/

Advanced Statistics: Theory and Methods

Probability and Statistics Baseline Assessment

Name Abhishek Thakor ID Number ______

1. With increased communication using the Internet, there has been an explosion of chat rooms. Here individuals can exchange opinions on various topics of interest. A side effect of such conversation is "flaming", i.e. negative criticism of others' contributions to the conversation.

M. Dsilva and colleagues studied this phenomenon ("Criticism on the Internet: An analysis of Participant Reactions", *Communication Research Reports*, 1998). The investigators were interested in the effect that personal criticism has on an individual. Would being criticized make one more likely to criticize others? The data are provided in the table below:

	Has Been Personally Criticized (O)	Has Not Been Personally Criticized O^c	Total
Has Criticized others (C)	19	8	27
Has Not Criticized others C^c	23	143	166
Total	42	151	193

Would being criticized make one more likely to criticize others?
O C

Conditional Probability

$$P(C|O) = \frac{P(C \cap O)}{P(O)} = \frac{19/193}{42/193} = \frac{19}{42}$$

$$P(C^c|O) = \frac{P(C^c \cap O)}{P(O)} = \frac{23/193}{42/193} = \frac{23}{42}$$

$P(C|O) < P(C^c|O)$ $\Rightarrow$ It is more likely that an individual does not criticize other given that he was personally criticized.

2. A manufacturer claims that its drug test will detect steroid use (that is, show positive for an athlete who uses steroids) 95% of the time. The company also releases information that 15% of all steroid-free individuals also test positive (the false positive rate). Based on anonymous surveys of amateur wrestlers, it is estimated that 10% use steroids. At a recent event, one of the wrestlers has tested positive. What is the probability that he uses steroids?

Bayes' Theorem

A: Wrestler uses steroids
B: Wrestler tests positive

$P(A) = 0.10$, $P(B|A) = 0.95$, $P(B|A^c) = 0.15$
$P(A^c) = 0.90$

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)}$$

$$= \frac{0.95 \times 0.10}{0.95 \times 0.10 + 0.15 \times 0.90}$$

$P(A|B) = 0.413$

3. The Public Health Department in Punjab is interested in understanding the effect of zinc supplements during pregnancy on birth weights.

(a) How would you design this study? Define the target population, sampling scheme, and data collection plan.

(b) Define the parameter(s) of interest.

(c) What are some confounding variables that may affect your results?

Sampling Theory

(a) Target pop. → Pregnant women in Punjab.

Sampling Scheme → A simple random sample of 1000 pregnant women from the target population shall be selected from each region in Punjab.

Data Collection Plan → Participants will be randomly assigned to either intervention group or the control group. The ~~data~~ intervention group will recieve Zinc supplements daily during pregnancy, while the control group will recieve a placebo.

Following data shall be further collected:-

- Maternal age
- Parity
- Gestational age at delivery
- Birth weights of the infant
- Zinc supplement status
- Medical history
- Dietary habits
- Demographic characteristics
- Socio-economic status

(b) Parameters: Mean birth weights of the intervention and control groups and their difference.

(c) Confounding Variables: Maternal age, parity, socio-economic status, demographic characteristics, maternal education, maternal heath status, multiple pregnancies, gestational age at delivery, infant sex, maternal nutrition, preterm birth, etc.

4. Here are results on a new experimental nasal spray vaccine: In a trial involving 1602 children, 1070 received the vaccine and 532 received a placebo. In the vaccine group, 14 children developed the flu compared to 95 in the placebo group.

(a) State the null and alternative hypothesis. (Define the parameters of interest)

(b) What are the consequences of making a Type I and Type II errors?

(c) Is there sufficient evidence to suggest that the vaccine is effective in reducing the chances of getting the flu?

Hypothesis Testing - Test of significance of differences of proportion.

(a) $H_0: p_1 = p_2$; the prob. of getting the flu in vaccine group and in control group are the same.

$H_a: p_1 < p_2$, the prob. of getting the flu in vaccine group is more effective (less likely) than getting the flu in control group.

(b) Type I error; reject H_0, when H_0 should be accepted. That is, we conclude that vaccine is effective when it's not. So then people may ~~..~~ be vaccinated for flu unnecessarily.

Type II error; accept H_0, when H_0 should be rejected. That is, we conclude that the vaccine ~~is~~ is not effective when actually it is effective. So then further trials may fail to be performed when in fact a beneficial treatment already exists.

(c) Yes there is sufficient evidence:

$n_1 = 1070,\ n_2 = 532,\ \hat{p}_1 = \frac{14}{1070} = 0.013,\ \hat{p}_2 = \frac{95}{532} = 0.179$

$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} = \frac{14+95}{1070+532} = \frac{109}{1602} = 0.068,\ \hat{q} = 1-\hat{p} = 0.932$

z score $= Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_1}\right)}} = \frac{0.013 - 0.179}{\sqrt{0.068 \times 0.932 \times \left(\frac{1}{1070} + \frac{1}{532}\right)}} = \frac{-0.166}{0.0134}$

$Z = -12.388$

Since $|Z| > 3$, we reject hypothesis H_0, so we conclude that the vaccine is significantly effective.