The University of Sydney
School of Mathematics and Statistics

Practice Quiz Week 7

MATH1905: Statistics (Advanced) Semester 2, 2017

Web Page: http://sydney.edu.au/science/maths/MATH1905
Lecturer: Michael Stewart

Full Name .... SID ....

Day .... Time .... Room...........................

Tutor........................... Signature...

Time allowed: 40 minutes

1. **This quiz is closed book. You may not use a computer.**
2. Full marks will only be given if you obtain the correct answer **and** your working is sufficient to justify your answer.
3. Partial marks may be awarded for working.
4. Please write carefully and legibly.
5. All of your answers should be written using ink and **not** pencil, with your final answer placed in the answer box.
6. All working must be done on the quiz paper in the indicated space.
7. Each question is worth **2 marks**.
8. Only University of Sydney approved calculators may be used (must have a sticker).
9. All pages (including working) of the quiz paper must be handed in at the end of the quiz.

This quiz paper has 12 pages (this cover sheet + 10 pages of questions + 1 page of statistical formulae) and 10 questions.

1. A vector `x` in R yields the following output:

```
length(x)
```

```
[1] 50
```

```
sum(x)
```

```
[1] 249
```

```
sum(x^2)
```

```
[1] 1453
```

Determine (to 2 decimal places) the mean and sample standard deviation of `x`.

Mean of `x` is	Sample SD of `x` is

Please show your working below this line

---

2. The chief accountant of a large company collected the following information on advertising expenditure (in thousands of dollars) and revenue (in ten thousands of dollars) for 7 of its popular products as shown below:

Revenue y	6	5.4	7.4	4.7	4.9	4.6	7
Expenditure x	2.8	2.5	2.5	2.6	2.6	2.7	2.5

Determine the correlation coefficient (to 3 decimal places). You may find the R output below useful:

```
y=c(6,5.4,7.4,4.7,4.9,4.6,7)
x=c(2.8,2.5,2.5,2.6,2.6,2.7,2.5)
```

```
var(x)
```

```
[1] 0.01333333
```

```
var(y)
```

```
[1] 1.268095
```

```
sum((x-mean(x))*(y-mean(y)))
```

```
[1] -0.32
```

The correlation coefficient is

Please show your working below this line

3. Two events A and B are such that $P(A) = 10/31$, $P(B) = 12/31$ and $P(A \cup B) = 16/31$. Determine $P(A|B)$, that is the conditional probability of A given B.

$P(A|B) =$

Please show your working below this line

---

4. An urn contains 13 balls: 4 are red, 6 are blue and 3 are white. A random sample of size 3 is taken **with** replacement. Determine the probability (to 4 decimal places, or as an exact ratio) that all balls in the sample are the same colour.

Probability all are the same colour is:

Please show your working below this line

---

5. It is known that 6% of the children in a particular community suffer from a particular blood disorder. A test performed in a clinic correctly diagnoses 97% of children with this disorder as "positive" for the disorder, but also misdiagnoses 9% of children who do not have the disorder as "positive" for the disorder.

A child (randomly chosen from the community) is diagnosed "positive" by the clinic. Write down (to 3 decimal places) $P(D|+)$, that is the (conditional) probability that they actually have the disorder, given they have a positive test result.

$P(D|+) =$

Please show your working below this line

---

6. A random variable X only taking values 0,1,. . . ,5 has the following probability distribution ($P(X = 5)$ is obscured by the $*$):

x	0	1	2	3	4	5
$P(X = x)$	0.2	0.05	0.25	0.35	0.1	$*$

Given that the mean or the expected value of X, $E(X) = 2.25$, determine $P(X = 5)$ and $Var(X)$.

$P(X = 5) =$	$Var(X) =$

Please show your working below this line

7. A fair six-sided die is thrown twice independently. Let A be the even that the sum of the two numbers showing face-up is strictly less than 6. Determine $P(A)$.

$P(A) =$

Please show your working below this line

---

8. 10 tickets of equal size, feel, each have a number written on them. The numbers are stored in the R vector `tickets`, whose summary statistics are given below:

```r
tickets
```

```
[1]  1  2  4  4  5  7  7  9 10 15
```

```r
sum(tickets)
```

```
[1] 64
```

```r
var(tickets)
```

```
[1] 17.37778
```

A ticket is drawn at random (so that each is equally likely). Let X denote the (random) number showing on the selected ticket. Determine $Var(X)$ (to 2 decimal places).

$Var(X) \approx$

Please show your working below this line

---

9. Emails arrive in an inbox at a rate of 1.5 per minute and the number over any time period is well modelled as a Poisson random variable. If X is the number of emails arriving in the next three minutes, determine $P(X = 3)$ (to 3 decimal places).

$P(X = 3) \approx$

Please show your working below this line

---

10. A random variable X only taking non-negative integer values has probability generating function given by

$$\pi_X(s) = E\left(s^X\right) = (4 - 3s)^{-1}\,.$$

Deduce the k-th derivative $\pi_X^{(k)}(s) = \frac{d^k}{ds^k}\pi_X(s)$ and hence determine the probability distribution of X i.e. write $P(X = x)$ as a function of x.

$\pi_X^{(k)}(s) =$	$P(X = x) =$

Please show your working below this line

Formula sheet for MATH1905 Statistics

- **Calculation formulae**:
  - *For a sample* $x_1, x_2, \ldots, x_n$

Sample mean $\overline{x}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
Sample variance s^2	$\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} x_i\Big)^2\right] = \frac{1}{n-1}S_{xx}$

  - *For paired observations* $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

S_{xy}	$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\Big(\sum_{i=1}^{n} x_i\Big)\Big(\sum_{i=1}^{n} y_i\Big)$
S_{xx}	$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} x_i\Big)^2$
S_{yy}	$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} y_i\Big)^2$
r	$\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

For the least-squares line $y = a + bx$:	
b	$\frac{S_{xy}}{S_{xx}}$
a	$\overline{y} - b\overline{x}$

- **Some probability results:**

For any two events A and B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and $P(A \cap B) = P(A)P(B\mid A)$
If A and B are mutually exclusive	$P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$
If A and B are independent	$P(A \cap B) = P(A)P(B)$

- If $Y \sim \text{Pois}(\lambda)$, $P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}$ for $y = 0, 1, 2, \ldots$, $E(Y) = \lambda$ and $Var(Y) = \lambda$.
- If $X \sim B(n,p)$, $P(X = x) = \binom{n}{x}p^x(1-p)^{n-x}$, for $x = 0, 1, \ldots, n$, $E(X) = np$ and $Var(X) = np(1-p)$.

- **Some test statistics** and sampling distributions under appropriate assumptions and hypotheses:

$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$\frac{\overline{X}-\overline{Y}}{S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \sim t_{n_x+n_y-2}$, where $S_p^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}$
$\hat{\alpha} \sim N\left(\alpha, \sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right]\right)$; $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$; $\hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n-2} \sim \frac{\sigma^2\chi^2_{n-2}}{n-2}$
$\sum_i \frac{(O_i-E_i)^2}{E_i} = \sum_i \frac{O_i^2}{E_i} - n \sim \chi^2_\nu$, for appropriate ν