Homework #5

1. **(12 points)** All cigarettes presently on the market have an average nicotine content of 1.6 mg per cigarette. A company that produces cigarettes want to test if the average nicotine content of a cigarette is 1.6 mg. To test this, a sample of 36 of the company's cigarettes were analyzed.

   a) If it is known that the standard deviation of a cigarette's nicotine content is 0.3 mg, what conclusions can be drawn, at the 1 percent level of significance, if the average nicotine content of the 36 cigarettes is 1.475?

   Answer: Two-tailed test $\begin{Bmatrix} H_0 : \mu = 1.6 \\ H_1 : \mu \neq 1.6 \end{Bmatrix}$. The z-statistic is $z^* = \dfrac{1.475 - 1.6}{0.3/6} = -2.5$.
   The rejection region is located between $[z_{-\alpha/2} = -2.58, z_{\alpha/2} = 2.58]$. Since $z_{-\alpha/2} < z^* < z_{\alpha/2}$, we cannot reject the null hypothesis at the 1% significance level.

   b) What is the P-value for the hypothesis test in (a)?

   Answer: P value = $2 * P(z < -2.5) = 0.0124 > 0.01$

2. **(12 points)** In a sample of 500 users of a computer program, 162 said they are satisfied. Construct a 95% confidence interval for the population proportion.

   Answer: The z-statistic $z = \dfrac{P - p}{\sqrt{\dfrac{P(1-P)}{n}}}$ where $P = 162/500 = 0.324$ is the sample mean estimate. The confidence interval is given by ($\alpha = 0.05$)

   $$\text{Prob}(P - z_{\alpha/2}\sqrt{P(1-P)/n} < p < P + z_{\alpha/2}\sqrt{P(1-P)/n}) = 0.95$$
   $$\Rightarrow \text{Prob}(0.283 < p < 0.365) = 0.95$$

3. **(12 points)** The table below shows the number of students absent from school on particular days in the week.

Day	M	Tu	W	Th	F
Number	125	88	85	94	108

   a) Find the expected frequencies (expected numbers of absent students) if it is assumed that students are equally likely to be absent on each working day.

Answer: If the number of absentees is independent of the day of the week, the expected number of students is 100 (a total of 500 absentees in 5 days).

Day	**M**	**Tu**	**W**	**Th**	**F**
Recorded #	125	88	85	94	108
Expected #	100	100	100	100	100

b) Test, at the significance level of 5%, the null hypothesis that students are equally likely to be absent on each working day.

Answer: The null hypothesis is that the number of absentees follows uniform distribution for the 5 days of a week and the alternative hypothesis is that the number of absentees does not follow the uniform distribution. For a chi-square goodness of fit test, the test statistic is given by

$$\chi_0^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = 10.94$$

The number of degrees of freedom is 5-1=4. Since $\chi_0^2 = 10.94$ (P-value 0.027) is larger than the critical value $\chi_{0.05,4}^2 = 9.49$, we will reject the null hypothesis and accept the alternative hypothesis.

4. **(10 points)** Cancer researchers have tested the effectiveness of a new drug on rats. A group of rats with tumors were given the drug and initially their tumors shrank to be undetectable. Listed below are the times (in days) until cancer in rats developed resistance to the drug so that tumors reappeared.

101, 104, 77, 104, 96, 82, 70, 89, 91, 103, 93, 85, 104, 104, 81, 67, 104, 87, 104, 89, 78, 104, 86, 76, 103, 102, 80, 45, 94, 104, 104, 76, 80, 72, 73

Use chi-squared distribution to calculate the 95% confidence interval on the population standard deviation of these times of tumor reappearance.

Answer: Let σ and S denote population and sample standard deviation, respectively. The sample size is 35 and the sample standard deviation is 14.41. Using the chi-square statistic, the confidence interval is given by ($\chi^2_{0.975,34} = 52$, $\chi^2_{0.025,34} = 20$)
/.÷≤≥÷

$$[\sqrt{\frac{(35-1)14.41^2}{\chi^2_{0.975,34}}},\ \sqrt{\frac{(35-1)14.41^2}{\chi^2_{0.025,34}}}] = [11.66, 18.87]$$

5. **(10 points)** Three identical six-sided dice, each with faces marked 1 to 6, are rolled 100 times. At each rolling, a record is made of the number of dice whose score is 5 or 6. The results are as follows.

# of dice with score 5 or 6	0	1	2	3
# of rolls	31	41	19	9

Please explain whether these results are consistent with the **unbiased dice** hypothesis. Use chi-squared goodness of fit test with α=3% to justify your conclusions.

Answer: Assuming the dice are unbiased, then $P(\text{score} = 5 \text{ or } 6) = \frac{1}{3}$. Then the distribution of the number of scores that are equal to 5 or 6 (denoted as X) in 3 trials is given by

$$X \sim B(3, 1/3)$$

Using the binomial distribution, we can find the expected count for each value of X

# of scores = 5 or 6	0	1	2	3
Recorded #	31	41	19	9
Expected #	29.63	44.44	22.22	3.71

The null hypothesis is that the number of scores equal to 5 or 6 follows binomial distribution and the alternative hypothesis is that the number of scores equal to 5 or 6 does not follow binomial distribution. For a chi-square goodness of fit test, the test statistic is given by

$$\chi_0^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = 8.34$$

The number of freedom is 4-1=3. Since $\chi_0^2 = 8.34$ is smaller than the critical value $\chi^2_{0.03,3} = 8.95$, we cannot reject the null hypothesis. Alternatively, we might believe that the observed distribution can be described by a binomial model.

6. **(14 points)** A polymer is manufactured in a batch chemical process and viscosity measurements are normally made on each batch. The sample mean and standard deviation of 50 batch viscosity measurements are 750.2 and 19.13, respectively. A process change is made which involves switching the type of catalyst used in the process. Following the process change, 30 batch viscosity measurements are taken and the sample mean and standard deviation are 756.88 and 21.28, respectively.

a) At the significance level of 10%, can you conclude that there is a significant difference between population means before and after the process was changed?

Answer: The null hypothesis is that there is no difference of the population mean before and after the process change. The test statistic is given by

$$t = \frac{750.2 - 756.88}{\sqrt{\frac{19.13^2}{50} + \frac{21.28^2}{30}}} = -1.411$$

The number of degrees of freedom is 50-1+30-1=78. Since $t = -1.411$ is outside of the rejection regions $[-\infty, -t_{0.05,78}] \cup [t_{0.05,78}, +\infty]$ where $t_{0.05,78} = 1.66$, we may conclude that there is no significant difference for the population mean at the significance level of 0.1.

Using the normal approximation, the rejection region is $[-\infty, -z_{0.05}] \cup [z_{0.05}, +\infty]$ where $z_{0.05} = 1.64$. We may still have the same conclusion.

b) Find the P-value of the null hypothesis that process change did not have any impact on the population mean.

Answer: Using t-distribution, the P-value is 0.1622 and using normal approximation, the P value is 0.1582.

c) Find a 90% confidence interval on the difference in mean batch viscosities resulting from the process change.

Answer: The t-statistic $t = \frac{750.2 - 756.88 - \Delta}{\sqrt{\frac{19.13^2}{50} + \frac{21.28^2}{30}}} = \frac{-6.68 - \Delta}{4.73}$ where Δ is the population mean difference. The confidence interval is given by

$[-6.68 - t_{0.05,78}4.73, \quad -6.68 + t_{0.05,78}4.73] = [-14.53, 1.17]$.

Using the normal approximation, the confidence internal is given by

$[-6.68 - z_{0.05} 4.73, \quad -6.68 + z_{0.05} 4.73] = [-14.44, 1.08]$.