THE UNIVERSITY OF SYDNEY

MATH1905 STATISTICS ADVANCED

Semester 2 **Tutorial Week 13** 2012

1. Using the R output below

```
> q=c(.75,.9,.95,.99)
> round(qt(q,13),3)
[1] 0.694 1.350 1.771 2.650
> round(qt(q,17),3)
[1] 0.689 1.333 1.740 2.567
```

Comment on whether or not the following $P-$values are significant:

(a) $P = \mathrm{P}(t_{17} \geq 1.8)$ (b) $P = \mathrm{P}(t_{13} < -0.5)$ (c) $P = \mathrm{P}(|t_{17}| > 0.9)$ (d) $P = \mathrm{P}(|t_{13}| > 1.71)$

2. Which of these statistics (mean, standard deviation, correlation coefficient) are unaffected by a change of origin in the data?

(a) all of them, (b) mean and correlation coefficient,

(c) correlation coefficient only, (d) none of them,

(e) standard deviation and correlation coefficient.

3. Four independent measurements are made on precipitate weight from a particular chemical reaction. The weights (in gms) are 7, 13, 4, 6. The summary statistics are $\bar{x} = 7.5$ and $s = 3.873$. Assume that the precipitate weight can be modelled by a $\mathcal{N}(\mu, \sigma^2)$ random variable.
A 95% confidence interval for the mean weight, μ, is

(a) (3.7, 11.3) (b) (4.4, 10.6) (c) (2.9, 12.1)

(d) (4.3, 10.7) (e) (1.3, 13.7)

4. The taste test for PTC (phenylthiourea) is a typical exercise for human genetics classes. It has been established that a single gene determines the characteristic and that 70% of the Australian population are 'tasters' and 30% are 'non-tasters'. Let X represent the number of tasters in a human genetics tutorial class of 10 unrelated students. Which of the following statements is true?

(a) The expected value of X is 3. (b) The standard deviation of X is 2.1.

(c) $\mathrm{P}(X \leq 3) = 0.0106$. (d) $\mathrm{P}(X = 3) = 0.0106$.

(e) $\mathrm{P}(X < 3) = 0.0106$.

5. Use R to answer the following.

(a) If $\mathrm{P}(\chi^2_5 > a) = 0.10$, find a. (b) If $\mathrm{P}(\chi^2_{10} \geq a) = 0.05$, find a.

(c) $\mathrm{P}(\chi^2_{25} \geq 38.5) =$ .... (d) $\mathrm{P}(\chi^2_{12} \geq 22.1) =$ ....

6. One of Mendel's breeding trials had the following results:

Type of pea	Frequency
Smooth yellow	315
Wrinkled yellow	101
Smooth green	108
Wrinkled green	32
Total	556

What frequencies are expected under a model $9:3:3:1$?
Examine the model for goodness of fit.

7. In a backcross experiment to investigate the genetic linkage between two factors A and B in a species of flower, some researchers classified 400 offspring by phenotype as follows

AB	Ab	aB	ab
128	86	74	112

(a) Under the "no linkage" model, the four phenotypes are equally likely. Show that this model is a poor fit.

(b) If linkage is in the "coupling phase", the probabilities of the four phenotypes are

AB	Ab	aB	ab
$\frac{1}{2}(1-p)$	$\frac{1}{2}p$	$\frac{1}{2}p$	$\frac{1}{2}(1-p)$

where p is the "recombination fraction" and is estimated by the overall proportion of Ab and aB. Show that this "coupling phase" model fits the data well.

8. The following is the frequency table summarising the sulfur oxide emissions (in tons) from an industrial plant on 80 different days.

Tons of sulfur oxides	Frequency
5.0 - 12.9	13
13.0 - 16.9	14
17.0 - 20.9	25
21.0 - 24.9	17
25.0 - 32.9	11
Total	80

Using the complete data set we have that $\bar{x} = 18.85$ and $s = 5.55$.

(a) Use the above information to calculate the expected cell frequencies if the size of the oxides emissions follow a normal distribution.

(b) Test the goodness of fit of the normal model.

1. A study of grand juries in a county of California compared the demographic characteristics of jurors with the general population, to see if the jury panels were representative. The breakdown by age for the county is known from Public Health Department data, and the breakdown by age for a random sample of 66 jurors (over 21) is also given. Are the data consistent with the theory that these 66 jurors were selected at random from the population (over 21) of the county?

Age	County percentage	Number of jurors
21 to 40	42	5
41 to 50	23	9
51 to 60	16	19
61 and over	19	33

2. The number of dust particles in coal gas was counted by illuminating the field of an ultramicroscope for a fixed period. This was repeated 143 times. A model for X, the number of particles illuminated, is proposed. This model is $\mathrm{P}(X = i) = \dfrac{1}{i!}e^{-\theta}\theta^i$ for $i = 0, 1, \ldots$

Number of particles:	0	1	2	3	4	5	Total
Frequency:	33	46	38	19	4	3	143

(a) Is this model a good fit if $\theta = 2$?

(b) Is this model a good fit if θ is unknown but is estimated from the data using the sample mean?

3. (Review) An analyst, unaware that he is being tested, is asked to make a standard measurement on each of 16 chemical samples that weigh 1.50 gm. It is suspected that he tends to over-read. His readings are:

1.50	1.53	1.54	1.48	1.49	1.50	1.51	1.51
1.52	1.53	1.51	1.50	1.52	1.51	1.53	1.52

(a) Find the mean and standard deviation of the data.

(b) Construct a box-plot to check the normal assumption.

(c) Analyse the data for evidence of over-reading using the t-test.

4. (Review) The following results were obtained for the clotting time (in minutes) observed in 10 pairs of blood samples. One of each pair was chosen at random and treated with paraffin, the other with methacrylate.

Sample:	1	2	3	4	5	6	7	8	9	10
Paraffin:	10	27	11	18	19	16	16	18	22	26
Methacrylate:	13	20	9	12	11	14	19	12	11	18

Use a t-test to determine if the sample provides evidence that blood treated with methacrylate clots faster than blood treated with paraffin, on average.