The
University
Of
Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS   **Spring Semester 2013–2014**

MAS113 Introduction to Probability and Statistics II   **2 hours**

*Attempt **ALL** questions. The allocation of marks is shown in brackets. Total marks 100.*

1 Two independent estimators X and Y are used to estimate an unknown parameter θ. Two new estimators Z and W are defined as

$$Z = 3X - Y - 1 \quad \text{and} \quad W = \frac{X}{3} + \frac{2Y}{3}.$$

If the sampling distributions of X and Y are $X \sim N(1, 2)$ and $Y \sim N(1, 7)$ and $\theta = 1$, then

(i)   investigate whether Z and W are unbiased estimators for $\theta = 1$;

(4 marks)

(ii)   find the standard errors of Z and W;   *(5 marks)*

(iii)   which of the two estimators Z and W is the best?   *(1 mark)*

2 Measurements of weight taken with a particular set of scales are known to be normally distributed with mean equal to the true weight of the object being weighed and standard deviation 0.5 grammes. Eight independent measurements are taken of the weight of the same object, and the values obtained (in grammes) are

10.87, 14.01, 18.02, 15.11, 15.39, 17.79, 14.33, 15.08.

Given the R output

```
> qnorm(0.995)
[1] 2.575829
> qnorm(0.95)
[1] 1.644854
```

(i)   calculate a 99% confidence interval for the true weight of the object being weighed.   *(7 marks)*

(ii)   if 20 measurements were taken, what would the length of a 90% confidence interval for the true weight be?   *(3 marks)*

3 (i) Let Z be a random variable that follows the standard normal distribution $N(0, 1)$. If $\Phi(z)$ denotes the c.d.f. of Z at point z, show that

$$\Phi(0) = \frac{1}{2}.$$

(5 marks)

(ii) Let X be a random variable that follows the normal distribution with mean 1 and variance 4. Given the R output

```
> pnorm(1/2)
[1] 0.6914625
```

calculate the following probabilities

(a) $P(1 \leq X < 2)$; *(5 marks)*

(b) $P\left(|X| < \frac{1}{2}\right)$. *(5 marks)*

4 In a survey about how many hours per week do grandfathers spend sending and answering email, the results of 8 grandfathers age at least 75 were

$$0, \ 1, \ 2, \ 2, \ 7, \ 10, \ 14, \ 15.$$

Given the R output

```
> qt(0.995, 7)
[1] 3.499483
> qchisq(0.05, 7)
[1] 2.16735
> qchisq(0.95, 7)
[1] 14.06714
```

(i) find a 99% confidence interval for the mean hours grandfathers spend sending and answering email; *(5 marks)*

(ii) find a 90% confidence interval for the standard deviation of hours grandfathers spend sending and answering email. *(5 marks)*

5 When Vincenzo Baranello's blood pressure is in control, the systolic blood pressure reading has a mean of 130. For the last six times he has monitored his blood pressure, the recordings of which are shown in the table below

$$140, \ 150, \ 135, \ 155, \ 155, \ 140.$$

Given the R output

```
> pt(4.503, 5)
[1] 0.9968091
```

assess the strength of evidence this data provides on whether or not his blood pressure has changed. *(10 marks)*

6   5 patients were offered two treatments A and B and their effects were measured in a % scale of efficacy, as shown in the following table.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
Treatment A	70	65	67	58	45
Treatment B	83	56	69	70	62

Given the R output

```
> pt(1.489,4)
[1] 0.8946394
```

use an appropriate test to assess the evidence of difference between the two treatments.   *(15 marks)*

7   A financial analyst trading regularly shares in the stock market, is interested in monitoring the number of shares, or trading volume, which are bought and sold each day. In the absence of any events that trigger heavy trading, volume tends to be the highest early in the week and lightest late in the week. To assess this claim data received on the trading volume of General Electric shares on Mondays (variable X) and Fridays (variable Y). The data are summarised in the table below

Sample size	Mean	Standard deviation
$n_X = 11$	$\bar{x} = 51.82$	$s_X = 17.39$
$n_Y = 12$	$\bar{y} = 50.00$	$s_Y = 11.34$

Given the R output

```
> pt(0.294, 10)
[1] 0.6126151
```

assess the claim that the trading volume of Monday is the same as that of Friday.   *(15 marks)*

8   For murderers in the United States in 2009, the table below cross-tabulates the sex of the victim by the sex of the offender.

	Female Victim	Male victim
Female offender	182	484
Male offender	1719	4078

Given the R output

```
> pchisq(1.558, 1)
[1] 0.7880417
```

use a suitable test to test whether there is a difference between the gender of offender and the gender of victim.   *(15 marks)*

End of Question Paper