Statistics Answers January 2005

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

Mu Alpha Theta January Regional 2005 Statistics Solutions----

1. B. Coefficient of determination is a square of correlation and all squares are positive. Standard deviation is the square root of the variance and all square roots are positive. The Variance is always positive.

2. E. None of these statements are true. Consider a data set consisting of all of the same number and it can be seen that A – D are all false.

3. B. Assuming A and B are mutually exclusive implies P(A u B) = P(A) + P(B). However this means P(A u B) = 1.2 which is impossible so the assumption must not be valid. The range of P(A n B) has a lower bound of 0.2 (that would bring the P(A u B) down to 1) and an upper bound of 0.5 (the intersection cannot be larger than the smaller of the two probabilities). Since zero is not contained in this range A and B cannot be mutually exclusive.

4. B. 1 – (P no Rain on Sat AND no rain on Sun) = .615

5. E. NOTA. Each of A, B, and C <u>are</u> the important principles of experimental design; however, the definitions of each are incorrect. The definitions used are all of the common mistakes students often make and should know.

6. E. NOTA Adding or subtracting the same number to all the values of the same variable or multiplying by a positive number has no effect on the correlation. Interchanging the variables also has no effect.

7. C . By definition, events A and B are independent if P(A n B) = P(A)*P(B). As seen in # 3, the P(A n B) cannot be calculated exactly but $0.2 \leq P(A \text{ n } B) \leq 0.5$. Since P(A)*P(B) falls within this range it is possible that two events are independent (It is not certain though because the P(A n B) could equal ANY value in that range.

8. D. A residual is defined as y – yhat, with y-hat being the predicted value. If y – y-hat is positive, it means that y-hat is smaller than y. This means that the LSRL has under-predicted the y. Two of the distractors list under and over predictions of the X to see if students know the difference.

9. A. The probability of A is $(.5)^4$ and the probability of B, C, and D are all $(.5)^5$

10. B. The p-value is NOT the probability that the null hypothesis is correct. The p-value is the probability that we will get a value this extreme given that the null hypothesis is correct.

11. A. The probability that there are at least 3 left-handers can be calculated by:
    1 – (Prob of 0 or 1 or 2 left-handers)
    So.... 1 – (.04239 + .1413 + .2276) = 0.5586
    The distractors are: the complement of the correct answer, and the answer calculated using 3
12. D. This is either a very easy or very hard question. "You and twelve friends" means a total of 13 people. If there are 13 people and only 12 months to be born in, there is a 100% chance that at least two people were born in one month.

13. C. II and III only. A residual plot cannot have all values above the line y = 0, so either the LSRL was calculated incorrectly or the residual plot was drawn incorrectly. Statement #1 is not correct because there is nothing in the description to imply a distinct pattern and a linear model certainly could be appropriate for the data.

14. B. P (A or B) = P(A) + P(B) – P (A and B) so we substitute...
    0.6 = 0.5 + P(B) - 0.1 P(B) = 0.2
15. C. The equation for slope is: slope = r (Sy/Sx) The standard deviations are positive. Therefore, if the slope is negative, the correlation must be negative.

16. C. Taking the linear regression of log x, log y will create an equation which can be transformed into a power regression

17. D. It is expected that students will blindly calculate this without checking to see if it is possible.
The value of the probabilities exceeds 1.0, so this problem is clearly not possible.

18. D. There are 2 factors (temperature and amount of yeast) with 12 treatments (4 levels X 3 levels) and the factor with the smallest number of levels has 3 levels.

19. B. If they are both disjoint, meaning they have no outcomes in common, P(A or B) has to have at least 2 numbers (at least one from A and one from B.)

20. D. The probability is .44 All other probabilities are smaller

21. D. The analogy compares the progression of single sample to multiple sample tests. The Progression is: One Sample proportion, Two sample proportion, Chi Square Goodness of Fit (which is a multiple sample proportion.) The correct answer has the same progression: One Sample t-test, Two Sample t-test, ANOVA (multiple same t-test). I labeled it as a One Way ANOVA, since a Two Way ANOVA really muddies the water for the comparison as it is not a pure multiple sample t-test.

22. A. 1/3
P($20 and $10) = P($20 and $10) or P($10 and $20)
= (5/10 X 3/9) + (3/10 X 5/9)
= 1/6 + 1/6
= 1/3

23. B. I has 4 df III has 5df II has 6df

24. A. Parametric --The AP course covers many parametric tests, but only one non-parametric test: The Chi Square test.

25. C. The solution is found by ${}_{53}C_5$ multiplied by 42. The order of the first 5 don't count, and then there are 42 times more ways to get what the same combinations with the powerball.

26. B. This is a definition.

27. B. The probability of II is 1/230,300 which is calculated by ${}_{50}C_4$. The probability of III is 1/ 1,860,480 which is calculated by ${}_{20}P_5$. The probability of I is (1/2.43E18) which is calculated by 20!

28. D. Our confidence is in the method and we are 90% confident that the mean of the population, not of the sample is in our interval.

29. C. 396 The LSRL equation of y-hat is 207.4 + 45.3x. Substituting 4 for x, we get 388.6 or 389. The residual is 7 meaning that Actual – Predicted = 7. The predicted is 389, so the actual is 7 larger than this, or 396.

30. B. The t-ratio is found by taking the Coef divided by the St.dev. So 207.4/8.3 = 24.988