FOR 6520 Winter 2003

Alliant International University

Research II: Data Analysis & Advanced Statistics

http://www.stat.ucla.edu/~dinov/courses_students.html

HOMEWORK 2

Due Date:

Saturday, Feb. 15, 2003

Please, submit your homework right after lecture on the due date. See the HW submission rules. On the front page include the following header.

- (**HW_2_1**) Using the following data for the Mean Absorbance Ration (MAR) scores from the ELISA HIV antibody test answer the following questions (assume the prevalence of HIV in the US is 1%):

MAR	Healthy Donor	HIV
< 2	202	0
2.0 - 2.99	73	2
(test threshold)	———	———
3.0 - 3.99	3	7
4.0 - 4.99	15	7
5.0 - 5.99	2	15
6.0 - 11.99	2	36
> 12.0	0	21
Total:	297	88

  - Identify the false-positive (Type I) **The probability of making a type I (α) error with this set of data is 22/297 = .074 or 7.4% This means a person without HIV is falsely**

identified as having the virus. A false rejection of the null hypothesis has occurred. ***(4 points)***

-
- and false-negative (Type II) errors. **The probability of making a Type II (β) error is 2/88 = .0227 or 2.27%. This means a person with HIV is tested as not having the virus. This is a failure to reject the null hypothesis.** ***(4 points)***
-
- What is the power of the test? **The power of the test is 1-β or 1-.0227 = .9773** ***(4 points)***
-
- P(Negative Test | Not-HIV)=? **275/297 = .9259 or 93% Question reads what is the probability that a person tests negative given that he is not HIV positive?** ***(6 points)***

	Positive Test	Negative Test	Total
HIV	(.98)(.01) = .0098	.9209-.9207 = .0002	.01
Non-HIV	.0791-.0098 = .0692	(.93)(.99) = .9207	.99
Total	.0791	.9209	1.0

-
-
- P(HIV & Positive Test)=? **86/88 = .98 (probability that he has a positive test given that he's HIV) The probability that he has HIV is .01 The intersection between the two probabilities is (.98)(.01) = .0098 or .98% Probability of A&B = P(A|B)*P(B) or can also be P(B|A)*P(A). Both formulas give the same answer.** ***(7 points)***
-
- How do we choose the **best threshold-value** for a test like that? **In this case it is better to have a type I error than a type II error. People are more likely to seek a second opinion if they test positive, and it is best to begin treatment as soon as possible. The current threshold-value of < 3 creates a test with good power and limits the amount of false positives. If the threshold were lowered any more, there would be too large a portion of the population falsely identified as positive, and it would cast doubts on the reliability of the test.** ***(2 points)***
-
- Suppose I chose a random subject from the pool of individuals that participated in this study. What is the probability that the subject is:
  - positive on the MAR test? **108/385 = .2805** ***(4 points)***
  - of Hispanic race? **No way to determine from the data set presented. If the population from which the sample is drawn is known, the probability that a subject is Hispanic will be the same as that of the original population.** ***(2 points)***
  - either HIV-positive & has negative test, **.9209-.9207 = .0002** ***(5 points)***
  - or Non-HIV and has positive test? **.0791-.0098 = .0692** ***(5 points)***
  - **The probability that the person is either positive and has a negative test or is Non-HIV and has a positive test: .0002 + .0692 = .0694 or 6.94%.** ***(7 points)***

- (**HW_2_2**) The Los Angeles Mayor is interested in assessing the population satisfaction with the performance of the Police Department (PD) and the District Attorney (DA) office in one county. You are contracted to do the study and report to the Mayor.
    - Who should you poll? Do you attempt a census or do a well-designed controlled poll/survey? **Various answers are acceptable here as long as the reasoning is sound.** ***(5 points)***
    -
    - What sample size(s) would you use? What criteria you need to satisfy to argue about the validity of your final conclusions? **Large sample size is best. Also you would want to make sure that you are collecting data from a representative sample (truly random and covering all portions of the county).** ***(5 points)***
    -
    - How to collect the survey responses? Do you ask the same people if they are satisfied (S) or dissatisfied (D) with the LAPD and separately the LADA office? **Again, if the reasoning is sound, various answers will be applicable here.** ***(5 points)***
    -
    - Suppose you split your pool of participants into two groups, each expressing their opinion on the PD or DA work. Let your sample sizes be nPD=1,000 and nDA=1,200 and proportion of satisfied citizens be pPD=0.6 and pDA=0.7, respectively. The Mayor wants to commit more resources to the law enforcement agency that does the best job, and replace the Head of the other agency that has not done well in the public's eyes. How to evaluate if there is evidence for statistical differences between these satisfaction rates?

Satisfaction Type	PD	DA
Satisfied	600	840
Dissatisfied	400	360
Total	1000	1200

    - **Independent t-test.** $p_1 = 0.6$, $p_2 = 0.7$
    - **Ho:** $p_1 - p_2 = 0$
    - **Ha:** $p_1 - p_2 < 0$
    - **The assignment did not ask to evaluate the problem, but just how one would evaluate it (method). In order to use the t-test we would need more information.** ***(5 points)***

    - How do you communicate your results to the Mayor's office, to avoid misinterpretation of your analysis? **Explain how the data was collected, depending on level of confidence, explain how certain you are that one group is perceived to be more satisfactory, if that is the result. Or explain that even though one group has a higher percent satisfactory rating, the groups are not different enough to be significant.** ***(5 points)***

- (**HW_2_3**) Suppose the probability of acquittal (A) given DNA match (M) is 0.245. What is the probability that someone has a DNA matching the sample taken at the crime scene, given that the subject was convicted (A^c)? Assume a DNA match could occur randomly only with probability of 0.001 of population and P(conviction | No DNA match) = 0.1.
-

DNA/Conviction?	Acquittal	Conviction	Total
Match	(0.245)(.001)	.	.001
No Match		(0.1)(.999)	.999
Total			1.0

P(Match and Acquittal) = P(Acquittal|Match)P(Match) = (.245)(.001) = .000245

P(Conviction and Match) = .001 - .000245 = .000755

P(Conviction and No Match) = (0.1)(0.999) = .0999

P(No Match and Acquittal) = .999-.0999 = .8991

DNA/Conviction?	Acquittal	Conviction	Total
Match	.000245	.000755	.001
No Match	.8991	.0999	.999
Total	.899345	.100645	1.0

P(Match|Conviction) = P(Match and Conviction)/P(Conviction)

.000755/.100645 = .0075 or .75% ***(15 points work, 10 points correct answer = 25 points)***

Last modified on 02/16/2003 10:02:12 by dinov@stat.ucla.edu.

Ivo D. Dinov, Ph.D., Departments of Statistics and Neurology, UCLA School of Medicine