COMP245: Probability and Statistics 2016 - Problem Sheet 9 Solutions

Hypothesis Testing

S1) We wish to know if the two binary variables, manufacturer A/B and Faulty/Not faulty are independent. We can use a chi-squared test to explore this.

The observed numbers are

	A	B
Faulty	6	10
Not Faulty	44	40

Under the null hypothesis of independence, the expected values are

	A	B
Faulty	$\frac{16\times 50}{100}$	$\frac{16\times 50}{100}$
Not Faulty	$\frac{84\times 50}{100}$	$\frac{84\times 50}{100}$

=

	A	B
Faulty	8	8
Not Faulty	42	42

The test statistic is thus

$$X^2 = \frac{(6-8)^2}{8} + \frac{(10-8)^2}{8} + \frac{(44-42)^2}{42} + \frac{(40-42)^2}{42} = 1.19.$$

This is to be compared with a χ^2_1 distribution. This is not significant, even at the 10% level, so we have no reason for supposing that the underlying faulty PC rates of the two manufacturers are different.

S2) The null hypothesis is that the probability of getting a card right, p, is $\frac{1}{2}$. That is, $H_0 : p = \frac{1}{2}$. A possible alternative hypothesis would be $H_1 : p \neq \frac{1}{2}$, but a better one would be $H_1 : p > \frac{1}{2}$ since we would probably really be interested in the possibility that the subject gets more than half of the guesses right and not less than half.

Under the null hypothesis, the proportion right would be expected to follow the Binomial$(100, \frac{1}{2})$ distribution. This has mean $100 \times \frac{1}{2} = 50$ and variance $100 \times \frac{1}{2} \times \frac{1}{2} = 25$.

Adopting the one-sided alternative hypothesis, we will use the upper tail of this distribution as the rejection region. Since the sample size, 100, is large and the Binomial$(100, \frac{1}{2})$ is symmetric, we can approximate it by a normal distribution, N(50,25).

Our test statistic is therefore given by

$$z = \frac{34-50}{\sqrt{25}} = -3.2,$$

which should then be compared with the upper tail of a standard normal. Since the test is one-sided, the rejection region for a test at the 5% level is given by

$$R = \{z | z > 1.64\}.$$

In fact, our observed test statistic is negative, so is certainly not in R. Thus we have no evidence to reject the null hypothesis.

S3) (a) $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$, where μ is the mean difference between the heights of a pair of cross-fertilised and self-fertilised plants whose parents were grown from the same seed.

An appropriate test statistic is

$$T = \frac{\bar{D}}{S_{n-1}/\sqrt{n}}$$

with null distribution t_{n-1}, where $\bar{D}$ is the sample mean of the differences and S_{n-1} is their bias-corrected sample standard deviation.

(b) For a two-sided test at the 10% significance level, the rejection region is defined by

$$R = \{t||t| > t_{14,0.95}\} = \{t||t| > 1.761\}.$$

(c) For these data we have

$$t = \frac{\bar{d}}{s_{n-1}/\sqrt{n}} = \frac{20.93}{37.74/\sqrt{15}} = 2.15.$$

Thus we have $t \in R$ and so the hypothesis of zero difference is rejected in favour of the alternative hypothesis that there is a difference in the mean height of cross-fertilised and self-fertilised plants.

S4) (a) $H_0 : \mu = 0.6$; $H_0 : \mu \neq 0.6$.

(b) Poisson(μ).

(c) The mean number of cells in a square is 0.6825. This number estimates both the mean and variance of the population, since the mean and variance of a Poisson distribution are the same.

(d) The variance of a Poisson(μ) distribution is μ. So the mean of a sample of size n from Poisson(μ) has variance $\frac{\mu}{n}$, and so the standard deviation of the sample mean is $\sqrt{\frac{\mu}{n}}$.

The 1.96 arises because the sample size is quite large, and we can approximate the distribution of the mean by a normal distribution (by the Central Limit Theorem), and this is the 5% critical region for a normal distribution with mean μ and standard deviation $\sqrt{\frac{\mu}{n}}$; this is a union of two terms, because the alternative hypothesis is two-sided. The appropriate test statistic would clearly be the sample mean number of cells in a square.

(e) The confidence limits are

$$\left(0.6 - 1.96\sqrt{\frac{0.6}{400}}, 0.6 + 1.96\sqrt{\frac{0.6}{400}}\right) = (0.5241, 0.6759).$$

(f) The observed value of the test statistic is 0.6825, and this lies in the critical region (it is greater than 0.6759). Thus we reject the null hypothesis that the true mean number of cells per square is 0.6 at the 5% level.

S5) We assume that the sexes of children in the same family are independent. Then the distribution hypothesised is Binomial$(5, \frac{1}{2})$. Thus the expected number of families with 1 boy and 4 girls, for example, is

$$320 \times \binom{5}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 320 \times \frac{5}{32} = 50.$$

Continuing in this way, we get

Number of boys	O	E	$O-E$	$\frac{(O-E)^2}{E}$
0	18	10	8	6.4
1	56	50	6	0.72
2	110	100	10	1
3	88	100	-12	1.44
4	40	50	-10	2
5	8	10	-2	0.4
$\sum$				11.96

Comparing this to the χ^2_5 distribution (no parameters have been estimated from the data) we see that this is greater than 11.07, the 5% level, but less than 15.09, the 1% level. Thus we can reject the null hypothesis that the births of males and females are equally likely at the 5% level but not at the 1% level.

S6) Compute marginal numbers: 345+135+20 = 500 etc for rows, 345+222 = 563 etc for cols.

Compute expected values 500x563/750 etc.

Compute $\displaystyle\sum_i \frac{(O_i - E_i)^2}{E_i}$.

Compare with a χ^2 distribution with (2-1)(3-1) = 2 degrees of freedom.

S7) (a) Power will increase.

(b) Power would decrease.

(c) Either increase sample size or increase α.

(d) Use a large sample size.

S8) A simple chi-squared test is appropriate. The null hypothesis is that the balls have equal probabilities of appearing. The sum of all the ball frequencies (the number of lottery balls drawn since November 1994) is 8,154.

The expected frequency of each ball, under the null hypothesis is thus $\frac{8154}{49} = 166.4082$.

The test statistic

$$\sum_{i=1}^{49} \frac{(O_i - 166.4082)^2}{166.4082}$$

turns out to be 46.475. Comparing with the χ^2_{48} distribution, we see the null hypothesis cannot be rejected at any reasonable significance level.