Math 213 ‡ Practice Test #3 ANSWERS ‡ Fall 2021

For the curious, data for this practice test are on Canvas in the file `practice3.RData`.

(1) Ebay sellers wonder if the type of photo posted with an item affects the selling price of that item. One hundred and forty three MarioKart packages were analyzed, which were classified as having a "stock" photo or not. A t distribution is used to compute a 95% confidence interval for the average difference in selling price between those without and with "stock" photos ($\mu_{no} - \mu_{yes}$) is (-\$7.20, -\$1.14). Which of the following are correct interpretations of this interval?

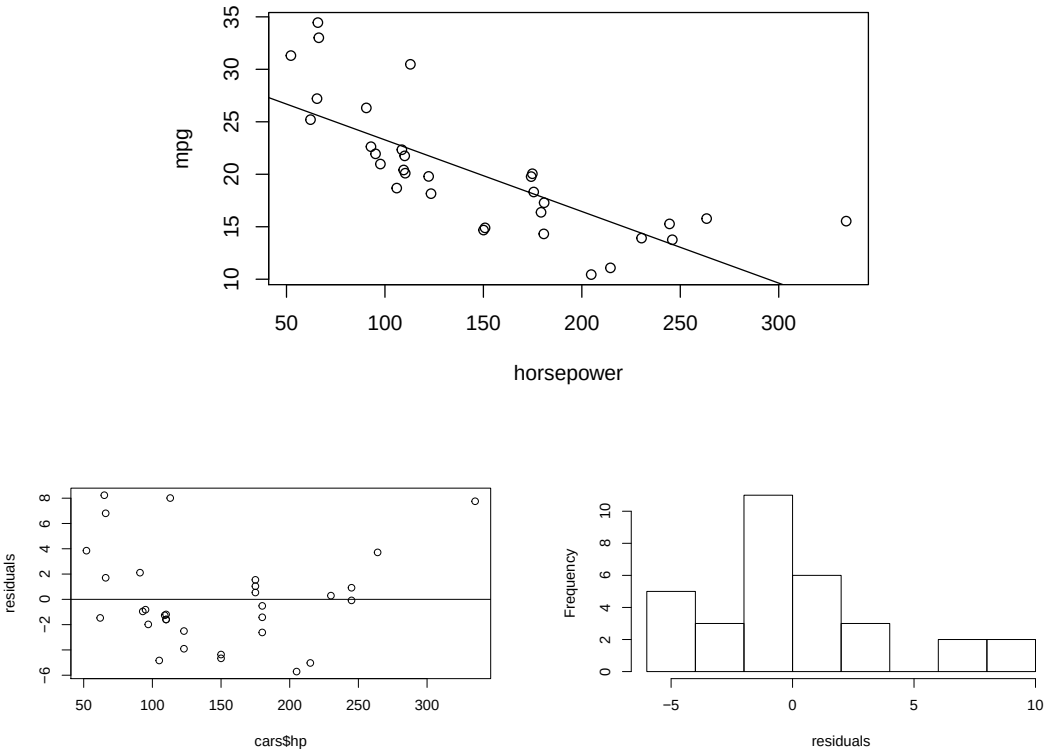
(a) There is no evidence that photo type is associated selling price.
(b) We have evidence that packages with stock photos sell, on average, for more than packages without stock photos.
(c) We have evidence that packages with stock photos sell, on average, for less than packages without stock photos.
(d) In general, the average selling price of the MarioKart packages is less than \$10.
(e) More than one statement is correct.

(2) A researcher plans to study levels of Seasonal Affective Disorder in the Gunnison Valley. He recruits 30 people for his study and collects survey data from each of them twice, once in summer and once in winter. From each survey, he computes a happiness index, which is a number on a continuous scale between 0 and 1, with 0 indicating severe unhappiness and 1 indicating complete happiness. Which of the following procedures would be most appropriate for analyzing the relationship between season and happiness index?

(a) Analysis of variance
(b) Linear regression
(c) Two sample t test
(d) Paired data t test
(e) None of the above are appropriate.

(3) Road tests of 32 cars[1] produced data about the relationship between horsepower and gas mileage. The output of a linear model is shown. A scatterplot with the fitted line from the linear model, as well as a histogram of the residuals and scatterplot of the residuals compared to the input variable.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  <0.0001 ***
cars$hp     -0.06823    0.01012  -6.742  <0.0001 ***
```



(a) What was the null hypotheses tested for the `cars$hp` coefficient? The null hypothesis is that the slope is zero, that the variables are independant and that knowing horsepower will not help predict mpg.

(b) What can you conclude about the relationship between horsepower and gas mileage? There appears to be a negative correlation. As horespower goes up, mpg goes down.

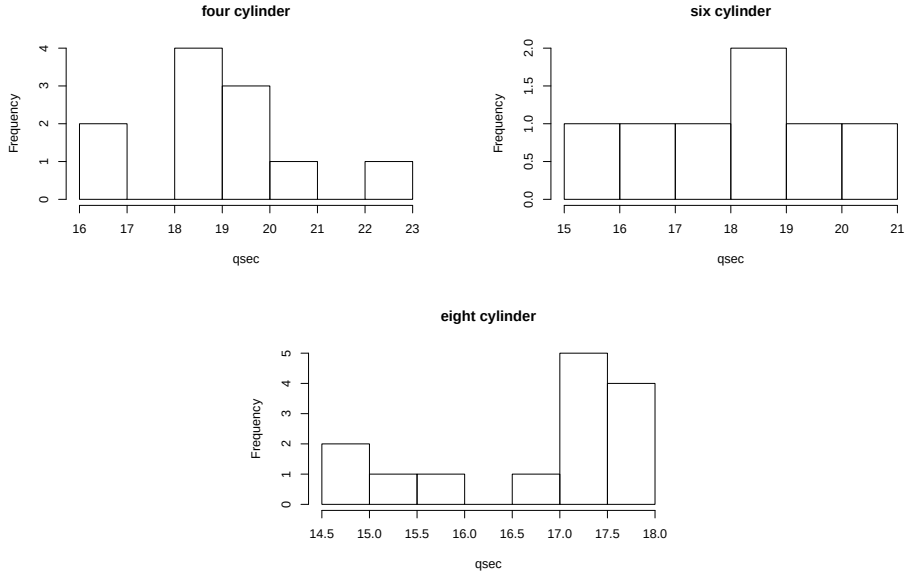
(c) Do the assumptions that underlie linear regression appear satisfied by this model? Explain. The histogram does not show significant departure from normality, so the assumption is valid. There appear to be more positive residuals at the low

[1]These data were published by Motor Trend magazine in 1974, analysed in *Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391-411* and are included in R as part of the `mtcars` dataset.

and high ends of the horsepower scale, and more negative residuals in the middle, so the assumption of a linear relationship might be questioned.

(4) Data are in `cars`. The same data set also contains the number of cylinders for each car and the time in seconds for each car to travel a quarter mile, starting from a stop, in seconds. Here are relevant data and graphics for analyzing the relationship between quarter mile times and number of cylinders:

cylinders	mean	sd	n
4	19.1	1.7	11
6	18.0	1.7	7
8	16.8	1.2	14



(a) One of the conditions for ANOVA appears questionable for these data. What is this condition, and why is it in question? Distributions in each group should be roughly normal, but the distribution for the 8 cylinder engines looks like it might have some skew to the left, or perhaps even bimodality.

(b) ANOVA is somewhat robust in regards this condition. Here is the result of running ANOVA for quarter mile times in terms of cylinders.

```
            Df Sum Sq Mean Sq F value   Pr(>F)
cars$cyl     2 34.606 17.3029  7.7938 0.001955 **
Residuals 29 64.382  2.2201
```

At the $\alpha = 0.05$ significance level, does this indicate a difference between quarter mile times for cars with different numbers of cylinders? Yes, $p = 0.001955 < 0.05$.

(c) A significant result from ANOVA leads to testing each pair of groups to determine which differences are significant. The following output lists the comparison performed, the expected difference in group means, a 95% confidence interval for those differences, and a p value for a test of no difference.

```
comparison difference lower     upper      p adj
6-4        -1.16013   -2.939271  0.6190113  0.2574564
8-4        -2.36513   -3.847748 -0.8825122  0.0013300
8-6        -1.20500   -2.908398  0.4983980  0.2053353
```

According to this output, is there a difference between quarter mile times between cars with 6 cylinder and 4 cylinder engines? No. Between 8 cylinder and 4 cylinder engines? Yes. Between 8 cylinder and 6 cylinder engines? No.

(5) In a study on Monarch butterflies, the relationship between wing weight in milligrams and wing area in square millimeters is plotted. A scatterplot of the data along with a linear regression line are shown below. This line has equation $y = 395.01 + 38.23x$ and correlation coefficient $R = 0.802$

(a) What is the biological meaning of the intercept 395.01? Do you trust this? Why or why not?
A butterfly wing weighing 0 mg would have an area of 395.01 mm^2. This is an extrapolation to values far from the observed data, and a wing with aero weight is impossible, so on its own this number is not useful.

(b) What is the biological meaning of the slope 38.23? Do you trust this? Why or why not?
A one mg section of butterfly wing on average has an area of 38.23 mm^2. This is a reasonable number, the line appears to fit the data, so this coefficient can be trusted.

(c) What information does the correlation coefficient 0.802 provide? There is a moderately strong correlation between wing weight and wing area.

(d) What do you expect that the area of a wing weighing 8 milligrams to be?
$395.01 + 38.23 \times 8 = 700.85$.