University of Texas at Austin, Department of Mathematics
M358K - Applied Statistics

In-term exam III

Name:

UTeid:

Free-response problems.

Problem 3.1. (15 points) Every jug of *Cinnamon Cider* is labeled to contain 1 gallon of sweet holiday drink. You are the quantity control expert in charge of testing whether the actual amount of cider deviates from the label. You do some research and you are comfortable assuming that the amount of cider per jug is normally distributed with a known standard deviation of 10 oz. Your plan is to randomly collect 64 jugs and measure their contents. What is your rejection region with the significance level of 4%?

Note: One gallon is 128 oz.

Solution: We are testing

$$H_0 : \mu = \mu_0 = 128 \quad vs. \quad H_a : \mu \neq \mu_0 = 128.$$

So, remembering that the amount of cider in a jug cannot be negative, the form of our rejection region is, in our usual notation and in real units,

$$RR = \left[0, \mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] \cup \left[\mu_0 - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \infty\right]$$

where $z_{\alpha/2} = \Phi^{-1}(0.02) = -2.053749$. We conclude that

$$RR = \left[0, 128 - 2.053749\left(\frac{10}{\sqrt{64}}\right)\right] \cup \left[128 + 2.053749\left(\frac{10}{\sqrt{64}}\right), \infty\right) = [0, 128 - 2.567186] \cup [128 + 2.567186, \infty)$$
$$= [0, 125.4328] \cup [130.567186, \infty).$$

Problem 3.2. (15 points) An eggnog fountain is supposed to be calibrated to dispense 200 ml of delicious libation per mug. Of course, the amount dispensed is not exact. You model the amount actually dispensed using the normal distribution with standard deviation 16 ml. Periodically, the fountain is tested to see if it's correctly calibrated. Each time, a sample of 16 mugs is taken and measured carefully. With a particular significance level, the rejection region is the complement of the interval $(190, 210)$. What is the power of the test at the alternative mean $\mu_a = 212$ ml?

Solution: Under the alternative $\mu_a = 212$, the distribution of the sample mean is

$$\bar{X} \sim Normal\left(mean = 212, sd = \frac{16}{\sqrt{16}} = \frac{16}{4} = 4\right).$$

The probability of making a Type II Error is

$$\beta = \mathbb{P}_{\mu_a}[190 < \bar{X} < 210]$$

In 'R', we use

$$pnorm(210, 212, 4) - pnorm(190, 212, 4)$$

to get $\beta = 0.3085375$ So, the power of the test is $1 - \beta = 0.6914625$.

Problem 3.3. (15 points) We want to compare the proportions of people from town A (pop. 25000) and people from town B (pop. 50000) whose favorite music genre is country music. Two polls are conducted. In town A, 120 out of the total of 200 people like country music best. In town B, 240 out of the total of 500 people like country music best. You conduct a hypothesis test for whether there is a difference in population proportions. Which $p-$value would you report?

Solution: Let p_A be the probability that a randomly chosen person from town A prefers country music and let p_B be the probability that a randomly chosen person from town B prefers country music. We are testing

$$H_0 : p_A = p_B \quad vs. \quad H_a : p_A \neq p_B.$$

The point estimates for the two proportions are $\hat{p}_A = \frac{120}{200} = 0.60$ and $\hat{p}_B = \frac{240}{500} = 0.48$. The pooled estimate of the proportion is

$$\hat{p} = \frac{120+240}{200+500} = 0.5143$$

The observed value of the $z-$statistic, under the null, is

$$z = \frac{0.60 - 0.48}{\sqrt{(0.5143)(1-0.5143)\left(\frac{1}{200}+\frac{1}{500}\right)}} = 2.87$$

So, the $p-$value is $2\Phi(-2.87) = 2(0.0021) = 0.0042$.

Problem 3.4. (5 points) You want to test whether a six-sided die is fair. Here are the observed counts in 120 rolls of the die:

Side	1	2	3	4	5	6
Count	20	22	17	19	24	18

What is your $p-$value? *Give bounds from the χ^2-table if not using R.*

Solution: Let $p_i, i = 1, \dots, 6$ be the probability that the die falls on i. We are testing

$$\begin{aligned} H_0: &\quad p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6} \\ &\quad vs. \\ H_a: &\quad \text{At least one of the probabilities } p_i \text{ is different from } \frac{1}{6}. \end{aligned}$$

The observed value of the χ^2-statistic is

$$q^2 = \frac{1}{20}((20-20)^2 + (22-20)^2 + (17-20)^2 + (19-20)^2 + (24-20)^2 + (18-20)^2) = 1.7.$$

With $6-1 = 5$ degrees of freedom, using the χ^2-table, we see that the $p-$value is more than 0.30.

Problem 3.5. (10 points) A survey of 1000 students in Austin found that 274 chose Pikachu as their favorite Pokemon. In another survey of 760 students in Pittsburgh, 240 chose Pikachu. Find the 95%-confidence interval for the difference in the two population proportions.

Solution: Let the proportion of Pikachu fans in Austin be p_A and let the proportion of Pikachu fans in Pittsburgh be p_P. Then, the point estimates of the two proportions equal

$$\hat{p}_A = \frac{274}{1000} = 0.274 \quad \text{and} \quad \hat{p}_P = \frac{240}{760} = 0.3158.$$

The critical value associated with the 95%-confidence level is 1.96. The standard error equals

$$\sqrt{\frac{0.274(1-0.274)}{1000} + \frac{0.3158(1-0.3158)}{760}} = 0.022.$$

So, the confidence interval is $(0.274 - 0.3158) \pm 1.96(0.022)$, i.e., $(-0.0849, 0.0013)$

Problem 3.6. (5 points) *Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*

A study was conducted to determine whether there is an association between being for or against tax reform and income level. The results are displayed in the following table:

	Low	Medium	High	Total
For	182	213	203	598
Against	154	138	110	402
Total	336	351	313	1000

Your goal is to test whether being for or against the tax reform is independent from income level. The observed value of the relevant test statistic is 7.85. What is your $p-$value for this hypothesis test (if using tables, provide bounds)?

Solution: The distribution of the test statistic is approximately χ^2 with $(3-1)(2-1) = 2$ degrees of freedom. Consulting the χ^2-table, we see that the given observed value of the test statistic is between the critical values $\chi^2_{0.01}(df = 2)$ and $\chi^2_{0.025}(df = 2)$. So, the $p-$value is between 0.01 and 0.025.

"1-pchisq(7.85, df=2)" in 'R', gives us 0.01974214.