Discussion 5 - Solution

1. A sample of size n = 9 from a large population that is thought to be normal has a sample mean of 13.01 and sample standard deviation of 4.85.

   (a) Create a 80% CI for the population mean μ, assuming that σ^2 is unknown.

   Solution: If we call the random variable here Y, then the mean of the observed data is $\bar{Y}$. Since we need to estimate σ^2, the appropriate distribution for $\bar{Y}$ is a t distribution, and since there were 9 data points, the t distribution has 8 degrees of freedom.\

   You are asked for a 80% confidence interval, so use $\alpha = 1 - 0.80 = 0.2$. In order to get a symmetric confidence interval (with $\bar{Y}$ at the center), go to the table with $\alpha/2 = 0.10$. We need to look up the value of t_8 closest to 0.1, which turns out to be $t_{8,0.1} = 1.4$.\

   Now the probability statement for the confidence interval is:

$$P\left(-t_{8,0.1} \leq \frac{\bar{Y} - \mu}{S/\sqrt{n}} \leq t_{8,0.1}\right) = 0.8.$$

   To get the 80% C.I. for μ, manipulate the probability statement until μ is alone in center. Then the 80% C.I. for μ is $\bar{Y} \pm t_{8,0.1} \times S/\sqrt{n} = 13.01 \pm 1.4 \times 4.85/\sqrt{9} = 13.01 \pm 2.26 = (10.75, 15.27)$

   (b) If it is now assumed that the variance is known to be $\sigma^2 = 16$, create a 80% CI for the population mean μ. Is it larger or smaller than the one created in (a)?

   Solution: Now σ^2 is known, so it is not necessary to estimate σ^2 and therefore we can assume the distribution of $\bar{Y}$ is Normal, not t_8. Once again, $\alpha/2 = 0.1$. Since the Z-table is formatted opposite of the t-table, we need to look up the value of $z_{0.1}$, which is the value in the Z-table closest to $1-\alpha/2 = 0.9$. That turns out to be $z_{0.1} = 1.28$.\

   Now the probability statement for the confidence interval is:

$$P\left(-z_{0.1} \leq \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \leq z_{0.1}\right) = 0.8.$$

   To get the 80% C.I. for μ, manipulate the probability statement until μ is alone in center. Then the 80% C.I. for μ is $\bar{Y} \pm z_{0.1} \times \sigma/\sqrt{n} = 13.01 \pm 1.28 \times 4/\sqrt{9} = 13.01 \pm 1.71 = (11.30, 14.72)$.

   This is **smaller** than the t confidence interval.

   (c) What is the minimum sample size n in part (a) (assuming the sample mean and variance don't change) so that the 80% C.I. half-width calculated using the T-distribution would be less than the half-width you calculated in part (b)? **Solution**: As a first approximation, we will keep the value from the t-table we used in part (a) and find n to make the half-widths equal:

$$\pm 1.4 \times 4.85/\sqrt{n} = 1.71$$

   Which results in $n = 15.76$, rounded up to 16. But by changing the sample size we should also change which column of the t table we work from. If n is 16 then the degrees of freedom are 15 and we need to look up $t_{15,0.1} \approx 1.35$. Then the half width is $\pm 1.35 \times 4.85/\sqrt{16} = 1.637$, which is a lot smaller than 1.71. Try calculating the half-width with $n = 15$ and $n = 14$ (keeping the same value from the t-table for now) and we see $\pm 1.35 \times 4.85/\sqrt{15} = 1.69$ and $\pm 1.35 \times 4.85/\sqrt{14} = 1.75$.\

   It looks like $n = 15$ is big enough (because $1.69 < 1.71$) but $n = 14$ is too small (because $1.75 > 1.71$), but now we need to once again check against the correct value from the t-table. As far as I can tell, 1.35 is as good an approximation to $t_{14,0.1}$ as to $t_{15,0.1}$. So conclude that the minimum sample size is 15.

2. An engineer is designing an experiment to estimate the stiffness of beams from a certain factory. A random selection of beams is to be tested by putting a load on each beam and measuring its deflection. She knows that she must have a 90% confidence interval for the population-mean deflection μ where the half-width of the interval (margin of error) is less than 1cm.

   (a) If the population variance were known to be 4.2, what is the minimum sample size in order to attain the required half-width?

Solution: Since we are asked for a 90% confidence interval with a known variance, we need to look up in the Normal table (Z-table) the value closest to $\alpha/2 = (1 - 0.9)/2 = 0.05$. This turns out to be $z_{0.05} = 1.645$. The half-width is $z_{0.05} \times \sigma/\sqrt{n} = 1\text{cm}$, where n is the sample size. Plugging in the known $z_{0.05} = 1.645$ and $\sigma = \sqrt{4.2} = 2.05$, we have:

$$1.645 \times 2.05/\sqrt{n} = 1\text{cm}$$

So $n = 11.4$ but since sample size has to be an integer we round up to $n = 12$ so as to make sure we still have the necessary half-width.

   (b) Suppose that the variance is still 4.2, but it was estimated from the sample (which is of the size calculated in part (a)). What is the 90% C.I.'s half-width? Is it acceptable according to the engineer's standard?

Solution: Here variance was estimated from the data and the sample size was 12, so we need to look up the value closest to $\alpha/2 = 0.05$ in the 11-degrees-of-freedom column of the t-table. That turns out to be $t_{11,0.05} = 1.8$. The half-width is

$$t_{11,0.05} \times S/\sqrt{n} = 1.8 \times \sqrt{4.2}/\sqrt{12} = 1.065\text{cm}$$

.

This half width is 1.065cm, greater than the allowed limit of 1cm, so it is not acceptable for this experiment.

   (c) Supposing she has collected a sample sized according to the part (a), where the sample's mean deflection was 7.21 cm and its variance is $s^2 = 3.69$, what is the 90% confidence interval?

Solution: Here variance was estimated from the data and the sample size was 12, so we need to look up the value closest to $\alpha/2 = 0.05$ in the 11-degrees-of-freedom column of the t-table. That turns out to be $t_{11,0.05} = 1.8$. Also, let's denote the data by random variable X.\

Now the Probability statement for the 90% C.I. is

$$P\left(-t_{11,0.05} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{11,0.05}\right) = 0.9.$$

Manipulate the probability statement until μ is alone in center. Then the 90% C.I. for μ is $\bar{X} \pm z_{0.05} \times \sigma/\sqrt{n} = 7.21 \pm 1.8 \times 1.92/\sqrt{12} = 7.21 \pm 0.998\text{cm} = (6.21\text{cm}, 8.21\text{cm})$.

   (d) Suppose that the engineer plans to repeat the study described in part (c) 20 times, generating an independent 90% confidence interval each time. How many of those confidence intervals are expected to contain the true value of μ? What is the variance in the number that contain the true value of μ? What is the probability that all of the intervals contain the true μ?

Solution: The interval's confidence level is defined so that the proportion of intervals that contain the true value is equal to the confidence level. So prior to looking at any data, the probability that each one of her intervals interval contains the true value is 90%. Since the intervals are independent and the probability of containing the true μ is equal each time, this is a binomial random variable (call it B) with $p = 0.9$ and $n = 20$.\

By definition of a binomial random variable, the expected number of intervals that cover the true μ is $np = 20 * 0.9 = 18$ intervals.\

By definition of a binomial random variable, the variance in the number that cover the true μ is $n \times p \times (1 - p) = 20 * 0.9 * 0.1 = 1.8$\

As far as we know before collecting any data, the probability that all 20 of the intervals will cover the true μ is calculated from the binomial pmf:

$$P(B = 20) = \binom{20}{20} \times 0.9^{20} \times 0.1^{0} = 0.122$$

.