1.A recent study published by the New England Journal of Medicine assigned hospitalized adult patients with severe Covid-19 pneumonia to receive convalescent plasma or placebo. A total of 228 patients were assigned to receive convalescent plasma and 105 to receive placebo. After 30 days of treatment, mortality was 10.96% in the convalescent plasma group and 11.43% in the placebo group. If you want to test the difference in mortality rates of these two medical treatments, what is the P-value? (Round z-statistic up to the nearest z value in the normal table)

|  |  |  |
| --- | --- | --- |
|  | A. | 44.04% |
|  | B. | 36.32% |
|  | C. | 42.08% |
|  | D. | 38.21% |
|  | E. | 40.13% |

2.R users with a keen sense of aesthetics are using the ggplot() command to create scatter diagrams instead of the plot() command.

Watch this YouTube video: https://www.youtube.com/watch?v=Z\_xn5R0ifxM&t=252s

If you are predicting mpg from hp using the **mtcars**data, which commands below can successfully create a scatter diagram?

|  |  |  |
| --- | --- | --- |
|  | A. | ggplot(data = mtcars, aes(x = hp, y = mpg))  geom\_point(col = "blue") |
|  | B. | ggplot(data = mtcars, aes(x = mpg, y = hp))  geom\_point(col = "blue") |
|  | C. | scatter <- ggplot(data = mtcars, aes(x = mpg, y = hp))  geom\_point(col = "red") |
|  | D. | scatter <- ggplot(data = mtcars, aes(x = hp, y = mpg))  scatter + geom\_point(col = "red") |
|  | E. | scatter <- ggplot(data = mtcars, aes(x = mpg, y = hp))  scatter + geom\_point(col = "red") |
|  |  |  |

3.Following Question 2, which commands below can add a regression line to the scatter diagram?

|  |  |  |
| --- | --- | --- |
|  | A. | scatter + stat\_smooth() |
|  | B. | stat\_smooth(method = "lm", se = FALSE) |
|  | C. | scatter + stat\_smooth(method = "lm", se = FALSE) |
|  | D. | stat\_smooth(se = FALSE) |
|  | E. | stat\_smooth() |

4. A university has 30,000 registered students. As part of a survey, 900 of these students are chosen at random. The average age of the sample students turns out to be 22.3 years, and the SD is 4.5 years. Find a 80% confidence interval for the average age of all 30,000 registered students.

|  |  |  |
| --- | --- | --- |
|  | A. | (12.10, 12.50) |
|  | B. | (22.50, 22.10) |
|  | C. | (22.10, 22.50) |
|  | D. | (12.10, 32.50) |
|  | E. | (21.10, 23.50) |

5. A larger sample offers no protection under bias.

 True

 False

6. Simple random sampling means drawing subjects at random with replacement.

 True

 False

7. A statistic computed from the sample is always equal to the parameter it estimates.

 True

 False

8. When drawing from a 0-1 box, the expected value for the percentage of 1’s in the sample equals the percentage of 1’s in the box.

 True

 False

9. The number of individuals in the population has a significant effect on the accuracy of the sample percentage.

 True

 False

10. Flower color in roses is controlled by one gene pair. There are two variants of such gene, r (for red) and w (for white). The rules are: rr, rw, wr make red; ww makes white. Of 1600 such seeds, what is the chance that 1230 or more of them will grow red roses?

|  |  |  |
| --- | --- | --- |
|  | A. | 2.35% |
|  | B. | 4.01% |
|  | C. | 5.45% |
|  | D. | 3.95% |
|  | E. | 6.30% |

11. In a study on traffic violations, motorists at a certain intersection were recorded. The number of times a red light was run and the number of accidents at the intersection were recorded each week. The red light violations and accidents were approximately normally distrbuted, and the resulting scatter diagram is football-shaped. The following results were obtained:

Red Light Violations Average = 10     SD = 2

                 Accidents Average = 14     SD = 3      r=0.75

Suppose the percentile rank of number of accidents in a given week is 95%. Predict how many red light violations occurred that week.

|  |  |  |
| --- | --- | --- |
|  | A. | 13.5 |
|  | B. | 15.5 |
|  | C. | 12.5 |
|  | D. | 14.5 |
|  | E. | 11.5 |