18.05 Problem Set 10, Spring 2022

Problem 1. (15: 10,5 pts.) **Z**
Suppose we have n data points drawn from a $\mathrm{N}(\theta, 9^2)$ distribution, where the value of the mean θ is unknown.

(a) Suppose that $n = 30$ and that the sample mean is $\overline{x} = 15$. Give the 95% confidence interval for the mean θ. Also give the 70% confidence interval.

In this problem you should use R to give **precise** confidence intervals. That is, you should find the critical values to several decimal points of accuracy.

(b) Explain why the 95% confidence interval for the mean θ is always narrower when $n = 90$ when $n = 30$.

Problem 2. (10 pts.) **(Soda)**
Consider a machine that is known to fill soda cans with amounts that follow a normal distribution with (unknown) mean μ and standard deviation $\sigma = 3$ mL. We measure the volume of soda in a sample of bottles and obtain the following data (in mL):

$$352, 351, 361, 353, 352, 358, 360, 358, 359$$

For this problem, you will probably want to use R to make the calculations.

(a) Construct a precise 95% confidence interval for the mean μ.

(b) Now construct a 98% confidence interval for the mean μ.

(c) Suppose now that σ is not known. Redo parts (a) and (b), and compare your answers to those above.

Problem 3. (10 pts.) **(Cholesterol)**
In a study on cholesterol levels a sample of 12 men and women was chosen. The plasma cholesterol levels (mmol/L) of the subjects were as follows:

$$6.0, 6.4, 7.0, 5.8, 6.0, 5.8, 5.9, 6.7, 6.1, 6.5, 6.3, 5.8.$$

(a) Estimate the variance of the plasma cholesterol levels with a 95% confidence interval.

(b) What assumptions did you make about the sample in order to make your estimate?

Problem 4. (10 pts.) **(Candy)**
A candy manufacturer produces packs of candy targeted to weigh 52 grams. A quality control manager was concerned that the variation in the actual weights was larger than acceptable. That is, They were concerned that some packs weighed significantly less than 52-grams and some weighed significantly more than 52 grams. In order to estimate σ, the standard deviation of the weights of the nominal 52-gram packs, they took a random sample of n = 10 packs off of the factory line. The random sample yielded a sample variance of 4.2 grams^2.

(a) Use the random sample to derive a 95% confidence interval for σ. *Note: the interval is for σ not σ^2.*

(b) What assumptions did you make about the sample in order to make your estimate.

Problem 5. (33: 5,10,18 pts.) **Probabilities for hypotheses**
When given a 95% confidence interval many people often mistake this to mean that there is a 95% probablility that the true value of the parameter is in the confidence interval. This violates the maxim that frequentists don't give probabilities for hypotheses, but the reasoning is subtle. This problem is aimed at understanding the subtlety.

(a) What 95% does mean. Imagine running the following experiment.

- Pick a value μ.
- Draw a sample of size n from a $\mathrm{N}(\mu, 5^2)$ distribution.
- Compute the 95% z-confidence interval for μ.
- Check if μ is in the confidence interval.
- Record the result.

If you repeat this experiment many times, what percentage of the time will the chosen μ be in the confidence interval? (Hint, this is not a question about the probability of hypothesized values of μ. It's about the probability of random intervals.)

(b) What 95% doesn't mean. Now let's modify this a bit. Suppose we have the following choices and prior probabilities for μ

μ	0	1	2	3
$p(\mu)$	0.94	0.02	0.02	0.02

Suppose I pick a random μ from this prior distribution and draw a sample of size 64 from a $\mathrm{N}(\mu, 5)$ distribution. I do not tell you the value of μ, but I do tell you that the sample mean is $\overline{x} = 1.6$.

We both know the sample size ($n = 64$) and the variance used in drawing the sample ($\sigma^2 = 5^2$). So we both know that the 95% confidence interval for μ is

$$\overline{x} \pm z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} \approx 1.6 \pm 1.96 \cdot \frac{5}{8} \approx [0.375,\ 2.825].$$

(i) Compute the prior probability that μ is in the above confidence interval.

(ii) Update the probabilities for μ based on the data and compute the posterior probability that μ is in the above confidence interval.

Are either of these probabilities close to 95%? Why is this not surprising?

(c) A medical study shows that patients receiving a new cancer treatment survived an average of 15 months longer than patients receiving the conventional treatment. The study also found that the 95% confidence interval for this average was 10 to 20 months.

Say whether each of the following statements is true or false. False means that the statement does not follow logically from the confidence-interval result.

(i) There is a 95% probability that the true increase in survival time lies between 10 and 20 months.

(ii) The probability that the new treatment increases survival time is at least 95%.

(iii) The confidence interval provides an estimate of the true value of the average increase in survival time.

(iv) The null hypothesis that there is no change in survival time due to the new treatment is likely to be incorrect.

(v) If we ran the same study 100 times, then in approximately 95 of the studies the 95% confidence intervals would contain the true value of the change in survival time.

(vi) We would reject the null hypothesis of no improvement in survival time at the 5% significance level.