# MAST20005/MAST90058: Week 8 Solutions

1. In a truly random sequence of numbers, the probability of the next digit being the same as the preceding one is 1/10 and the probability of the next one differing by 1 from the preceding is 2/10. The null hypothesis is:

$$H_0\colon p_1 = 0.1,\ p_2 = 0.2,\ p_3 = 0.7$$

Suppose we obtain the following observed counts:

| | Observed | Expected |
|---|---|---|
| Same | 0 | $50 \times 0.1 = 5$ |
| Differ by 1 | 8 | $50 \times 0.2 = 10$ |
| Other | 42 | $50 \times 0.7 = 35$ |

The chi-squared statistic is:

$$\frac{(0-5)^2}{5} + \frac{(8-10)^2}{10} + \frac{(42-35)^2}{35} = 6.8 > 5.991 \quad (0.95 \text{ quantile of } \chi^2_2)$$

Thus we conclude that the string of 51 digits is unlikely to have been randomly generated.

2. (a) The table is:

| $i$ | $x_i$ | $x_i - m$ | Rank | Sign |
|---|---|---|---|---|
| 1 | 41.195 | 1.195 | 5 | 1 |
| 2 | 39.485 | −0.515 | 1 | −1 |
| 3 | 41.229 | 1.229 | 6 | 1 |
| 4 | 36.840 | −3.160 | 10 | −1 |
| 5 | 38.050 | −1.950 | 8.5 | −1 |
| 6 | 40.890 | 0.890 | 4 | 1 |
| 7 | 38.345 | −1.655 | 7 | −1 |
| 8 | 34.930 | −5.070 | 11 | −1 |
| 9 | 39.245 | −0.755 | 2 | −1 |
| 10 | 31.031 | −8.969 | 12 | −1 |
| 11 | 40.780 | 0.780 | 3 | 1 |
| 12 | 38.050 | −1.950 | 8.5 | −1 |
| 13 | 30.906 | −9.094 | 13 | −1 |

and

$$W = 5 - 1 + 6 - 10 - 8.5 + 4 - 7 - 11 - 2 - 12 + 3 - 8.5 - 13 = -55.$$

Recall that

$$\mathbb{E}(W) = 0, \quad \operatorname{var}(W) = \frac{n(n+1)(2n+1)}{6} = 819$$

so

$$z = \frac{-55}{\sqrt{819}} = -1.922$$

which is less than −1.645 (0.05 quantile of a standard normal distribution), so we reject $H_0$ at a 5% level of significance.

(b) Bounding $z$ against known quantiles from a standard normal gives: $-1.96 < z < -1.645$. Therefore, we deduce that, $0.025 <$ p-value $< 0.05$.

(c) There are 4 positive signs. Therefore, the p-value is

$$\Pr(Y \leqslant 4 \mid p = 0.5) = 0.1334.$$

This is greater than $\alpha$ so we cannot reject $H_0$.

(d) The null hypothesis is rejected using the signed-rank test but cannot be rejected using the sign test.

3. The observed and expected frequencies are:

| | Red | Brown | Scarlet | White |
|---|---|---|---|---|
| $O$ | 254 | 69 | 87 | 22 |
| $E$ | 243 | 81 | 81 | 27 |

and

$$\chi^2 = 3.646 < 7.815 \quad (0.95 \text{ quantile of } \chi^2_3)$$

so we cannot reject $H_0$ at the 5% level of significance.

4. This a problem where we want to do a goodness-of-fit test of a particular model but where we need to first estimate some of the parameters. We can set it up in one of two ways.

The first way is to think about the null distribution and work out which parameters need to be estimated. Under $H_0$ we have $p_{i1} = p_{i2}$, so let's call both of them $p_i$ (since they are equal). These define the probabilities of each category (columns) that apply to each group of nurses (rows). Note that these are conditional probabilities, $p_i = \Pr(\text{category } i \mid \text{group I}) = \Pr(\text{category } i \mid \text{group II})$. To complete the model we also need to estimate the marginal probabilities of the two groups, let's call these $g_j = \Pr(\text{group } j)$, for $j = 1, 2$. The null model, therefore, is that the probability of an observation for category $i$ in group $j$ is $g_j p_i$. Note that there are 6 *independent* parameters to estimate (5 conditional column probabilities and one row probability), so ultimately we'll end up with a test with $12 - 6 - 1 = 5$ degrees of freedom.

The other way is to note that this model is equivalent to the usual test of independence of a contingency table, we end up estimating the same parameters and apply the same test as described above.

Under either setup, the observed and expected frequencies are:

| | | Category | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Group I | $O$ | 95.0 | 36.0 | 71.0 | 21.0 | 45.0 | 32.0 |
| | $E$ | 88.8 | 37.2 | 68.4 | 23.4 | 46.2 | 36.0 |
| Group II | $O$ | 53.0 | 26.0 | 43.0 | 18.0 | 32.0 | 28.0 |
| | $E$ | 59.2 | 24.8 | 45.6 | 15.6 | 30.8 | 24.0 |

and as there are 5 df

$$\chi^2 = \frac{(95 - 88.8)^2}{88.8} + \cdots + \frac{(28 - 24)^2}{24} = 3.23 < 11.07 \quad (0.95 \text{ quantile of } \chi^2_5)$$

so we cannot reject $H_0$.