Copenhagen Business School

Exam in Statistics

HA-IB 2nd year

Tuesday 24 January 2012 9:00-13:00

Solutions

Problem 1

1.

$$\begin{aligned}
&P\{\textit{child develops disease}\}\\
&= P\{\textit{child gets "bad" allele from mother } \textbf{and} \textit{ child gets "bad" allele from father}\}\\
&= P\{\textit{child gets "bad" allele from mother}\} \times P\{\textit{child gets "bad" allele from father}\}\\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\end{aligned}$$

2. Define events as follows:

$$\begin{aligned}
C &= \textit{child does not develop disease}\\
M &= \textit{mother is a carrier}\\
F &= \textit{father is a carrier}
\end{aligned}$$

Note that (by question 1)

$$P(C|M \textbf{ and } F) = 1 - P(\textbf{not } C|M \textbf{ and } F) = 3/4$$

and (independence)

$$P(M \textbf{ and } F) = P(M)P(F) = 1/10^2.$$

Hence

$$\begin{aligned}
P(C) &= P(C|M \textbf{ and } F)P(M \textbf{ and } F) + P(C|M \textbf{ but not } F)P(M \textbf{ but not } F)\\
&\quad + P(C|F \textbf{ but not } M)P(F \textbf{ but not } M)\\
&\quad + P(C|\textbf{neither } M \textbf{ nor } F)P(\textbf{neither } M \textbf{ nor } F)\\
&= 3/4 \cdot 1/10^2 + 1/10 \cdot 9/10 + 1/10 \cdot 9/10 + 9/10 \cdot 9/10 = 3/400 + 99/100 = 399/400
\end{aligned}$$

or (as developing the disease requires both parents to be carriers)

$$P(C) = 1 - P(\textbf{not } C) = 1 - P(M \textbf{ and } F, \textbf{ but not } C)$$
$$= 1 - P(\textbf{not } C|M \textbf{ and } F)P(M \textbf{ and } F) = 1 - 1/4 \cdot 1/10^2$$
$$= 1 - 1/400 = 399/400$$

3.

$$P(M \textbf{ and } F|C) = \frac{P(C|M \textbf{ and } F)}{P(C)} \cdot P(M\textbf{and } F) = \frac{3/4}{399/400} \cdot 1/100 = \frac{3}{399} \approx 0.00752$$

Problem 2

The average is $\bar{x} = 42$ and the standard deviation is

$$\sqrt{\left((31-42)^2 + (40-42)^2 + 2(41-42)^2 + 4(44-42)^2 + (46+42)^2\right)/8} = \sqrt{\tfrac{102}{8}} \approx 3.5707$$

so the empirical rule suggests that approximately

- 68% of the observations are in the interval $42 \pm 3.5707 = [38.429; 45.571]$
- 95% of the observations are in the interval $42 \pm 2 \cdot 3.5707 = [34.859; 49.1414]$
- almost all of the observations are in the interval $42 \pm 2 \cdot 3.5707 = [31.288; 52.712]$

Problem 3

As the observations are "paired" (all 101 students were asked to solve both problems) we use McNemar's test.

The test statistic is

$$z = \frac{20-14}{\sqrt{20+14}} = \frac{6}{\sqrt{34}} \approx 1.028$$

which is (approximately) standard normally distributed. The p-value is 0.3035 (or, using tables 2(1-0.8461)=0.3078), so there is no evidence to conclude that the problems differed in difficulty.

Problem 4

The two probabilities are estimated by

$$\hat{p}_1 = \frac{148}{353} \approx 0.419 \qquad \hat{p}_2 = \frac{88}{467} \approx 0.188$$

so the difference is

$$\hat{p}_1 - \hat{p}_2 \approx 0.231$$

The standard error of this difference is

$$\sqrt{\tfrac{\hat{p}1(1-\hat{p}_2)}{353} + \tfrac{\hat{p}_2(1-\hat{p}_2)}{467}} \approx 0.0319$$

leading to a 95%-confidence interval of

$$]0.168; 0.293[$$

As the confidence interval does not include 0 we conclude that left voters are more in favour of going on strike than right voters are.

Problem 5

We get a test statistic of

$$\chi^2 = 9.0945$$

with (2-1)(3-1)=2 degrees of freedom and a p-value of 0.0106 (the 0.99-quantile is 9.21, the 0.95-quantile is 7.38). We conclude that there is a significant difference in the amount of capital held by the Japanese partner in joint ventures in developing countries and industrialized countries.

A look at the expected values suggests that "minority" is less common and "equal" is more common with developing countries than expected, whereas "majority" is very close to the expected.

Problem 6

1. The slope is estimated by

   $$\hat{\beta} = r\frac{s_y}{s_x}) = 0.315$$

   and the intercept by

   $$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 2.215$$

   so that the fitted regression line is

   $$\widehat{\mathit{Danisco}} = 2.215 + 0.315 \times \mathit{Market}$$

   with

   $$R^2 = \mathit{Cor}\,[\mathit{Danisco}, \mathit{Market}]^2 = 0.0650$$

2. The "overall F-test" is in a simple linear regression a test of the effect of the covariate. With 19 observations, the residual (or error) degrees of freedom are $19 - 2 = 17$ and the model degrees of freedom equal 1. Thus the test statistic is

   $$F = \frac{34.81288/1}{501.77413/17} = 1.179$$

   which is $F(1, 17)$-distributed, leading to a p-value of 0.293 (the upper 5%-quantile is 4.45, far greater than the value of the test statistic). Hence, we cannot reject the hypothesis of no effect of the Market return, so it seems that there may be no effect.

3. The test of the effect of the risk free return is given by

   $$t = \tfrac{24.107851}{11.15593} = 2.161$$

   which is t-distributed with $19 - 3 = 16$ degrees of freedom leading to a p-value of 2(1-0.9846)=0.046 (using tables: the upper 2.5%-quantile is 2.120, so the p-value is just below 5%), so there is sufficient evidence to conclude that there is an effect of the risk-free interest rate on the return of Danisco shares after accounting for the market return.