CGS602A: Basic Statistics, Data Analysis and Inference

Quiz #2:

Max marks: 50 27-11-2020
Time: 1hr

1. *Answer all 4 questions.*
2. *Please do not collaborate.*
3. *You can use any online tables or statistical calculators to get actual values where needed. For example, https://www.danielsoper.com/statcalc/default.aspx.*

1. Fill in the following table where: $\bar{x}$ is the sample mean, σ is the population standard deviation, s the sample standard deviation, $s_{\bar{x}}$ is an estimate of the sampling distribution std. dev., $\sigma_{\bar{x}}$ is the standard deviation of the sampling distribution, $z_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2},\nu}$ have their usual meanings where α is the confidence level and ν is the degrees of freedom for the t-distribution. If there is no formula (exact or approx.) to calculate the CI (confidence interval) write NA.

   The first line in the table shows a filled in example. For other rows fill in the missing (blank) entries.

Pop. distr.	σ	**Samp. size (N)**	**CI**	**Exact/Approx.**
Normal	Known	$N \geq 30$	$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma_{\bar{x}}$	Exact
Normal	Known	$N < 30$		
Not normal	Known	$N < 30$		
Normal	Unknown	$N < 30$		
Not normal	Unknown	$N \geq 30$		
Not normal	Unknown	$N < 30$		

Solution:

Pop. distr.	σ	**Samp. size (N)**	**CI**	**Exact/Approx.**
Normal	Known	$N < 30$	$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma_{\bar{x}}$	Exact
Not normal	Known	$N < 30$	NA	NA
Normal	Unknown	$N < 30$	$\bar{x} \pm t_{\frac{\alpha}{2},\nu}s_{\bar{x}}$	Exact
Not normal	Unknown	$N \geq 30$	$\bar{x} \pm z_{\frac{\alpha}{2}}s_{\bar{x}}$ or $\bar{x} \pm t_{\frac{\alpha}{2},\nu}s_{\bar{x}}$	Approx.
Not normal	Unknown	$N < 30$	NA	NA

[10]

2. (a) Given a sample of size N chosen from an infinite normal distribution what can you say about the confidence interval $a \leq \mu \leq b$ as α changes from 0.1 to 0.05 to 0.01. Justify your answer.

Solution:

We have $P(\bar{X} - z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le \mu \le \bar{X} + z_{\frac{\alpha}{2}}\sigma_{\bar{x}}) = 1 - \alpha$ where $\sigma_{\bar{x}}$ is the std. deviation of the sampling distribution. So, as α becomes smaller the value of $z_{\frac{\alpha}{2}}$ increases and the CI becomes systematically larger. We know $z_{\frac{\alpha}{2}}$ for $\alpha = 0.1,\ 0.05,\ 0.01$ are 1.645, 1.96. 2.576. So, CI keeps increasing.

Intuitively, if we want the mean to fall within an interval of the sample mean more frequently (that is higher $1 - \alpha$) then the interval has to be wider.

(b) If for a given confidence level you wish to shrink the confidence interval what can you do? Justify your answer.

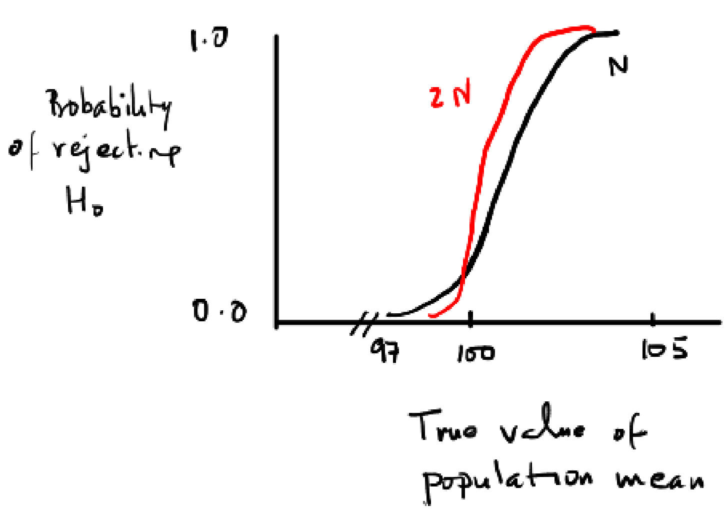
Solution:

From the equation in the part above. To shrink the CI we must reduce $z_{\frac{\alpha}{2}}$ or $\sigma_{\bar{x}}$. Since α is fixed we must reduce $\sigma_{\bar{x}}$. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ can be decreased by increasing sample size N.

[6,4=10]

3. (a) Suppose $H_0 : \mu = 100$. You do a right tailed test at $\alpha = 0.05$ for some sample size N. Assume the population distribution is normal. Draw a graph of *probability of rejecting* H_0 on the Y-axis versus *possible values for the true population mean* on the X-axis (choosing values on either side of 100 i.e. values that are less than and greater than 100). Give a brief justification for the shape of your graph.

Solution:



Since the test is right tailed the alternate hypothesis should be $\mu_A > 100$ and the probability of rejection increases as the population's true μ becomes larger, that is moves to the right of $\mu = 100$. Also, the rejection probability reduces and approaches 0 for μ values less than 100.

Approx. schematic curves are shown both for sample size N (black) and sample size $2N$ (red) above. The point at which the two cross-over corresponds to $\alpha = 0.05$.

(b) How will the graph in part (a) change if the sample size is doubled to $2N$? Justify.

Solution:

The changed curve is shown in red in the graph for part a). Doubling the sample size to $2N$ will make the CI smaller and so the probability of rejecting will be higher for smaller values of the population mean beyond 100 and similarly lower for values less than 100. So, it is to the left of the curve for N beyond 100 and to its right for values less than 100.

[(6,4),5=15]

4. For a sample of $N = 19$ write the full decision rules to reject $H_0 : \sigma^2 = \sigma_0^2$ where $\alpha = 0.05$ and the alternate hypotheses are:

   (a) $H_a : \sigma^2 > \sigma_0^2$.

   (b) $H_a : \sigma^2 < \sigma_0^2$.

   (c) $H_a : \sigma^2 \neq \sigma_0^2$.

   First decide what statistic you will use and then write the decision rules using the statistic.

Solution:

Since we are looking at hypotheses involving variance so the sampling distribution is chi-square and the statistic is $\chi^2 = \frac{(N-1)s^2}{\sigma_0^2}$. The alternate hypotheses are right-sided (or upper-tailed), left-sided(or lower-tailed) and two-sided(or two-tailed) respectively. The critical region is $[0, x_1)$ and (x_2, ∞) where $x_1 < x_2$. Note that the chi-square distribution is not symmetric.

So, for an upper-tailed H_a the decision rule is: $\chi^2 > \chi^2_{\alpha,\eta} = \chi^2_{0.05,18} = 28.87$ where $\eta = N - 1$ is the degree of freedom for the chi-square distribution and the rhs is the critical value. Similarly, for lower-tailed H_a the decision rule will be: $\chi^2 < \chi^2_{0.95,18} < 9.39$ and for the two-tailed H_a it is: $\chi^2 > \chi^2_{\frac{\alpha}{2},\eta} = \chi^2_{0.025,19} = 31.53$ or $\chi^2 < \chi^2_{\frac{1-\alpha}{2},\eta} = \chi^2_{0.975,18} < 8.23$.

[5×5=15]