| Question Number | | |
|---|---|---|
| **Blue** | **White** | **Answer** |
| 3 | 1 | (B). Most watched TV show. |
| 5 | 2 | (C). 22,855 = (0.17/0.30)×41,416 |
| 6 | 3 | (D). Option (A) has ON too high and SK and MB too low. (B) has ON and BC on par with each other while (C) has AB and QC on par with each other. |
| 8 | 4 | (A). The interquartile range is a resistant measure of spread. The sample mean is not appropriate for skewed distributions and is on the same scale as the standard deviation, but not as the sample variance. |
| 1 | 5 | (D). 483.3 and 550. The mean can be calculated by (2×200 + 300 + 2×400+…+750)/15 = 483.33. Since there are 15 data points, the median corresponds to the $8^{th}$ observation which is \$550K. |
| 2 | 6 | (B). To detect outliers, we need to look 1.5×IQR below Q1 and above Q3, or 1.5 (600 – 400) = 300 below 400 or 300 above 600. Since 950 is the only observation outside the interval (100,900), the answer is (B). |
| 7 | 7 | (D). Only statement (ii) is TRUE; the rest are FALSE. |
| 10 | 8 | (A). We have Z = (2050 – 1800)/120 = 2.08, and according to the Standard normal table, the area below 2.08 is 0.9812, so the area above 2.08 is 1 – 0.9812=0.0188. Given the 68-95-99.7 rule this is reasonable since our answer should be less than 0.025. |
| 9 | 9 | (D). Let $Z_1$ = (1750 – 1800)/120 = –0.42 and $Z_2$ = (2050 – 1800)/120 = 2.08. Using the answer from Q8, the area below 2.08 is 0.9812. From the normal table, the area below -0.42 is 0.3372. Therefore the proportion between 1650 hrs and 1800 hrs is (0.9812 – 0.3372) = 0.6440, or 64.40%. |
| 13 | 10 | (B). Let $Z = (95 – 130)/\sigma = –35/\sigma$. The area below Z on the standard normal curve is 0.8% = 0.008, so from the table we see that Z is approximately equal to –2.41. Then the standard deviation is $\sigma = –35/Z = –35/(–2.41) = 14.52 \approx 15$. |
| 15 | 11 | (D). There were 404 employed adults, of which 105+121 = 226 attained a post-secondary education. The desired proportion is 2226/404 = 0.559. |
| 16 | 12 | (B). We are asked to find the proportion of students that are **jointly** unemployed **and** highest attained level education was senior high school or lower, which is (9 + 16 + 43)/522 = 0.130. |
| 14 | 13 | (A). We are asked to find the proportion of students unemployed **conditional** on highest attained level of education being senior high school, which is 43/188 = 0.229. |

| Blue | White | **Answer** (cont'd) |
|---|---|---|
| 19 | 14 | (C). Since $b_0 = \bar{y} - b_1\bar{x}$ = 371.58 + 67.76(8.44)= 943.37. |
| 18 | 15 | (A). For 8.7 minute run, we have $\hat{y} = 7.47 - 0.184(8.7) = 5.8692$. So the residual is $y - \hat{y} = 5.93 - 5.8692 = 0.0608$. This is closest to 0.06. |
| 17 | 16 | (E). Recall that $b_1 = r\, s_y/s_x$ from which we can get $r = b_1\, s_x/s_y = -0.184(0.395)/0.108 = -0.673$, which is closest to –0.7. |
| 20 | 17 | (C). Statement (i) is TRUE because the residuals appear to be randomly distributed, but with very high variance for the original data. Although the variance of the residuals for the logged data has been highly reduced, both models have the same $R^2$ value and other than scale, the two plots look almost identical. So logging the data did not improve the fit. There may be one suspected outlier in the residual plot. **THIS QUESTION WAS NOT GRADED DUE TO INTERPRETATION ISSUES**. |
| 4 | 18 | (E) The value of a statistic DOES NOT always increase as the sample size increases. |
| 11 | 19 | (E) A double-blind study. |
| 12 | 20 | (C). Statements (i) and (iii) are TRUE. |

**Answer Key:**

| **QUESTION:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **WHITE:** | B | C | D | A | D | B | B | A | D | B | D | B | A | C | A | E | B | E | E | C |

| **QUESTION:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BLUE:** | D | B | B | E | C | D | B | A | D | A | E | C | B | A | D | B | E | A | C | B |

**Midterm1Resp**: '**.**' means you answered correctly,
a number means you got the question wrong, and
you submitted the answer given as shown below:

1 = A
2 = B
3 = C
4 = D
5 = E