A N S W E R K E Y

STAT 1051: QUIZ 6 Take Home

1. The grade point averages (GPA) of a sample of 60 students were obtained. Denote the GPA of a student by X_i. From the data, it was found that $\bar{x} = 3.3$ and $s = .5$. Find the 90% confidence interval about the population mean.

   (a) **Answer 1** *(5 pts)* $\alpha = .10$ *so* $\frac{\alpha}{2} = .05$ *and n-1=59.*

   (b) **Answer 2** *(5 pts)* $t_{59,.05} = 1.67591$ *(using n=50 degrees of freedom in approximating the table.)*

   (c) **Answer 3** *(5 pts) lower limit* $a = \bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 3.3 - \frac{.5}{\sqrt{60}} 1.67591 = 3.19$

   (d) **Answer 4** *(5 pts) upper limit* $b = \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 3.3 + \frac{.5}{\sqrt{60}} 1.67591 = 3.41$

   (e) **Answer 5** *(5 pts) 90% CI is: (3.19,3.41)* ***Parentheses and comma are required for correct notation****.*

2. Only one answer is correct in this multiple choice problem.

   A test engineer wants to estimate the mean gas mileage, μ (in miles per gallon), for a particular model of automobile. Eleven of these cars are subjected to a road test, and the gas mileage is computed for each car. A dot plot of the eleven gas mileage values is roughly symmetrical and has no outliers. The mean and standard deviation of these values are 25.2 and 3.01, respectively. Assuming that these eleven automobiles can be considered a simple random sample of cars of this model. which of the following is a correct statement?

   (a) A 95 percent confidence interval for μ is $25.2 \pm 2.228\frac{3.01}{\sqrt{11}}$

   (b) A 95 percent confidence interval for μ is $25.2 \pm 2.201\frac{3.01}{\sqrt{11}}$

   (c) A 95 percent confidence interval for μ is $25.2 \pm 2.228\frac{3.01}{\sqrt{10}}$

   (d) A 95 percent confidence interval for μ is $25.2 \pm 2.201\frac{3.01}{\sqrt{10}}$

   (e) The results cannot be trusted; the sample is too small.

   Answer 6 *(10 pts) The following steps are not needed for the answer, but show that A is the correct answer.*

   (a) $\alpha = .05$ *so* $\frac{\alpha}{2} = .025$ *and n-1=10.*

   (b) $t_{10,.025} = 2.228139$

   (c) lower limit $a = \bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 25.2 - \frac{3.01}{\sqrt{10}} 2.228139 = 23.07916$

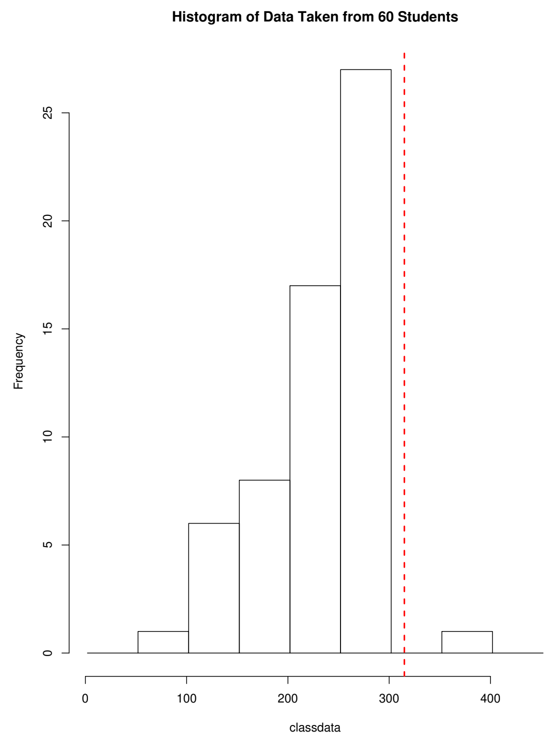
   (d) upper limit $a = \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 25.2 + \frac{3.01}{\sqrt{10}} 2.228139 = 27.32084$

   (e) 95% CI is: (23.07,27.32)

Part (a) is the correct answer, because n=11 and t-quantile=2.228 is the easy answer.

I will philosophize about the conditions, though this is not a part of the answer. That the data assumes a symmetric shape when plotted with no outliers and that the elements of the sample were drawn at random (that is i.i.d. observations), we may assume by the Central Limit Theorem that $\overline{x}$ is approximately distributed as a Normal distribution, so that our formula for a confidence interval is adequate even though the sampling size is rather small.

3. From 60 STAT51 students counted the number of letter e's on a page which was copied from a statistics textbook in five minutes, the average reported number of e's was 234.9 with a standard deviation of 56.8118. A histogram of the experimental data is shown below.



(a) Construct a 95% confidence interval for the actual number of e's that were printed on that page.

i. **Answer 7** *(5 pts)* $\alpha = .05$ *so* $\frac{\alpha}{2} = .025$ *and n-1=59.*

ii. **Answer 8** *(5 pts)* $t_{59,.025} = 1.99210$ *(Using 75 degrees of freedom for reading the table.)*

iii. **Answer 9** *(5 pts) lower limit* $a = \bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 234.9 - \frac{56.8118}{\sqrt{60}} 1.99210 = 220.2$

iv. **Answer 10** *(5 pts) upper limit* $b = \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} = 234.9 + \frac{56.8118}{\sqrt{60}} 1.99210 = 249.57$

v. **Answer 11** *(5 pts) 95% CI is: (220.2,249.57)* ***Parentheses and comma are required for correct notation***.

(b) (10 pts) The dotted line represents the actual number of e's on the page. Why does the 95% confidence interval not cover even closely the exact number of e's that appeared on the page? (Hint: Draw the confidence interval on the histogram).

Answer 12 *The confidence interval might be one of the 5% which will not, on the average, cover the population. Another correct and better answer is that the set of data is actually biased.*

4. (10 pts) On the average, how many 40 85% confidence intervals will contain μ?

Answer 13 *On the average,* $40(.85) = 34$ *confidence intervals will contain* μ*.*

Only one answer is correct in this multiple choice problem.

5. (10 pts) A simple random sample produces a sample mean, $\bar{x}$, of 15. A 95 percent confidence interval for the corresponding population mean is 15 ± 3. Which of the following statements must be true?

(a) Ninety-five percent of the population measurements fall between 12 and 18.

(b) Ninety-five percent of the sample measurements fall between 12 and 18.

(c) If 100 samples were taken, 95 of the sample means would fall between 12 and 18.

(d) $P(12 \leqslant 15 \leqslant 18) = .95$

(e) If $\mu = 19$, this $\bar{x}$ of 15 would be unlikely to occur.

Answer 14 *E is the correct answer. Answers A-D are obviously incorrect and that leaves E by the process of elimination.*

(a) A refers to a special boxplot defined by the 1^{st} *of the* 40^{th} *percentile =12 for left edge and the* 39^{th} *of the* 40^{th} *percentile =18 for the right edge of the box of the population measurements. This boxplot is not a confidence interval.*

(b) B refers to a special boxplot defined by the 1^{st} *of the* 40^{th} *percentile =12 for left edge and the* 39^{th} *of the* 40^{th} *percentile =18 for the right edge of the box of the sample measurements. This boxplot is not a confidence interval.*

(c) For choice C, we are not given the 100 sample means to determine if 95 of them fall between 12 and 18. Like choices A and B, this choice refers to a special boxplot where the left edge of the box = 12 and the right edge of the box =18. This boxplot is not a confidence interval.

(d) The probability stated in choice D is 1: 15 is always between 12 and 18.

(e) E is correct. Because the confidence interval (12,18) does not contain 19 we reject the hypothesis $H_0 : \mu = 15$ *vs* $H_1 : \mu \neq 15$ *at* $\alpha = .05$*; therefore, 15 is unlikely to occur.*

(10 pts) Explain your answer.

Answer 15 *E is correct. Because the confidence interval (12,18) does not contain 19 we reject the hypothesis* $H_0 : \mu = 15$ *vs* $H_1 : \mu \neq 15$ *at* $\alpha = .05$*; therefore, 15 is unlikely to occur.*

Another explanation would be that because we are 95% confident that μ *is in (12,18) and since we are told to the contrary that* $\mu = 19$ *the confidence interval is therefore located in the wrong place, that is, 15 is unlikely to occur.*

As an aside, I will philosophize:

The wording of E is not quite a satisfactory answer. because 15 did occur. This is a problem in which the random variable $\bar{x}$ *and its realized value 15 are easily confused. We know, too, that for a continuous random variable, P(X=15)=0. Part E should have been written as: Given* $\bar{x} \sim N(19, \sigma^2)$*, , then the event that* $\bar{x}$ *is 15 or less is unlikely to occur. Since* $2\sigma = 3$ *from the statement of the problem,* $\sigma = 1.5$*.* $P(\bar{x} \leqslant 15) = P(\frac{\bar{x}-19}{1.5} \leqslant \frac{15-19}{1.5}) = P(z \leqslant -2.666)$ *draw a picture and use table in book to conclude that* $P(z \leqslant -2.666) = .0038$*; therefore the event is unlikely.*

A more precise calculation is: $t_{n-1,.025}\frac{s}{\sqrt{n}} = 3$*; therefore,* $\frac{s}{\sqrt{n}} = \frac{3}{t_{n-1,.025}}$.

$P(\bar{x} \leqslant 15) = P(\frac{\bar{x}-19}{\frac{s}{\sqrt{n}}} \leqslant \frac{15-19}{\frac{s}{\sqrt{n}}}) = P(T_{n-1} \leqslant \frac{-4}{\frac{3}{t_{n-1,.025}}})$

In R *the probability is: p=pt(-4/(3/qt(.975,n-1)),n-1). For n=10, p=0.007012687, for n=100, p=0.00474005. For a very large sample,* $p = P(z \leqslant -2.613285) = .00448$

Bonus (10 pts) A radio talk show invites listeners to express their opinions about a proposed pay increase for members of the city council. A total of 958 listeners called the radio station in response to the station's solicitation for opinions. A statistical consultant calculated the 95% confidence interval to be ($9669, $9811) for the mean pay, μ, of a councilman which reflected the opinion of those who called the station. Is this result trustworthy? Explain your answer.

Answer 16 *The result is untrustworthy. This is a biased sample, because the elements of the sample were not selected at random from the list of the population which is presumably the listening area of the radio station. That is, the sample not being a probability based survey is not representative of the population. The formulas for the confidence interval are predicated on unbiased data. See 1936 election. FDR vs Alf Landon. Literary Digest predicted that Alf Landon would win, but FDR won with 60.8% of the vote.*