

Stats 314, Data Analysis #4

Cody Malick
malickc@oregonstate.edu

November 14, 2016

Part I

a

A total of 256 students were sampled. While 45 were female, 211 were male. Based off this data from the graph, there is evidence that less than 25% of students identify as female.

b

$$p = .25$$

$$p < .25$$

c

I think that it is an accurate measure as it is an engineers only class. For that reason, we could take it as a sample of the engineer population.

d

It is NOT a random sample.

$$n * \hat{p} \geq 10 \text{ and } n * (1 - \hat{p}) \geq 10$$

$$256 * .1757 = 44.979$$

$$256 * (1 - .1757) = 211.0208$$

It is of sufficient size.

e

The test statistic uses the following formula:

$$zstat = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Plugging in the appropriate values:

$$zstat = \frac{.1757 - .25}{\sqrt{\frac{.25(1-.25)}{256}}} = -2.7454$$

f

Lower one sided p-stat, equal to .003022

g

The confidence interval formula for a one proportion z test is:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.1757 \pm 1.96 * \sqrt{\frac{.1757(1-.1757)}{256}}$$

$$= .2223 \text{ and } .1291$$

95% CI from .1291 to .2223

h

There is moderate evidence that the proportion of students who identify as female is less than .25. The sample estimates the average proportion of females to be .1767, with a 95% confidence interval from .1291 to .2223. The null hypothesis is rejected at significance value .05. The university should receive funding to actively market to women to get an engineering or computer science degree.

Part II

a

The parameter of interest is the proportion of "guzzlers" models in American and international car companies.

b

$$p = 0$$

$$p \neq 0$$

c

Conditions:

Random Sample: No, the EPA site says: "The test data used to determine fuel economy estimates is derived from vehicle testing done at EPA's National Vehicle and Fuel Emissions Laboratory in Ann Arbor, Michigan, and by vehicle manufacturers who submit their own test data to EPA.

Manufactures submit their own voluntary data.

Number of successes and failures in each group must be greater than 10: From our data set, only 7 of the 267 vehicles from America are considered guzzlers. This does not pass the required level. International does meet the requirement with 48 success, and plenty more failure.

d

Test statistic:

$$zstat = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\hat{p} = \frac{7+48}{267+581} = \frac{55}{848} = .064858$$

$$\hat{p}_1 = \frac{7}{267} = .0262$$

$$\hat{p}_2 = \frac{48}{581} = .0826$$

$$z_{stat} = \frac{.0262 - .0826}{\sqrt{.06485(1 - .06485)(\frac{1}{267} + \frac{1}{581})}} = -3.0976$$

e

$$p - value = 0.000975$$

f

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

$$CI = (.0262 - .0826) \pm 1.645 * \sqrt{\frac{.0262(1 - .0262)}{267} + \frac{.0826(1 - .0826)}{581}}$$

$$= -.03167 \text{ and } -.08112$$

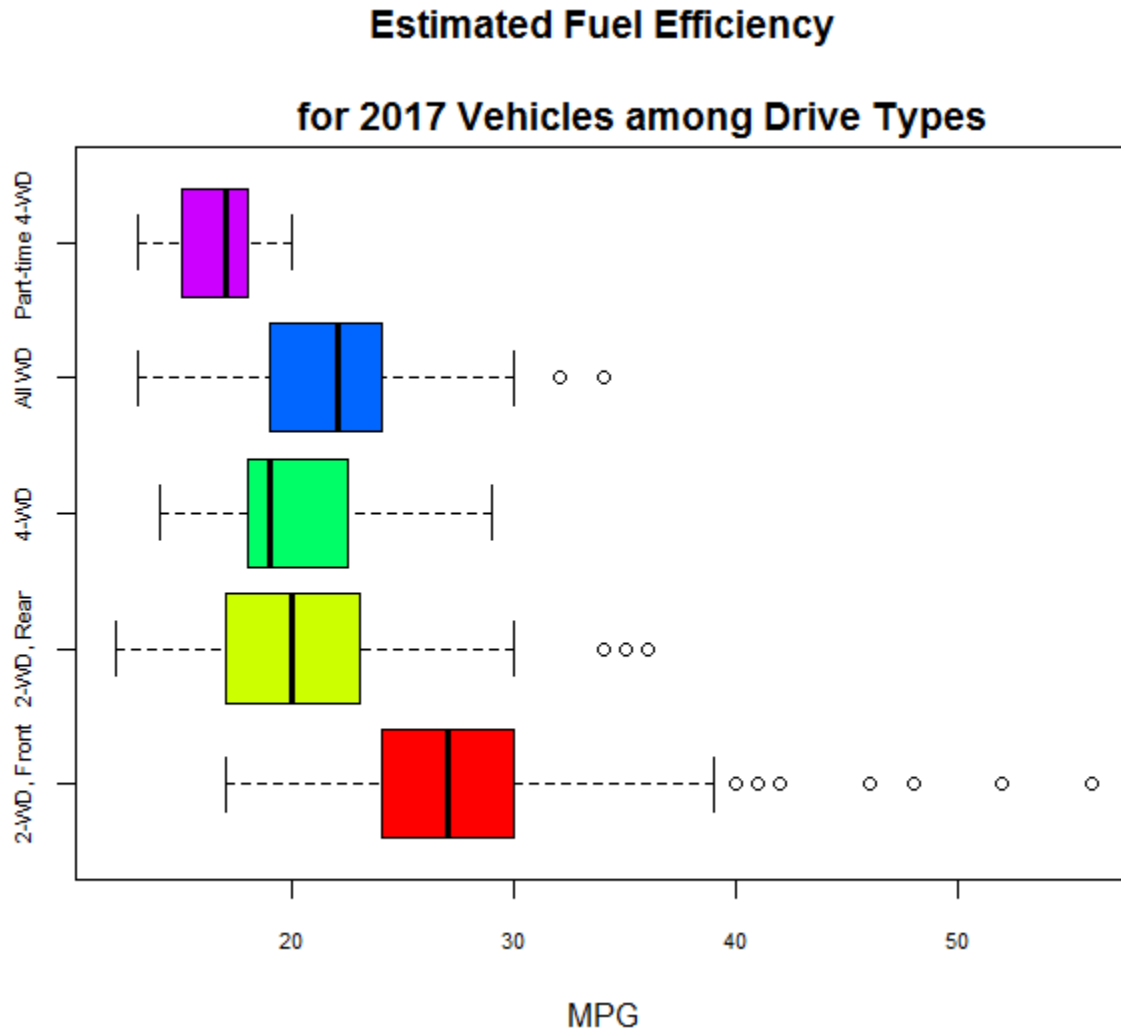
90% CI from -.08112 to -.03167

g

There is slightly convincing evidence that the proportion of gas guzzling models in American made cars is different than international models. The 90% confidence interval estimates the proportion of American models that are gas guzzlers is -.08112 to .03167 less than international model cars with a point estimate of .06485 less. The null hypothesis is rejected at a significance level of .10 (Z=-3.0976, two-sided p-value = .000975).

Part III

a



There does seem to be a difference in the average combined fuel efficiency. We can see by the different inter-quartile ranges that the average does vary from drive type to drive type.

b

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_a : at least two μ_i s differ

c

Conditions:

Samples are obtained using random mechanism: As stated in part II, they are not random samples.

Populations are independent: Populations are independent

Populations are normal: Based off the visual data, the distributions are not all normally distributed.

Population standard deviations are the same: Not all the standard deviations are the same.

d

a

	Df	Sum Sq	F value	Pr(>F)	
fueldata\$Drive	4	9221	2305	115.4	2e-16 ***
Residuals	843	16845	20		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

b

The significance value is rejected at .05, as the f-value is very low, close to zero.

c

The evidence to reject is moderate at a .05 significance value.

e

a

```
> TukeyHSD(mod, conf.level = 0.95)
  Tukey multiple comparisons of means
  95% family-wise confidence level

Fit: aov(formula = fueldata$CombFE ~ fueldata$Drive)

$`fueldata$Drive`
              diff          lwr          upr
2-WD, Rear-2-WD, Front -7.3512697 -8.4372850 -6.2652544
4-WD-2-WD, Front      -7.6333246 -9.1453813 -6.1212678
All WD-2-WD, Front     -5.7571424 -6.8525887 -4.6616960
Part-time 4-WD-2-WD, Front -11.1140548 -14.1724433 -8.0556663
4-WD-2-WD, Rear       -0.2820549 -1.8071328  1.2430230
All WD-2-WD, Rear      1.5941273  0.4807766  2.7074779
Part-time 4-WD-2-WD, Rear -3.7627851 -6.8276321 -0.6979381
All WD-4-WD           1.8761822  0.3443741  3.4079903
Part-time 4-WD-4-WD    -3.4807302 -6.7211274 -0.2403331
Part-time 4-WD-All WD  -5.3569124 -8.4251139 -2.2887108

              p adj
2-WD, Rear-2-WD, Front 0.0000000
4-WD-2-WD, Front      0.0000000
All WD-2-WD, Front    0.0000000
Part-time 4-WD-2-WD, Front 0.0000000
4-WD-2-WD, Rear       0.9868227
All WD-2-WD, Rear     0.0009288
Part-time 4-WD-2-WD, Rear 0.0073573
All WD-4-WD           0.0075598
Part-time 4-WD-4-WD    0.0280918
Part-time 4-WD-All WD  0.0000211
```

b

```
Part-time 4-WD-2-WD, Front -11.1140548 -14.1724433 -8.0556663
2-WD, Rear-2-WD, Front -7.3512697 -8.4372850 -6.2652544
4-WD-2-WD, Front -7.6333246 -9.1453813 -6.1212678
Part-time 4-WD-All WD -5.3569124 -8.4251139 -2.2887108
```

Part-time 4-WD-2-WD has the largest difference in MPG. 8.055 MPG difference.

f

Two Seaters Minicompact Cars
Minicompact Cars Standard SUV 4WD
Standard SUV 4WD Standard SUV 4WD
Standard SUV 4WD Standard Pick-up Trucks 4WD
Standard Pick-up Trucks 4WD Small Pick-up Trucks 4WD
Small Pick-up Trucks 4WD Small Pick-up Trucks 4WD
Small Pick-up Trucks 4WD Small Pick-up Trucks 4WD
Standard Pick-up Trucks 4WD Standard Pick-up Trucks 4WD
Standard Pick-up Trucks 4WD

This make great sense for fuel efficiency as 4-WD tend to have lower MPG.