STAT7203: Applied Probability and Statistics
Quiz 9

1. To compare the antioxidant capacities of green and black teas, a researcher bought 43 different brands of tea bags that were widely available in supermarkets across Europe. A random tea bag was selected from each brand and brewed in a cup with 200 mL of purified water at 100^c for 2 minutes. After 2 minutes, the antioxidant capacity (in milligrams) of each cup of tea was measured. The results are summarised below:

	sample size	sample mean	sample standard deviation
Black tea	24	239 mg	102 mg
Green tea	19	336 mg	139 mg

Construct an approximate 90% confidence interval for the true mean difference in antioxidant capacities between green and black teas.

Solution: Let μ_G and μ_B be the antioxidant capacities of green tea and black tea, respectively. The confidence interval for $\mu_B - \mu_G$ is given by $\bar{x}_B - \bar{x}_G \pm t_{1-\alpha/2;\, n_G+n_B-2} s_p \sqrt{\frac{1}{n_G} + \frac{1}{n_B}}$. We need the pooled sample standard deviation and critical value from the t-distribution.

$$s_p = \sqrt{\frac{23 \times 102^2 + 18 \times 139^2}{24 + 19 - 2}} = 116.129$$

$$t_{0.95;\,41} = 1.683 \quad \text{if not available in tables use } t_{0.95;\,40} = 1.684$$

The 90% confidence interval is

$$(239 - 336) \pm 1.683 \times 116.129\sqrt{1/24 + 1/19} = -97 \pm 60.013$$

2. A study was conducted to compare the effects of using an existing drug (Prochlorperazine) with using THC (the active ingredient in marijuana) for nausea relief in patients undergoing chemotherapy for cancer. A total of 157 patients being treated in a cancer clinic were divided into two groups in a randomised, double-blind, comparative experiment. The table below shows how many patients in each group found the treatment to be effective or not effective.

	Effective	Not Effective
THC	36	43
Prochloroperazine	16	62

Compute a 99% confidence interval for the true difference in nausea relief rates between THC and Prochloroperazine.

Solution: Let p_{THC} and $p_{proch.}$ be the proportion of people who would get relief from nausea from THC and prochloroperazine, respectively. The confidence intervale for $p_{THC} - p_{proch.}$ is

$$\widehat{p}_{THC} - \widehat{p}_{proch.} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_{THC}(1-\widehat{p}_{THC})}{n_{THC}} + \frac{\widehat{p}_{proch.}(1-\widehat{p}_{proch.})}{n_{proch.}}}.$$

The proportion of subjects in the study who found relief with THC is $\widehat{p}_{THC} = 36/79 = 0.456$. The proportion of subjects in the study who found relief with Prochloroperazine is $\widehat{p}_{proch.} 16/78 = 0.205$. The standard error of $\widehat{p}_{THC} - \widehat{p}_{proch.}$ is

$$se(\widehat{p}_{THC} - \widehat{p}_{propch.}) = \sqrt{\frac{0.456 \times 0.544}{79} + \frac{0.205 \times 0.795}{78}} = 0.0723$$

We need the critical value from the normal distribution $z_{0.995} = 2.5758$. So the 99% confidence interval for the difference in proportions is

$$(0.456 - 0.205) \pm 2.5758 \times 0.0723 = 0.251 \pm 0.1858$$

3. A total of 200 Brisbane suburban trains were randomly selected. The arrival time of each train at various train stations across the city network was compared to the published timetable to see if they arrived late. 32 trains were observed to arrive late. Construct a 95% confidence interval for the true proportion for Brisbane suburban trains that arrive late at their station.

   Solution: Let p be the proportion of trains that are late. The confidence interval for p is $\widehat{p} \pm z_{1-\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}$.

$$\begin{aligned} \widehat{p} &= \frac{32}{200} = 0.16 \\ z_{0.975} &= 1.96 \\ sd(\widehat{p}) &= \sqrt{\frac{0.16 \times 0.84}{200}} = 0.02592. \end{aligned}$$

   So the 90% confidence interval is

$$0.16 \pm 1.96 \times 0.02592 = 0.16 \pm 0.0508.$$

4. A random sample of 25 cattle trucks was selected and weighed at a highway weighing station. From this sample, it was found that the average truck weight was 28 tonnes and the sample standard deviation of 0.3 tonnes. The critical value needed to construct a 95% confidence interval for the mean weight (in tonnes) of cattle trucks using this data is -

   (a) 1.708
   (b) 2.064
   (c) 2.485
   (d) 2.492

   Solution: Sample size $n = 25$ so the degrees of freedom of the t-distribution is $n - 1 = 24$. As we want a 95% confidence interval the critical value is $t_{0.975;\,24} = 2.063899$. The answer is (b).

5. A quality control worker in a factory wishes to estimate the mean weight for boxes of strawberries being packed at the factory so as to be within 50 grams. The distribution of the weight of boxes of strawberries is known to have a standard deviation of 120 grams from past experience. For a 95% level of confidence, the number of boxes of strawberries the quality control worker requires to sample would be -

(a) 22

(b) 23

(c) 24

(d) 25

Solution: We want the margin of error to be no more than 50g. When the standard deviation is known the margin of error is given by

$$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

The standard deviation is *known* to be 120g and the level of confidence is 95%. So we want n such that

$$\begin{aligned} 1.96 \times \frac{120}{\sqrt{n}} &\leqslant 50 \\ \sqrt{n} &\geqslant \frac{1.96 \times 120}{50} = 4.704 \\ n &\geqslant 22.12762 \end{aligned}$$

So we need $n = 23$. (Taking $n = 22$ would lead to a margin of error that is a little larger than 50g)