The University of Sydney
School of Mathematics and Statistics

Quiz 2a

MATH1905: Statistics Advanced — Semester 1, 2017

Full Name .... SID ....

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

Tutor .... Signature ....

Time allowed: 40 minutes

1. **This quiz is closed book. You may not use the 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. Let $X_1, X_2, \ldots, X_{39}$ be a random sample of geometric random variables with expectation 4 and variance 12. A normal approximation to $P\left(113 \leq \sum_{i=1}^{39} X_i < 188\right)$, with continuity correction, is of the form $P(a \leq Z \leq b)$ (where $Z \sim N(0,1)$) for some constants a and b. Write these down (to 2 decimal places) where indicated in the boxes below.

$a =$

$b =$

---

*Working for Question **1.***

The next two questions relate to the following scenario: A random sample of size 20 is taken from a population with variance known to be 2.6 and unknown mean μ. The following R output may be useful

```
> qnorm(c(0.8,0.9,0.95,0.975,0.98,0.99,0.995))

[1] 0.8416212 1.2815516 1.6448536 1.9599640 2.0537489 2.3263479 2.5758293
```

2. The p-value for a two-sided Z-test of $H_0\colon \mu = 46$ is given by 0.646. Given that $P(Z \leq 0.46) = 0.677$ (where $Z \sim N(0,1)$), determine to 3 decimal places the value taken by the sample mean *given that it is greater than 46.*

Answer is

Working for Question ***2.***

3. A 90% confidence interval for μ is of the form $\bar{x} \pm a$ where $\bar{x}$ is your answer to the previous question. Write down the value of a (to 3 decimal places) in the box below.

Ans

---

*Working for Question **3.***

The following two questions relate to the following scenario: A random sample of size 9 is taken from a population, assumed to be $N(\mu, \sigma^2)$ for μ and σ^2 both unknown. The sample mean takes the value 13.37 while the sample standard deviation takes the value 5.686. The following R output may be useful:

```
> n
[1] 9
> qt(c(0.9,0.95,0.975,0.99,0.995),df=n-1)
[1] 1.396815 1.859548 2.306004 2.896459 3.355387
> qt(c(0.9,0.95,0.975,0.99,0.995),df=n)
[1] 1.383029 1.833113 2.262157 2.821438 3.249836
```

4. Suppose it is desired to test the null hypothesis $H_0\colon \mu = 15$. Compute the value taken by the appropriate t-statistic to 3 decimal places.

Ans

---

*Working for Question **4.***

5. Compute a 95% upper confidence limit for μ (to 2 decimal places).

Ans

Working for Question ***5.***

The next three questions relate to the following scenario: Two random samples are drawn independently from two normal populations with unknown means and variances. The first sample (sample "A") is of size 5, has mean 7.811 and standard deviation 1.95. The second sample (sample "B") is of size 11, has mean 19.591 and standard deviation 7.81.

6. If it is assumed that the two populations have the same variance σ^2, compute and write down the pooled estimate of σ^2 (to 3 decimal places).

Ans

---

*Working for Question **6.***

7. Still assuming equal population variances, the appropriate t-statistic for testing the hypothesis that the population means are equal is of the form b/s_p where s_p is the square root of your answer to the previous question. Write down the value of b to 3 decimal places in the box below.

Ans

Working for Question ***7.***

8. Suppose now that it is *not* assumed that the population variances are equal (so a Welch test is to be performed). Write down to 3 decimal places the standard error of the mean difference.

Ans

Working for Question ***8.***

The final two questions relate to the following scenario. Bivariate data $(x_1, y_1), \ldots (x_{14}, y_{14})$ is collected and each y_i is modelled as the value taken by a normal random variable $Y_i \sim N(\alpha+\beta x_i, \sigma^2)$ for unknown parameters α, β and $\sigma > 0$. All random variables are assumed independent.

After computing the least-squares regression line the residual sum of squares is 1.869 and the sample standard deviation of the x_i's is 0.434. The scatterplot appears roughly linear and there is no obvious pattern in the residual plot.

The following R output may be useful:

```
> n

[1] 14

> qt(c(0.9,0.95,0.98,0.975,0.99,0.995),df=n)

[1] 1.345030 1.761310 2.263781 2.144787 2.624494 2.976843

> qt(c(0.9,0.95,0.98,0.975,0.99,0.995),df=n-1)

[1] 1.350171 1.770933 2.281604 2.160369 2.650309 3.012276

> qt(c(0.9,0.95,0.98,0.975,0.99,0.995),df=n-2)

[1] 1.356217 1.782288 2.302722 2.178813 2.680998 3.054540
```

9. Write down to 3 decimal places the value of $\hat{\sigma}^2$, the estimate of the error variance.

Ans

Working for Question ***9.***

10. A 99% lower confidence limit for β is of the form $\hat{\beta} - c\hat{\sigma}$ where $\hat{\sigma}$ is the square root of your answer to the previous question. Write the value of c to 3 decimal places in the box below.

Ans

*Working for Question **10.***

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 regression 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 $P(A \cap B) = P(A)P(B\|A)$
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 $Y \sim \text{Pois}(\lambda)$, then $P(Y = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. Furthermore, $\mathbb{E}(Y) = \lambda$ and $\text{var}(Y) = \lambda$.
- If $X \sim B(n, p)$, then, $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, \ldots, n$. Furthermore, $E(X) = np$ and $\text{var}(X) = np(1-p)$.
- **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 $S_p^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}$
$\hat{\alpha} \sim N\left(\alpha, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]\right)$; $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$; $\hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n-2} \sim \frac{\sigma^2\chi^2_{n-2}}{n-2}$
$\sum_i \frac{(O_i-E_i)^2}{E_i} = \sum_i \frac{O_i^2}{E_i} - n \sim \chi^2_\nu$, for appropriate ν