國立清華大學102學年度碩士班考試入學試題

系所班組別：服務科學研究所

考試科目（代碼）：統計學（4602）

For each question, please only pick the one answer that seems most appropriate.

A computer scientist is developing a new algorithm that processes fixed-size images of x-ray scans. Based on an analysis of the algorithm, she estimates that its processing time should generally follow a normal distribution with a mean value of 3500 milliseconds.

1. [9 pts] Our computer scientist now tests her algorithm on a sample of 400 scanned x-ray images. The results of the sample suggest that the general mean processing time is 3505 with a variance of 100. Should she reject her earlier belief that the mean processing time is 3500 milliseconds?
(a) Can reject at 0.1% significance
(b) Can only reject at 1% significance
(c) Can only reject at 5% significance
(d) Cannot reject earlier belief
(e) Not enough information to answer this question

2. [9 pts] What are the units of the variance?
(a) √milliseconds
(b) milliseconds
(c) milliseconds^2
(d) milliseconds^3
(e) none of the above

In a separate test, the memory usage of our computer scientist's image processing algorithm is tested against the memory usage of other algorithms, and against memory usage when different types of image content are considered. The results of an ANOVA test are shown below. The factor `Algorithm` refers to the choice of algorithm used, and the factor `Content` refers to different types of image content.

	DF	Sum-Sq	Mean-Sq	F-value	Pr(>F)
Algorithm	2	45300	22650	10.270	0.00476
Content	2	6100	3050	1.383	0.29944
Algorithm*Content	4	11200	2800	1.270	0.35033
Residuals	9	19850	2206		

3. [9 pts] How many different types of image processing algorithms were tested in this ANOVA?
(a) 2
(b) 3
(c) 4
(d) 5
(e) Not enough information to answer this question

[9 pts] 4. From the ANOVA results, which factors would you say were significantly related to memory usage at 5% significance or less?
(a) The choice of algorithm
(b) The choice of image content
(c) The interaction of algorithm and content
(d) More than one of the above
(e) None of the above

[9 pts] 5. From the ANOVA results, how much of the total variation is due to differences in the algorithm (to the nearest percent)?
(a) 22%
(b) 33%
(c) 55%
(d) 74%
(e) Not enough information to answer this question

A sports coach is monitoring the performance (recorded as points) of an athlete every day at practice, and has also kept record of how much sleep (recorded as hours) he got the night before practice, and how much caffeine (recorded in milliliters) he consumed before practice each day. Each row in the coach's dataset represents one day's entry of performance, sleep, and caffeine intake.

The coach has conducted a regression where the dependent variable was performance, and the independent variables sleep and caffeine. An interaction between sleep and caffeine was also included. All variables were fully standardized before running the regression.

```
                  Estimate Std. Error t-value Pr(>|t|)
(Intercept)        0.01109    0.09832   0.113   0.9105
sleep              0.43004    0.09986   4.307 4.89e-05
caffeine          -0.01669    0.10433  -0.160   0.8734
sleep*caffeine    -0.18082    0.08386  -2.156   0.0342

Residual standard error: 0.8782 on 76 degrees of freedom
Multiple R-squared: 0.258,  Adjusted R-squared: 0.2287
```

[9 pts] 6. We are only interested in factors that are significant at 5% significance or less. Which independent terms should we consider to be significant?
(a) Only sleep
(b) Only caffeine
(c) Only Sleep*Caffeine
(d) Sleep and also Sleep*Caffeine
(e) Caffeine and also Sleep*Caffeine

[9 pts] 7. What are the units of the estimate of the sleep coefficient reported in the results above?
(a) no units
(b) points/hour
(c) hours/point
(d) hours^2
(e) None of the above

[9 pts] 8. How would you interpret the regression coefficient of the interaction term?
(a) Interaction is not significant – we can ignore it.
(b) The relationship between sleep and performance decreases as caffeine intake increases
(c) The relationship between caffeine and performance decreases as sleep intake increases
(d) Both [b] and [c] are valid
(e) The interaction is significant but neither [b] nor [c] are valid

[9 pts] 9. How many days worth of data would you say the coach has recorded?
(a) 50 days
(b) 60 days
(c) 70 days
(d) 80 days
(e) Not enough information to answer this question

[9 pts] 10. If the athlete in question kept his caffeine intake to his average, but increased his amount of sleep by one standard deviation, by how would you predict his performance would change?
(a) Performance would increase by 0.23 standard deviations
(b) Performance would increase by 0.25 standard deviations
(c) Performance would increase by 0.43 standard deviations
(d) Performance would increase by 1 standard deviation
(e) Not enough information to answer this question

[10 pts] 11. From our regression, how much of the variance of performance should we report is predicted by sleep and caffeine?
(a) 22.87%
(b) 25.8%
(c) 53.66%
(d) 87.82%
(e) Not enough information to answer this question