A *two sample Z-test* can be conducted to *compare two population proportions* with $H_0 : p_1 = p_2$ based on two random samples of binary variables with sample totals X_1 and X_2 and sample sizes n_1 and n_2 of which both are large. The test statistic is $Z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$

where the pooled sample proportion $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$.

A *two sample t-test* can be conducted to *compare two population means* with $H_0 : \mu_1 = \mu_2$ based on two independent samples of sizes n_1 and n_2 with means $\bar{x}_1$ and $\bar{x}_2$ and s.d. s_1 and s_2 respectively from two normal populations with equal variance. The test statistic is $t_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

where the pooled sample variance is $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

Tutorial discussion: Q5, Q6 and Q8-10 marked with *

1. **Multiple choice**   A sample of size $n = 27$ is used to conduct a significance test for $H_0 : \mu = 75$ versus $H_1 : \mu > 75$. The test statistic is $t_{\text{obs}} = 3.45$ and $\alpha = 0.05$. The rejection region for this test is:

   (a) $t_{\text{obs}} > 2.056$ (b) $t_{\text{obs}} > 1.706$ (c) $t_{\text{obs}} > 3.450$

   (d) $t_{\text{obs}} > 75.000$ (e) none of these.

2. **Multiple choice**   In a study of the effect of oral contraceptives (OCs) on blood pressure, measurements were made on 12 OC users (x) and 15 non-users (y). From these we calculate $\bar{x} = 119.40, \bar{y} = 116.25, s_x = 8.75, s_y = 10.00$. The pooled estimate of the standard deviation, s_p, is (2 dp):

   (a) 89.69   (b) 83.04   (c) 9.47   (d) 9.11   (e) 9.84

3. **Multiple choice**   To test $H_0 : \mu_x = \mu_y$ against the alternative $H_1 : \mu_x \neq \mu_y$ in Q2, the degrees of freedom of a suitable t-statistic are:

   (a) 27   (b) 54   (c) 12   (d) 25   (e) 50

4. Test the null hypothesis in Q2 using the two-sample t-test. State any assumptions you may require to do this test.

5. *In an effort to compare the durability of two different types of sandpaper, 10 pieces of type A sandpaper were subjected to treatment by a machine which measures abrasive wear. Eleven pieces of type B sandpaper were subjected to the same treatment. Assuming normality and equality of variance, test for equality of mean abrasive wear using the following summary.

Type	Size	Mean	St. Dev
A	10	27.4	2.3
B	11	24.1	3.1

6. *Suppose a Drug Company develops a new drug, designed to prevent flu. The company states that the drug is equally effective for men and women. To test this claim, they choose simple random samples of 100 women and 200 men. At the end of the study, 38% of the women caught a flu and 51% of the men caught a flu. Based on these findings, can we reject the company's claim that the drug is equally effective for men and women? Use a 0.05 level of significance.

7. Two hundred and forty dieters volunteered for a study of chromium picolinate, a fat-reducing dietary supplement. They were randomly assigned to receive placebo or chromium picolinate. One side effect is reduction in iron, a key component of hemoglobin. Here are the data on the number of subjects with lower iron after 8 weeks of treatment:

	Placebo	Chromium Picolinate
n	120	120
Number with reduced iron	19	35

Is the proportion of reduced iron higher among the treated patients? Answer the question using significance level of $\alpha = 0.05$.

Use R to answer Q8 to Q10

---

8. *Find the *exact* P-value for the test in Q1 using the `pt` command. Also, find the critical value using the `qt` command.

9. *The file `marks.txt` contains two lines of data. The first line are the marks of quiz 1 from students, enrolled in an early morning MATH1015 tutorial while the second line are the marks from students, enrolled in a midmorning tutorial.

   (a) Read the early morning marks into a vector called `x` by typing `x=scan()` in a R session window and hit `enter`. Then open `marks.txt` in the course website under `Data Sets` and `Marks`. This is a text file which should be easier to copy and paste than the pdf file of ast week. After copy and paste, hit `enter` twice. Alternatively, copy and paste the command below:

   `x=scan(file=url("http://www.maths.usyd.edu.au/math1015/r/marks.txt"),nlines=1)`

   (b) Read the midmorning marks into a vector called `y` as described above or copy and paste the command below:

   `y=scan(file=url("http://www.maths.usyd.edu.au/math1015/r/marks.txt"),skip=1)`

   Note that optional arguments `nlines` and `skip` of the `scan` command read only the first or only the second row of data.

   (c) Perform a test to decide if the mean marks in the two tutorials are equal or not. What is your conclusion?

   Hint: use `t.test(x,y,var.equal=T)`. The default of `t.test` is a two-sided alternative and `mu=0`. If you drop `var.equal=T`, you will get a different result from a different test which does not assume equality of variance.

   (d) Draw side-by-side boxplots of the two groups. Do you think the *normality* and *equality of variance* assumptions for the test might be violated?

10. *Redo Q7 using the command `prop.test`. Notice that it performs a chi-square test (studied next week). However the P-value should be the same (don't forget to use the appropriate alternative option `alt="`"). Check that the reported `X-squared` statistic is equal to the square of the z-statistic computed in Q7. Also find the P-value of the z-test using the `pnorm` command.

   Hint: try `prop.test(c(35,19),c(120,120),alt="greater",correct=F)`. That means you need to specify the vector of observed counts and then the vector of sample sizes in `prop.test`. The option `correct=F` means that you don't want continuity correction. Note that you haven't learned it yet.

---

1. Book P.153, Q8.22-8.24 on two-sample t-test.

2. Book P.153-154, Q8.31-8.36 on paired sample t-test.

3. Book P.183, Q10.25-10.28 on two-sample z-test.