Probability and Statistics for JMC
Solutions 7 — Hypothesis Testing

1. To decide from which manufacturer to purchase 2000 PCs for its undergraduates, a university decided to carry out a test. It bought 50 PCs from manufacturer A and 50 from manufacturer B. Over the course of a six month period, 6 of those from manufacturer A experienced problems, as did 10 from manufacturer B. This suggested that, to be on the safe side, the university should buy machines from manufacturer A. However, A's machines are more expensive than B's. Before making the decision to buy from A, the university wanted to be confident that the difference was a real one, and was not merely due to chance fluctuation. Carry out a test to investigate this.

   You may wish to use this extract from the chi-squared table:

Degrees of freedom	Upper tail area		
	.10	.05	.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086

One approach is to consider two binary random variables, one which tells us whether the computer was from company A or B and the other which says if it was faulty or not. The null hypothesis is that the two variables are independent (i.e. probability of faultiness has nothing to do with which company the PC is from). The contingency table looks like:

(observed)

	A	B	
Faulty	6	10	16
Not faulty	44	40	84
	50	50	100

Under the null hypothesis of independence we estimate the expected counts in each cell of the table to be:

(expected)

$$\begin{array}{r|cc} & \text{A} & \text{B} \\ \hline \text{Faulty} & \frac{50\times16}{100} & \frac{50\times16}{100} \\ \text{Not faulty} & \frac{50\times84}{100} & \frac{50\times84}{100} \end{array} = \begin{array}{|cc} \text{A} & \text{B} \\ \hline 8 & 8 \\ 42 & 42 \end{array}$$

The χ^2 test statistic is $X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(6-8)^2}{8} + \frac{(10-8)^2}{8} + \frac{(44-42)^2}{42} + \frac{(40-42)^2}{42} = 1.19$. Under the null hypothesis this should be distributed as a χ^2 RV with $(2-1)\times(2-1) = 1$ degree of freedom. Using the table this is not significant even at the 10% level (since we expect 10% of the time that a χ^2_1 RV is greater than 2.706). Therefore, the data is not sufficient to reject the hypothesis that company A and B produce computers with the same rate of being faulty.

2. In an ESP experiment, a subject in one room is asked to state the colors of 100 cards chosen by someone in another room. The chooser has a pack of 25 red and 25 blue cards and selects the cards at random, with replacement. If the subject gets 64 right, determine whether the results are significant at the 1% level. Clearly write down your null hypothesis, alternative hypothesis, the test statistic you use, the rejection region, and interpret your conclusions.

Hint: you may regard 100 as a large sample. You may wish to use this table giving quantile function of the standard normal distribution for various upper tail areas.

Upper tail area α	0.1	.05	.025	.01	.005	.001
$\Phi^{-1}(1-\alpha)$	1.28	1.64	1.96	2.33	2.58	3.09

We are observing Bernoulli RVs X_i for $i = 1, \dots, 100$, where X_i is 1 or 0 depending on if the subject guesses correctly. The null hypothesis is that the Bernoulli parameter p is equal to 0.5 and the alternative is that $p > 0.5$ (i.e. we only want to reject the null if the subject is doing much better than random chance, not worse). We can write this as $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$.

A convenient test statistic is simply based on the total number of correct guesses (i.e. $\sum_{i=1}^{100} X_i$), which under the null hypothesis is a Binomial$(100, p)$ RV. We can "standardize" this quantity by subtracting its mean and dividing by its standard deviation (assuming the null hypothesis) to get

$$T = \frac{\sum X_i - 50}{\sqrt{100 \times 0.5 \times (1 - 0.5)}}.$$

Since T is based on the sum of 100 independent RVs we invoke the central limit theorem and assume that under the null hypothesis T should be distributed as a standard normal. We want to reject if T is too large. Therefore, for a test at 99% significance the rejection region should be $R = \{t \,|\, t > 2.33\}$. The observed value of T is $t = \frac{64-50}{\sqrt{25}} = 2.8$, inside the rejection region. Therefore, we can reject the null hypothesis that the subject does no better than chance at 99% significance.

3. Charles Darwin measured differences in height for 15 pairs of plants of the species *Zea mays*. (Each plant had parents grown from the same seed – one plant in each pair was the progeny of a cross-fertilisation, the other of a self-fertilisation. Darwin's measurements were the differences in height between cross-fertilised and self-fertilised progeny.) The data, measured in eighths of an inch, are given below.

49, -67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, -48

(a) Supposing that the observed differences $\{d_i \,|\, i = 1, \dots, 15\}$ are independent observations of a normally distributed random variable D with mean μ and variance σ^2, state appropriate null and alternative hypotheses for a two-sided test of the hypothesis that there is no difference between the heights of progeny of a cross-fertilised and self-fertilised plant, and state the null distribution of an appropriate test statistic.

$H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$, where μ is the average difference in heights between the two types of progeny.

We do not know what the variance σ^2 is, so an appropriate test statistic would be $T = \frac{\bar{D}}{\sqrt{S_{n-1}^2/n}}$, where $n = 15$ and S_{n-1}^2 is the bias-corrected estimate of the variance based on the 15 observed differences. Under the null hypothesis T would be distributed as a Student-t RV with $15 - 1 = 14$ degrees of freedom.

(b) Obtain the form of the rejection region for the test you defined in part (a), assuming a 10% significance level.

We are doing a two sided test so the rejection region should be both the upper and lower tails of the Student-t distribution. For a 10% test we want 5% area in the upper tail and 5% in the lower tail. Using the table we get $R = \{t \mid |t| > 1.7613\}$.

(c) Calculate the value of the test statistic for this data set, and state the conclusions of your test.

The sample mean is $\bar{D} = 20.933$, the bias-corrected variance estimate is $S^2_{n-1} = 1424.6$ so the observed test statistic is $t = \frac{20.933}{\sqrt{1424.6/15}} = 2.15$. Therefore, $t \in R$ so we reject the null hypothesis of zero difference between cross and self-fertilised progeny at 90% significance.

You may want to use the following extract from a t-distribution, giving the point x corresponding to specified areas under the upper tail of a t-distribution with df degrees of freedom.

df	5%	10%
13	1.7709	1.3502
14	1.7613	1.3450
15	1.7531	1.3406

4. Some of "Student's" original experiments involved counting the numbers of yeast cells found on a microscope slide. The results of one experiment are given in the table below, which shows the number of small squares on a slide which contain 0, 1, 2, 3, 4, or 5 cells. We want to use these data to see if the mean number of cells per square is 0.6, using the 5% significance level.

Number of cells in square	0	1	2	3	4	5
Frequency	213	128	37	18	3	1

(a) State the null hypothesis and the alternative hypothesis.

If μ is the mean number of cells per square we are testing $H_0 : \mu = 0.6$ vs. $H_1 : \mu \neq 0.6$.

(b) The distribution of the numbers of cells is far from normal; it can take only positive integer values, it is very far from symmetric, and dies away very quickly. Which familiar distribution might be appropriate as a model for these data?

Poisson(μ). (The null hypothesis will include the assumption that the data are Poisson.)

(c) Estimate the mean and the variance of the distribution you suggested in part (b).

We have $213 + 128 + 37 + 18 + 3 + 1 = 400$ observations. The MLE of the mean of a Poisson happens to be the sample mean, so $\hat{\mu} = \frac{1}{400}(213 \times 0 + 128 \times 1 + 37 \times 2 + \ldots) = 0.6825$. The variance of a Poisson distribution is equal to its mean so the estimate of the variance is also 0.6825.

(d) Explain why a critical region with the form

$$\left\{\bar{x} < \mu - 1.96\sqrt{\frac{\mu}{n}}\right\} \cup \left\{\bar{x} > \mu + 1.96\sqrt{\frac{\mu}{n}}\right\}$$

would be a reasonable region, making sure you explain the 1.96, the term $\sqrt{\frac{\mu}{n}}$ and the implications of the union. What is the test statistic in mind?

The test statistic in mind is the sample mean of number of cells in a square since we are looking at a rejection region that is some subset of possible values for the sample mean. We are doing a two-sided test so the rejection region will be formed of the union of both tails of the distribution of the sample mean.

In this case the sample mean is based on $n = 400$ observations so we use the central limit theorem to assume it is normally distributed under the null hypothesis. The expected value of the sample mean is the population mean μ and the variance of the sample mean is the population variance σ^2 divided by the number of samples n. For a Poisson, $\sigma^2 = \mu$. So under the null hypothesis we have $\bar{X} \sim \mathrm{N}(\mu, \frac{\mu}{n})$. Here, μ is the mean under the null hypothesis: $\mu = 0.6$.

The 1.96 comes from doing a two-sided 95% hypothesis test: The area under a normal distribution further than 1.96 standard deviations away from the mean is 95% (2.5% in each tail; see table from Q2).

(e) Compute the limits of the critical region.

$\mu \pm 1.96\sqrt{\mu/n} = 0.6 \pm 0.076$. So the rejection region is $R = \{\bar{x} \mid \bar{x} < 0.524\} \cup \{\bar{x} \mid \bar{x} > 0.676\}$.

(f) Draw a conclusion about whether or not the null hypothesis can be rejected.

For the observed data the sample mean is 0.6825, which is in the rejection region. Therefore, the hypothesis that the cells per square is a Poisson RV with mean of 0.6 can be rejected at 95% significance.

5. A survey of 320 families with 5 children each, gave the distribution shown below. Is this table consistent with the hypothesis that male and female children are equally probable? Obtain results for both the 1% and 5% levels. Work through the details of the test – don't just hit the chi-squared button on a statistical calculator.

[You may wish to use the table extract from Q1.]

boys : girls	5 : 0	4 : 1	3 : 2	2 : 3	1 : 4	0 : 5
Number of families	18	56	110	88	40	8

We are working with binned counts data so a χ^2 will probably be appropriate. We need the expected number of counts in each bin (i.e. the expected number of families with each ratio of boys to girls). The null hypothesis is that the number of male children in a family of 5 is a Binomial RV with $n = 5$ and $p = 0.5$. Under the null, the expected number of 5-child families (out of 320) with i boys is therefore $320 \times \binom{5}{i}(0.5)^i(0.5)^{5-i}$. We work out the χ^2 test statistic as

boys : girls	5 : 0	4 : 1	3 : 2	2 : 3	1 : 4	0 : 5
O_i	18	56	110	88	40	8
E_i	10	50	100	100	50	10
$O_i - E_i$	8	6	10	-12	-10	-2
$\frac{(O_i-E_i)^2}{E_i}$	6.4	0.72	1.0	1.44	2.0	0.4

and find the observed value of the test statistic to be $\chi^2 = \sum_i \frac{(O_i-E_i)^2}{E_i} = 11.96$. There are no parameters estimated from the data so the test statistic should be distributed, under the null hypothesis, as a χ^2 RV with $6-1=5$ degrees of freedom. We reject if the observed χ^2 is too large. Using the table from Q1 we find that the null hypothesis is rejected at the 5% level (since $11.96 > 11.07$) but cannot be rejected at the 1% level (since $11.96 \not> 15.086$).

6. As part of a telephone interview, a sample of 500 executives and a sample of 250 MBA students were asked to respond to the question "Should corporations become more directly involved with social problems such as homelessness, education, and drugs?" The results are shown below. Test the hypothesis that the patterns of response for the two groups are the same.
[You may wish to use the table extract from Q1.]

		More involved	Not more involved	Not sure	
(observed)	Executives	345	135	20	500
	MBA students	222	20	8	250
		567	155	28	750

The null hypothesis is that executive vs. MBA student is independent from which survey response they picked. We can estimate the marginal probability of picking each response under the null hypothesis. E.g. probability of "more involved" is $(345+222)/750 = 0.756$.
Thus the expected values in the contingency table for the null hypothesis are

		More involved	Not more involved	Not sure
(expected)	Executives	378.0	103.3	18.67
	MBA students	189.0	51.67	9.33

The χ^2 test statistic is $\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$, where the sum is over all bins in the table. The observed value is $\chi^2 = 38.0$. Under the null hypothesis this test statistic should be a χ^2 RV with $(3-1)\times(2-1) = 2$ degrees of freedom. The observed value is extremely far into the right tail of the χ^2_2 distribution. In fact, the p value is about 5×10^{-9} (area under the χ^2_2 pdf to the right of 38). So we reject the null hypothesis at extremely high significance. The data is highly inconsistent with the hypothesis that the executives and the MBA students have the same pattern of responses.

7. (a) For a test at a fixed significance level, and with given null and alternative hypotheses, what will happen to the power as the sample size increases?

   Typically the power will increase as sample size increases, all else being held equal.

   (b) For a test of a given null hypothesis against a given alternative hypothesis, and with a fixed sample size, describe what would happen to the power of the test if the significance level was changed from 5% to 1%.

   Power would decrease. Changing the type I error rate would decrease the size of the rejection region. Therefore, the probability of rejecting under the alternative (i.e. the power) would decrease.

   (c) A test of a given null hypothesis against a given alternative hypothesis, with a sample of size n and significance level of α, has a power of 80%. What change could I make to the test to increase my chance of rejecting a false null hypothesis?

Either increase the sample size n or increase α, the type I error rate.

(d) How can we attain a test which has a very low probability of Type I error and also a very low probability of Type II error?

Typically this can be achieved by having a large sample size.

8. The data below show the frequency with which each of the balls numbered 1 to 49 have appeared in the main draw in the National Lottery between its inception in November 1994 until December 2008.

This table shows that there are substantial differences between the numbers of times different balls have appeared — for an extreme example, the number 20 has been drawn just 134 times, whereas 38 has appeared 197 times, almost 50% more often.

Should we conclude from this table that the balls have different probabilities of appearing?

Ball	Freq.	Ball	Freq.	Ball	Freq.	Ball	Freq.	Ball	Freq.
1	160	11	185	21	148	31	180	41	137
2	168	12	178	22	169	32	170	42	167
3	163	13	147	23	183	33	176	43	182
4	163	14	160	24	164	34	159	44	184
5	155	15	157	25	182	35	173	45	162
6	174	16	147	26	161	36	148	46	164
7	168	17	158	27	173	37	150	47	177
8	158	18	164	28	165	38	197	48	177
9	176	19	166	29	164	39	166	49	169
10	171	20	134	30	177	40	178		

[To save you a lot of data entry you might use the following way of rewritting a particular polynomial:

$$\sum_{i=1}^{49}(n_i - x)^2 = 49x^2 - 16308x + 1364626,$$

where n_i is the frequency of ball i in the above table.]

We want to test the null hypothesis that each number has an equal probability of appearing. Under the null hypothesis the expected number of appearances of each number will be $1/49$ times the total of all the counts in the table. We will form the test statistic $\chi^2 = \sum_{i=1}^{49}(O_i - E_i)^2/E_i$. Under H_0 this will be distributed as a χ^2 RV with $49 - 1 = 48$ degrees of freedom (we did not need to estimate any parameters from the data).

We can either compute E_i using the table data or use the given quadratic by realizing that the linear term is $-2x\sum_i n_i$, so $\sum_i n_i = 8154$. Thus the observed value of χ^2 is

$$\chi^2 = \sum_{i=1}^{49} \frac{(n_i - 8154/49)^2}{8154/49} = 46.5,$$

again making use of the quadratic above.

We can either look up the quantiles of a χ^2_{48} RV or we can use the fact that the mean value of a χ^2 RV is equal to the number of degrees of freedom and the variance is equal to twice the number of degrees of freedom. Since the observed χ^2 is less than its mean value under the null hypothesis we can be certain that the observed χ^2 is not out in the right tail of the null distribution. Thus we cannot reject the null hypothesis of equal probabilities for the numbers at any interesting significance level (p value turns out to be 0.54).