Problem Set 6 **Prof. Oke**

CEE 260/MIE 273: Probability & Statistics in Civil Engineering *10.21.2025*

Due Tuesday, October 28, 2025 at 1:00 PM as PDF uploaded on Canvas. *If it helps and if possible, you can write your responses directly on this document and upload it instead.* ***Show as much work as possible in order to get FULL credit.*** *There are 6 problems with a total of 38 points available.* ***Important:*** *If you use Python for any probabilistic or statistical computations, briefly write/include the statements you used to arrive at your answers. If instead you use probability tables, note this in the respective solution, as well.*

Problem 1 *(3 points)*

Respond "T" (*True*) or "F" (*False*) to the following statements. Use the boxes provided. Each response is worth 1 point.

(i) $\boxed{T}$ The result that a binomial distribution with n large enough (and p reasonably less than 1 or greater than 0) can be approximated by a normal distribution with parameters $\mu = np$ and $\sigma^2 = np(1-p) = npq$ can be derived via the Central Limit Theorem.

(ii) $\boxed{F}$ The household income of the full population of a certain country follows the lognormal distribution. The mean household income from a large random sample ($n = 10,000$) will also be lognormally distributed.

(iii) $\boxed{T}$ The standard error of the mean decreases as the sample size increases.

Problem 2 *(4 points)*

Choose the option that best fills in the blank.

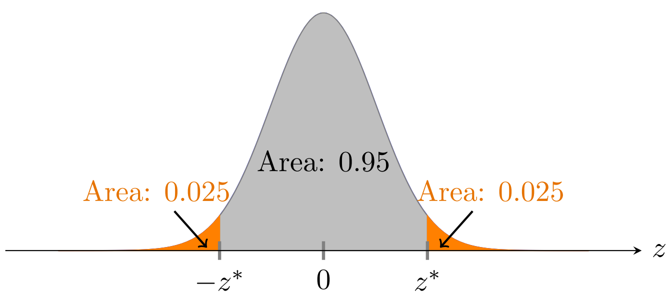
[1] **(a)** If an estimator is unbiased, then ____.

- **(i)** the estimator is equal to the true value.
- **(ii)** the estimator is usually close to the true value.
- **(iii)** **the mean of the estimator is equal to the true value.**
- **(iv)** the mean of the estimator is usually close to the true value.

[1] **(b)** The property of an estimator associated with the bias converging to 0 as the sample size increases (or tends to infinity) is called ________________.

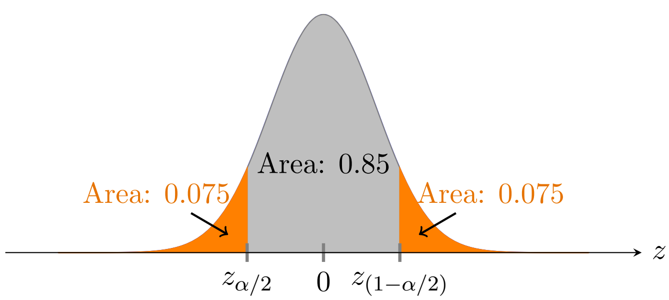
- **(i)** **consistency**
- **(ii)** unbiasedness
- **(iii)** efficiency
- **(iv)** sufficiency

[1] **(c)** The figure below depicts the PDF of a standard normal distribution. What is the value of z^* in the figure?



- **(i)** 0
- **(ii)** 1.65
- **(iii)** **1.96**
- **(iv)** 2.58

[1] **(d)** The figure below depicts the PDF of a standard normal distribution. What is the value of $z_{\alpha/2}$ in the figure?



- **(i)** 0
- **(ii)** -1.04
- **(iii)** -1.28
- **(iv)** $\mathbf{-1.44}$

Problem 3: Standard error of a proportion *(5 points)*

An article reports that when each football helmet in a random sample of 37 suspension-type helmets was subjected to a certain impact test, 24 showed damage. Let p denote the proportion of all helmets of this type that would show damage when tested in the prescribed manner.

(a) Estimate the value of the proportion p. *[1]*

$$p = \frac{24}{37} \approx \boxed{0.65}$$

(b) Compute the standard error of the estimate $\hat{p}$. *[2]*

$$\begin{aligned} SE(\hat{p}) &= \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.65(0.35)}{37}} \\ &= \boxed{0.078} \end{aligned}$$

(c) Using the 68–95–99.7 (empirical) rule, estimate the interval that would contain the sample proportion $\hat{p}$ in 95% of all possible samples of a similar size. *[2]*

According to the 68–95–99.7 empirical rule, approximately 95% of all sample proportions will lie within ± 2 standard errors of the population proportion p.

$$\hat{p} \pm 2\,SE(\hat{p})$$

Substitute the known values:

$$\hat{p} = 0.65, \quad SE(\hat{p}) = 0.078$$

$$0.65 \pm 2(0.078) = 0.65 \pm 0.156$$

$$\boxed{(0.494, 0.806)}$$

Problem 4: Point estimate and sampling variability *(12 points)*

(Adapted from Ex 5.3 in OS) As part of a quality control process for computer chips, an engineer at a factory randomly samples 212 chips during a week of production to test the current rate of chips with severe defects. She finds that 27 of the chips are defective. Match the labels corresponding to the rows in the table below that correctly answers questions **(a)** through **(d)**. Then work out the questions that follow.

Label	Answer
(i)	All computer chips produced during the week
(ii)	Sample proportion of defective chips, $\hat{p}$
(iii)	Standard error of the sample proportion, $SE_{\hat{p}}$
(iv)	Population proportion of defective chips

[1] **(a)** What population is under consideration in the data set? $\boxed{i}$

[1] **(b)** What parameter is being estimated? $\boxed{iv}$

[1] **(c)** What is the name of point estimate of the parameter? $\boxed{ii}$

[1] **(d)** What is the name of the statistic that measures the uncertainty of the point estimate? $\boxed{iii}$

[2] **(e)** Compute the value from part **(c)** for this context.

The sample proportion is:

$$\hat{p} = \frac{x}{n} = \frac{27}{212} = 0.12736 \boxed{\approx 0.127.}$$

Thus, about 12.7% of the sampled chips were defective.

[2] **(f)** Compute the value from part **(d)** for this context.

The standard error of $\hat{p}$ is:

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.12736)(1-0.12736)}{212}} \boxed{= 0.0229.}$$

[2] **(g)** The historical rate of defects is 10%. Should the engineer be surprised by the observed rate of defects during the current week? Explain briefly.

Use the $\pm 2SE$ (about 95%) rule around $\hat{p}$ to check consistency with the historical rate:

$$\hat{p} = 0.12736, \quad SE_{\hat{p}} = 0.0229$$

$$\hat{p} \pm 2SE = 0.12736 \pm 2(0.0229) = \boxed{(0.0816, 0.1732).}$$

The historical value $p_0 = 0.10$ lies within this interval. Therefore, the observed rate is $\boxed{\text{not surprising}}$ given a historical rate of 10%.

(h) Suppose the true population value was found to be 10%. If we use this proportion to recompute the value in part **(f)** using $p = 0.1$ instead of $\hat{p}$, does the resulting value change much? (Show your work.) *[2]*

If we use the hypothesized population value $p = 0.1$ in the standard error,

$$SE_{\hat{p}}(p{=}0.1) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.1 \cdot 0.9}{212}} \boxed{\approx 0.0206.}$$

This is around 10% smaller than the previous 0.0229. $\boxed{\text{Hence, using } p \text{ instead of } \hat{p} \text{ makes little difference.}}$

Problem 5: Estimating probability from a random sample *(8 points)*

Among the adults in a large city, 30% have a college degree. A simple random sample of $n = 100$ adults is chosen. Let p be the proportion of the full adult population that have a college degree. Thus, $p = 0.30$. Now, this proportion (of college degree holders) can also be estimated from a random sample. Let the estimate be denoted by $\hat{p}$.

[2] **(a)** Show that $\hat{p}$ satisfies the success-failure condition.

The *Success-Failure Condition* requires:

$$np \geq 10$$
$$n(1-p) \geq 10$$

Calculating each term:

$$np = 100 \times 0.30 = \boxed{30 > 10}$$
$$n(1-p) = 100 \times 0.70 = \boxed{70 > 10}$$

Since $\boxed{\textbf{both values are greater than 10}}$, the Success-Failure Condition is satisfied.

[3] **(b)** According to the Central Limit Theorem, $\hat{p}$ follows a normal distribution. What are the parameters μ and σ of this distribution? [1 point for μ, 2 points for σ]

Given:

$$p = 0.30, \quad n = 100,$$

we find:

$$\mu_{\hat{p}} = p = 0.30,$$

and

$$\sigma_{\hat{p}} = \sqrt{\frac{0.30(1-0.30)}{100}} = \sqrt{\frac{0.21}{100}} = \sqrt{0.0021} \approx 0.0458.$$

Therefore,

$$\boxed{\mu_{\hat{p}} = 0.30, \quad \sigma_{\hat{p}} \approx 0.0458.}$$

[3] **(c)** Estimate the probability that over 40% of the full adult population have a college degree?

$$\begin{aligned} P(\hat{p} > 0.40) &= 1 - P(\hat{p} \leq 0.40) \\ &= 1 - \Phi\left(\frac{0.40 - 0.30}{\sqrt{0.0021}}\right) \\ &= \boxed{0.0145} \end{aligned}$$

Problem 6: Confidence interval of a proportion *(6 points)*

Recall the example of the solar energy poll we examined in class, with $p = 0.88$ and $n = 1000$.

(a) Compute the standard error $SE_{\hat{p}}$. *[2]*

$$\begin{aligned} SE_{\hat{p}} &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= \sqrt{\frac{0.88 \times (1-0.88)}{1000}} \\ &= \sqrt{\frac{0.88 \times 0.12}{1000}} \\ &= \sqrt{\frac{0.1056}{1000}} \\ &= \sqrt{0.0001056} \\ &\approx 0.0103. \end{aligned}$$

Thus, the standard error is:

$$\boxed{SE_{\hat{p}} \approx 0.0103}$$

(b) Find the 95% confidence interval of $\hat{p}$. (Express your final answer as an interval.) *[2]*

For a 95% confidence level, the critical z-score z^* is 1.96. The confidence interval is given by:

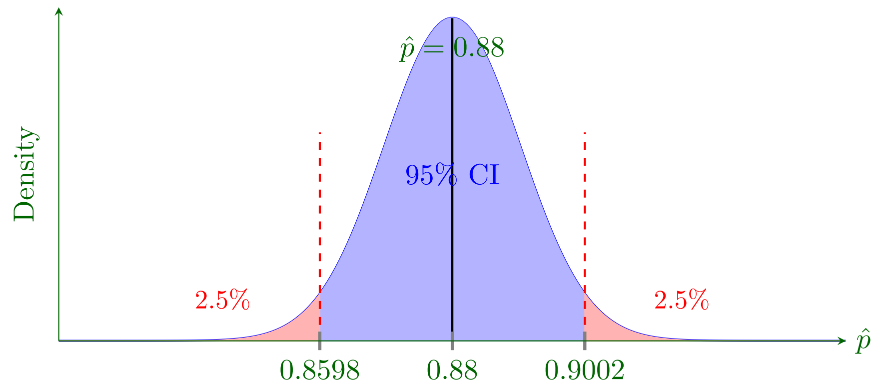
$$\begin{aligned} CI &= \hat{p} \pm z^* \times SE_{\hat{p}} \\ &= 0.88 \pm 1.96 \times 0.0103 \\ &= 0.88 \pm 0.0202. \end{aligned}$$

This gives the interval:

$$\boxed{(0.8598, 0.9002)}$$

(c) Sketch the distribution of $\hat{p}$ and indicate the confidence interval found in part **(b)** on the plot. *[2]*

The distribution of $\hat{p}$ follows a normal distribution with mean $\mu = 0.88$ and standard error $SE_{\hat{p}} = 0.0103$. The 95% confidence interval is $[0.8598, 0.9002]$.



The plot shows the sampling distribution of $\hat{p}$ centered at 0.88, with the 95