Midterm Exam
Statistical Methods and Models - Math 410, Fall 2011
October 28, 2011

You may use a calculator, and you may bring in one sheet (8.5" by 11" or A4) of notes. Otherwise closed book. You may use the backs of the exam sheets if you need more space for answers.

(1) Let E and F be possible outcomes of an experiment. First state what it mean for E to be independent of F. Then prove that if E is independent of F then E is independent of F'.

SOLUTION: *E independent of F means*

$$P(E|F) = P(E).$$

We assume E is independent of F, so

$$P(E) = P(E|F) = \frac{P(E \cap F)}{P(F)} \text{ so } P(E \cap F) = P(E)P(F).$$

Then we see

$$P(E) = P(E \cap F) + P(E \cap F') = P(E) \cdot P(F) + P(E \cap F')$$

so solving for $P(E \cap F')$, we get

$$P(E \cap F') = P(E) - P(E \cap F) = P(E)(1 - P(F)) = P(E)P(F').$$

It follows that

$$P(E|F') = \frac{P(E \cap F')}{P(F')} = P(E).$$

(2) Suppose you have a screening test for the Andromeda virus. Suppose the prevalence of Andromeda is .008, the sensitivity of the test is .97 and the specificity is .95. What is the preditictive value positive?

In plain language, what does this number mean?

SOLUTION:

$$PV^+ = P(\text{infected} \mid \text{positive}) = \frac{.008 \cdot .97}{.008 \cdot .97 + .992 \cdot (1 - .95)} = .135286.$$

This tells us that roughly 14% of the positive results correspond to people who are actually infected with the virus.

(3) Rain is considered "acid rain" if the pH is below 5.2. The following observations of pH levels of rain were made in the Washington Cascades on 16 dates.

$$4.73, 4.79, 4.87, 4.88, 5.04, 5.06, 5.07, 5.09, 5.11, 5.16, 5.18, 5.21, 5.23, 5.24, 5.25, 5.25$$

Use a sign test to determine if there sufficient evidence at the .05 signifancance level to conclude rain is acid more often than not.

SOLUTION: *There are 16 subjects. 11 of the 16 are below the threshold, and 5 are above the threshold.*

(a) *H_0 is that the acid level is just as likely to be above the threshold as below the threshold. That is, on any given observation, the probability of being under the threshold is $\pi = .5$.*
(b) *H_a is that the pH is more likely to be below the threshold. (the probability of being below the threshold is $\pi > .5$).*
(c) *The significance level is $\alpha = .05$*
(d) *The test statistic is that out of the 16 measurements, 11 were below the pH threshold.*

(e) *The p-value is the probability that 11 or more measurements are below the threshold. This is:*

$$\sum_{i=11}^{16} \binom{16}{11} (.5)^{16} = (1+16+120+560+1820+4368)(.5)^{16} = .1020.$$

(f) *The p-value is greater than the significance level, so we retain the null hypothesis that rain with acidity above the threshold is just as likely as rain with acidity below the threshold.*

(4) In a vaccine study, 3,000 subjects were divided randomly into two. One group of 2000 subjects was given the vaccine, and the other group of 1000 subjects was given a placebo. After 5 years, 10 subjects came down with the disease. 8 were in the placebo group and 2 were in the vaccine group.

Is this sufficient evidence to conclude that the vaccine is protective against this disease at the 1% significance level? (Use a binomial exact test.)

SOLUTION:

(a) *H_0 is that the vaccine has no effect, and thus the subjects developing the disease have probability $\pi = 2/3$ of being in the vaccinated group.*

(b) *H_a is that the vaccine has an effect and the subjects developing the disease have probability $\pi < 2/3$ of being in the vaccinated group.*

(c) *The significance level is $\alpha = .01$.*

(d) *The test statistic is that, out of our sample of 10 subjects developing this disease, 2 are in the vaccinated group.*

(e) *The p-value is the probability (assuming H_0) that 2 or fewer of our sample ended up in the vaccinated group. This is*

$$\binom{10}{0}(1/3)^{10}+\binom{10}{1}(2/3)(1/3)^9+\binom{10}{2}(2/3)^2(1/3)^8 = (1\cdot\frac{1}{9}+10\cdot\frac{2}{9}+45\cdot\frac{4}{9})\frac{1}{3}^8 = \frac{201}{9}\frac{1}{3}^8 = .0034.$$

(f) *Our p-value is smaller than the significance level, so we reject H_0 and conclude that the vaccinated group is less likely to have develop the disease.*

(5) Let $f(x)$ be a function defined by

$$f(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 < x < \sqrt[3]{3} \\ 0 & x > \sqrt[3]{3}. \end{cases}$$

(a) Show that $f(x)$ is a pdf.

(b) Find the corresponding cdf, $F(x)$.

(c) Find the expected value of the random variable X associated to $f(x)$.

SOLUTION:

(a) *We need to calculate*

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\sqrt[3]{3}} x^2\, dx = [\frac{x^3}{3}]_0^{\sqrt[3]{3}} = \frac{3}{3} - 0 = 1$$

and $f(x)$ is non-negative. So $f(x)$ is a pdf.

(b)

$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x^3}{3} & 0 \le x \le \sqrt[3]{3}. \\ 1 & x \ge \sqrt[3]{3}. \end{cases}$$

(c)

$$E(X) = \int_{-\infty}^{\infty} x f(x)\ dx = \int_0^{\sqrt[3]{3}} x^3\ dx = [x^4/4]_0^{\sqrt[3]{3}} = 3^{4/3}/4 = 1.08169.$$