

ASSIGNMENT: TWO-WAY ANOVA WITH INTERACTION AND COVARIATES

1. TWO-WAY ANOVA WITH INTERACTION AND COVARIATES

1.1. **overview.** The belief that overall a manual transmission is better for gas mileage than an automatic is widely held. In this assignment we will determine whether the EPA mileage data supports that claim, and if it does, produce an estimate of the actual difference.

We will consider only cars (`car.truck='C'`) and city driving (`C.H='C'`) because city driving requires many more gear changes than highway driving.

We will select four types of transmissions for which there are enough measurements in the EPA data:

- Five speed manual (`trns='M5'`)
- Five speed automatic (`trns='L5'`)
- Six speed manual (`trns='M6'`)
- Six speed automatic (`trns='L6'`)

We will create two factors, one (`auto.manual`) to indicate whether the transmission is automatic or manual, and another (`trns.gears`) to indicate whether it is a 5 or 6 speed.

Because the six speed transmissions tend to occur in vehicles with more horsepower, and horsepower has a negative effect on mileage, we will include rated horsepower (`rhp`) as a covariate.

This will allow us to interpret the estimated differences due to the factors (if any) as being between cars with the same horsepower rating.

1.2. **R setup.** The R statements to download the EPA data and create the necessary variables is posted as a text file on the website.

You can include R programming statements from a URL using the following statement:

```
source('url')
```

The URL for the setup code for this problem is::

http://www.sandquinn.org/stonehill/MA225/Spring2010/R_takehome1_setup.txt

Copy this URL and paste it into the following statement between the adjacent quotes:

```
]vspace0.3 cm source'
```

If all goes well, this will:

- download the EPA test data to a data frame called `epa`
- Create a subset data frame called `trnsdata2` which:
 - contains only city mileage data for cars
 - contains only records with transmission codes L5, L6, M5, and M6
 - has a new variable called `auto.manual` which is L for an automatic, M for a manual
 - has a new variable called `trns.gears` which is 5 or 6 depending on the number of speeds
- attach the new `trnsdata2` data frame

You can check the results by listing the structure of the `trnsdata2` data frame with the command:

```
str(trnsdata2)
```

1.3. Analysis. Once the `trnsdata2` data frame is built, use it to perform a 2-factor ANOVA with interaction (see the lecture foils) predicting `mpg` from `auto.manual`, `trns.gears`, and the interaction of these two factors.

You should use the `aov()` procedure and perform Tukey's Honest Significant Difference test with `drop1()` (see the lecture foils for examples).

Note that when we entered `str(trnsdata2)`, `auto.manual` and `trns.gears` were listed as type `chr` or character. Since some R procedures require a factor, we will sometimes have to tell R to treat them as factors. The mechanism for this is called a *cast*. To cast `auto.manual` to a factor, the command is `as.factor(auto.manual)` Use this in place of `auto.manual` if a factor is required.

For example, the `TukeyHSD` procedure requires factors and will give an error unless `auto.manual` and `trns.gears` are cast to factors when the linear model is run. So instead of

ASSIGNMENT: TWO-WAY ANOVA WITH INTERACTION AND COVARIATES 3

```
lm0<-aov(mpg ~ auto.manual*trns.gears)
```

use

```
lm0<-aov(mpg ~ as.factor(auto.manual)*as.factor(trns.gears))
```

To get the estimated difference between manual and automatic, rerun the linear model with `lm()` instead of `aov()` and use the `summary()` command to view the results.

Finally, repeat the analysis incorporating rated horsepower (`rhp`) as a covariate to adjust for differences in average horsepower among the categories.

Use the results of the analyses to answer the questions on the "Final Takehome1" eLearn assessment.

There are examples in the R cribsheet posted on the website that should copy and paste correctly. If not, you should be able to download the cribsheet source, which is a text file, and copy from your local version.

Once you have pasted them in, change the variable names and, where necessary, add the `as.factor()` to cast character or numerical values to factors.