STAC51 TUT02

Feb 11, 2021

$$\underline{X} \sim \text{Multinomial}\,(n=9,\ p_1=0.2,\ p_2=0.4,\ p_3=0.3,\ p_4=0.1)$$

$$SD(X_1) = \sqrt{9 \times 0.2 \times 0.8}$$

$$Var(X_1) = np_1(1-p_1)$$

A large introductory statistics class has students from all levels. 20% of them are first-year students, 40% second-year students, 30% third-year students and the remaining 10% are fourth-year students. We are going to select a random sample of 9 students from this class.

What is the standard deviation of the number of first-year students in the sample?

1.20	32 respondents	48 %	✓
Something Else	31 respondents	46 %	
No Answer	4 respondents	6 %	

$\frac{x}{n} = \hat{\pi} = 0.75$

$x = 27 \quad n = 36$

$\boxed{\hat{\pi}} \pm \underbrace{z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}_{}$

binom. confint

ME

0.6 0.75 0.9

$\hat{\pi} - ME$ $\hat{\pi}$ $\hat{\pi} + ME$

Based on a sample of 36 patients, a researcher calculated a 95 percent Wald confidence interval for the success rate of a new vaccine for a disease to be (0.6, 0.9). Calculate the **95 percent score confidence interval** for the success rate of this vaccine. You can use the built-in R function. Just report the lower limit of the interval. Round the number to the second decimal place. For example, 2.4756 will be **2.48**.

0.59	29 respondents	**43 %**	✓
Something Else	31 respondents	46 %	
No Answer	7 respondents	10 %	

For years, brand awareness for Big Red chewing gum has been stuck at about 6%, meaning that about 6% of consumers who chew gum say they remember hearing about Big Red gum. The marketing department is planning an advertising campaign to increase brand awareness, in the hope that increased brand awareness will lead to increased sales. After the campaign was running for a few weeks, they wanted to test whether the brand awareness is now different than 6%. To test this, they interviewed a random sample of **200** gum chewers and found that **twenty** had heard of Big Red.

1) What are the appropriate null hypothesis and alternative hypothesis corresponding to the main question in this study? (1 point)

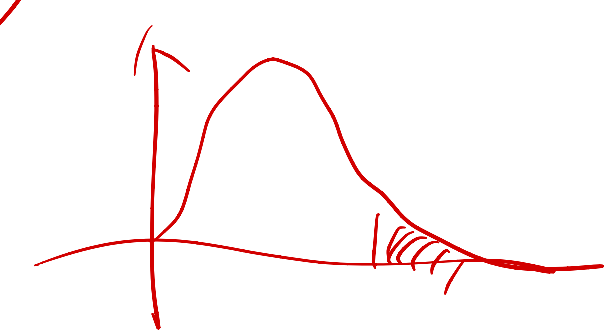
$H_0: \pi = 0.06$ $\quad$ $H_a: \pi \neq 0.06$

2) Compute the likelihood ratio test statistic value (2 points) (show the R codes to compute this test statistic value)

$$2 \times \Sigma \left(obs \times \log\left(\frac{obs}{exp}\right)\right) = \boxed{4.7783}$$

3) Compute the p-value and make a conclusion based on the p-value (1 point)

pchisq(q = 4.7783, df = 1, lower.tail = F)



For years, brand awareness for Big Red chewing gum has been stuck at about 6%, meaning that about 6% of consumers who chew gum say they remember hearing about Big Red gum. The marketing department is planning an advertising campaign to increase brand awareness, in the hope that increased brand awareness will lead to increased sales. After the campaign was running for a few weeks, they wanted to test whether the brand awareness is now different than 6%. To test this, they interviewed a random sample of **200** gum chewers and found that **twenty** had heard of Big Red.

Construct a likelihood-ratio-based 90 percent confidence interval for π .

Submit your R codes (**do not use "binom" package**).

ALWAYS > 0.95

when n > 40, the "true confidence" or "coverage" level of 95% Clopper Pearson CI is lower than 95%.

True	20 respondents	30 %	
False	45 respondents	**67 %**	✓
No Answer	2 respondents	3 %	

wider CI.

. The following data consist of a random sample of 793 individuals who were involved in bicycle accidents during a specified one-year period.

	Wearing Helmet		
Head Injury	Yes	No	Total
Yes	17	218	235
No	130	428	558
Total	147	646	793
	n_1	n_2	

(a) Construct a 90% confidence interval for the difference of proportions, and interpret.
(b) Construct a 90% confidence interval for odds ratio, and interpret.
(c) Conduct a test of statistical independence. Interpret.

RR..

$p_1 = P(\text{Head Injury} = \text{YES} \mid \text{Helmet})$

$p_2 = P(\quad \mid \text{No Helmet})$

$\hat{p}_1 = \frac{17}{147}, \quad \hat{p}_2 = \frac{218}{646}$

$\hat{p}_1 - \hat{p}_2 = -0.2218$

$\hat{\sigma}(p_1 - p_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.0323$

C.I. : $(\hat{p}_1 - \hat{p}_2) \pm \hat{\sigma}(p_1 - p_2) \underbrace{Z_{0.05}}_{1.645} = (-0.2749, -0.1687)$

1-OR ✓

X	Y

① ② logit(x) ∈

A	B
C	D

$\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}$

$\frac{AD}{BC} = \hat{OR}$.

$p \in [0, 1)$

$Var[\log(\hat{OR})]$, Delta Method $\hat{RR}$

$\hat{OR} = \frac{17 \times 428}{218 \times 130} = 0.2567$ $\Rightarrow$ $\frac{p}{1-p}$

$X \sim N(\mu, \sigma^2) \Rightarrow$ 1-to-1 function $g(x)$

$g(x) \sim N(g(\mu), [g'(x)]^2 \sigma^2)$ ∈

$\hat{SE}(g(x)) = \sqrt{\frac{1}{17}+\frac{1}{218}+\frac{1}{130}+\frac{1}{428}} = 0.2710$

90% CI log OR

0

$\log(\hat{OR}) \pm$

$\exp\left(\log(0.2567) \pm 1.645 \times 0.2710 = [-1.8055, -0.9139]\right)$

$\boxed{OR} \Rightarrow \mp$ H_0 $\theta=1$ $\log(\theta)=0$, Let's assume that a university consists only two professional schools: Law school and business school and we are interested in studying the association between the two variables X = "gender" and Y = "admission decision". Z = "school" is another variable that might influence the admission decision. X has two levels: 1 = Male, 2 = Female. Y has two levels: 1 = accepted, 2 = rejected. Z has two levels: 1 = law school, 2 = business school. The following are the marginal tables by school and gender,

Law	Accepted	Rejected	Total
Males	10	90	100
Females	100	200	300
Total	110	290	400

Business	Accepted	Rejected	Total
Males	480	120	600
Females	180	20	200
Total	660	140	800

(a) Calculate the crude (Marginal) odds ratio $\hat{\theta}_{XY}$. [4 Marks]

(b) Let the conditional odds ratios for the kth table be $\hat{\theta}_k = \frac{a_k d_k}{b_k c_k}$. The Cochran-Mantel-Haenszel odds ratio is calculated as $\hat{\theta}_{CMH} = \frac{\sum_k w_k \hat{\theta}_k}{\sum_k w_k} = \frac{\sum_k \frac{a_k d_k}{n_k}}{\sum_k \frac{b_k c_k}{n_k}}$.

Show that, $w_k = \frac{b_k c_k}{n_k}$ [3 Marks]

(c) Calculate the Cochran-Mantel-Haenszel odds ratio $\hat{\theta}_{CMH}$ for the table above [4 Marks]

(d) Calculate the CMH test statistics

$W_k \hat{\theta} = \frac{a_k d_k}{n_k}$

(a) 490 210
280 220

$\hat{\theta} = \frac{490 \times 220}{210 \times 280} = 1.833.$

$W_k = \frac{a_k d_k}{n_k \hat{\theta}_k}$

(b) $\hat{\theta}_{CMH} = \frac{\sum_k W_k \hat{\theta}_k}{\sum W_k}$

$\hat{\theta}_k = \frac{a_k d_k}{b_k c_k}$

$W_k = \frac{b_k c_k}{n_k}$

If Y_1 and Y_2 are independent Poisson random variable with parameters μ_1 and μ_2 respectively. Then find the conditional distribution of Y_1 given $Y_1 + Y_2 = n$. That is calculate $P(Y_1 = k|Y_1 + Y_2 = n)$. [5 Marks]

binom$\left(n = n,\ \pi = \frac{\mu_1}{\mu_1 + \mu_2}\right)$

Fisher Exact

If $Y_1 \sim \text{Bin}(n_1, \pi)$ and $Y_2 \sim \text{Bin}(n_2, \pi)$ are independent. Then find the conditional distribution of Y_1 given $Y_1 + Y_2 = m$. That is calculate $P(Y_1 = k | Y_1 + Y_2 = m)$. [5 Marks]