12

COMPARING TWO POPULATIONS

QUESTION 1.

In the box below write down what you think 'associations between variables' means and what 'differences between groups' means.

Associations between variables means:
Differences between groups means:

QUESTION 2.

In the box below write down the statistical analysis techniques offered by Excel for two-sample testing. Also give a brief explanation of when each test would be most appropriate.

QUESTION 3.

Write down the statistical analysis techniques offered by SPSS for comparing means and give a brief explanation of when each test would be most appropriate.

QUESTION 4.

The table below gives the heights of 10 boys and 10 girls. Using a suitable statistical technique, test the assumption that the boys, on average, are taller than the girls.

Boys	Girls
72	67
72	67
79	67
74	62
71	64
66	62
71	64
65	69
67	69
72	66

Complete the following:

Due to the size of the sample, the most appropriate test would be _____. If the sample had been larger, a more appropriate test would be _____.

In the box below, write down 2 assumptions you would make concerning this test.

Complete the following:

The null hypothesis H_0: _____

The alternative hypothesis H_1: _____

A _____ tailed test can be used since the _____ hypothesis states that the mean height of the boys is ___ than the mean height of the girls.

The test will be conducted with an α value of _____

The mean height of the boys is _____

The mean height of the girls is _____

The standard deviation for the boys is _____

The standard deviation for the girls is _____

The number of degrees of freedom is _____

The value of the test statistic is _____

The *p*-value is _____

The results from this analysis show that H_0 can be rejected/accepted (delete as appropriate) which means it can be inferred that, on average, _____ are taller than _____.

Complete the following:

The analysis can be interpreted to mean that there is a probability of _____ that the sample means could be _____ if the null hypothesis was accepted.

QUESTION 5.

200 people who have been diagnosed with a chronic illness have been divided into two groups to test the effectiveness of a new medicine. Group A have been given the medicine and Group B have been given a 'sugar' pill (a placebo). The results show that 75 people from Group A reported beneficial effects from taking the medicine and 65 people from Group B reported similar beneficial effects.

Before the medicine can be sold in large quantities, the manufacturer needs to know if these results are significant. Your task is to select and carry out a suitable test to find out if the medicine is indeed effective.

You decide to test at three levels of significance (1) 0.01, (2) 0.05 and (3) 0.10.

The null hypothesis is that the observed frequencies are due to _____ which implies the medicine is effective/ineffective (delete as appropriate).

The alternative hypothesis is the medicine is effective/ineffective (delete as appropriate).

If p_1 represents the population proportion who reported benefits from using the medicine and p_2 the population proportion who reported no benefits from using the medicine, the hypotheses can be written as:

$$H_0: p_1 \_\_\_\_ p_2$$
$$H_1: p_1 \_\_\_\_ p_2$$

p is the estimated average of people who reported beneficial effects in the two sample groups and is given by _____

In the box below evaluate the following formula:

$$\sigma_{p_1 - p_2} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$\sigma_{p_1 - p_2} =$

The test statistic $Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sigma_{p_1 - p_2}} = \frac{\quad - \quad}{\quad} =$

Complete the following:

On the basis of a _____ tailed test at a _____ level of significance, H_0 would be accepted/rejected (delete as appropriate). This means the medicine is effective/ineffective (delete as appropriate) and the results are due to chance/reliable (delete as appropriate).

On the basis of a _____ tailed test at a _____ level of significance, H_0 would be accepted/rejected (delete as appropriate). This means the medicine is effective/ineffective (delete as appropriate) and the results are due to chance/reliable (delete as appropriate).

On the basis of a _____ tailed test at a _____ level of significance, H_0 would be accepted/rejected (delete as appropriate). This means the medicine is effective/ineffective (delete as appropriate) and the results are due to chance/reliable (delete as appropriate).

From the above results decide whether a Type I error has been committed. In the box below, write down what a Type I error is, its relevance to this question and explain how it is related to the level of risk the company is prepared to take.

A Type I error is when the:
It is relevant to this question because:
The level of risk is inferred in the:

From the above results decide whether a Type II error has been committed. In the box below, write down what a Type II error is, its relevance to this question and explain how it is related to the level of risk the company is prepared to take.

A Type II error is when the:
A type II error is related to the level of risk because:

QUESTION 6.

The company you work for plan to install new equipment which the supplier claims will improve productivity by reducing the design, implement and manufacture times. The table below shows the completion times of 24 engineers, 12 of whom used the existing equipment and the other 12 used the new equipment.

Current equipment	Proposed equipment
300	274
280	220
344	308

Current equipment	Proposed equipment
385	336
372	198
360	300
288	315
321	258
376	318
290	310
301	332
283	263

You have been asked to verify the supplier's claim of improved productivity.

_____ is the population mean for the current equipment.

_____ is the population mean for the proposed equipment.

Evidence is required to show that _____ is less than _____.

The null hypothesis can be written in statistical terms as _____ – _____ _____ 0 and the alternative hypothesis as _____ – _____ > _____.

H_0: ____ – ____ ≤ 0

H_1: ____ – ____ > 0

You decide to use a level of significance of $\alpha = 0.05$

The sample sizes are n_1 = ____ and n_2 = ____.

The sample means are $\bar{x}_1$ = ____ and $\bar{x}_2$ = ____.

The sample standard deviations are s_1 = ____ and s_2 = ____.

The test statistic, t = ____.

The degrees of freedom, df = ____.

Using an upper tail test, the p-value is between _____ and _____.

H_0 is therefore _____ since the p-value is less than ____ = 0.05.

The statistical evidence suggests that the new equipment does/does not (delete as appropriate) improve productivity.

QUESTION 7.

Currently you are employed as the Marketing Manager for a medium-sized company. You are now looking for an opportunity to earn more money. You are considering moving to either London or Birmingham. Being a competent statistician you have investigated average salaries in both cities.

The table below shows your summary statistics.

Birmingham	London
$n_1 = 40$	$n_2 = 50$
$\overline{x}_1 = £56,100$	$\overline{x}_2 = £59,400$
$s_1 = £6000$	$s_2 = £7000$

Complete the following:

H_0: ____

H_1: ____

The value of the test statistic is ____.

The p-value is ____ which is between ____ and ____.

The conclusion is that salaries in _____ are lower than salaries in _____, therefore reject the _____ hypothesis.

MINI PROJECT

You have been employed by a company who are investigating the best medium for advertising their products. You decided to compare the amount of time people spent watching satellite television and/or terrestrial television. You used a sample of 15 people.

The table below shows the data you collected:

Person	Sat TV	Ter TV
1	22	25
2	8	10
3	25	29
4	22	19
5	12	13
6	26	28
7	22	23
8	19	21
9	21	21
10	23	23
11	14	15
12	14	18
13	14	17
14	16	15
15	24	23

You are expected to produce a report which uses statistical analysis to make a recommendation on the best medium to use for advertising your company's products.

And finally...

At Christmas time, who's the meanest, Scrooge or the Grinch?