Psyc 3101, Homework 10

Name:____________________________ **TA:** _______________________

County	Population	Elevation	Longitude	Obama%	Prediction	Squared Error
Broomfield	57	5269	105.1	51.7	44.45	52.56
Gilpin	5	9141	105.5	56.7	49.44	52.71
Denver	620	5245	104.9	73.5	73.86	0.13
Lake	7	11580	106.3	60.4	55.50	24.01
San Juan	1	10791	107.7	52.6	56.49	15.13
Clear Creek	9	9676	105.7	54.2	51.01	10.18
Ouray	4	10236	107.8	51.5	55.85	18.92
Sedgwick	2	3802	102.4	31.3	33.63	5.43
Teller	23	9150	105.2	32.1	49.83	314.35
Summit	28	10959	106.1	61.0	55.11	34.69

Above is a table of presidential election results for Colorado counties. I ran a regression to find out whether population (in 1000s), elevation (in feet), and longitude predict anything about the percentage of people voting for Obama. My input and result in R were

```
> lm(obamaPercent ~ population + elevation + longitude)

Coefficients:
(Intercept)      Population      Elevation    Longitude
 -173            .053            .0018        1.95
```

1. Write the regression equation, using the numbers above.
Yhat = -173 + .053·Population + .0018·Elevation + 1.95·Longitude

2. Fill in the columns for prediction and squared error.

3. Find the residual sum of squares.
sum(Squared Error) = 528.11

4. Find the total sum of squares of the outcome.
Mean(Obama%) = 52.5
Sum $(\text{Obama\%} - 52.5)^2 = 1463.44$

5. Find the sum of squares explained by the regression.
$SS_{regression} = 1463.44 - 528.11 = 935.33$

6. What proportion of the total variance does the regression explain?
$R^2 = 935.33/1463.44 = .64$

7. Convert the regression and residual sums of squares to mean squares. The degrees of freedom are $df_{regression} = m = 3$ and $df_{residual} = n - m - 1 = 6$.
$MS_{regression} = 935.33/3 = 311.78$
$MS_{residual} = 528.11/6 = 88.02$

8. Calculate the F statistic for deciding whether the regression explains meaningful variance in the outcome.
F = 311.78/88.02 = 3.54

9. The critical value (with α = 5%) is 4.76. What do you conclude? (Don't just write null or alternative.)
There is not enough evidence to conclude the regression explains meaningful variation in counties' voting.

10. The standard error for $b_{population}$ is .019. Calculate the t statistic for testing whether this regression coefficient is reliably different from zero.
t = .053/.019 = 2.79

11. The two-tailed critical value is 2.45. Write a sentence (about counties and voting, not about hypotheses) describing what you conclude.
Larger counties are reliably more in favor of Obama.

Here are data from an experiment comparing memory of people given lists of 20 items to remember. One group of subjects was given words, another was given faces, and a third was given pictures. Instead of giving you all the data, for each group I just give you the mean and a measure of variability in the sum of squares.

Group	Size	Mean	Sum of Squares
Words	20	9.5	205.7
Faces	24	12.1	231.2
Pictures	19	10.2	211.8

12. Find the grand mean. Hint: The sum of all the scores in each group equals the group mean times the number of people in that group.
sum(X) = sum(word group) + sum(faces group) + sum(pictures group) = 20·9.5 + 24·12.1 + 19·10.2 = 674.2
mean(X) = sum(X)/n = 674.2/63 = 10.70

13. Find the sum of squares for the differences among groups ($SS_{treatment}$).
$SS_{treatment} = sum(n_{group}(M\text{-}Mbar)^2) = 20 \cdot 1.2^2 + 24 \cdot 1.4^2 + 19 \cdot .5^2 = 80.59$

14. Find the residual sum of squares.
$SS_{residual}$ = SS(word group) + SS(faces group) + SS(pictures group) = 205.7 + 231.2 + 211.8 = 648.7

15. Convert both sums of squares to mean squares. The degrees of freedom are $df_{treatment} = k - 1 = 2$ and $df_{residual} = \sum n_i - k = 60$.
$MS_{treatment} = 80.59/2 = 40.30$
$MS_{residual} = 648.7/60 = 10.81$

16. Calculate the F statistic for deciding whether the group means differ.
F = 40.30/10.81 = 3.73

17. The critical value (with $\alpha = 5\%$) is 3.15. What do you conclude? (Don't just write null or alternative.)
Average memory reliably differs depending on whether the material is words, faces, or pictures.