UNIVERSITY OF NSW, SCHOOL OF MATHEMATICS AND STATISTICS
MATH2089 MATLAB test

STATISTICS COMPONENT. Sample Test questions

Note that in your test there will only be 2 questions on the statistics component, however there are more questions here to help with your preparation.

Note that $X \sim N(\mu, \sigma)$ denotes a **normal** random variable with mean μ and standard deviation σ;
$X \sim Exponential(\mu)$ denotes an **exponential** random variable with mean μ;
$X \sim Poisson(\mu)$ denotes a **Poisson** random variable with mean μ; and
$X \sim Binomial(n, \pi)$ denotes a **binomial** random variable with n trials and chance π of success.

Question 1 The times (in mins) taken by two independent processes, X and Y can be modelled by

$$X \sim N(12, 4) \qquad Y \sim Exponential(1).$$

1 − normcdf(6, 12, 4).

i) Determine $P(X > 6)$ (give answer to 4 decimal places). $\boxed{0.9332}$

ii) The total time is $T = X + Y$. Determine the mean of T? (This part of question does not require the use of Matlab) $\boxed{\mu_T = \underbrace{12}_{\mu_X} + \underbrace{1}_{\mu_Y} = 13}$

iii) Use Matlab to generate 1000 random values of T.

a) *Write down the required matlab command* used to generate these values:

```
RT= normrnd (12,4,1000,1) +exprnd(1,1000,1);
```

b) Use Matlab to generate to determine the mean of your sample (to 4 decimal places) and verify that this value supports your answer to part ii):

- Write down the mean of your sample: Sample mean = $\boxed{13.1405}$

this was for my random sample: your's may be different

- *Write down the required matlab command* used to generate the mean of your sample:

```
mean (RT)
```

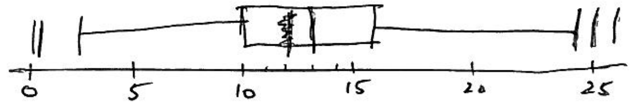
- Compare your sample mean with the answer to part (ii).

$\mu_T = 13$ and the sample mean $\overline{RT} = 13.1405$ these are close, as expected.

c) Determine the maximum value of your sample: Maximum= $\boxed{27.1685}$

yours may be different

d) Use Matlab to produce a horizontal boxplot for your sample of 1000 values, and sketch the resulting graph here.



e) Comment on the key features of the boxplot.

fairly symmetric and unimodal
Centred about 13
ranged between about 0 and 27.1685
about 4 outliers.

Question 2

i) Suppose that $X \sim Binomial(25, 0.5)$.

a) Determine (to 4 decimal places) the probability $P(10 \leq X \leq 15)$ $\boxed{0.7705}$

b) *Write down the Matlab command* required to answer part a).

```
binocdf(15,25,0.5) - binocdf(9,25,0.5)
```

ii) Geotechnical engineers use water-level surveys to assess the levelness of concrete slabs. Of interest is the *maximal-difference* (in cm) between elevations. Data for a random sample of slabs is in the file *SAMPLEDATA.txt* with a header for the column 'maxdiff'. Locate the data file on the Desktop and import it into matlab using the **Import data** command on the file menu. Make sure that on the last interactive screen you tick the button that says "create vectors from each column using column names"
Check that the mean of the *maximal-differences* is 1.0546. If it isn't, check you have imported the dataset correctly.

a) For what proportion of the sample values is the 'maxdiff' bigger than 1.2 cm? $\boxed{\frac{24}{50}}$

b) Calculate a 90% confidence interval for the proportion of all slabs with *maximal-difference* greater than 1.2cm. (Include the necessary working.)

$(n = 50 \quad p = \frac{24}{50})$ ~~(p = sum(maxdiff > 1.2)~~

```
n= length (maxdiff);
p= sum (maxdiff > 1.2) /n;
zstar = norminv (0.95);
[p - zstar * sqrt(p(1-p)/n), p+ zstar *sqrt (p(1-p)/n)]
```

.05 .05 −z* 0 z*

[0.3638, 0.5962].

3

Question 3

i) Suppose that $X \sim Poisson(14)$.

a) Determine (to 4 decimal places) the probability $P(10 \leq X \leq 15)$ 0.4982.

b) *Write down the Matlab command* required to answer part a).

poisscdf(15,15) - poisscdf(9,15)

ii) Suppose that $Y \sim Poisson(10)$.

a) Determine (to 4 decimal places) the probability $P(Y \geq 15)$ 0.0835

b) *Write down the Matlab command* required to answer part a).

1- poisscdf(14, 10)

iii) Suppose that $W \sim N(2, 1)$.

a) Determine (to 4 decimal places) the probability $P(W \leq 3.2)$ 0.8849

b) *Write down the Matlab command* required to answer part a).

normcdf (3.2, 2, 1)

iv) Geotechnical engineers use water-level surveys to assess the levelness of concrete slabs. Of interest is the *maximal-difference* (in cm) between elevations. Data for a random sample of slabs is in the file *SAMPLEDATA.txt* with a header for the column 'maxdiff'. Locate the data file on the Desktop and import it into matlab using the **Import data** command on the file menu. Make sure that on the last interactive screen you tick the button that says "create vectors from each column using column names"
Check that the mean of the *maximal-differences* is 1.0546. If it isn't, check you have imported the dataset correctly.
a) Determine the five-number summary for the data:

0.2000 ~~1.4000~~ 0.5000 1.1000 1.4000 1.9000

b) Determine the interquartile range of the data. IQR=

c) What is the size of the sample? n = 50

d) *Write down the Matlab command* which you used to answer part ~~d~~ c. i meant to ask part c,

length (maxdiff).

4.

Question 4

i) Suppose that $Z \sim N(0, 1)$.

a) Determine (to 4 dp) the value of z^* such that $P(-z^* < Z < z^*) = 0.75$) $z^* =$ 1.1503

b) *Write down the Matlab command* required to answer part a).

norminv (0.875)

ii) Suppose that $T \sim t_4$.

a) Determine (to 4 dp) $P(T < 2)$ 0.9419.

b) *Write down the Matlab command* required to answer part a).

tcdf(2,4)

c) Determine (to 4 dp) the value of t^* such that $P(-t^* < T < t^*) = 0.75$) $t^* =$ 1.3444

d) *Write down the Matlab command* required to answer part ~~d~~ c. this should have said c)

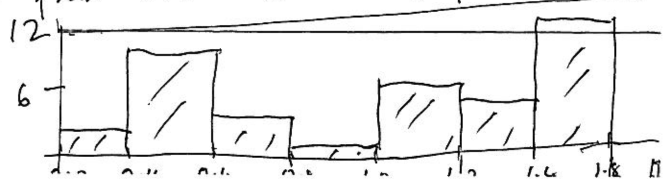
tinv(0.875, 4)

iv) Geotechnical engineers use water-level surveys to assess the levelness of concrete slabs. Of interest is the *maximal-difference* (in cm) between elevations. Data for a random sample of slabs is in the file *SAMPLEDATA.txt* with a header for the column 'maxdiff'. Locate the data file on the Desktop and import it into matlab using the **Import data** command on the file menu. Make sure that on the last interactive screen you tick the button that says "create vectors from each column using column names"
Check that the mean of the *maximal-differences* is 1.0546. If it isn't, check you have imported the dataset correctly.
Use Matlab to make a frequency histogram of the data set. Sketch the histogram below. Describe the shape of the data as reflected in the frequency histogram, - include mention of spread, centre, symmetry andor skewness if appropriate, existence of possible outliers etc.

The answer here depends on whether you use the default bins. or determine them yourself

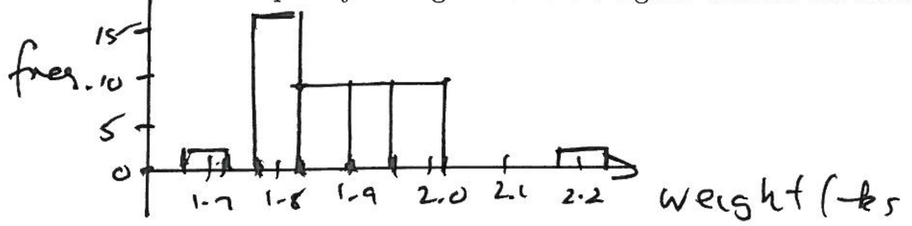
No particular pattern evident - values range from 0.2 to 1.9

eg



Question 5 A part of an environmental study into a river delta many variables are recorded, including the weights of a certain species of fish. The data file *WEIGHTS* gives the weights (in kg) for a random sample of fish caught (and released) one weekend. Locate the data file on the Desktop and import it into matlab using the **Import data** command on the file menu. Make sure that on the last interactive screen you tick the button that says "create vectors from each column using column names"

i) Use Matlab to make a frequency histogram of the weights. Sketch the histogram below.

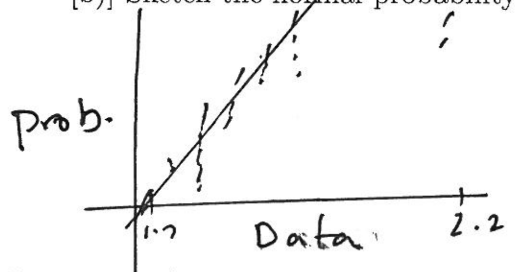


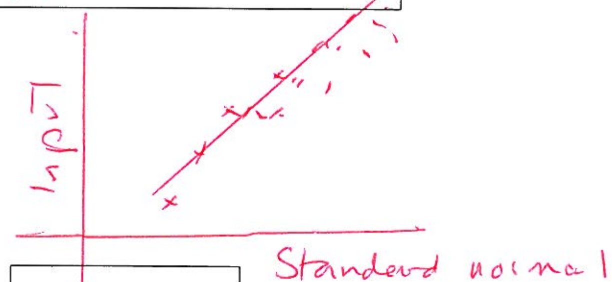
ii) Use Matlab to make a normal probability plot of the weights.

[a)] write down the matlab command you use to do the normal probability plot.

either normplot(weight) or qqplot(weight)

[b)] Sketch the normal probability plot below.





iii) Determine the mean of your sample: Sample mean = $\boxed{1.8761 \text{ kg}}$.

iv) Suppose that you can assume that the weights of fish of this species has a distribution with a standard deviation of $\sigma = 0.15$ kg. Determine a 99% confidence interval for the true mean weight (μ) of the fish. (Include necessary working.)

$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ [h p ci99] = ztest(weight, mean(weight), 0.15, 0.01)

ci99.

[1.8179, 1.9344]. 99% CI is (1.82, 1.93) kg.

v) Suppose now that you can cannot assume that you know σ. Determine a 99% confidence interval for the true mean weight (μ) of the fish. (Include necessary working.)

$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ [h p ci99] = ttest(weight, mean(weight), 0.01)

[1.8336, 1.9187]. 99% C.I is (1.83, 1.92)

vi) What assumptions do you need to make for your confidence ~~inteval~~ interval in parts (~~vi~~) iv and v to be valid? Are these plausible here? explain your answer.

we need assume random sample - can't tell if reasonable

for (v) we need that the weights are approximately normal. This looks plausible from histogram and normplot.