# Problem Set 9

Statistics - NYU, Summer 2016
Ercan Karadas

## Section 1

**[1]** A test is normally distributed with a mean of 70 and a standard deviation of 8.

a) What score would be needed to be in the $85^{th}$ percentile?

b) What score would be needed to be in the $22^{nd}$ percentile?

**[2]** A college admissions officer for an MBA program has determined that historically applicants have undergraduate grade point averages that are normally distributed with standard deviation 0.45. From a random sample of 25 applications from the current year, the sample mean grade point average is 2.90.

a) Find a 95% confidence interval for the population mean?

b) Without doing calculations, explain whether a 99% confidence interval for the population mean would be wider than, narrower than that of found in part (a).

c) Suppose that population standard deviation is 0.56 (instead of 0.45). Without doing calculations explain whether a 95% confidence interval for the population mean would be wider than, narrower than, or the same width as found in part (a).

d) Suppose that sample mean that is given in the problem is calculated from a sample of size 40 (instead of 25). Without doing calculations explain whether a 95% confidence interval for the population mean would be wider than, narrower than, or the same width as found in part (a).

e) Based on these sample results, a statistician computes for the population mean a confidence interval extending from 2.81 to 2.99. Find the confidence level associated with this interval.

f) Suppose that we know that grade point averages that are normally distributed, but we don't have the population standard deviation. From a random sample of 25 applications from the current year, the sample mean grade point average is 2.90, and sample standard deviation is found to be 0.40. Find a 95% confidence interval for the population mean?

**[3]** Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. During a two-month period (44 weekdays), daily fees collected averaged \$126, with a standard deviation of \$15.

a) Construct a 90% confidence interval for the mean daily income this parking garage will raise.

b) Interpret the confidence interval in this context. Explain what "90% confidence" means in this context.

c) The consultant who advised the city on this project predicted that parking revenues would average \$130 per day. Based on the confidence interval you computed, do you think the consultant was correct? Why?

**[4]** Suppose that for a random sample of size $n$, $\{x_1, x_2, \ldots, x_n\}$, which comes from a normal distribution $N(\mu, \sigma^2)$, the sample variance is computed as $s^2$. In this problem, we will carry out the details of how we can construct a $100(1-\alpha)\%$ confidence interval for the population variance, $\sigma^2$.

a) Claim that we would like to find random variables $a$ and $b$ satisfying

$$P(a < \sigma^2 < b) = 1 - \alpha$$

b) Manipulate the expression inside the probability to get

$$P\Big(\frac{(n-1)s^2}{b} < \frac{(n-1)s^2}{\sigma^2} < \frac{(n-1)s^2}{a}\Big) = 1 - \alpha$$

c) Note that in the middle we have the Chi-square distribution with $(n-1)$ degrees of freedom, $\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$. Assuming that we would like leave the same probability $\alpha/2$ in both tails we should have

$$P\big(\chi^2_{n-1,1-\frac{\alpha}{2}} < \chi^2_{n-1} < \chi^2_{n-1,\frac{\alpha}{2}}\big) = 1 - \alpha$$

Conclude that comparing these last two equations and solving for $a$ and $b$ yields

$$a = \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \quad \text{and} \quad b = \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}$$

as desired. We say $a$ is the lower bound of the confidence interval (LCL) and $b$ is the upper bound of the confidence interval (UCL).

**[5]** A natural food company is marketing a new yogurt that it has been advertising as having **less than** half of the fat of regular yogurt. The average amount of fat in a cup of regular yogurt is 1 unit. The Food and Drug Administration has asked us to investigate the product to see whether the company has engaged in false advertising. The test results are as follows:

| | |
|---|---|
| Amount of yogurt tested | 400 cups |
| Average amount of fat contained per cup | .52 units |
| Standard deviation of the amount of fat per cup | .2 units |
| Number of cups containing more than 1.1 units of fat | 24 cups |

a) Compute the probability of observing more than .52 units of average fat, as shown in the report if the true population average fat were .5, as stated in the advertisement?

b) Construct a 95% C.I. and decide whether do you think the advertisement is accurate?

c) The company further claims that only 25% of the cups contain more than 0.6 units fat on average. What should be the population mean, $\mu$, so that the claim would be true. For this and the next part of the problem, assume that the population distribution is normal with the standard deviation 0.22.

d) For this part assume that the distribution of the fat content of yogurt, $X$, is $N(\mu, 0.22^2)$, and $\mu$ is uniformly distributed on [0.4, 0.9]. The report also finds that 24 cups, which corresponds to 24/400=6%, contained more than 1.1 units of fat in the investigation.

Find the probability that at least 6% of the yogurt produced will contain more than 1.1 units of fat. Also, interpret the first statement, or give an example for it.

**[6]** One of the variables that is used to model the number of accidents on highways is the density of cars, which is calculated as follows: at a random point along the road, the distance between consecutive cars is measured for the first $n$ cars and then the standard deviation of the measured distances is calculated. This is supposed to be a good approximation to measure the traffic density.

Data shows that the distance between consecutive cars is normally distributed and the density of traffic is related to the number of accidents as in the following table

| Density ($\sigma$), (per ft) | average # of accidents (per hr) |
|---|---|
| $\sigma < 32$ | 1.5 |
| $32 < \sigma < 60$ | 2.8 |
| $\sigma > 60$ | 4.5 |

Suppose that from a random sample of 20 consecutive cars, the traffic density is calculated as 40. Find the expected number of car accidents in an hour.

## Section 2

**[7]** A filling machine is set up to fill bottles with 35 oz of coke each. The standard deviation s is known to be 1.2 oz. The quality control department periodically selects samples of 20 bottles and measures their contents. Assume the distribution of filling volumes is normal. A production process is considered in control if no more than 3% of the items produced are defective.

a) Determine the upper and lower control limits and explain what they indicate.

b) The means of six samples were 37.8, 29.2, 41.9, 25.9, 32.1, and 43.8 oz. Indicate whether or not the process is in control for each case.

**[8]** Consider a random sample $\{x_1, x_2, \ldots, x_n\}$ from a normal distribution $N(\mu, \sigma^2)$. In this problem, we will review how we can construct a $100(1-\alpha)\%$ confidence interval for the population mean $\mu$.

a) Consider $a$ and $b$ satisfying

$$P(a < \mu < b) = 1 - \alpha$$

Explain why $a$ and $b$ are the bounds of the confidence interval that we are looking for?

b) Manipulate the expression inside the probability to get

$$P\Big(\frac{\overline{X} - b}{\sigma/\sqrt{n}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{\overline{X} - a}{\sigma/\sqrt{n}}\Big) = 1 - \alpha$$

c) Note that in the middle we have $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, and because of the symmetry of the standard normal distribution we should have

$$P\big(-z_{\alpha/2} < Z < z_{\alpha/2}\big) = 1 - \alpha$$

Show that solving for $a$ and $b$ in these last two equations yields

$$a = \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad b = \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

as desired. We say $a$ is the lower bound of the confidence interval (LCL) and $b$ is the upper bound of the confidence interval (UCL).

**[9]** A manufacturer is concerned about the variability of the levels of impurity contained in consignments of raw material from a supplier. A random sample of 15 consignments showed a standard deviation of 2.36 in the concentration of impurity levels. Assume normality.

a) Find a 95% confidence interval for the population variance.

b) Would a 99% confidence interval for this variance be wider or narrower than that found in part a?

**[10]** The trains scheduled to arrive at the New Brunswick train station at 7:35 A.M. every weekday do not always arrive at 7:35. A commuter carefully recorded the arrival time for the last 200 working days and found that a mean delay of 1 minute and a standard deviation of 2.3 minutes. Assume that late arrivals follow a normal distribution.

a) Estimate the average arrival time for the train.

b) Estimate the average arrival time using a 90% confidence interval.

c) Suppose that the true distribution of the late arrivals is $N(0, 1.23)$. If you plan to arrive at the train station at 7:34 regularly for the next 200 working days, how many trains should you expect to miss?

**[11]** What proportion of a normal distribution is

a) within one standard deviation of the mean?

b) more than 2.0 standard deviations from the mean?

c) between 1.25 and 2.1 standard deviations above the mean?