Stat 50, Short Answers to Quiz #4, Spring 2025 04/29/2025

1. The distribution of weights (in pounds) of male students at a college is approximately normal with an unknown mean μ and with an unknown standard deviation σ.

   (a) (1.5 pts) Assume that the average weight of a random sample of 20 male students in the college is 152.33 lb with a sample standard deviation of 15.24 lb. Determine a two-sided confidence interval for μ at 95% level of confidence using two-decimal places of rounding. Justify your steps including your choice of the interval estimation procedure (e.g. t-interval, z-interval etc.).

   Answer. Since σ is unknown and a random sample from an approximate normal population is used, a t-interval procedure applies with $\bar{x} = 152.33$, $s = 15.24$, $1-\alpha = 0.95, n = 20$ and $df = 19$. We also use $t_{19,0.025} =$ as the t-score. Then, the margin of error is $E = 2.093\frac{15.24}{\sqrt{20}} \simeq 7.1325$ lb, and the interval is 152.33 ± 7.1325 which is approximately $[145.20, 159.46]$ lb, meaning we are 95% certain that the actual mean weight is between 145.20 lb and 159.46 lb.

   (b) (1 pt) Assume that the population standard deviation of the weights of the male students is 15 lb. Based on this, determine the minimum sample size n needed to ensure that the margin of error at 98% level of confidence is not larger than 5 lb. Use an integer end result with proper units.

   Answer. If $\sigma = 15$ is given, we can use a z-interval approach with $z_{0.01} = 2.326$ (the level of confidence has increased here). Then, $n \simeq (2.326\frac{15}{5})^2 = 48.692$ an so we need to use at least $n = 49$ male students in the sample.

2. In a random sample of 64 cars registered in a certain state, 11 of them were found to have emission levels that exceed the state standard. Let p denote the actual (unknown) proportion of all cars registered in that state whose emission levels exceed the state standards.

   (a) (1.5 pts) Obtain a lower-bound confidence interval (CI) for p at 95% confidence level (use three decimal places). Justify the procedure that you apply.

   Answer. Since both $x = 11$ and $n - x = 53$ are larger than 10, the minimum criteria for a large-sample z-interval procedure applies (though the Agresti-Coull interval with $\tilde{p} = \frac{x+2}{n+4} = \frac{13}{68}$ is also an option) at level $1 - \alpha = 0.95$. With $\hat{p} = \frac{x}{n} = \frac{11}{64} = 0.17188$ and $z_{0.05} = 1.645$, the lower bound is

$$LB = \hat{p} - z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{11}{64} - (1.645)\sqrt{\frac{\frac{11}{64}\frac{53}{64}}{64}} \simeq 0.0943 \simeq 0.094.$$

   So, the lower-bound CI for p is $[0.094, 1)$.

   (b) (0.5 pts) Based on the interval above, can you reasonably conclude that p is larger than 0.1? Briefly explain.

   Answer. The CI constructed above also covers proportions between 0.094 and 0.10 as plausible values of p. So, we can NOT conclude that p is larger than 0.1.

3. (1 pt) Let $X \sim Bin(16, p)$ where p is unknown. Consider the estimator $\hat{\theta} = \frac{X+2}{20}$ for p. Determine $Bias(\hat{\theta}) = p - E[\hat{\theta}]$, $Var(\hat{\theta})$ and $MSE(\hat{\theta})$ in terms of the values of p.

   Answer. Using the mean and variance of $Bin(n = 16, p)$ distribution as well as the properties of linear functions, we get

$$\begin{aligned} E[\hat{\theta}] &= \frac{E[X]+2}{20} = \frac{16p+2}{20} = \frac{4}{5}p + \frac{1}{10} \text{ and} \\ Var(\hat{\theta}) &= \frac{Var(X+2)}{20^2} = \frac{Var(X)}{400} = \frac{16}{400}p(1-p) = \frac{p\,(1-p)}{25}. \end{aligned}$$

   Then, $Bias(\hat{\theta}) = p - (\frac{4}{5}p + \frac{1}{10}) = \frac{1}{5}p - \frac{1}{10}$ or $\frac{2p-1}{10}$ (which is 0 if $p = 0.5$) and

$$MSE(\hat{\theta}) = (\frac{2p-1}{10})^2 + \frac{p\,(1-p)}{25} = \frac{1}{100}.$$