DALHOUSIE UNIVERSITY FACULTY OF SCIENCE

Department of Mathematics and Statistics

STATISTICS 1060A(01,02,03) Final Examination

Date and Time: 1:30-4:30pm, 3 December, 2002

1. In class this fall, we collected some data on the time (in seconds) that it takes a marble to drop through a funnel. There were 30 measurements taken, and these were entered into column C1 of Minitab. Here is some Minitab output for these data:

```
MTB > stem c1

Stem-and-leaf of C1        N  = 30
Leaf Unit = 0.010

    1   46 9
    1   47
    2   48 4
    2   49
    3   50 0
    5   51 02
    5   52
   11   53 244478
   12   54 4
   14   55 00
   (5)  56 36666
   11   57 22555
    6   58 112
    3   59 4
    2   60 00

MTB > descr c1

                  N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
C1               30   5.5207   5.6450   5.5419   0.3306   0.0604

                MIN      MAX       Q1       Q3
C1           4.6900   6.0000             5.7500
```

(2) (a) Find Q1 for these data.

Q1=5.34

(3) (b) Are there any outliers in the data? If so, identify them. Justify your answer with the appropriate calculations.

1.5IQR= 1.5(5.75-5.34)=.615

Q1-1.5IQR = 4.725

Q3+1.5IQR = 6.365

4.69 is an outlier as it is below 4.725.

(2) (c) What would be the mean if the data were recorded in minutes rather than seconds?

5.5207/60 = .092

(2) (d) What would be the standard deviation if the data were recorded in minutes rather than seconds?

.3306/60 = .00551

(3) (e) If we deleted the smallest observation, what would the new mean of the data be?

(30*5.5207 - 4.69) / 29 = 5.549

2. A person has two meetings scheduled in a day. The probability she is late for the first meeting is .4, the probability she is late for the second is .5, and the probability she is late for both meetings is .3.

(2) (a) Is the event that she is late for the first meeting independent of the event that she is late for the second meeting? Show your work.

Let A=late for first B=late for second

No,because .4*.5=P(A)P(B) does not equal P(A and B)=.3

(2) (b) Are the two events disjoint? Explain.

No, because P(A and B) does not equal zero (ie there are outcomes in common to A and B).

(2) (c) What is the probability she is late for the second meeting if she was late for the first meeting?

P(B—A) = P(A and B) / P(A) = .3/.4 = .75

(3) (d) What is the probability she is late for at least one meeting?

P(A or B) = P(A) + P(B) - P(A and B) = .4+.5-.3=.6

3. The actual weight of almonds in a 1-pound tin produced by a company is approximately normally distributed with a mean of 16.00 ounces and a standard deviation of 0.05 ounce.

(2) (a) What are the mean and standard deviation of the sampling distribution of the mean weight of the almonds in a random sample of 49 tins?

mean = 16. standard deviation = $.05/\sqrt{49}$ = .007

(3) (b) How large a sample would we need to estimate the mean weight of the almonds in a tin with a margin of error of at most .0075 ounces, for a 99% confidence interval.

$(\frac{Z_{.005}\sigma}{.0075})^2$ = (2.576*.05/.0075)**2 = 294.92

require n=295

(3) (c) What is the probability that the mean weight of the almonds in 36 randomly selected tins would differ from the population mean by at most .01 ounce?

$P(-.01 < X - \mu < .01) = P(-.01/(.05/\sqrt{36}) < Z < .01/(.05/\sqrt{36}))$

$= P(-1.2 < Z < 1.2) = .770$

4. The number of births per day on the maternity ward in the Grace Maternity hospital is known to have a standard deviation of 1.2 births. The mean number of births in a sample of 36 days was 3.14 births.

(5) (a) The administrator of the Grace Maternity hospital claims that the mean number of births per day is greater than 3.5. What conclusion can you draw on the basis of the sample? State your hypothesis, give the test statistic, find the P-value and draw a conclusion at $\alpha = .01$

$H_0 : \mu = 3.5$ $H_A : \mu > 3.5$

Z= (3.14 - 3.5)/(1.2/$\sqrt{36}$) = -1.8

p-value = $P(Z > -1.8)$ = .964

do not reject at level .01 as p-value > .01.

(4) (b) Find a 95% percent confidence interval for the true mean number of births per day in the Grace Maternity hospital.

$3.14 \pm 1.96(1.2)/\sqrt{36}$

(2.748, 3.532)

(2) (c) If we increase the sample size to 49 days, will the 95% confidence interval become wider?

No, width of confidence interval is $2Z_{.025}\sigma/\sqrt{n}$, which decreases with n.

(2) (d) If we construct a 90% confidence interval based on a sample of size 36, will this interval be wider than the confidence interval in (b)?

No, the smaller the confidence coefficient (eg 90%), the narrower the interval.

5. A game consists of tossing two coins, first a dime and then a nickel. For each coin which comes up a head you win the value of the coin. For each coin which comes up a tail you get nothing.

(3) (a) List the possible outcomes, in terms of heads and tails, and give the probability and your winnings for each outcome.

TT - 0; TH - 5; HT - 10; HH - 15

each outcome has probability .25

(2) (b) Find the mean or expected winnings.

0*.25 + 5*.25 + 10*.25 + 15*.25 = 7.5

(2) (c) Find the standard deviation of your winnings.

(0-7.5)**2*.25 + (5-7.5)**2*.25 + (10-7.5)**2*.25 + (15-7.5)**2*.25 =31.25

standard deviation = $\sqrt{31.25}$ = 5.59

(2) (d) If you had to pay a dime to play the game, what would be the mean and standard deviation of your winnings?

mean = 7.5 - 10 = -2.5

standard deviation = 5.59

6. The following table summarizes voltage drop measurements for samples of connectors wired with alloy aluminum and EC aluminum.

Type	*Sample Size*	*Sample Mean*	*Sample SD*
Alloy	20	17.5	0.55
EC	20	16.9	0.49

Do these data indicate that the true voltage drop for alloy connections is higher than for EC connections?

(5) (a) Carry out an appropriate test at the 0.01 level of significance.

let μ_1 be mean voltage drop for alloy, μ_2 for EC

$H_0 : \mu_1 - \mu_2$ $H_A : \mu_1 > \mu_2$

$S_p^2 = (19 * .55 * *2 + 19 * .49 * *2)/(38)$=.2713

$S_p = \sqrt{.2713} = .521$

test statistic $t = (17.5 - 16.9)/(.521 * \sqrt{1/20 + 1/20}) = 3.64$

p-value $=P(t_{38} > 3.64) \approx .0004$

reject at level .01

(3) (b) Calculate a 99% confidence interval for the difference in mean voltage drop between alloy aluminum and EC aluminum connectors.

$(17.5 - 16.9) \pm t_{38,.005}.521 * \sqrt{1/20 + 1/20}$

$(17.5 - 16.9) \pm 2.71 * .521 * \sqrt{1/20 + 1/20}$

(.154, 1.05)

7. A city council is considering drafting legislation related to the exclusion of renters with children. Since implementation will be expensive they will only consider drafting legislation if the true proportion of apartments that prohibit children is more than 75%.

(5) (a) Suppose that out of 125 sampled apartments, 102 exclude renters with children. Should the city council draft legislation? Be sure to state the null and alternative hypothesis.

Let p proportion of apartments that prohibit children. $\hat{p} = 102/125 = .816$.

$H_0 : p = .75$ $H_A : p > .75$

test statistic $Z = (.816 - .75)/\sqrt{.75(.25)/125} = 1.70$

p-value = P(Z > 1.70) = .045.

(3) (b) In order to be sure to have a margin of error less than 0.01 for a 95% confidence interval, how many apartments would need to be sampled?

presumption here is that we don't have knowledge of true value of p, so use (1.96/.01)**2*.25 = 9604

(5) 8. Two cars A and B have quite different lengths, wheelbases and turning radii. Four subjects were asked to parallel park each car. The time in seconds required are given below.

	Subject			
	1	2	3	4
Car A	37.0	25.8	16.2	24.2
Car B	17.8	20.2	16.8	41.4

Does the data suggest that a person will more easily handle one car than the other? Test the relevant hypotheses by calculating and interpreting a P-value.

let difference be Car A time - Car B time

mean difference = 1.7 standard deviation of differences = 15.18

let μ be mean difference

$H_0 : \mu = 0$ $H_A : \mu \neq 0$

test statistic $t = 1.7/(15.18/\sqrt{4}) = .224$

p-value $= 2P(t_3 > .224) \approx .837$

p-value large, so little evidence against H_0.

(4) 9. A tension link in an engineering application can be constructed using either alloy A or alloy B. The ultimate strength (ksi) for a number of tension links is tabulated below.

	A	B
26-under 30	6	4
30-under 34	12	9
34-under 38	15	19
38-under 42	7	10

Compute a 95% confidence interval for the difference between the true proportions of all specimens of alloys A and B that have an ultimate strength of at least 34 ksi.

X1=22 n1=40 X2=29 n2=42

$\tilde{p}_1 = 23/42 = .548$ $\tilde{p}_2 = 30/44 = .682$

$(.548 - .682) \pm 1.96 * \sqrt{.548 * (1 - .548)/42 + .682 * (1 - .682)/44}$

(-.338, .070)

10. Only 40% of drivers properly clean their car windows after a snowstorm.

(2) (a) If you randomly observe 3 cars after a storm, what is the probability that exactly 2 have had their windows properly cleaned?

X=number with windows cleaned, has Bin(n=3,p=.4) distribution.

$P(X = 2) = \left(\frac{3}{2}.4^2.6^1\right)$

3*.16*.6 = .288

(4) (b) If you randomly observe 250 cars, and let X denote the number with properly cleaned windows, what are the mean and variance of X?

mean = 250(.4) = 100

variance = 250(.4)(.6) = 60

(4) (c) If you randomly observe 250 cars, what is the probability that 90 or fewer have their windows properly cleaned?

standard deviation $= \sqrt{60} = 7.75$

X approximately N(100,7.75)

$P(X \leq 90) = P(Z \leq (90 - 100)/7.75) = P(Z \leq 1.29) = .099$

11. A data set consists of 1000 observations. 500 of the observations are equal to -1, and 500 of the observations are equal to +1.

(2) (a) What is the mean of the data?

(500*(-1)+500*(1))/1000 = 0

(2) (b) What is the standard deviation of the data?

variance is (500*(-1-0)**2+500*(1-0)**2)/999 = 1.001

standard deviation $= \sqrt{1.001} = 1.0005$

12. The relationship between energy consumption Y (in units of 10^8 Btu/year) and household income X (in units of \$10000/year) was studied.

It was found that $S_x = 2.5$, $S_y = 27.1$, $\bar{Y} = 58.1$, $\hat{\beta}_1 = 10.4$, and $\hat{\beta}_0 = 1.2$.

(2) (a) Calculate the correlation coefficient between X and Y.

$r = \hat{\beta}_1 S_x / S_y = 10.4(2.5)/27.1 = .959$

(2) (b) If $x = 5$ (household income of \$50,000), estimate the mean energy consumed for households of this income.

1.2+10.4(5) = 53.2

(2) (c) How much would you expect the change in consumption to be if a household's income increases \$20000/year (2 units of \$10000)?

10.4(2) = 20.8

(2) (d) Calculate $\bar{X}$.

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$1.2 = 58.1 - 10.4\bar{X}$

$\bar{X} = (58.1 - 1.2)/10.4 = 5.47$

(2) (e) If X had been measured in units of \$1000/year, what would be the slope of the regression line?

X is multiplied by 10, so $\hat{\beta}_1$ will be divided by 10. $10.4/10 = 1.04$