475.101 / 102 / 107 / 108 Semester 1 2000
Assignment 4 Solutions

Question 1.

(a) Obtain a list of names of both female and male secondary school hockey players, from the organisation administering school hockey, such as the secondary school division of the New Zealand Hockey Federation. Using a method such as the random number tables randomly select from each list at least 50 students to allow for refusal or non-response. Contact these students either via phone or mail, informing them of the purpose of the study and asking for their co-operation.

(b) For the male hockey players $\bar{x}_M = 2.70$, $s_M \approx 0.3641$, $n_M = 26$
For the female hockey players $\bar{x}_F \approx 1.9897$ $s_F \approx 0.4045$, $n_F = 30$

(c) The parameter of interest here is the difference between the mean approval score for unsporting play for New Zealand male and female secondary school hockey players.
$\theta = \mu_M - \mu_F$, where μ_M is the mean approval score for unsporting play for New Zealand male secondary school hockey players and μ_F is the mean approval score for unsporting play for New Zealand female secondary school hockey players.
Estimate is $\bar{x}_M - \bar{x}_F = 2.70 - 1.9897 = 0.7103$

$$se(\bar{x}_M - \bar{x}_F) = \sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}} = \sqrt{\frac{0.3641^2}{26} + \frac{0.4045^2}{30}} = 0.102726864$$

$df = \min(n_M - 1, n_F - 1) = 25$
We want a 95% confidence interval: $t = 2.060$

A 95% confidence interval for $\mu_M - \mu_F$ is given by:

$$0.7103 \pm 2.060 \times 0.102726864 = 0.7103 \pm 0.211617341 \approx [\, 0.4987, 0.9219 \,]$$

We estimate the true mean approval score for unsporting play for New Zealand male secondary school hockey players to be somewhere between 0.50 and 0.92 units higher than the true mean score for unsporting play for New Zealand female secondary school hockey players. Statements such as this are correct, on average, 19 times out of 20.

(d) We do not know if the confidence interval in (c) contains the true difference. We only know that the method used will lead to a confidence interval that contains the true difference for 95% of samples taken. We do not know whether or not this sample is one of the 95% that do actually contain the true difference.

Question 2.

(a) **(i)** Hypothesis test. You wish to find out if there is any evidence that the mean number of e-mails received before the power crisis is different to the mean number of emails received after the power crisis.

(ii) Confidence interval. You are interested in estimating the actual mean level of nitrogen oxide.

(b) **(i)** Some evidence against H_0.

(ii) No evidence against H_0.

(iii) Weak evidence against H_0.

(iv) No evidence against H_0.

(c) **(i)** 2-sided test: $2\times0.025 < P\text{-}value < 2\times0.05 \Rightarrow 0.05 < P\text{-}value < 0.1$

(ii) 1-sided test: $0.0005 < P\text{-}value < 0.001$

(d) **(i)** The contention of no difference must be in the null hypothesis for all *t*-tests.

$H_0 : \mu_1 - \mu_2 = 0$ v $H_1 : \mu_1 - \mu_2 > 0$

(ii) The contention of no difference must be in the null hypothesis for all *t*-tests. Also $\hat{p}$'s are estimates, we form hypotheses about parameters. Also the alternative hypothesis should be denoted H_1.

$H_0 : p_1 - p_2 = 0$ v $H_1 : p_1 - p_2 > 0$

(e) **(i)** False. The size of the multiplier, *t*, depends on **both** the desired confidence level and the degrees of freedom (*df*).

(ii) True.

(iii) False. A value of t_0 equalling 2.15 implies that the estimated value is 2.15 **standard errors** above the hypothesised value.

(iv) True

(v) True

(vi) EITHER

False. Differences or effects seen in data that are **not** easily explainable in terms of sampling variation provide convincing evidence that real differences or effects exist.

OR

False. Differences or effects seen in data that are easily explainable in terms of sampling variation **do not** provide convincing evidence that real differences or effects exist.

(vii) False. If H_0 is rejected at the 5% level, it will **not necessarily** be rejected at the 1% level. Any *P-value* ranging between (but not inclusive of) 1% and 5% leads to rejection of H_0 at the 5% level but **not** at the 1% level.

Question 3.

(a) The parameter of interest here is $\mu_M - \mu_F$ where μ_M is the mean approval score for unsporting play for New Zealand male secondary school hockey players and μ_F is the mean approval score for unsporting play for New Zealand female secondary school hockey players.

$H_0 : \mu_F - \mu_M = 0$ vs $H_1 : \mu_F - \mu_M \neq 0$

Estimate is $\bar{x}_M - \bar{x}_F = 2.70 - 1.9897 = 0.7103$

$$se(\bar{x}_M - \bar{x}_F) = \sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}} = \sqrt{\frac{0.3641^2}{26} + \frac{0.4045^2}{30}} = 0.102726864$$

df=min$(n_M - 1, n_F - 1) = 25$

$$t_0 = \frac{0.7103}{0.1027252} = 6.9145633 \approx 6.915$$

P-value < 2× 0.0001 ⇒ *P-value* < 0.0002

We have very strong evidence (*P-value* < 0.0002) of a difference between the approval score for unsporting play for New Zealand male secondary school hockey players and the approval score for unsporting play for New Zealand female secondary school hockey players.

(b) The purpose of this study was to investigate if there is a difference between males and females in the extent of approval of unsporting play. Twenty six male and thirty female New Zealand secondary school hockey players were randomly chosen to complete a questionnaire that was used to arrive at an approval score with higher scores indicating greater approval rating for unsporting play.

We can conclude from our results that there is a significant difference in approval score between male and female New Zealand secondary school hockey players, with males approval score being on average higher than that of females by somewhere between 0.50 and 0.92.

As the university lecturer is interested in examining the extent of approval of unsporting play in New Zealand secondary schools as a whole we recommend that he or she randomly samples all New Zealand secondary school students.

Question 4.

(a) (i) Situation (b): One sample, several response categories.

(ii) Situation (c): One sample, two or more Yes/No items.

(iii) Situation (a): Two independent samples.

(b) (i) $se(\hat{p}_1 - \hat{p}_2) = \sqrt{\dfrac{0.145 + 0.035 - (0.145 - 0.035)^2}{2000}} = 0.0091624233 \approx 0.00916$

(ii) $se(\hat{p}_1 - \hat{p}_2) = \sqrt{\dfrac{0.385 + 0.295 - (0.385 - 0.295)^2}{2000}} = 0.0183289 \approx 0.01833$

(iii) $se(\hat{p}_1 - \hat{p}_2) = \sqrt{\dfrac{0.5 \times (1 - 0.5)}{1000} + \dfrac{0.39 \times (1 - 0.39)}{1000}} = 0.0220885 \approx 0.02209$

(c) The parameter of interest here is $p_M - p_F$ where p_M is proportion of male teenagers who have had serious discussions about the risks of using illegal drugs with both parents and p_F and the corresponding proportion of female teenagers.

$H_0 : p_M - p_F = 0$ vs $H_1 : p_M - p_F \neq 0$

***Minitab* Computer Output**

Test and Confidence Interval for Two Proportions

Sample	X	N	Sample p
Males	500	1000	0.500000
Females	390	1000	0.390000

Estimate for p(1) - p(2): 0.11
95% CI for p(1) - p(2): (0.0667074, 0.153293)
Test for p(1) - p(2) = 0 (vs not = 0): Z = 4.98 P-Value = 0.000

***Excel* Computer Output**

	A	B
1	Test of No Difference Between	
2	Two Population Proportions	
3		
4	Input data	
5	p1_ratio	0.5
6	p2_ratio	0.39
7	p_diff	0.11
8	X1_sample	500
9	X2_sample	390
10	n1_total	1000
11	n2_total	1000
12		
13	Alpha	0.05
14	Calcualated Value	
15	se	0.022088459
16	Zo	4.979976302
17		
18	Two-Tail Test	
19	pvalue	6.36907E-07
20		
21	Confidence Intervals (95%)	
22	Lower Limit	Upper Limit
23	0.066706621	0.153293379

We have very strong evidence (*P-value* ≈ 0.000) of a difference between the true proportion of male teenagers who have had serious discussions about the risks of using illegal drugs with both parents and the proportion of female teenagers who have had serious discussions about the risks of using illegal drugs with both parents.

With 95% confidence the true proportion of male teenagers who have had serious discussions about the risks of using illegal drugs with both parents is higher than the proportion of female teenagers who have had serious discussions about the risks of using illegal drugs with both parents by somewhere between 6.7% and 15.3%.

(d) $H_0 : p_M - p_F = 0$ vs $H_1 : p_M - p_F \neq 0$

From the ***Excel*** output we can see that we have very strong evidence (*P-value* ≈ 0.0000) of a difference between the proportion of teenagers who describe their relationship with their mother as excellent (p_M) and the proportion of teenagers who describe their relationship with their father as excellent (p_F). We estimate that the true proportion of teenagers who describe their relationship with their mother as excellent to be between 5.4% and 12.6% more than the proportion of teenagers who describe their relationship with their father as excellent. Statements such as this are correct, on average, 19 times out of 20.