475.101 / 102 / 107 / 108 Semester 2 2000
Assignment 4 Solutions

Question 1.

(a) The parameter of interest here is the mean number of nights spent at this hotel.
$\theta = \mu$.
Estimate is $\bar{x}$ = 3.75
$se(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{1.8316}{\sqrt{32}} = 0.3238$
$df = n - 1 = 31$
We want a 95% confidence interval: t = 2.040

A 95% confidence interval for μ is given by:
$3.75 \pm 2.040 \times 0.3238 = 3.75 \pm 0.6606 \approx [\,3.09, 4.41\,]$

We estimate the true mean number of nights spent at this hotel to be 3.75 nights with a margin of error of 0.66. Statements such as this are correct, on average, 19 times out of 20.

(b) We do not know if the confidence interval in (a) contains the true mean. We only know that the method used will lead to a confidence interval that contains the true mean 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 mean.

Question 2.

(a) **(i)** No evidence or weak to no evidence.
(ii) Weak evidence or some to weak evidence.
(iii) No evidence.
(iv) Very strong evidence.

(b) **(i)** 2-tailed test: $2 \times 0.005 < P\text{-}value < 2 \times 0.01 \Rightarrow 0.01 < P\text{-}value < 0.02$
(ii) 1-tailed test: $P\text{-}value < 0.0001$

(c) **(i)** Confidence interval. You are interested in estimating the actual mean execution time.
(ii) Hypothesis test. You wish to find out if there is any evidence that the containers were being overfilled or underfilled thus requiring the production line to be shut down.

(d) **(i)** The contention of no difference must be in the null hypothesis for all 2 sample *t*-tests.
$H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$

(ii) $\bar{x}$'s are estimates, we form hypotheses about parameters.
$H_0: \mu_1 - \mu_2 = 0$ vs $H_1: \mu_1 - \mu_2 \neq 0$

(e) **(i)** False. A *test statistic* is a measure of discrepancy between what we see in the data **and what we would expect to see if H_0 is true**.

(ii) True.

(iii) False. The margin of error is the quantity added to and subtracted from the **estimate** to construct a confidence interval.

(iv) False. A two-sided *t*-test will reject (at the 5% level) values for the null hypothesis that are **outside** the corresponding 95% confidence interval.

(v) True.

(vi) True.

Question 3.

(a) The parameter of interest here is the mean number of nights spent at this hotel.

$\theta = \mu$.

$H_0 : \mu = 3.4$ vs $H_1 : \mu \neq 3.4$

Estimate is $\bar{x} = 3.75$

$se(\bar{x}) = \dfrac{s}{\sqrt{n}} = \dfrac{1.8316}{\sqrt{32}} = 0.3238$

$df = \min(n-1) = 31$

$t_0 = \dfrac{3.75 - 3.4}{0.3238} = \dfrac{0.35}{0.3238} = 1.080914 \approx 1.081$

$2 \times 0.1 < P\text{-}value < 2 \times 0.15 \Rightarrow 0.2 < P\text{-}value < 0.3$

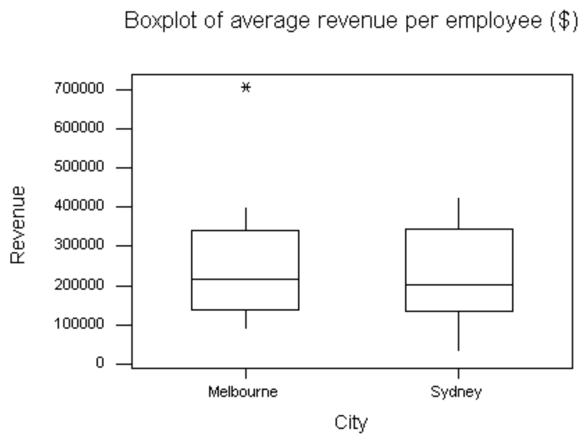
We have no evidence ($0.2 < P\text{-}value < 0.3$) that the mean number of nights spent at this hotel has changed since the recent fall in the New Zealand dollar.

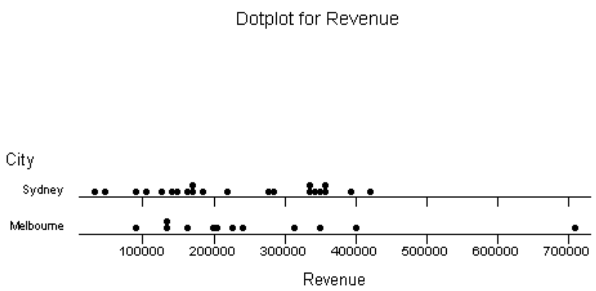
(b) No we cannot justify carrying out a one-tailed *t*-test as there is no indication (prior to conducting the study) as to whether the mean number of nights spent at this hotel will increase or decrease.
Note: It is not legitimate to base the choice of hypotheses on what we see in the sample data.

(c) A 90% confidence interval for μ will contain 3.4 as the *P-value* for the two-tailed test is greater than 0.1 (10%) so 3.4 (θ_0) lies inside the corresponding 90% confidence interval.

Question 4.

(a) The plots show that Melbourne revenue is centred slightly higher than Sydney revenue. The "Melbourne" data set appears to be symmetric with short tails and has an outlier at just over \$700,000. The "Sydney" data set also appears to have short tails .





(b) The parameter of interest here is $\mu_S - \mu_M$ where μ_S is the mean revenue per employee for companies with head offices in Sydney and μ_M the mean revenue per employee for companies with head offices in Melbourne.

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

E	F	G
t-Test: Two-Sample Assuming Unequal Variances		
	Sydney	*Melbourne*
Mean	228635	262382.3333
Variance	14037454013	28287533374
Observations	22	12
Pearson Correlation	#N/A	
Pooled Variance	18935918793	
df	17.10440392	
t	-0.61661617	
P(T<=t) one-tail	0.272828847	
t Critical one-tail	1.739606432	
P(T<=t) two-tail	0.545657695	
t Critical two-tail	2.109818524	

We have no evidence (*P-value* = 0.546) of a difference in the mean revenue per employee for companies with head offices in Sydney and the mean revenue per employee for companies with head offices in Melbourne. With 95% confidence, the mean revenue per employee for companies with head offices in Sydney is somewhere between $149,217 less than that of the mean revenue per employee for companies with head offices in Melbourne or up to $81,723 more. Since zero is contained in this interval there may be no difference in mean revenue per employee for companies with head offices in Sydney or Melbourne.

Question 5.

(a) The parameter of interest here is $p_S - p_T$ where p_S is proportion of travellers using Singapore Airlines who would definitely recommend the airline to a friend and p_T the corresponding proportion of travellers using Thai Airlines.

$H_0 : p_S - p_T = 0$ vs $H_1 : p_S - p_T \neq 0$

```
Test and Confidence Interval for Two Proportions

Sample       X      N  Sample p
1          194    316  0.613924
2           29     66  0.439394

Estimate for p(1) - p(2):  0.174530
95% CI for p(1) - p(2):  (0.0433107, 0.305750)
Test for p(1) - p(2) = 0 (vs not = 0):  Z = 2.61  P-Value = 0.009
```

We have very strong evidence (*P-value* = 0.009) of a difference between the true proportion of travellers using Singapore Airlines who would definitely recommend the airline to a friend and the proportion of travellers using Thai Airlines who would definitely recommend the airline to a friend.

With 95% confidence, the true proportion of travellers using Singapore Airlines who would definitely recommend the airline to a friend is higher than the proportion of travellers using Thai Airlines who would definitely recommend the airline to a friend by somewhere between 4.3% and 30.6%.

(b) This data set should not be treated as 2 independent samples as the samples may not be independent. Some Consumer members may have travelled on both Singapore Airlines and Thai Airlines in the last twelve months (and therefore could be in both samples).

Question 6.

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

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

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

(b) **(i)** $$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.35 + 0.31 - (0.35 - 0.31)^2}{1000}} = 0.02565931 \approx 0.02566$$

(ii) $$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.05(1-0.05)}{1000} + \frac{0.08(1-0.08)}{1000}} = 0.01100454 \approx 0.01100$$

Question 7.

(a) Dr David Murray thinks that the CNN headline "Poll: Gore Closes Gap Against Bush," abuses the results from a recent poll as point estimates only were reported. There was no mention of the margins of error, sampling variability or other possible values.

(b) The point made about sub-samples in polls is that the margin of error becomes larger.

(c) In media reports of polls any reported estimates should also be accompanied by a margin of error statement.