. It will be a close book exam. (Most of our quizzes will be on R programming and will be open book exams – only the first few quizzes will be close book exams. At least the 3 lowest 3 quiz scores will be dropped for each student when computing the final quiz scores – so no worries if you did not do well in a few quizzes.)

1. To determine whether glaucoma affects the corneal thickness, measurements were made in 8 people affected by glaucoma in one eye but not in the other. The corneal thickness data (in microns) were as follows:

Person	1	2	3	4	5	6	7	8
Eye affected	488	478	480	426	440	410	458	460
Eye not affected	484	478	492	444	436	398	464	476

(a) According to the data, can you conclude, at the significance level of 0.10, that the corneal thickness is not equal for affected versus unaffected eyes?

(b) Calculate a 90% confidence interval for the mean difference in thickness.

Solution: This problem involves a paired sample – and once we take the paired difference, it will reduce to the inference on one population mean, here being the mean difference of corneal thickness for those affected versus those not affected.

The paired difference is calculated as follows:

Person	1	2	3	4	5	6	7	8
Eye affected	488	478	480	426	440	410	458	460
Eye not affected	484	478	492	444	436	398	464	476
Difference (d)	4	0	-12	-18	4	12	-6	-16

(a) Using $\bar{d} = -4$ and $s_d = 10.744$, the test statistic is

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{-4 - 0}{10.744 / \sqrt{8}} = -1.053$$

Since $|t| \prec t_{8-1,0.05} = 1.895$, we can NOT reject H_0 at $\alpha = 0.10$. That is, we do NOT have enough evidence to support the claim that the average corneal thicknesses are affected by glaucoma.

(b) A 90% CI for $\mu_d = \mu_1 - \mu_2$ is given by

$$\bar{d} \pm t_{n-1,\alpha/2} \cdot s_d / \sqrt{n} = -4 \pm 1.895 \times 10.744 / \sqrt{8}$$

That is, $[-11.198,\ 3.198]$

2. Thanksgiving was coming up and Harvey's Turkey Farm was doing a land-office business. Harvey sold 100 gobblers to Nedicks for their famous Turkey-dogs. Nedicks found that 90 of Harvey's turkeys were in reality peacocks.
(a) Estimate the proportion of peacocks at Harvey's Turkey Farm and find a 95% confidence interval for the true proportion of turkeys that Harvey owns.

Solution: Let p be the proportion of peacocks and q be the proportion of turkeys.

$\hat{p} = \frac{90}{100} = 0.9$, $\hat{q} = 1 - \hat{p} = 0.1$ and $n = 100$.

The 95% C.I on q is

$$\hat{q} \pm z_{\alpha/2}\sqrt{\frac{\hat{q}(1-\hat{q})}{n}} = 0.1 \pm 1.96\sqrt{\frac{0.1 \cdot 0.9}{100}} = [0.0412, 0.1588].$$

(b) (extra credit) How large a random sample should we select from Harvey's Farm to guarantee the length of the 95% confidence interval to be no more than 0.06? (Note: please first derive the general formula for sample size calculation based on the length of the CI for inference on one population proportion, large sample situation. Please give the formula for the two cases: (i) we have an estimate of the proportion and (ii) we do not have an estimate of the proportion to be estimated. (iii) Finally, please plug in the numerical values and obtain the sample size for this particular problem.

Solution: $L = 0.06$

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

$$n = \frac{4\left(Z_{\alpha/2}\right)^2}{L^2}\hat{q}(1-\hat{q}) = 4\left(\frac{1.96}{0.06}\right)^2 * 01 * 0.9 \approx 385$$

if $\hat{q}$ is unknown,

$$\hat{q}(1-\hat{q}) = -\left(\hat{q} - \frac{1}{2}\right)^2 + \frac{1}{4} \le \frac{1}{4}$$

$$n = \frac{4\left(Z_{\alpha/2}\right)^2}{L^2}\hat{q}(1-\hat{q}) \le \frac{4\left(Z_{\alpha/2}\right)^2}{L^2} * \frac{1}{4} = \left(\frac{1.96}{0.06}\right)^2 \approx 1068$$

3. Treatment of Kidney Cancer

Historically, one in five kidney cancer patients (i.e. 20%) survive 5 years past diagnosis An oncologist using an experimental therapy treats n = 40 kidney cancer patients and 16 of them survive at least 5 years. Is there evidence that patients receiving the experimental therapy have a higher 5-year survival rate? Please test at the significance level of $\alpha = 0.05$.

Solution 1. (Large sample approximate test, required for all)

Let p = the proportion of kidney cancer patients receiving the experimental therapy that survive at least 5 years.

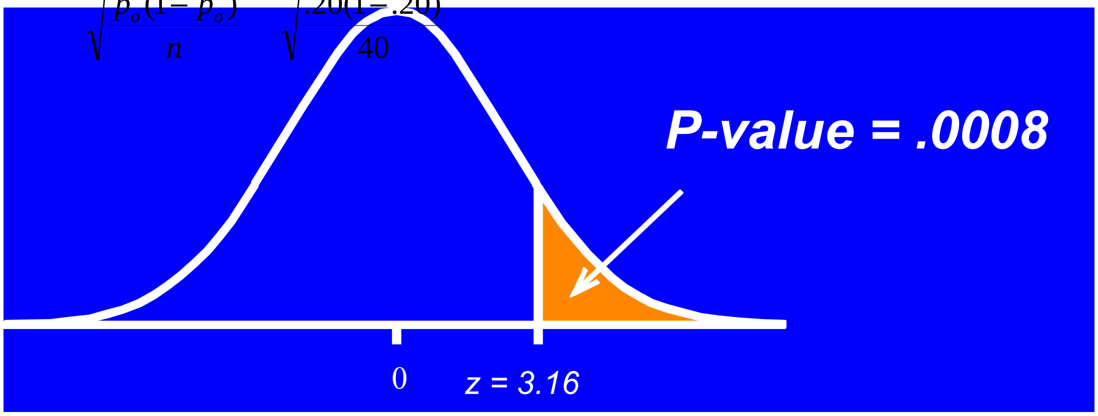
$H_0: p = 0.2$ versus $H_a: p > 0.2$

Since np = (40)(.20) = 8 > 5 and *n(1-p) = (40)(.80) = 32 > 5*

We conclude that this sample is a large sample

$$\hat{p} = \frac{16}{40} = .40 \text{ or a 40\% 5-yr. survival rate}$$

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{.40 - .20}{\sqrt{\frac{.20(1-.20)}{40}}} = 3.16$$



Since p = .0008 < $\alpha = 0.05$, *we conclude that the 5-year survival rate for kidney cancer patients undergoing the experimental therapy is greater than the current 5-yr. survival rate of 20%.*

Solution 2. (Exact test; required only for AMS graduate students and Data Science PhD students)

In our example we had n = 40 patients and if we assume the experimental therapy is no better than current treatments then probability of 5-year survival is p = .20.

Thus the number of patients in our study surviving at least 5 years has a binomial distribution, i.e. X ~ BIN(40,0.2).

The exact ***p-value = P(X ≥ 16) = P(X = 16) + P(X = 17) + … + P(X = 40) = .002936***

Since p = .0002396 < $\alpha = 0.05$, *we conclude that the 5-year survival rate for kidney cancer patients undergoing the experimental therapy is greater than the current 5-yr. survival rate of 20%.*