MATH 476 Statistics

Fred J. Hickernell **Spring 2010**
Test **Thursday, April 15**

Instructions:

i. *This test consists of* THREE *questions. Answer all of them.*

ii. *The time allowed for this test is 75 minutes.*

iii. *The data and situations portrayed in this test are fictitious, but realistic.*

iv. *This test is closed book, but you may use 2 double-sided letter-size sheets of notes.*

v. *Calculators, even of the programmable variety, are allowed. Computers using JMP or MATLAB, are also allowed. No internet access, web browsing, email, chat, etc. is allowed.*

vi. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*

1. (40 marks)
There is concern that proportion of Chinese living in Chinatown that are turning in their census forms, θ, is lower than rest of Chicagoland, which has a compliance of 65%.

a) What null and alternative hypothesis would you propose if you are the Chinatown Service Center, and you hope to identify low compliance among the Chinese that needs urgent attention?

Answer:

$$H_0 : \theta = 65\%, \quad H_a : \theta < 65\%.$$

b) You find a way to survey people and accurately determine the number who have turned in their census forms. If n is the (large) number of people surveyed at random, N is the random number that turned in their census forms, and $\hat{\theta} = N/n$ is the estimated proportion of those complying, then how small must $\hat{\theta}$ be (in terms of n) to reject your null hypothesis and maintain a Type I error, α, equal to 5%?

Answer: Let θ_c be the critical value. The probability of rejecting H_0 even if H_0 is true is

$$5\% = \alpha = \Pr(\hat{\theta} < \theta_c | \theta = 65\%) = \Pr\left(\frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)/n}} < \frac{\theta_c - \theta}{\sqrt{\theta(1-\theta)/n}} \middle| \theta = 65\% \right),$$

which holds provided

$$-1.645 = \frac{\theta_c - \theta}{\sqrt{\theta(1-\theta)/n}},$$
$$\theta_c = \theta - 1.645\sqrt{\theta(1-\theta)/n} = 0.65 - \frac{0.7846}{\sqrt{n}}.$$

If $\theta < \theta_c$, then the null hypothesis is rejected. The Type I error is 5%.

c) How large must n be so that the Type II error, β, will be no greater than 20% if $\theta = 55\%$ for the test described in the previous part?

Answer: The probability of not *rejecting H_0 if $\theta = 55\%$ is*

$$\beta = \Pr(\hat{\theta} \geq \theta_c | \theta = 55\%) = \Pr\left(\frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)/n}} \geq \frac{\theta_c - \theta}{\sqrt{\theta(1-\theta)/n}} \middle| \theta = 55\%\right) \leq 20\%,$$

which holds provided

$$\begin{aligned} 0.8416 &\leq \frac{\theta_c - 0.55}{\sqrt{0.55 \times 0.45/n}} = \frac{0.1 - 0.7846/\sqrt{n}}{\sqrt{0.55 \times 0.45/n}} \\ &= 0.2010\sqrt{n} - 1.5771. \\ n &\geq \left(\frac{0.8416 + 1.5771}{0.2010}\right)^2 \approx 145. \end{aligned}$$

d) For the sample size in the previous part, what is the critical value of θ below which you may reject the null hypothesis?

Answer:

$$\theta_c = 0.65 - \frac{0.7846}{\sqrt{145}} = 58.48\%.$$

2. (30 marks)
The pomelos at Hong Kong Market, Mayflower, and Richland have the following masses:

Grocery Store	Pomelo masses (in grams)								
Hong Kong Market	256	342	512	441	484	291	356	412	378
Mayflower	277	355	675	447	548	456	399	385	514
Richland	461	355	343	541	289	290	578	543	441

Is there evidence that the different stores carry pomelos with different average masses?

Answer: Input all the pomelo masses in one column in JMP. Label them by grocery store in the second column. Choose `Fit Y by X` *to do a oneway ANOVA. This tests the null hypothesis of all widgets produced by the three factories having the same average lifetimes. The p-value is 43%, which leads one to not reject the null hypothesis and conclude that the mean pomelo masses from the three factories are not necessarily different.*

3. (30 marks)
Sam Ho's Chinese BBQ, open Monday through Saturday, sells roasted pork rice boxes. They would like to sell the same number each day on average. One week of data shows the following sales totals:

Day of the week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Number Sold	123	115	132	118	127	119

Does this data indicate a difference in average rice box sales for different days of the week?

Answer: In JMP input the six days of the week into a column labeled `Day of the Week`*. In the second column, labeled* `Number Sold` *input the corresponding number of rice boxes sold. Then choose* `Analyze/Distribution` *and put* `Day of the Week` *into* `Y` *and* `Number Sold` *into* `Frequency` *and submit. Then choose* `Test Probabilities` *and input* 1 *for each day's probability. Since the* $p = 90\%$*, there is no compelling reason to reject the null hypothesis of no difference in average number sold for different day of the week.*