

# Analysis of Variance (ANOVA) with R

Steve UHLIG

# Example

- What aspects matter for BGP paths geographic length?
- Approach:
  - perform full-factorial simulations
  - Analyze the sensitivity of path choices to routing parameters

# One-way analysis of variance

- R command:  
> aov.length = aov(depend~indep,data)
- Parameters:
  - Dependent variable, e.g., path length
  - Independent variable(s), e.g., routing policies, AS size, iBGP mesh, IGP costs,...
  - Data structure
- Summary of the anova analysis:  
> summary(aov.length)
- Table of means:  
> print(model.tables(aov.length,"means"),digits=3)

# One-way analysis of variance

```
> data.geo = read.table("anova.geo.best")
```

```
> aov.geo = aov(Y~POL,data.geo)
```

```
> summary(aov.geo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
POL	1	7.9350	7.9350	9235.2	< 2.2e-16 ***
Residuals	2398	2.0604	0.0009		

---

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

- Routing policies affect geographic path length

# Multi-way analysis of variance

- In practice many parameters influence choice and length of BGP paths: policies, AS size, iBGP, IGP, peerings,...

```
> aov.geo =  
  aov(Y~(POL*ASSIZE*HIER*IGP*CONN),data.geo)
```

```
> summary(aov.geo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
POL	1	2.6279e+09	2627911820	78.8766	< 2.2e-16	***
ASSIZE	3	1.0815e+10	3604949017	108.2024	< 2.2e-16	***
HIER	1	2.7481e+09	2748059743	82.4829	< 2.2e-16	***
IGP	2	1.5599e+08	77994587	2.3410	0.096462	.
CONN	3	1.4506e+09	483542089	14.5135	2.299e-09	***
POL:ASSIZE	3	3.9356e+08	131186354	3.9376	0.008158	**
ASSIZE:CONN	9	8.2605e+08	91783497	2.7549	0.003324	**
Residuals	2280	7.5962e+10	33316733			

---

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