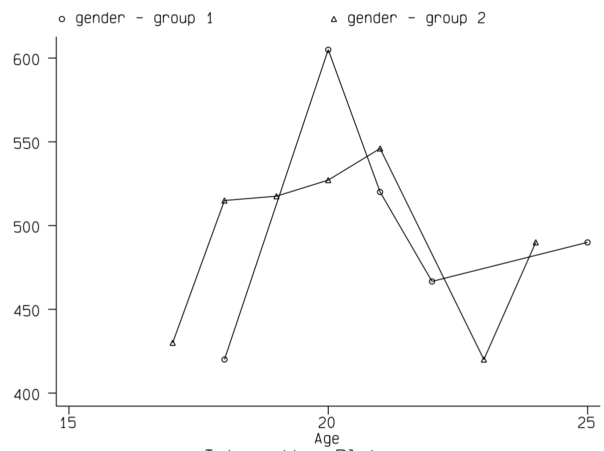
STAT302: Secs 102 & 103
Summer I 1999
Exam #4
Form A

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.
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!



Interaction Plot

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

   A. The lines cross, therefore, there is significant interaction between *Age* and *Gender*.
   B. The lines cross, therefore, there is NOT significant interaction between *Age* and *Gender*.
   C. The lines cross, therefore, *Age* and *Gender* have a significant effect.
   D. The lines cross, therefore, *Gender* has a significant effect.
   E. The lines cross, therefore, *Age* has a significant effect.

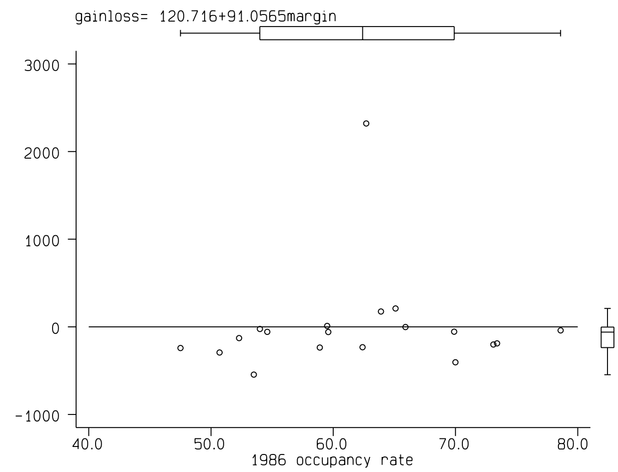
Source	Partial SS	df	MS	F	Prob > F
gend	.008638264	1	.008638264	0.03	0.8753
age	3.36985862	10	.336985862	0.98	0.4825
gend*age	2.42676946	2	1.21338473	3.52	0.0427
Residual	9.98636364	29	.344357367		
Total	14.7906977	42	.352159468		

2. (NOTE: this output is NOT related to the graph above, so don't let it affect your answer to that question.) We must assume that the variance is constant within each *Age* and *Gender* combination (they are all equal) for the F-tests to be valid. What is our estimate of this variance?

   A. We need a residual plot to tell if it is safe to assume that the variance is constant.
   B. 1.213
   C. 0.344
   D. 0.352
   E. 3.52

3. What is the best conclusion based on the output above?

   A. The p-value $= 0.0427$, so we cannot assume that the variances are equal.
   B. The p-value $= 0.0427$, so we can assume that the variances are equal.
   C. The p-value $= 0.0427$, so we can *Age* and *Gender* are significant effects.
   D. The p-value $= 0.0427$, so we can *Age* and *Gender* are NOT significant effects.
   E. The p-value $= 0.0427$, so we can the interaction between *Age* and *Gender* is significant.



4. What does the residual plot above indicate?

   A. The assumption of equal variances (constant) has been violated.
   B. The assumption of normality may have been violated.
   C. The assumption of all points being equally weighted has been violated.
   D. The assumption of linearity has been violated.
   E. All of the assumptions for simple linear regression are valid.

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

   A. We want to increase the chance of a Type I error.
   B. We want to reduce the chance of a Type I error.
   C. We want the p-value for the F-test to be smaller.
   D. We want to be sure we are doing a valid procedure.
   E. We want to be sure that the means are equal.

```
  Source |       SS       df       MS
---------+------------------------------
   Model |  47.5154184     1  47.5154184
Residual |   8.0485833     8  1.00607291
---------+------------------------------
   Total |  55.5640017     9  6.17377797

------------------------------------------------
fertile| Coef.  Std. Err.  t   P>|t|  [95%ConfInt]
-------+----------------------------------------
fschool|-.06872 .009998 -6.872 0.000 -.0918 -.0456
  _cons| 7.3465 .528655 13.897 0.000 6.1274  8.566
------------------------------------------------
```

6. According to the output above, is *fschool*, the percent of females enrolled in secondary schools, a viable means of prediciting the fertility rate, *fertile*?
   A. Of course, the more educated the population, the lower the risk of unwanted pregnancies.
   B. The p-value = 0.000, so the true slope, $\beta_1 = 0$.
   C. The p-value = 0.000, so yes, we should use *fschool* to predict *fertile* rather than just the average fertility rate.
   D. The sample slope, $b_1 = -0.06872$, which is practically 0, so no, we shouldn't use *fschool* to predict *fertile*.
   E. The sample slope, $b_1 = -0.06872$, which is not 0, so, we should use *fschool* to predict *fertile*.

7. What is the correlation coefficient, r, for the output above?
   A. -0.06872
   B. 0.06872
   C. 0.2604
   D. 0.9247
   E. -0.9247

8. Again using the output above, what is the total variance of *fertile*, the fertility rate?
   A. 47.515
   B. 8.0486
   C. 55.564
   D. 1.0061
   E. 6.1738

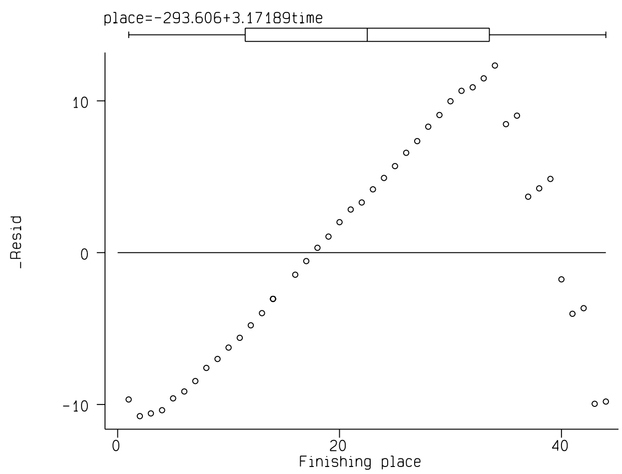
```
     Spouse| Present family type
      abuse|
    present|
   marriage| intact  remarry   reconst. |  Total
-----------+----------------------------+-------
  no abuse |    743       92         78 |    913
     abuse |     36        9         11 |     56
-----------+----------------------------+-------
      Total|    779      101         89 |    969

        Pearson chi2(2) =  10.8147   Pr = 0.004
```

9. Using the Two-way table above, which represents the counts of families with or without spousal abuse by the type of family, would you say the two categories are independent?
   A. It is obvious *family type* and *spousal abuse* are related, since 'broken' families tend to be more violent.
   B. It is obvious *family type* and *spousal abuse* are related, since families tend split because of violence.
   C. The p-value for testing whether *family type* and *spousal abuse* are related is 0.004, so there is only a 0.4% chance that the two are related.
   D. The p-value for testing whether *family type* and *spousal abuse* are related is 0.004, so the hypothesis that the two are independent would be rejected.
   E. The p-value for testing whether the variances are equal is 0.004, so the assumption of equal variances is invalid.

10. What is the probability of being in a *remarried* and *abusive* family if the two categories are independent?
   A. 9/969
   B. 101/969 * 56/969
   C. (101*56)/969
   D. 9/101
   E. 9/56

---

11. What does it mean to be 'statistically significant'?
   A. It means that you rejected the null hypothesis.
   B. It means that you found the means to be significant.
   C. It means there really is some reason to do a hypothesis test.
   D. It means there is an influential point in the data.
   E. It means that you passed STAT302.



12. What does the plot above tell us?

    A. The relationship between *Finishing place* and *time* is not strictly linear.
    B. The relationship between the residuals and *time* is not strictly linear.
    C. The residuals are not strictly linear.
    D. The residuals don't have constant variance.
    E. The residuals are skewed, therefore, not normally distributed.

```
Variable |      Obs        Mean   Std. Dev.
---------+--------------------------------
     age |       13    3.753846    1.637893
  estage |       13    3.861538    1.823247
---------+--------------------------------
   diff. |       13   -.1076923    .9059631

             Ho:  mean difference = 0  (paired data)
                       t = -0.43 with 12 d.f.
                 Pr > |t| = 0.6758
95% CI for difference = (-.65516029,.43977571)
```

13. The above output is a paired t-test for testing whether there is a difference in the true *age* and the *estimated age*. The p-value = 0.6758. Had we done a two-sample t-test instead, would we still have found no difference in the ages?

    A. Yes, the paired t-test has the most power, so if it couldn't find a difference, neither could a two-sample t.
    B. It makes no sense to NOT do a paired t-test, since we're comparing how good the estimates are.
    C. The p-value would be larger for the two-sample t, so it would be even harder to get a rejection.
    D. All of the above are correct.
    E. Exactly two of the above are correct (excluding D.).

```
   Political|   Summary of Mathematics SAT score
       party|
  preference|        Mean   Std. Dev.       Freq.
------------+-----------------------------------
   Democrat |   512.85714   93.102803          14
   Republic |   538.57143   55.205245           7
   Independ |   492.14286   60.533344          14
------------+-----------------------------------
      Total |   509.71429   74.537228          35

                 Analysis of Variance
Source         SS      df      MS       F    Prob > F
-----------------------------------------------------
Between    10290.00     2  5145.00    0.92   0.4081
 Within   178607.143   32  5581.473
-----------------------------------------------------
  Total   188897.143   34  5555.798

Bartlett's test for equal variances:
              chi2(2) =   3.2232  Prob>chi2 = 0.200
```

14. Does one's *Political party preference* affect the true mean Math SAT score?

    A. Yes, the p-value is 0.20, so we reject the null and conclude that there is a difference in the means.
    B. No, the p-value is 0.20, so we fail to reject the null and conclude that there is not a difference in the means.
    C. No, the p-value is 0.4081, so we fail to reject the null and conclude that there is not a difference in the means.
    D. Yes, the p-value is 0.4081, so we reject the null and conclude that there is a difference in the means.
    E. No, the p-value is 0.92, so we fail to reject the null and conclude that there is not a difference in the means.

15. What is the correct alternative hypothesis for the output above?

    A. $H_0 : \mu_D = \mu_R = \mu_I$
    B. $H_A : \mu_D = \mu_R = \mu_I$
    C. $H_A : \mu_D \neq \mu_R \neq \mu_I$
    D. H_A: all of the means are not the same
    E. H_0: all of the means are the same

16. What is the consequences of making a Type I error in One-way ANOVA?

    A. You conclude that the variances are equal, but they really aren't so you use an invalid F-test.
    B. You conclude that there is an effect due to the different groups, but there really is not.
    C. You conclude that the variances are not all equal, but they really are different.
    D. You conclude that the means are equal when they really are not.
    E. Exactly two of the above are Type I errors.

```
                     Number of obs =      11
                     F(  1,     9) =   21.27
                     Prob > F      =  0.0013
                     R-squared     =  0.7027
                     Adj R-squared =  0.6696
                     Root MSE      =  38.873

  Source |       SS       df       MS
---------+------------------------------
   Model |  32140.9628     1  32140.9628
Residual |  13599.7645     9  1511.08495
---------+------------------------------
   Total |  45740.7273    10  4574.07273

-----------------------------------------------
 msat |  Coef. Std.Err. t   P>|t|[95% Conf. Int]
------+----------------------------------------
 vsat | 1.0351  .2244 4.612 0.001   .5274  1.543
_cons |12.1834 123.77 0.098 0.924 -267.81 292.18
-----------------------------------------------
```

17. If the point (400,326) were added to the model above which of the following would happen?

    A. The R^2 would decrease and the standard deviation of the regression line, s_e, would increase.
    B. The R^2 would increase and the standard deviation of the regression line, s_e, would decrease.
    C. The R^2 would decrease and the standard deviation of the regression line, s_e, would decrease.
    D. The R^2 would increase and the standard deviation of the regression line, s_e, would increase.
    E. Neither R^2 nor s_e would change.

18. If instead of adding a point, we decided to multiply the Math SAT scores, *msat*, by 100, which of the following would happen?

    A. Nothing would change since you haven't actually changed the relationship.
    B. The R^2 and s_e would both increase by a factor of 100.
    C. The s_e would increase by a factor of 100.
    D. The R^2 would increase by a factor of 100.
    E. The R^2 would decrease by a factor of 100.

```
            |    Summary of Actual age in years
         Sex|        Mean   Std. Dev.       Freq.
------------+-----------------------------------
       male |   4.5833334   1.9772878           6
     female |   3.0428571   .91078198           7
------------+-----------------------------------
      Total |   3.7538462   1.6378927          13

                        Analysis of Variance
Source       SS        df       MS        F  Prob > F
------------------------------------------------------
Between   7.66683197    1    7.66683   3.44  0.0907
 Within  24.5254776    11    2.22959
------------------------------------------------------
  Total  32.1923095   12    2.682692

Bartlett's test for equal variances:
             chi2(1) =   2.8612  Prob>chi2 = 0.091
```

19. What is the correct conclusion for the output above?

    A. The p-value for Bartlett's test is less than the p-value for the F-test, so we reject the null hypothesis.
    B. The p-value for Bartlett's test is not greater than 10%, so we conclude we cannot say the variances are equal.
    C. The p-value for Bartlett's test is not greater than 10%, so we fail to reject and conclude the variances are equal.
    D. The p-value for F-test is not greater than 10%, so we fail to reject and conclude the variances are equal.
    E. The p-value for F-test is not greater than 10%, so we fail to reject and conclude the means are equal.

20. How much of the total variation of *actual age* was explained by the difference in the sexes?

    A. 7.667/24.525
    B. 7.667/2.230
    C. 7.667/32.192
    D. 7.667/2.683
    E. 2.230/2.683

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