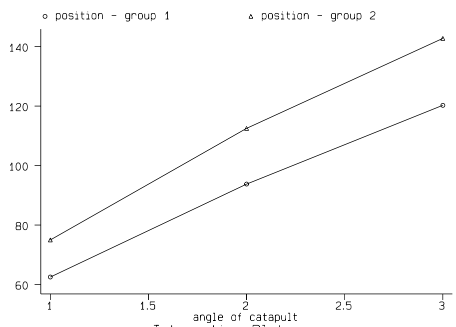
STAT302: Secs 102 & 103
Summer I 1997
Exam #4
Form A

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 Tuesday 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 20% of your final grade.
8. Good luck!

1. Which of the following is NOT!!!! an advantage of Two-way ANOVA over One-way ANOVA?

   A. If another effect is causing added variability, including it in the Two-way ANOVA will reduce MSE $= s_p^2$.
   B. If there are replications, then we can also test for interaction between the 2 effects.
   C. We can assume the variances are equal.
   D. All of the above are advantages.
   E. None of the above are advantages.



Interaction Plot

2. What conclusions may be drawn from the interaction plot above?

   A. Since it is an interaction plot, the interaction must be significant which means we cannot say anything about *angle* nor *position*.
   B. Since the lines are relatively parallel, the interaction is not significant, but both *angle* and *position* are.
   C. Since the lines are relatively parallel, the interaction is significant which means we cannot say anything about *angle* nor *position*..
   D. Since the lines are relatively parallel, the interaction is not significant, and neither is *angle*, but *position* is.
   E. Since the lines are relatively parallel, the interaction is not significant, and neither is *position*, but *angle* is.

3. A small s_e, the average deviation about the regression line, means we have a 'good' fit IF

   A. it is small relative to 0.
   B. it is small relative to s_y.
   C. it is small relative to R^2.
   D. it is large relative to R^2.
   E. Exactly two of the above

Source	SS	df	MS
Model	11948.9887	1	11948.9887
Residual	4172.82949	20	208.641475
Total	16121.8182	21	767.705628

weight	Coef.	Std. Err.	t	P>|t|	[95% C. I.]	
length	2.202	.2909564	7.568	0.000	1.595	2.81
_cons	-105.74	23.15161	-4.567	0.000	-154.0	-57.45

4. Based on the data represented above, what is the correlation coefficient, r, for *weight* and *length*?

   A. You can't determine r from the output above.
   B. 0.74
   C. -0.74
   D. 0.86
   E. -0.86

5. Again using the output above, what is the predicted regression line?

   A. $\widehat{weight} = 2.202 length - 105.74$
   B. $\widehat{weight} = -105.74 length + 2.202$
   C. $\widehat{length} = 2.202 weight - 105.74$
   D. $\widehat{length} = -105.74 weight + 2.202$
   E. $\widehat{weight} = -105.74\_cons + 2.202 length$

6. What is the appropriate conclusion for the output above?

   A. Since the p-value is 0, we can assume the true slope, β_1, is 0.
   B. Since the p-value is 0, we can assume the model using the x's is better at predicting the y's than $\overline{y}$.
   C. Since the R^2 is 0.74, which is considered only moderate linearity, the relationship between x and y is only moderate.
   D. We need the plot of the fitted model before we can determine the correct conclusion.
   E. None of the above are correct conclusions.

7. If the point (75, 100) were added to the data, the associated $\hat{y}$ would be 59.4 and

   A. s_e would increase since $e = 40.6$.
   B. s_e would increase since $e = -40.6$.
   C. s_e would decrease since $e = 40.6$.
   D. s_e would decrease since $e = -40.6$.
   E. s_e would remain unchanged since s_y would change.

8. In Two-way ANOVA with interaction, we must test the interaction first because

   A. it appears before the 2 main effects in the ANOVA table.
   B. it is the hardest to detect.
   C. it adds the least amount of variability.
   D. it affects the interpretation of the 2 main effects.
   E. Exactly two of the above are correct.

```
   Source | Partl SS   df      MS        F     Pr > F
 ---------+----------------------------------------
     rate |   536.00    2   268.00    15.03   0.0138
     year |  3658.67    2  1829.33   102.58   0.0004
 Residual |    71.33    4    17.83
 ---------+----------------------------------------
    Total |  4266.00    8   533.25
```

9. What is the proportion of variation that was explained by the model represented above?

   A. 16.7%
   B. 1.67%
   C. 12.6%
   D. 85.8%
   E. 98.3%

10. What conclusions may be made from the output above?

    A. Since we don't have a test for interaction, we cannot make conclusions about either main effect.
    B. Assuming there is no interaction, both *rate* and *year* are significant effects at the 5 and 10% levels.
    C. Assuming there is no interaction, both *rate* and *year* are significant effects at the 10% level only.
    D. Assuming there is no interaction, both *rate* and *year* are significant effects at the 1, 5 and 10% levels.
    E. Assuming there is no interaction, both *rate* and *year* are significant effects at the 5% level only.

11. What would we have to do to check if there is interaction between *rate* and *year*?

    A. Run the same output but include the interaction line.
    B. Look at the interaction plot.
    C. Gather more data since there are not enough degrees of freedom to test the interaction here.
    D. Exactly two of the above would work.
    E. None of the above will tell us conclusively since a hypothesis test can never prove H_0 true.

```
           Analysis of Variance
 Source     SS      df    MS      F    Prob > F
------------------------------------------------
Between   114.580    3  38.19   2.17   0.0991
 Within   1340.71   76  17.64
------------------------------------------------
  Total   1455.29   79  18.42

Bartlett's test for equal variances:
      chi2(3) =   2.2953  Prob>chi2 = 0.513
```

12. What is the appropriate *alternative* hypothesis for the output above?

    A. $H_A : \mu_1 = \mu_2 = \mu_3$
    B. $H_A : \mu_1 = \mu_2 = \mu_3 = \mu_4$
    C. $H_A : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$
    D. $H_A : \mu_1 \neq \mu_2 \neq \mu_3$
    E. None of the above are the correct alternative hypothesis.

13. What is the appropriate conclusion for the output above?

    A. At the 10% level, there is significant interaction, but not at the 1 or 5% levels.
    B. At the 10% level, we can conclude that the means are NOT all equal, but not at the 1 or 5% levels.
    C. At the 10% level, we can conclude that the means ARE all equal, but not at the 1 or 5% levels.
    D. Since the p-value for Barlett's test is not less than 10%, the F test is invalid.
    E. Since the p-value for Barlett's test is not less than the p-value for the F test, we cannot conclude that the means are different.

14. What would be the consequence of a Type I error in the hypothesis above?

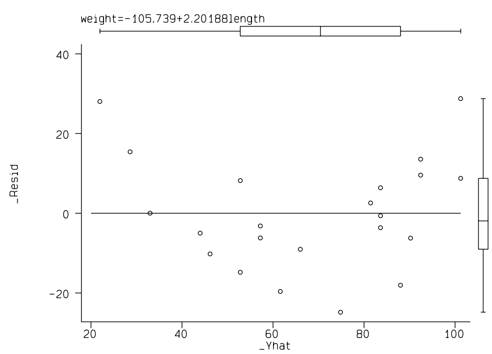
    A. We are only concerned about Type I and II errors in 1 or 2-sample testing, not ANOVA.
    B. We conclude that there is an effect when there isn't one.
    C. We conclude that there is an effect when there is one.
    D. We conclude that there is NOT an effect when there isn't one.
    E. We conclude that there is NOT an effect when there is.

15. The assumptions necessary for inference (hypothesis testing) concerning the true slope of the simple linear regression line, β_1, are

    A. the y's are linear, *i.e.*, they fall along a straight line.
    B. the e's are linear where $e = y - \hat{y}$.
    C. the y's are independent and normally distributed with mean = 0.
    D. All of the above are necessary assumptions.
    E. Exactly two of the above are correct assumptions.

16. Which of the following statements about the intervals based on the estimated regression line is/are TRUE?

    A. Prediction intervals will always include all of the data points.
    B. Confidence intervals will always include the true regression line.
    C. Confidence intervals are always narrower than prediction intervals if $s_e \neq 0$.
    D. All of the above are true.
    E. None of the above are true.



17. What does the residual plot above tell us?

    A. The mean of the residuals is not 0.
    B. The data is not linear.
    C. The variance of the residuals is not constant.
    D. There is an outlier in the data.
    E. There is an influential point in the data.

18. If the variability about the regression line is large, *i.e.*, s_e is close to s_y, what can we do to improve the model?

    A. We can't do anything but find another variable to use to predict y.
    B. Divide the y's by a large number since that would reduce both s_y and s_e.
    C. Multiply the y's by a large number since that would reduce both s_y and s_e.
    D. Gather more data.
    E. None of the above will help.

19. What is the point of looking at *residual plots* when making inferences about the true regression line?

    A. We can see if the residuals have constant variance.
    B. We can see if the residuals are normal.
    C. We can see if the residuals are linear.
    D. Exactly two of the above
    E. All of the above

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

    A. We want to be confident that an F test is valid for the data.
    B. We want to reduce the chance of a Type I error.
    C. The Barlett's test isn't very accurate, so we use a large α to compensate.
    D. We don't care if there is a chance of a Type II error.
    E. The Barlett's test is a two-sided test of hypotheses, so dividing α in two gives us the usual $\alpha = 0.05$.

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