

Present your method of solution in a clear, well-labeled manner. Show the details of your calculations; explain and justify your answers fully. Use clear notation, and define the symbols that you introduce. These take-home problems are due by 10am on Friday, March 21.

You are to work *completely independently* on all parts of this exam. You are to *discuss no aspect* of the exam with anyone other than me. I will answer only questions of clarification and software use. You may use your class notes, text, and assignments. You may use computer software (statistics packages, spreadsheets, computer algebra systems) as long as you indicate clearly how you use it. You may not consult other aids (such as other books or websites) unless you get permission from me.

1. Let Y_1, Y_2, \dots, Y_n be a random sample (i.i.d.) of size n from a normal distribution with mean μ and variance σ^2 . Consider estimators of σ^2 having the form $T = c \sum_{i=1}^n (Y_i - \bar{Y})^2$. Determine the value of c (as a function of n) for which the MSE of T is minimized.

2. Suppose that X and Y have independent and identically distributed normal distributions with mean 0 and standard deviation σ . Show that the random variable $\frac{X+Y}{X-Y}$ has a t -distribution with 1 degree of freedom. [Be very clear about justifying every step in your derivation.]

3. Let X_1, X_2, \dots, X_{10} be i.i.d. according to an exponential distribution with parameter β (and so mean $1/\beta$). Consider testing the hypotheses $H_0: \beta \geq 1$ vs. $H_1: \beta < 1$. Let $Y = \min\{X_i\}$, and suppose that the test procedure is to reject H_0 when $Y \geq k$, for some constant k .

[Note: If you use Excel or Minitab to do any calculations or graphing for this problem, be aware that they both regard the exponential parameter to be the mean; in other words, they regard $1/\beta$ as the exponential parameter. Similarly, both Excel and Minitab regard the second parameter of a Gamma distribution as the reciprocal of what our text regards as the parameter.]

a) Determine the value of k so that this test has size .05.

b) Determine the power function of this test procedure from a). Also calculate the power at $\beta = .5$ and at $\beta = 2$.

c) Produce a graph of this power function from b).

Continue to test the hypotheses $H_0: \beta \geq 1$ vs. $H_1: \beta < 1$. But now suppose that the test procedure is to reject H_0 when $\bar{X} \geq c$, for some constant c , where $\bar{X} = \sum_{i=1}^{10} X_i / 10$.

- d) Determine the value of c so that this test has size .05.
- e) Determine the power function of this test procedure from d). Also calculate the power at $\beta = .5$ and at $\beta = 2$.
- f) Produce a graph of this power function from e).
- g) Produce a graph with these two power functions on the same set of axes. Comment on how the two power functions compare to each other, and what this graph reveals about the relative merits and disadvantages of these two test procedures.

4. Suppose that a new machine is designed to fill boxes of cereal by weight and that you want to estimate μ , the mean weight of boxes produced by this machine. We have studied a confidence interval procedure for μ based on the sample standard deviation. However, plant managers often want their employees to be able to calculate such intervals immediately on the factory floor, without the benefit of more technology than a simple calculator. It would be much simpler if they could use the sample range R as the measure of variability instead of the sample standard deviation S .

Suppose (for now) that the sample size is $n = 5$, and consider using $\bar{X} \pm R/2$ as a 95% confidence interval for μ . Conduct a simulation analysis to investigate how well this procedure performs, assuming that the cereal box weights follow i.i.d. normal distributions.

- a) First generate 10,000 random samples of size $n = 5$ from a normally distributed population. (Feel free to use whatever mean μ and standard deviation σ you like. Be sure to report these values.) Then calculate the sample mean \bar{X} and sample range R for each of the 10,000 simulated samples. Submit a histogram of the sample range values. [*Hint*: The Minitab command for calculating the range across rows is `rrange`. You can also find this using the menu option `Calc > Row Statistics`.]
- b) Apply the $\bar{X} \pm R/2$ procedure to calculate the 10,000 confidence intervals for μ . Determine the length of each interval, and report the average of these lengths.
- c) Determine and report the percentage of your 10,000 intervals that succeed in capturing the population mean μ . Comment on how well this interval procedure performs.
- d) Change the sample size to $n = 10$, and repeat a)-c).
- e) Use your simulation results to investigate how you might tweak the $\bar{X} \pm R/2$ procedure to improve it for the $n = 10$ case, still basing it on the sample range R and keeping it simple for employees to implement on the factory floor. Summarize your recommendation for tweaking this procedure and explain the merits of your suggestion.