Statistics Workshop - Second Semester Revision Solutions

1. (a) What is the area below $z = 1.74$?

   ANSWER: Area $= 0.9591$

   (b) A population has a NORMAL distribution with a mean of 230 and a standard deviation of 46. What is the area under the curve between 190 and 260?

   ANSWER: Area$= 1 - 0.1922 - 0.2578 = 0.55$.

   (c) Samples of size 9 are taken from a normal poulation with $\mu = 49$ and $\sigma = 12$. What is the probability that a sample will have a mean of 54 or more?

   ANSWER: $z = 1.25$ and so probability =0.1056.

   (d) A normal population has a mean of 76 and a standard deviation of 16. Find the 24th percentile of this distribution.

   ANSWER: $z = -0.71$ and so $y = 64.64$.

   (e) 80 samples of size 50 are taken from a population with a mean of 24 and a standard deviation of 8.

   i. What is the shape of the distribution of sample means?
      ANSWER: Normal - CLT applies.

   ii. How many samples would you expect to have a mean of greater than 24.8?
      ANSWER: $z = 0.71$ so required area $= .2389$ therefore 19 samples.

   (f) What is the interquartile range of the standard normal distribution?

   ANSWER: Upper quartile when $z = 0.67$, Lower quartile when $z = -0.67$ so $IQR = .67 - -0.67 = 1.34$.

2. A sample of size 14 is drawn from a population with a known standard deviation of 23. The sample mean is 95.4. We wish to test whether the true mean of the population is 86.

   (a) What test should you perform?

   ANSWER: z-test for the mean

   (b) What assumptions are your making?

   ANSWER: Distribution of population is normal.

   (c) Is the true mean of the population

   i. possibly 86?

   ii. greater than 86?

   iii. less than 86?

   ANSWER: Confidence interval is $(83.35, 107.45)$ so the answer is possibly 86.

3. A sample of size 20 is drawn from a normal population. The sample mean is 24.5 and the sample standard deviation is 8.

   (a) What is a point estimate for the mean of this population?

   ANSWER: 24.5

   (b) Give a 95% confidence interval for the true mean of the population.

   ANSWER: $t\ crit = 2.093$. Confidence interval $(20.76, 28.24)$.

(c) Suppose I now take a sample of 80 with the same mean and standard deviation. What happens to the confidence interval?

ANSWER: Confidence interval almost half, slightly different from half because of change in t *crit*.

(d) What if the sample size is 5?

ANSWER: Confidence interval almost double, slightly different from double because of a change in t *crit*.

(e) Suppose I take 200 samples and construct confidence intervals for each one. How many would you expect to contain the true mean?

ANSWER: $.95 \times 200 = 190$.

4. Looking at unemployment rates in 6 suburbs of Sydney in 2000 and 2003. Has there been a change in unemployment rates? Assume that the differences come from a normal population.

Suburb	A	B	C	D	E	F
2000	10	12	15	20	18	13
2003	8	11	16	16	15	11

ANSWER: There has been a change in unemployment rates. (t-test on differences, $\bar{y}_d = 1.833$, $s_d = 1.722$, test statistic $t = 2.607$ and $.02 < p < .05$.) 95% sure unemployment rates are somewhere between 0.03 and 3.64 lower in 2003 than they were in 2000.

5. We know that the scores on a maths test follow a normal distribution. We wish to know whether students from school A do better in the test than students from school B. A sample of students from school A and school B were selected and the following summarises the data obtained from these samples.

School	sample size	mean	standard deviation
A	12	74.8	12.3
B	15	71.6	13.7

(a) What is an estimate for the difference of means?

ANSWER: $74.8 - 71.6 = 3.2$

(b) What is the estimated standard error of the difference in sample means?

ANSWER: $s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 13.10244252 \times \sqrt{\frac{1}{12} + \frac{1}{15}} = 5.075$.

(c) Is there a significant difference in the mean scores of the two schools?

ANSWER: 2-sample t-test , $t = 0.63$, $p > 0.5$, therefore no significant difference.

(d) What assumptions am I making here? How can I tell if my assumptions are reasonable?

ANSWER: Both samples come from normal populations and $\sigma_1 = \sigma_2$.

6. A sample of size 42 is drawn from a population with a mean of 174. The sample mean is 181.3 and the sample standard deviation is 12.4. Is this particular sample different from the population?

ANSWER: One sample t-test, $t = 3.815$, $p < 0.0005$ so this sample is different from the population.

7. People who harvest wild mushrooms sometimes accidentally eat the toxic death cap mushroom, Amanita phalloides. In reviewing 205 European cases of death cap poisoning from 1971 through to 1980, researchers found that 45 of the victims had died. Compare this mortality to the 30% mortality that was recorded before 1970; is it the same?

   ANSWER: z-test on proportions, $p = \frac{45}{205}$, $\pi = .3$ and then we get $z = -2.51$ and $p = .0121$. So the mortality rate is different. Based on this sample we can be 95% certain that the mortality rate is somewhere between 16.3% and 27.5%.

8. The distribution of the number of sternopleural bristles in Drosophilia melanagaster is as follows:

Number of bristles	7	8	9	10	11	12	13	14
Proportion	.004	.093	.335	.355	.167	.037	.008	.001

   The mean number of bristles is 9.736 and the standard deviation is 1.0375.

   (a) A new strain of Drosophilia was found to have a mean number of bristles equal to 10.0 in a sample of size 100. Does this new strain have more bristles than the original population?

   ANSWER: z-test for the mean, $z = 2.54$ and $p = .0111$, so there is evidence to suggest this strain has more bristles than the previous strain.

   (b) In a sample of size 200 from the new strain, it was found that 50 of the flies had more than 10 bristles. Could we say that this new strain is similar to D. melanagaster in the proportion of flies with more than 10 bristles?

   ANSWER: z-test for population proportion with $\pi_0 = .167 + .037 + .008 + .001 = .213$, $z = 1.28$ and $p = .2005$ so this strain is similar to the old strain in the proportion of flies with more than 10 bristles.

9. We know that the amount of money spent annually on clothes by university students follows a normal distribution. 8 sets of fraternal twins were chosen randomly from the Macquarie University student body and asked to estimate the amount of money they spent on clothes a year. The results of this survey are given below.

Twins	Male	Female
A	\$550	\$600
B	\$400	\$380
C	\$450	\$550
D	\$520	\$500
E	\$750	\$750
F	\$680	\$600
G	\$600	\$650
H	\$550	\$580

   (a) We wish to test whether there is a difference in the expenditure of male and female university students on clothes. What is the appropriate test here?

   ANSWER: paired t-test (one sample t-test on the differences).

   (b) What assumptions are we making, if any? How can we check our assumptions?

   ANSWER: The differences arise from a normal distribution. Histogram.

   (c) Calculate a confidence interval for the difference in expenditure.

   ANSWER: $(-60.18, 32.68)$ doing the Males minus the Females. (If you did the other way you should just get the same numbers with the signs reversed.

10. Some data were collected on the temperature (in degrees Celsius) at noon for a small island. A random sample of these temperatures yielded the following values:

25	21	21	18	27	28	26
20	19	23	23	17	28	23
24	26	24	29	24	18	30
25	27	19	24	23	23	24

Test the claim that the median temperature at noon for this small island is 22 degrees Celsius. Justify any assumptions required to answer this question.

ANSWER: This is a z-test on proportions. Null hypothesis is that 50% of the days are warmer than 22° C. The assumptions we need to test are that $n\pi_0 \geq 5$ and $n(1-\pi_0) \geq 5$. Since $28 \times 0.5 = 14 \geq 5$ and $28(1-0.5) = 15 \geq 5$, the CLT applies. The proportion of days with temperatures greater than 22° is sample proportion $p = \frac{20}{28}$. The test statistic $z = 2.27$ and so the probability is 0.0232. The median temperature is unlikely to be 22° C and the confidence interval tells us that the proportion of days with temps higher than 22° C is probably between .55 and .88.