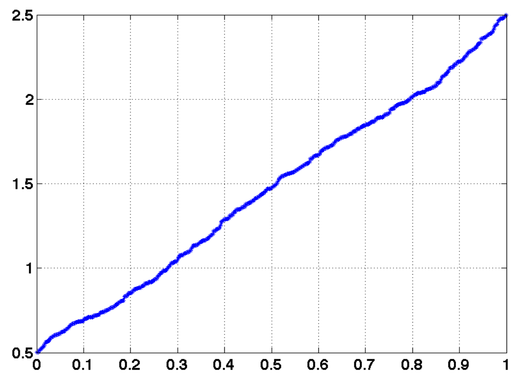
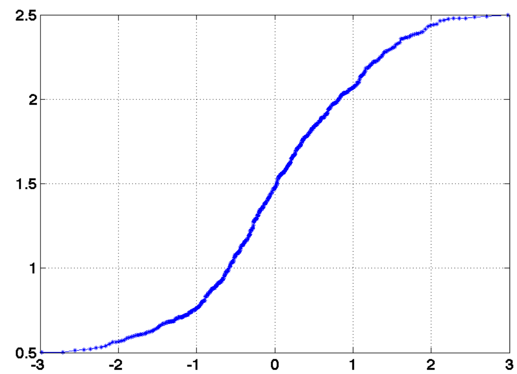
ECE 3530 PRACTICE FINAL 2 Solutions

1. A random sample with size n=200 is taken from a population with a known standard deviation $\sigma = 0.56$. The sample has the following quantile and normal-quantile plots:

Quantile Plot



Normal-Quantile Plot



Based on this information, answer the following questions:

(a) What is the approximate value of the median for the sample?

By definition a quantile plot is plotting q(f) vs f. The median is q(0.5). The value for q(0.5) can be found from the quantile plot as approximately 1.5.

(b) Is the population distribution from which this sample is collected approximately a normal distribution? Justify your answer.

No. The normal-quantile plot shows a nonlinear curve which means that the population distribution is not a normal distribution.

(c) Is the sampling distribution for the sample mean $\bar{X}$ approximately a normal distribution? Justify your answer.

Yes. Even though the population distribution is not normal, the sample size is larger than 30 and the population σ is known; therefore, using the central limit theorem we can say that the sampling distribution of $\bar{X}$ will be approximately normal.

2. A power company wants to study the impact of air conditioners on electricity usage during the summer in Salt Lake City. Let x be day time high temperatures in degrees Fahrenheit. Let Y be the daily total power use of a small neighborhood in kWh. Over the summer the following data are collected on 6 different days:

i	1	2	3	4	5	6
x_i	85	90	100	80	95	90
y_i	350	400	500	350	400	400

(a) **Find the equation of the fitted regression line.**

The equation for the slope and intercept estimates are

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$a = \bar{y} - b\bar{x}$$

Therefore, we start by computing $\bar{x}$ and $\bar{y}$:

$$\bar{x} = \frac{1}{6}(85 + 90 + 100 + 80 + 95 + 90) = 90$$

$$\bar{y} = \frac{1}{6}(350 + 400 + 500 + 350 + 400 + 400) = 400$$

Now we can compute S_{xx} and S_{xy}:

$$S_{xx} = \sum_{i=1}^{6}(x_i - 90)^2 = 25 + 0 + 100 + 100 + 25 + 0 = 250$$

$$\begin{aligned} S_{xy} &= \sum_{i=1}^{6}(x_i - 90)(y_i - 400) \\ &= -5 \times -50 + 0 \times 0 + 10 \times 100 + -10 \times -50 + 0 \times 0 \\ &= 250 + 0 + 1000 + 500 + 0 + 0 = 1750 \end{aligned}$$

So $b = S_{xy}/S_{xx} = 1750/250 = 7$ and $a = \bar{y} - b\bar{x} = 400 - 7 \times 90 = -230$.
The equation for the fitted regression line is $\hat{y} = -230 + 7x$.

(b) **Estimate the power usage of the neighborhood for a 110^o Fahrenheit day.**

We simply plug in $x = 110$ to the equation found in the previous part.
So $\hat{y} = -230 + 7 \times 110 = 540kWh$.

3. A company manufacturing pacemakers is testing a new electrode. The electrodes must adhere to a silicone substrate for at least 20 years. The company is going to test the hypothesis that the mean adherence time is 20 years vs. the alternative that it is less than 20 years at the significance level $\alpha = 0.05$. The experiment will be conducted with a sample of 25 volunteers. Assume that the population distribution for the adherence time is approximately normally distributed.

   The average adherence time for the pacemakers in the 25 volunteers is found to be 18.8 years and the standard deviation of the sample is found to be 3 years.

   (a) **Is the null hypothesis rejected?**

   The hypothesis are

$$H_o : \mu = 20$$

$$H_1 : \mu < 20$$

Since the population variance is unknown, the sampling distribution for $\bar{X}$ is the t-distribution. The degrees of freedom is $v = n - 1 = 24$. Therefore the null hypothesis (H_o) is rejected when

$$\bar{X} < \mu_o - t_\alpha \frac{S}{\sqrt{n}} = 20 - 1.711 \times \frac{3}{5} = 18.97$$

Since $\bar{x} = 18.8$ is less than the critical value 18.97 H_o is rejected.

(b) **If the company wants to decrease the probability of making a type I error without increasing the sample size, should the critical value be increased or decreased? Justify your answer.**

Decrease the critical value. The probability of a type I error is the probability of rejecting H_o when it is true. Therefore it can always be decreased by decreasing the size of the critical region (the region where H_o is rejected). Decreasing the critical region corresponds to decreasing the critical value in this case since we reject H_o when $\bar{X}$ is less than the critical value.

(c) **Find the 95% confidence interval for the population variance σ^2.**

The confidence interval for σ^2 is given as

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}$$

The sample standard deviation is given as 3 so $S^2 = 9$. The degrees of freedom is $v = n - 1 = 24$. For 95% confidence interval $\alpha/2 = 0.025$. We find that $\chi^2_{0.025} = 39.364$ and $\chi^2_{0.975} = 12.401$. So

$$\frac{24 \times 9}{39.364} < \sigma^2 < \frac{24 \times 9}{12.401}$$

$$5.487 < \sigma^2 < 17.418$$

4. (a) **An electrical engineer wants to study the mean melting point of a certain metal alloy used in soldering. Based on his knowledge of the population standard deviation, the engineer computes that a minimum sample size of 50 is needed if he wants to be 95% confident that the error between the sample mean and the population mean will not exceed $\pm 2^o C$. What should the minimum sample size be if he wants to be 95% confident that the error between the sample mean and the population mean will not exceed $\pm 1^o C$?**

The minimum number of samples required to achieve a maximum error $e = 2$ is given by

$$n = \left(\frac{z_{\alpha/2}\sigma}{2}\right)^2 = 50$$

When we change the maximum error to $e = 1$ we have

$$n = \left(\frac{z_{\alpha/2}\sigma}{1}\right)^2 = 4 \times \left(\frac{z_{\alpha/2}\sigma}{2}\right)^2 = 4 \times 50 = 200$$

Therefore if e is halved (changed from 2 to 1) then n will be quadrupled. So we would need $n = 4 \times 50 = 200$. Note that we don't need to know the value of σ for this computation.

Another electrical engineer wants to compare the mean melting points for two different metal alloys used in soldering. He collects a sample of size $n_1 = 36$ for the melting point of alloy 1 and finds that $\bar{x}_1 = 185^o C$. He collects a sample of size $n_2 = 64$ for the melting point of alloy 2 and finds that $\bar{x}_2 = 185^o C$. Assume that the population standard deviation for the melting point of the first alloy is $\sigma_1 = 3^o C$ and the population standard deviation for the melting point of the second alloy is $\sigma_2 = 4^o C$. Compute a 99% confidence interval for the difference of the means $\mu_1 - \mu_2$.

Since the population variances are known the confidence interval for the difference of the means is given as

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For a 99% confidence interval $\alpha/2 = 0.005$ and we find that $z_{0.005} = 2.575$ from Table A.3 (average of the two z-values closest to 0.005). So

$$(185 - 185) - 2.575 \times \sqrt{\frac{9}{36} + \frac{16}{64}} < \mu_1 - \mu_2 < (185 - 185) + 2.575 \times \sqrt{\frac{9}{36} + \frac{16}{64}}$$

$$-1.82 < \mu_1 - \mu_2 < 1.82$$