*# Dr. M. Baron, Statistical Machine Learning class, STAT-427/627*

# # Partial F-tests and Lack-of-Fit Tests

**# PARTIAL F-TEST.**
**# Consider full and reduced models**

```
> load("Auto.rda")
> attach(Auto)

> reg_full = lm(mpg ~ year + acceleration + horsepower + weight)
```

**# How to test significance of year and acceleration?**

```
> reg_reduced = lm(mpg ~ horsepower + weight)

> anova( reg_full, reg_reduced )

Analysis of Variance Table

Model 1: mpg ~ year + acceleration + horsepower + weight
Model 2: mpg ~ horsepower + weight
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    387 4558.0
2    389 6993.8 -2   -2435.8 103.41 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**# The p-value comparing these two models is very significant, so the two variables make a significant contribution for the prediction of mpg, in addition of weight and horsepower.**

**# LACK-OF-FIT.**
**# Here we test linearity by comparing the linear model (reduced) with the model with dummy variables, one for each value of X (full model that does not assume linearity).**

```
> reg_reduced = lm(mpg ~ cylinders)
> reg_full = lm(mpg ~ as.factor(cylinders))

> anova( reg_full, reg_reduced )
Analysis of Variance Table

Model 1: mpg ~ as.factor(cylinders)
Model 2: mpg ~ cylinders
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    387 8544.5
2    390 9415.9 -3   -871.42 13.156 3.383e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**# Low p-value shows that the relation of mpg to the number of cylinders is non-linear.**

**# Continuous case – what to do if the X-variable has no repeated values?**

```
> round(horsepower/10)*10
  [1] 130 160 150 150 140 200 220 220 220 190 170 160 150 220 100 100 100
80
 [19]  90  50  90  90 100 110  90 220 200 210 190  90  90 100 100 100 100
90
     < truncated >

> reg_reduced = lm(mpg ~ horsepower)
> hp_rounded = round(horsepower/10)*10
> reg_full = lm( mpg ~ as.factor(hp_rounded) )

> anova( reg_full, reg_reduced )
Analysis of Variance Table

Model 1: mpg ~ as.factor(hp_rounded)
Model 2: mpg ~ horsepower
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    373 7101.9
2    390 9385.9 -17     -2284 7.0565 6.662e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**# The full model is significantly better. So, mpg is a non-linear function of horsepower.**