# *SAMPLE STA255 Final Exam Questions (WINTER 2018)*

**PART I (*WRITTEN QUESTION*)**

Random variable $X$ follows a distribution defined by

$$f(x) = \begin{cases} (\theta + 1)x^{\theta}, & 0 \le x \le 1 \\ 0, & otherwise \end{cases},$$

where $\theta > -1$. Use this information to answer parts **a-d** of this question. Be sure to clearly show all your steps.

**a.** Derive expressions for $E(X)$ and $V(X)$ in terms of parameter $\theta$.

**b.** Suppose we have an SRS of size *n* from a population where X follows this probability distribution. Find the method of moments estimator of $\theta$ as a function of the sample observations $x_1, x_2, x_3, \dots, x_n$.

**c.** Suppose we have an SRS of size *n* from a population where X follows this probability distribution. Find the maximum likelihood estimator of $\theta$ as a function of the sample observations $x_1, x_2, x_3, \dots, x_n$.

**d.** Find expressions for the maximum likelihood estimators of $\mu_X$ and $\sigma_X^2$ as a function of the sample observations $x_1, x_2, x_3, \dots, x_n$.

**PART II (*MULTIPLE CHOICE*)**

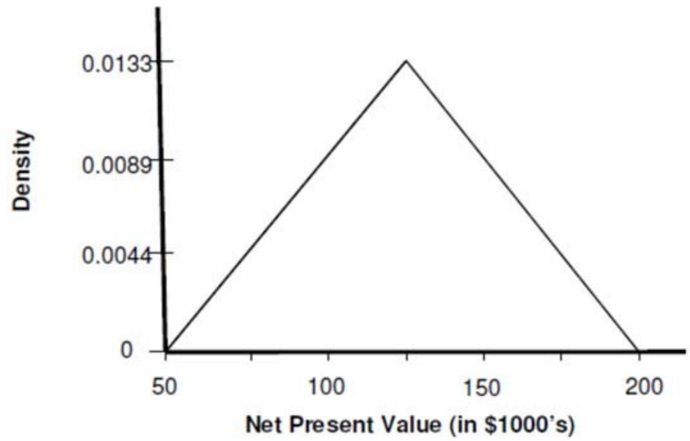
**Question 1**

It has been reported that the average student graduating with a bachelor's degree left school with a debt of \$19,500. The student council at a large university feels that the school's recent graduates have a debt much less than this, so it conducts a study on a large SRS of recent graduates, and finds the mean debt to be \$17,900.

What would the alternative hypothesis be for the appropriate hypothesis test?

A. $\mu < 19{,}500$
B. $\bar{x} < 17{,}900$
C. $\mu < 17{,}900$
D. $\mu = 19{,}500$
E. $\mu \neq 19{,}500$

**<u>Question 2</u>**
The net present values of a large company's assets can be described by a triangular density curve shown in the following figure:




What is the probability that a randomly chosen asset from this company's holdings will have a net present value of more than **$100,000**?

A. 0.22
B. 0.33
C. 0.67
D. 0.78
E. Not enough information to determine.

**<u>Question 3</u>**
A car dealership sells 0, 1, or 2 luxury cars per day. During the purchase, the dealer tries to encourage the new owner to buy the extended warranty. Let $Y_1$ and $Y_2$ be the number of cars and extended warrantees sold on a randomly selected day, respectively.

$$p(y_1, y_2) = \begin{cases} 1/6 \ for \ y_1 = y_2 = 0 \\ 1/12 \ for \ y_1 = 1, \ y_2 = 0 \\ 1/6 \ for \ y_1 = y_2 = 1 \\ 1/12 \ for \ y_1 = 2, \ y_2 = 0 \\ 1/3 \ for \ y_1 = 2, \ y_2 = 1 \\ 1/6 \ for \ y_1 = y_2 = 2 \end{cases}$$

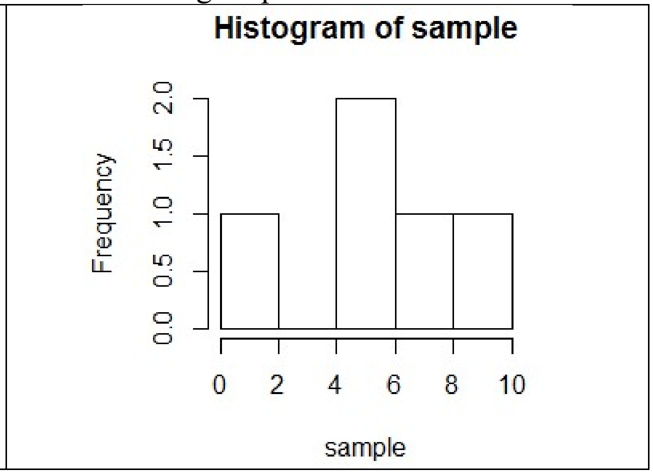
What is the variance of the number of luxury cars sold per day?

A. 0.47
B. 0.58
C. 0.83
D. 1.42
E. None of the above.

***Questions 4-6 are based on the following information:***

A simulation that was conducted in R yielded the following output:

```
> sample<-runif(5,1,10)
> mean(sample)
[1] 5.615777
> sd(sample)
[1] 2.924695
> mean(sample<4)
[1] 0.2
```




**<u>Question 4 (refer to information at top of this page)</u>**
What quantity does the number **5.615777** circled in the R output above estimate? Select the best answer.

A. $E(X)$
B. $P(X \leq 4)$
C. $P(\bar{X} < 4)$
D. $E(X|X < 4)$
E. $E(\bar{X})$

**<u>Question 5 (refer to information at top of this page)</u>**
This R code simulates taking a simple random sample from which distribution?

A. Uniform distribution over the interval (1,10)
B. Normal distribution with $\mu$=1 and $\sigma$=10
C. Discrete Uniform distribution over values 1, 2, 3…,10
D. Gamma distribution with $\alpha = 1$ and $\beta = 0.1$
E. None of the above.

**<u>Question 6 (refer to information at top of this page)</u>**
Which distribution does the histogram represent? Select the best answer.

A. The distribution of the variable in a population
B. The distribution of a variable in a sample
C. The distribution of a statistic across samples
D. The normal distribution
E. None of the above.

***Questions 7-8 are based on the following information:***

When answering a question on a multiple choice exam, Jon either thinks he knows the answer or just guesses. Suppose the probability that Jon thinks he knows the answer is 0.75 and there's a 90% chance that he'll get the question correct if he thinks he knows the answer. There are 5 choices for each multiple-choice exam question.

**Question 7 (refer to information at top of this page)**

Suppose Jon does not leave any answers blank so he guesses the answer (at random) if he does not think he knows the answer. What is the probability that Jon knew the answer to a question he answered correctly?

A. 0.750
B. 0.800
C. 0.675
D. 0.900
E. 0.931

**Question 8 (refer to information at top of this page)**

Each question is worth 2 marks (no part marks), but blank answers earn 0.4 marks. Suppose Jon always answers the question if he thinks he knows the answer, but if he does not, he guesses the answer (at random) half the time and the other half of the time, he leaves the answer blank.

What is the expected value of Jon's score on a randomly selected question?

A. 1.35
B. 1.45
C. 1.50
D. 1.52
E. 1.80

**Question 9**

An SRS of size n is drawn from a distribution with probability density function:

$$f(x) = \begin{cases} \frac{1}{2}(1+\theta x), & -1 \le x \le 1 \\ 0, & otherwise \end{cases}$$

What is the method of moments estimator of $\theta$?

A. $\bar{X}$
B. $2\bar{X}$
C. $1/\bar{X}$
D. $3\bar{X}$
E. $\bar{X}/2$

**Question 10**
The joint probability distribution function for random variables X and Y is

$$f(x,y) = \begin{cases} 2e^{-x-y}, & if\ x \geq 0,\ y \geq 0,\ y \geq x \\ 0, & otherwise \end{cases}$$

How can we determine E(X) from this?

A. $\int_0^\infty \int_0^y 2e^{-x-y}dxdy$
B. $\int_0^\infty \int_0^\infty 2xe^{-x-y}dxdy$
C. $\int_0^\infty 2xe^{-x-y}dx$
D. $\int_0^\infty 2e^{-x-y}dy$
E. $\int_0^\infty \int_x^\infty 2xe^{-x-y}dydx$

**Question 11**
A physician uses a diagnostic test to decide whether or not Mary (his patient) has rheumatoid arthritis. The doctor will prescribe treatment only if he really thinks Mary has rheumatoid arthritis. In a sense, the doctor is using a null and an alternative hypothesis to decide whether or not to administer treatment.

Which of the following would be a consequence of type II error in this situation?

A. Mary's rheumatoid arthritis will be left untreated.
B. Mary's rheumatoid arthritis will be treated.
C. Mary will be treated for rheumatoid arthritis unnecessarily.
D. Mary will not be treated for a condition that she does not have.
E. None of the above.

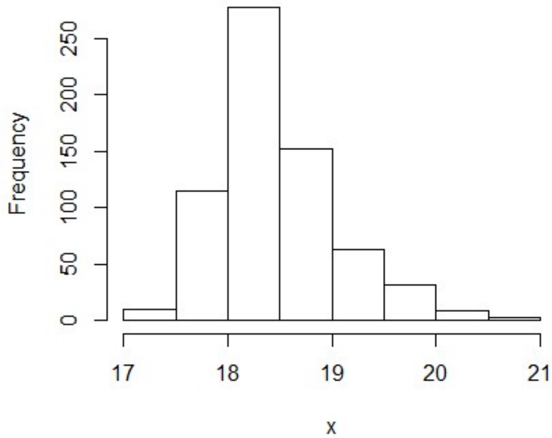
**Question 12**
The volume of fruit juice that a dispensing machine pours into a 200 mL box follows a normal distribution with a standard deviation of 10 mL. If more than 220 mL of juice is poured into a box, the juice spills over and the juice box needs to go through a special cleaning process before it can be sold.

What is the mean volume of juice that the machine should dispense in order to limit the proportion of juice boxes that need to be cleaned due to spillage to no more than 0.01?

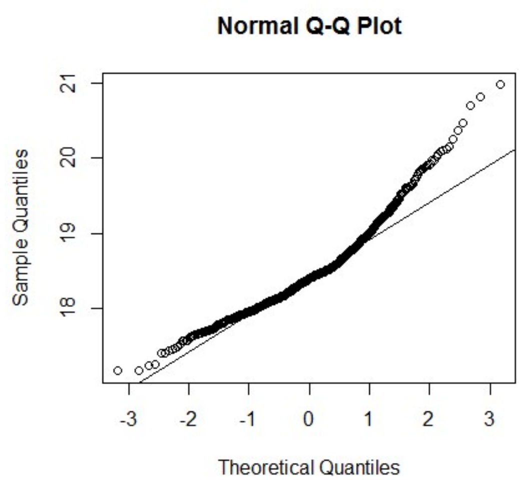
A. 196.7 mL
B. 200.0 mL
C. 220.0 mL
D. 243.3 mL
E. Not enough information to determine.

## Question 13

A population consists of N=661 individuals. Select R output summarizing the distribution of variable X for these 661 individuals is included below. What is the mean of the sampling distribution of sample means based on SRS's of size n=9 from this population?




```
> length(x)
[1] 661
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  17.18   18.09   18.40   18.48   18.75   20.97
> sd(x)
[1] 0.573725
> sqrt(var(x)*(length(x)-1)/length(x))
[1] 0.5732909
```




A. 0.19
B. 0.57
C. 18.40
D. 18.48
E. Not enough information to determine.

## Question 14

Suppose X and Y are continuous random variables with joint pdf:

$$f(x,y) = \begin{cases} 24xy, & 0 < x < 1 \text{ and } 0 < y < 1 - x \\ 0, & \text{otherwise} \end{cases}$$

What is P(Y < X|X = 1/3)?

A. 0.04
B. 0.07
C. 0.25
D. 0.33
E. 0.44

**Question 15**

The gestation period of humans is approximately normally distributed with a standard deviation of 16 days. A random sample of 9 pregnancies yields a sample mean of 266 days. A 90% confidence interval for $\mu$, the population mean gestation period, is (257.2 days, 274.8 days).

Which of the following interpretations of this interval is correct?

1. 90% of all gestation periods fall between 257.2 days and 274.8 days.
2. $\mu$ is between 257.2 days and 274.8 days 90% of the time.
3. There's a 90% chance that $\bar{x}$ is between 257.2 days and 274.8 days.
4. We are 90% confident that the interval (257.2 days, 274.8 days) contains $\mu$.

A. 1 and 3
B. 2 and 4
C. 1, 2, and 3
D. 4 only
E. 1, 2, 3, and 4

**Question 16**

An insurance company is trying to estimate the mean number of sick days that food service workers use per year. A pilot study conducted on an SRS of 25 food service workers estimated the mean number of sick days, with 90% confidence, to be 12.88–14.52 days. Since the insurance company would like a more precise estimate of the mean in order to improve their policies, they are prepared to conduct a second study.

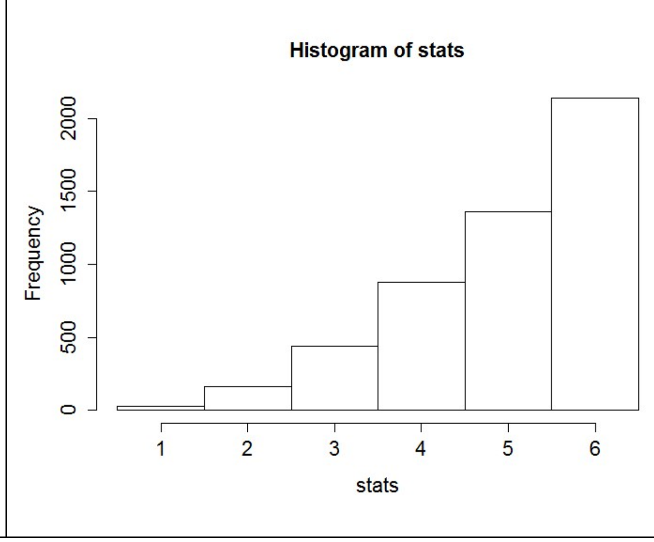
Which of the following changes would result in a more precise estimate of the mean number of sick days taken per year by food service workers?

Note: For each option, assume everything else (except for what is described in that option) remains the same as the original pilot study.

A. Conducting the second study, and estimating the population mean with a lower value for $\alpha$.
B. Conducting the second study, using an SRS of only full-time food service workers between the ages of 30-40 years old.
C. Conducting the second study, and estimating the population mean with a higher confidence level.
D. Conducting the second study, using an SRS of 20 food service workers.
E. None of the above.

***Questions 17-18 are based on the following information:***
A simulation that was conducted in R yielded the following output:

```
> x<-c(1,2,3,4,5,6)
> sample<-sample(x,1000,replace=TRUE)
> mean(sample)
[1] 3.433
> sd(sample)
[1] 1.688301
> mean(sample>4)
[1] 0.316
>
> stats<-rep(0,5000)
> for (i in 1:5000) {
+   sample<-sample(x,3,replace=TRUE)
+   stats[i]<-max(sample)
+ }
> mean(stats)
[1] 4.957
> sd(stats)
[1] 1.15543
> hist(stats)
```




**<u>Question 17 (refer to information at top of this page)</u>**
What quantity does the number **0.316** circled in the R output above estimate?

A. $E(X)$
B. $P(X \geq 5)$
C. $P(\bar{X} > 4)$
D. $E(\bar{X})$
E. None of the above.

**<u>Question 18 (refer to information at top of this page)</u>**
What does the histogram estimate?

A. Distribution of a variable in a population
B. Distribution of a variable in one sample
C. Distribution of a statistic across all samples
D. Distribution of a statistic in one sample
E. None of the above.

**Question 19**
The herb, cilantro, is very polarizing: Some people love it, and some people hate it. A genetic component is suspected to be at play. A survey based on a simple random sample of 120 American adults of European ancestry asked whether they like or dislike the taste of cilantro. A total of 31 stated that they dislike the taste.

What is the upper limit (rounded to 3 decimal points) of the large sample approximate 98% confidence interval for the proportion of American Adults of European ancestry who like the taste of cilantro?

A. 0.351
B. 0.649
C. 0.824
D. 0.835
E. Not enough information to determine.

**Question 20**
An SRS of size *n* is drawn from a distribution with probability density function:

$$f(x) = \begin{cases} \frac{1}{2}(1+\theta x), & -1 \leq x \leq 1 \\ 0. otherwise \end{cases}$$

What is the Mean Squared Error (MSE) of $\hat{\theta} = 3\bar{X}$?

A. $\frac{1}{3}\left(1 - \frac{\theta^2}{3}\right)$
B. $\frac{1}{n}(3 - \theta^2)$
C. $\frac{\theta}{3}$
D. $\frac{1}{n}\left(1 - \frac{\theta^2}{3}\right)$
E. None of the above.

**Question 21**
A publication entitled *Statistical Report on the Health of Canadians* indicated that 24% Canadians aged 12 years and over visited a physician one or more times in the previous year. If an SRS of 10 Canadians aged 12+ years are selected, find the probability that at least three of these people visited a physician at least once in the previous year.

A. 0.2012
B. 0.2429
C. 0.4442
D. 0.4721
E. 0.5558

**Question 22**
Suppose tornadoes hit major metropolitan areas in North America at a constant rate of two per year. What is the probability that one tornado will hit a major North American metropolitan area next month?

A.
```
> dbinom(1,1,1/6)
[1] 0.1666667
```
B.
```
> dpois(1,1/6)
[1] 0.1410803
```
C.
```
> dpois(1,2)
[1] 0.2706706
```
D.
```
> ppois(1,2)
[1] 0.4060058
```
E.
```
> ppois(1,1/6)
[1] 0.987562
```

**Question 23**
The times until the first car insurance claims are submitted by ‘good’ and ‘bad’ drivers are independent and exponentially distributed with mean 6 years for the good drivers and 3 years for the bad drivers.

What is the probability that the first claim from a good driver is submitted within 3 years and the first claim from a bad driver is submitted within 2 years?

A. $1 - e^{-2/3} - e^{-1/2} + e^{-7/6}$
B. $\frac{1}{18}e^{-7/6}$
C. $1 - e^{-2/3} - e^{-1/2} + e^{-1/3}$
D. $1 - \frac{1}{3}e^{-2/3} - \frac{1}{6}e^{-1/2} + \frac{1}{18}e^{-7/6}$
E. $\frac{1}{18}\left(1 - e^{-2/3} - e^{-1/2} + e^{-7/6}\right)$

**Question 24**
Suppose the cumulative distribution function (i.e., cdf) of X is given by:

$$F(x) = \begin{cases} 0, & x < 5 \\ 0.8, & 5 \le x < 10 \\ 1, & x \ge 10 \end{cases}$$

What is P(X = 9)?

A. 0.0
B. 0.2
C. 0.8
D. 1.0
E. None of the above.

**Question 25**
Let Y be a continuous random variable with probability density function:

$$f(y) = \begin{cases} \alpha y^{\alpha-1}, & 0 < y < 1 \; and \; \alpha > 0 \\ 0, & otherwise \end{cases}$$

What is the maximum likelihood estimator of $\alpha$ based on an SRS of size *n*?

A. $\frac{\sum_{i=1}^{n} Y_i}{n}$
B. $\frac{-n}{\sum_{i=1}^{n} log(Y_i)}$
C. $\frac{-\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} Y_i - n}$
D. $\frac{n}{\sum_{i=1}^{n} Y_i}$
E. $\frac{\sum_{i=1}^{n} log(Y_i)}{n}$

**Question 26**
Let Y be a continuous random variable with probability density function:

$$f(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y < 1 \\ 0, & otherwise \end{cases}$$

What is the $80^{th}$ percentile of this distribution?
A. 0.20
B. 0.36
C. 0.39
D. 0.64
E. 0.80

**Question 27**
Suppose 80% of all drivers always wear a seatbelt. A simple random sample of 100 drivers is observed while they are driving and their seatbelt use is recorded. What is the probability that more than 75 of the selected drivers will be wearing a seatbelt?
A. 0.87
B. 0.89
C. 0.92
D. 0.94
E. Not enough information to determine.

**Question 28**

X has follows a distribution given by the pdf $f(x) = \frac{1}{2}\left[e^{-x} + \frac{1}{\sigma}e^{-\frac{x}{\sigma}}\right]$, for x > 0 and $\sigma > 0$.

What is the moment generating function for this distribution?

A. $\frac{1}{2}\left[e^{-tx} + \frac{1}{\sigma}e^{-\frac{tx}{\sigma}}\right]$
B. $\frac{1}{2}\left[e^{(t-1)x} + \frac{1}{\sigma}e^{(t-\frac{1}{\sigma})x}\right]$
C. $\sigma/2$
D. $\frac{1}{2}\left[\frac{1}{1-t} + \frac{1}{1-\sigma t}\right]$
E. $\frac{1}{2}[t + \sigma t]$

**Question 29**

Suppose we plan to estimate $p$ with $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$, where $X_i = 1$ if the i^th sampled individual is a success, and *0* otherwise, based on a simple random sample of size *n*. Which of the following methods produce an estimate of the standard deviation of the estimator $\hat{p}$?

1. Take many SRSs of size n with replacement from the sample and compute sample standard deviation of resulting $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$ values.
2. Use the observed value of $\hat{p}$ based on the sample obtained to compute $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
3. If $p$ is known, generate many Y~BIN(n,p) values and compute sample standard deviation of resulting $\hat{p} = \frac{Y}{n}$ values.
4. Compute the sample standard deviation of the sample values, $x_1, x_2, x_3, \dots, x_n$.

A. 1, 2, and 3
B. 1 and 3
C. 2 and 4
D. 4 only
E. 1, 2, 3, and 4

**Question 30**

Suppose investment losses follow a Pareto distribution with pdf

$$f(x) = \frac{\alpha}{x^{\alpha+1}}, x > 1 \text{ and } \alpha > 1$$

A simple random sample of investment losses yields values 3, 15, 6, 10, and 25 units.

What is the method of moments estimate of $\alpha$ based on this sample?

A. 0.085
B. 0.450
C. 0.922
D. 1.035
E. 1.093

***Questions 31-32 are based on the following information:***

Suppose Y is a random variable with pdf $f(y) = \begin{cases} 2/y^2 & for\ y \geq 2 \\ 0 & otherwise \end{cases}$.

**Question 31 (refer to information at top of this page)**

Suppose this is a reasonable probability model for the lifetimes of a specific smartphone battery, with *y* measured in years since manufacturing. A used smartphone with this type of battery is randomly selected and its battery is determined to be still in working order at **3** years after manufacturing (i.e., we say its lifetime was right censored at 3 years).

What is the probability that this battery will last at least **4** years given that it lasted for at least **3** years?

A. 0.17
B. 0.50
C. 0.67
D. 0.75
E. Not enough information to determine.

**Question 32 (refer to information at top of this page)**

What is the pdf of $X = \frac{1}{Y-1}$ over the interval $0 < x \leq 1$?

A. $2/x^2$
B. $2/(x+1)^2$
C. $2/(x-1)^2$
D. $2x^2/(x+1)^2$
E. $2(x+1)^2/x^4$

**Question 33**

Independent simple random samples of $n_m$ male smokers and $n_f$ female smokers are selected and the number who smoke e-cigarettes (i.e., $x_m$ and $x_f$) are recorded. Let $p_m$ and $p_f$ denote the true probabilities that a randomly selected male and female, respectively, smoke e-cigarettes. Interest lies in the parameter $p_m - p_f$, the difference in population proportions of males and females who smoke e-cigarettes. What is $MSE\left(\frac{X_m}{n_m} - \frac{X_f}{n_f}\right)$?

A. 0
B. $p_m - p_f$
C. $p_m(1-p_m) + p_f(1-p_f)$
D. $\frac{p_m(1-p_m)}{n_m} + \frac{p_f(1-p_f)}{n_f}$
E. None of the above.

# ANSWER KEY

**PART I**

**a.** $E(X) = \frac{\theta+1}{\theta+2}$ and $V(X) = \frac{\theta+1}{(\theta+3)(\theta+2)^2}$

**b.** $\hat{\theta}_{mom} = \frac{2\bar{X}-1}{1-\bar{X}}$

**c.** $\hat{\theta}_{mle} = \frac{-n-\sum_{i=1}^{n}\log(x_i)}{\sum_{i=1}^{n}\log(x_i)}$

**d.** $\hat{\mu}_{mle} = \frac{\hat{\theta}_{mle}+1}{\hat{\theta}_{mle}+2}$ and $\hat{\sigma}^2_{mle} = \frac{\hat{\theta}_{mle}+1}{(\hat{\theta}_{mle}+3)(\hat{\theta}_{mle}+2)^2}$

| | |
|---|---|
| **1** | A |
| **2** | D |
| **3** | B |
| **4** | A |
| **5** | A |
| **6** | B |
| **7** | E |
| **8** | B |
| **9** | D |
| **10** | E |
| **11** | A |
| **12** | A |
| **13** | D |
| **14** | C |
| **15** | D |
| **16** | B |
| **17** | B |
| **18** | C |
| **19** | A |
| **20** | B |
| **21** | C |
| **22** | B |
| **23** | A |
| **24** | A |
| **25** | B |
| **26** | D |
| **27** | A |
| **28** | D |
| **29** | A |
| **30** | E |
| **31** | D |
| **32** | B |
| **33** | D |