STAT302: Secs 102 & 103
Summer I 1998
Exam #4
Form A

Instructor: Julie Hagen Carroll

1. **Don't even open this until you are told to do so.**
2. Be sure to mark your section number (102 or 103) and your test form (A, B, C or D) on the scantron!
3. Sign your name where indicated on your scantron and write your Wednesday section number and computer number beside it. Also, you must place your scantron in the correct section stack (for next Tuesday).
4. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly** on the scantron. Multiple marks will be counted wrong.
5. You will have 60 minutes to finish this exam.
6. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
7. This exam is worth 100 points, and will constitute 25% of your final grade.
8. Good luck!

```
                  Number of obs =      55
                  F(  1,    53) =    2.67
                  Prob > F      =  0.1082
                  R-squared     =  0.0480
                  Adj R-squared =  0.0300
                  Root MSE      =  1.7777

  Source |       SS       df       MS
---------+------------------------------
   Model |  8.43720904     1  8.43720904
Residual |  167.490064    53  3.16018988
---------+------------------------------
   Total |  175.927273    54  3.25791246

------------------------------------------------
     Age | Coef. S.E.  t     P>|t|  [95% C.I.]
---------+--------------------------------------
ShoeSize | .180 .110  1.634  0.108  -.041 .4003
   _cons | 19.5 1.12 17.400  0.000  17.21 21.69
------------------------------------------------
```

1. What is the best conclusion that may be made based on the output above?

   A. Since the p-value = 0.000, the regression equation is better at predicting the y's than the average y's.
   B. Since the $R^2 = 0.0480$, at the 5 and 10% levels the regression equation is better at predicting the y's than the average y's.
   C. Since the p-value = 0.000, the regression equation is 0.
   D. Since the p-value = 0.108, the regression equation is not any better at predicting the y's than the average y's.
   E. Since the p-value = 0.108, the regression equation is any better at predicting the y's than the average y's.

2. What is the prediction (least squares) equation for the output above?

   A. $\widehat{ShoeSize} = 0.180 + 19.5$
   B. $\widehat{ShoeSize} = 0.180 + 19.5 * Age$
   C. $\widehat{Age} = 0.180 + 19.5 * ShoeSize$
   D. $\widehat{Age} = 19.5 + 0.180 * ShoeSize$
   E. $\widehat{ShoeSize} = 19.5 + 0.180 * Age$

3. What would happen if we added the point (20,40)?

   A. s_e would increase and R^2 would decrease
   B. s_e would decrease and R^2 would increase
   C. s_e would increase and R^2 would not change
   D. It's is likely that a ShoeSize of 20 is outside the valid range of the model, and we cannot assume the relationship is still linear.
   E. None of the above

4. What is the total variance of *Age* in the output above?

   A. It is not given in this output.
   B. 175.93
   C. 8.44
   D. 3.16
   E. 3.26

5. Why should we run a Two-Way ANOVA rather than just two One-Way ANOVA's?

   A. We are likely to explain more of the variation.
   B. We can always test the interaction, too.
   C. Two-Way uses the less data.
   D. All of the above.
   E. Exactly two of the above.

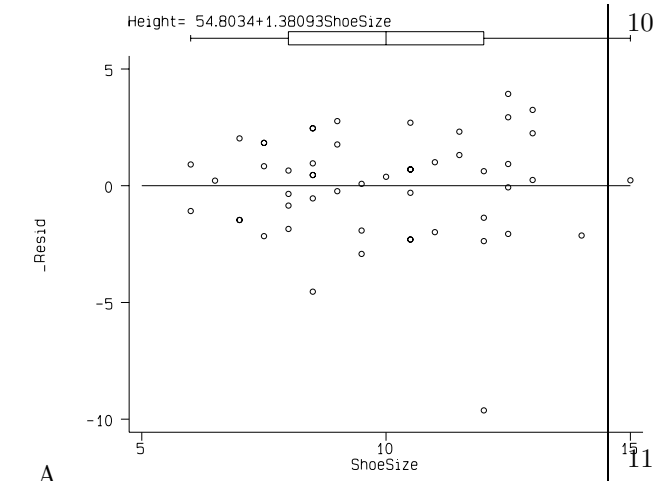
```
                Analysis of Variance
    Source          SS       df     MS      F  Prob > F
------------------------------------------------------------
Between groups  1.9152113    3 .638404  3.32  0.0274
 Within groups 9.43046516   49 .192458
------------------------------------------------------------
    Total      11.3456765   52 .218186

Bartlett's test for equal variances:
                chi2(3)= 4.2899  Prob>chi2 = 0.232
```

6. What should we say about the treatment (group) effect in the ANOVA table above?

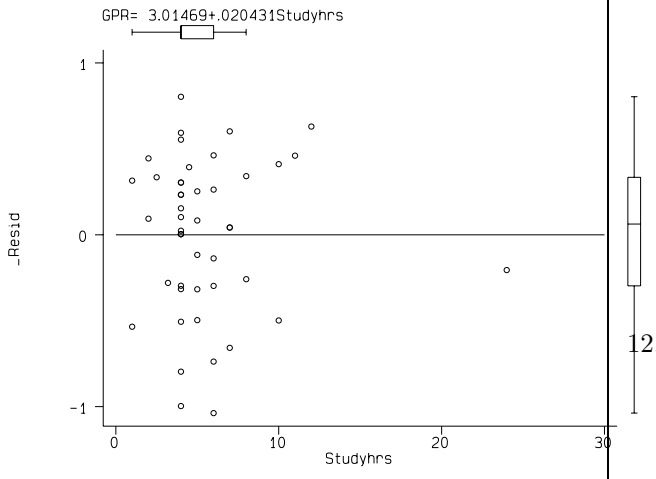
   A. There is only a 2.74% probability that the effect exists.
   B. The effect is only 2.74% significant.
   C. The effect is significant at the 5 and 10% levels.
   D. There is a 23.3% probability that the effect exists.
   E. The effect is not significant.

7. In reference to the ANOVA table above, we assume the variances within each group are equal. What is our estimate of this variance?

   A. We don't assume the variances are equal since the p-value is less than 0.10.
   B. 9.43
   C. 11.3
   D. 0.218
   E. 0.192



A



B

8. Which of the residual plots above indicate that there is a possible influential point in the data?

   A. A
   B. B
   C. Both plots show points that should be deleted.
   D. Both plots indicate an influential point.
   E. Both plots indicate an outlier in the data.

9. Why do we use $\alpha = 0.10$ for Barlett's test for equal variances?

   A. Because a Type I error would mean we did an invalid F-test.
   B. Because a Type I error would mean we claim there is an effect when there isn't any.
   C. Because a Type II error would mean we did an invalid F-test.
   D. Because a Type II error would mean we claim there is an effect when there isn't any.
   E. Because Barlett's test is very conservative.

10. Suppose you want to test the significance of the slope in a simple linear regression equation, but the only information you have are 3 confidence intervals for β_1: (.295,4.305), (-0.119,4.719), (-0.978,5.578). What is the range of the p-value for testing $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$?

    A. pv > 0.10
    B. $0.10 >$ pv > 0.05
    C. $0.05 >$ pv > 0.01
    D. $0.01 >$ pv
    E. Confidence intervals only provide information when testing means.

11. Suppose we are trying to predict y but the only x available has only a moderate correlation with y (the R^2 is around 50

    A. Multiply the y's by a large number so the slope will increase.
    B. Divide the x's by a large number so the slope will increase.
    C. Gather more data.
    D. All of the above.
    E. Exactly two of the above (excluding D.).

12. What is the consequence of making a Type II error when testing the interaction in a Two-way ANOVA?

    A. We claim that the interaction is significant when it is not.
    B. We claim that the interaction exists when it does not.
    C. We claim that there is no interaction when there is interaction.
    D. We claim that the variances are equal when they are not.
    E. Exactly two of the above.

13. What does it mean to be statistically significant in reference to testing the true slope, $\beta_1 = 0$?

    A. It means that the true slope is statistically equal to 0.
    B. It means that the estimated slope is statistically equal to β_1.
    C. It means that regression model is a statistically significant prediction equation for the y's.
    D. All of the above are correct.
    E. Exactly two of the above are correct.

14. What does it mean to have a coefficient of determination, $R^2 = 0$?

A. There is no relationship between the x and y variables.
B. The x's are of no use in predicting the y's with simple linear regression.
C. The standard deviation of the regression line, $s_e = 1$.
D. All of the above.
E. Exactly two of the above.

```
                Analysis of Variance
Source      SS       df      MS      F     Prob > F
-------------------------------------------------
Between 193.092867   1   193.09   129.08   0.0000
 Within  82.275554 55   1.4959
-------------------------------------------------
 Total 275.368421   56   4.9173

Bartlett's test for equal variances:
             chi2(1) =   0.1795  Prob>chi2 = 0.672
```

15. What is the proportion of variation, R^2, explained above?

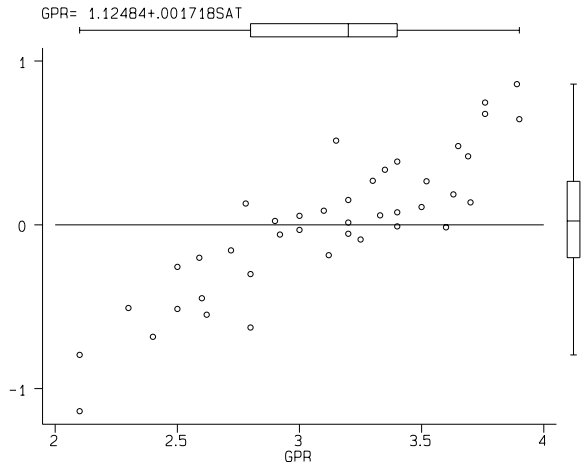
A. 42.6%
B. 69.6%
C. 30.4%
D. 29.9%
E. 70.1%

16. Based on the output above, we claim that the means are not all equal. Later we discover that there is actually no treatment effect. What happened?

A. We made a correct decision since no treatment effect says the means are not all equal.
B. We made a correct decision since no treatment effect says the means are all equal.
C. We made a Type I error since we rejected a true null hypothesis.
D. We made a Type II error since we rejected a false null hypothesis.
E. We made a Type II error since we failed to reject a false null hypothesis.

17. What is the correct alternative hypothesis for the ANOVA above?

A. $H_A : \mu_1 = 0$
B. $H_A : \mu_1 \neq 0$
C. $H_A : \mu_1 = \mu_2$
D. $H_A : \mu_1 \neq \mu_2$
E. None of the above are correct.



18. What does the residual plot above indicate?

A. The y's are linear, so that assumption is valid.
B. The residuals are not independent.
C. The residuals are not normally distributed.
D. The residuals don't have constant variance.
E. The residuals don't have a mean = 0.

```
  Source |  Partial SS  df       MS     F   Prob > F
---------+------------------------------------------
    CMus |  28960.2222   1  28960.222  3.34  0.0924
     Veh |  117067.111   2  58533.556  6.76  0.0108
CMus*Veh |  265507.111   2  132753.56 15.33  0.0005
         |
Residual |                ?????????
---------+------------------------------------------
   Total |  515443.778  17  30320.22
```

19. What is the Mean Square Error for the model above?

A. 103909.333
B. 132753.56
C. 8659.111
D. 411534.444
E. 51954.665

20. What is the conclusion based on the output above?

A. Since there is significant interaction, we cannot make any statements about either main effect.
B. $CMus$ is significant at the 10% level only.
C. Veh is significant at the 5 and 10% levels.
D. Both B and C are correct.
E. B and C are correct and the interaction is significant at the 1, 5 and 10% levels.

Answers: 1. D 2. D 3. D 4. E 5. A 6. C 7. E 8. B 9. C 10. B 11. C 12. C 13. C 14. B 15. E 16. C 17. D 18. B 19. C 20. A