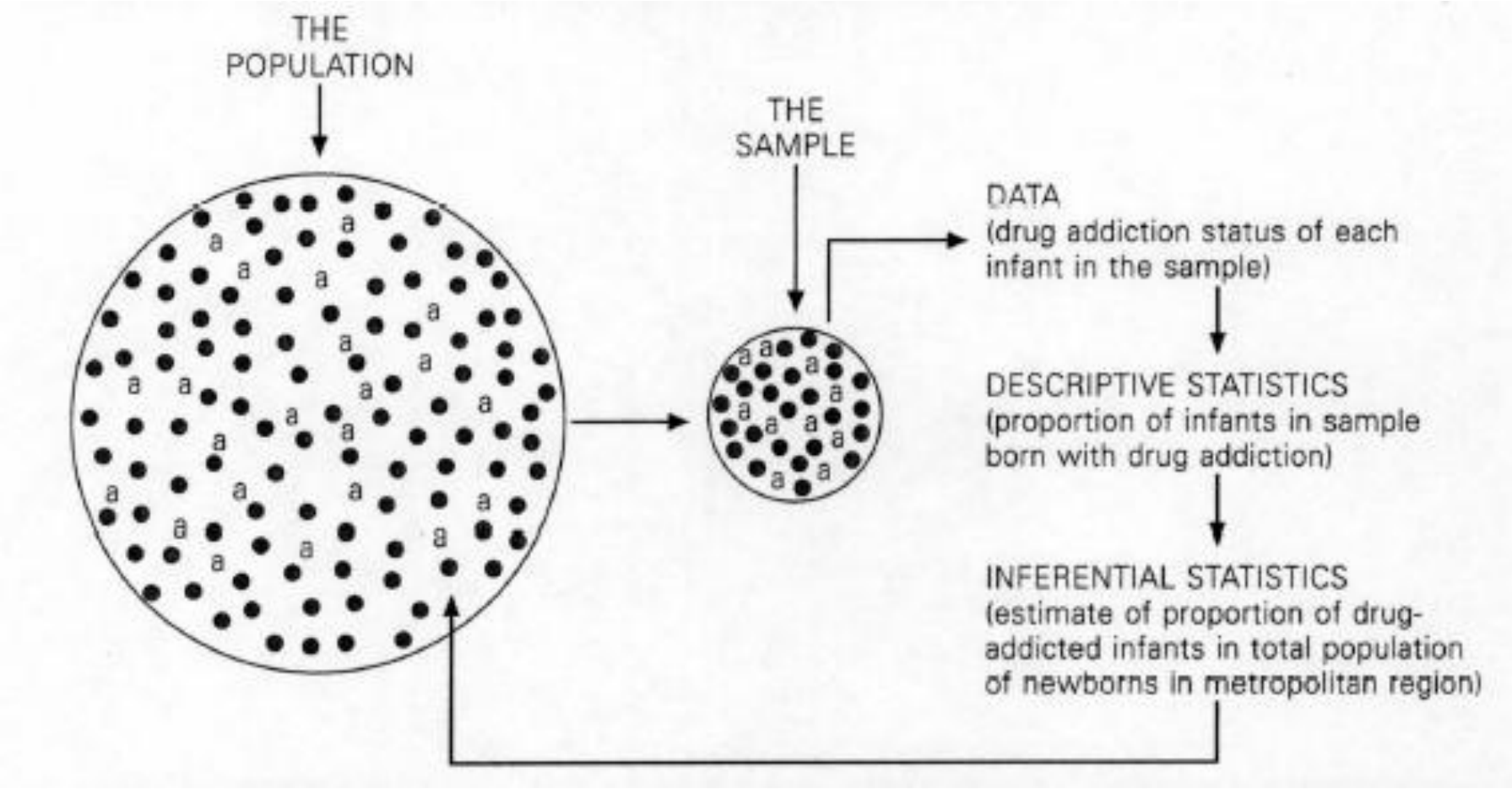
Inferential Statistics



A man walks into a bar…

- You are sitting in a bar and the man next to you says "I'll flip a coin. If it comes up heads, buy me a drink. If it comes up tails, I will buy you a drink."

- You agree…



The outcome

- H
- HH
- HHH
- HHHH
- HHHHH
- HHHHHH
- HHHHHHH
- HHHHHHHH

- P(H) = ½
- P(HH) = ½ x ½ = 0.25
- P(HHH) = $(½)^3$ = 0.125
- P(HHHH) = $(½)^4$ = 0.063
- P(HHHHH) = $(½)^5$ = 0.031 $p < .05$
- P(HHHHHH) = $(½)^6$ = 0.016
- P(HHHHHHH) = $(½)^7$ = 0.008 $p < .01$
- P(HHHHHHHH) = $(½)^8$ = 0.004

Are you being cheated?

A statistician's approach

- Null hypothesis – H_0: p(H) = ½
  the coin is fair (no cheating)
- Alternative hypothesis – H_1: p(H) ≠ ½
  the coin is not fair (cheating)

- Hypothesis test - if P(event) < .05 reject null hypothesis

Note: null hypothesis is needed to calculate probabilities

Short list of steps in Hypothesis testing

1. State your research question
2. Specify the null and alternative hypotheses (H_0 versus H_a)
3. Implement the appropriate test
4. Evaluate output and make decision: “reject” or “fail to reject” the null hypothesis based on p-value
5. Interpret the results in context

Interpreting a *P*-value

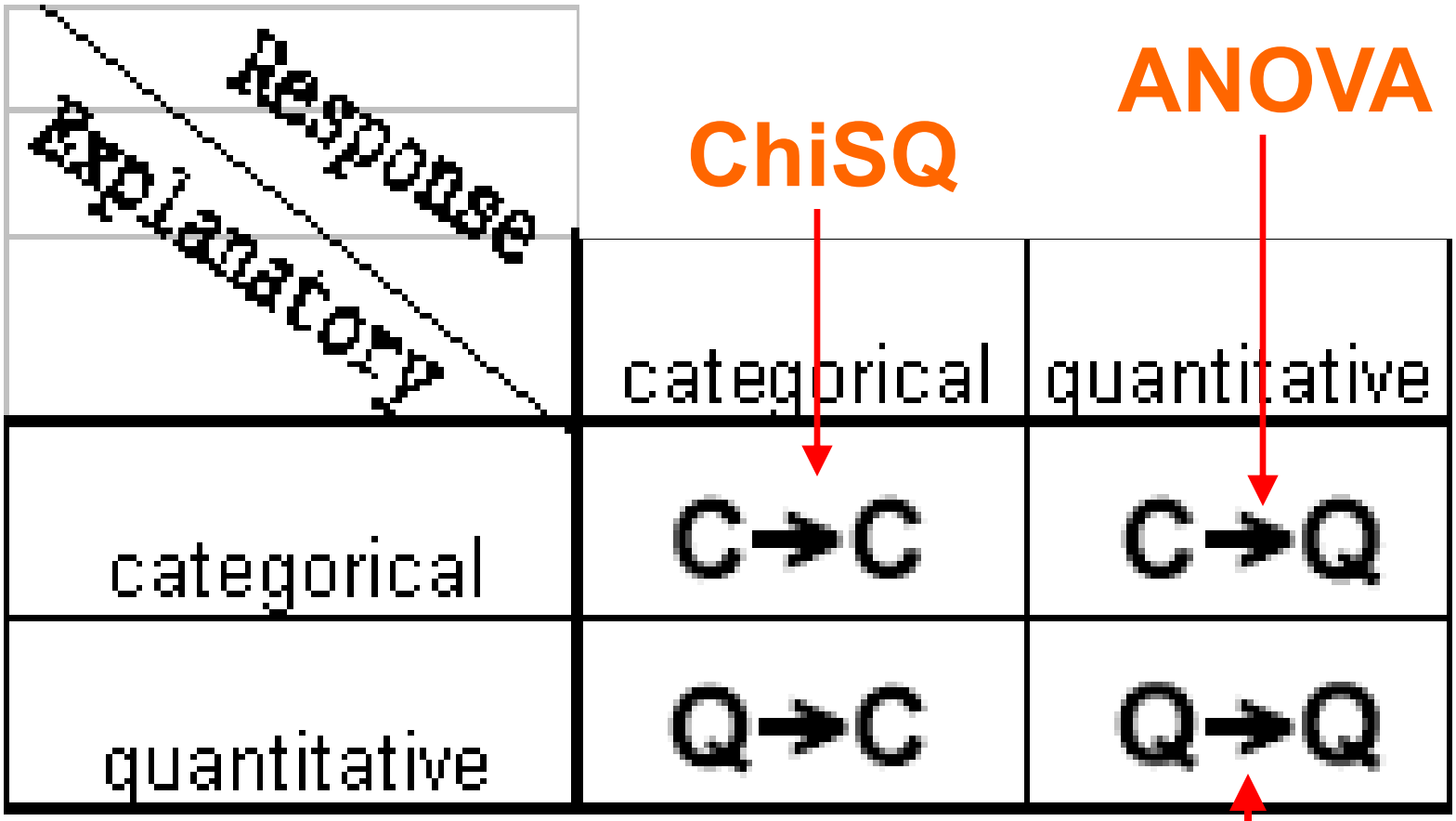
Could random variation alone account for the difference between the null hypothesis and observations from a sample?

- A small *P*-value (<0.05) implies that random variation (due to the sampling process alone) is not likely to account for the observed difference.

- With a small *P*-value (p<0.05), we **reject H_0**. The true mean or proportion of the population is **significantly** different from what was stated in H_0.

Thus small P-values are strong evidence AGAINST H_0.

Determine the appropriate test



Explanatory=Independent Variable
Response=Dependent Variable

Analysis of Variance

ANOVA

ANOVA

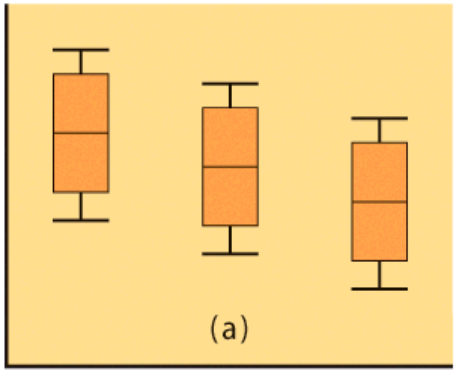
C → Q (comparing means, Q – DV, among two or more categories, C – IV)

The ANOVA F-test

The **analysis of variance F-test** compares the variation due to specific sources (levels of the factor) with the variation among individuals who should be similar (individuals in the same sample).

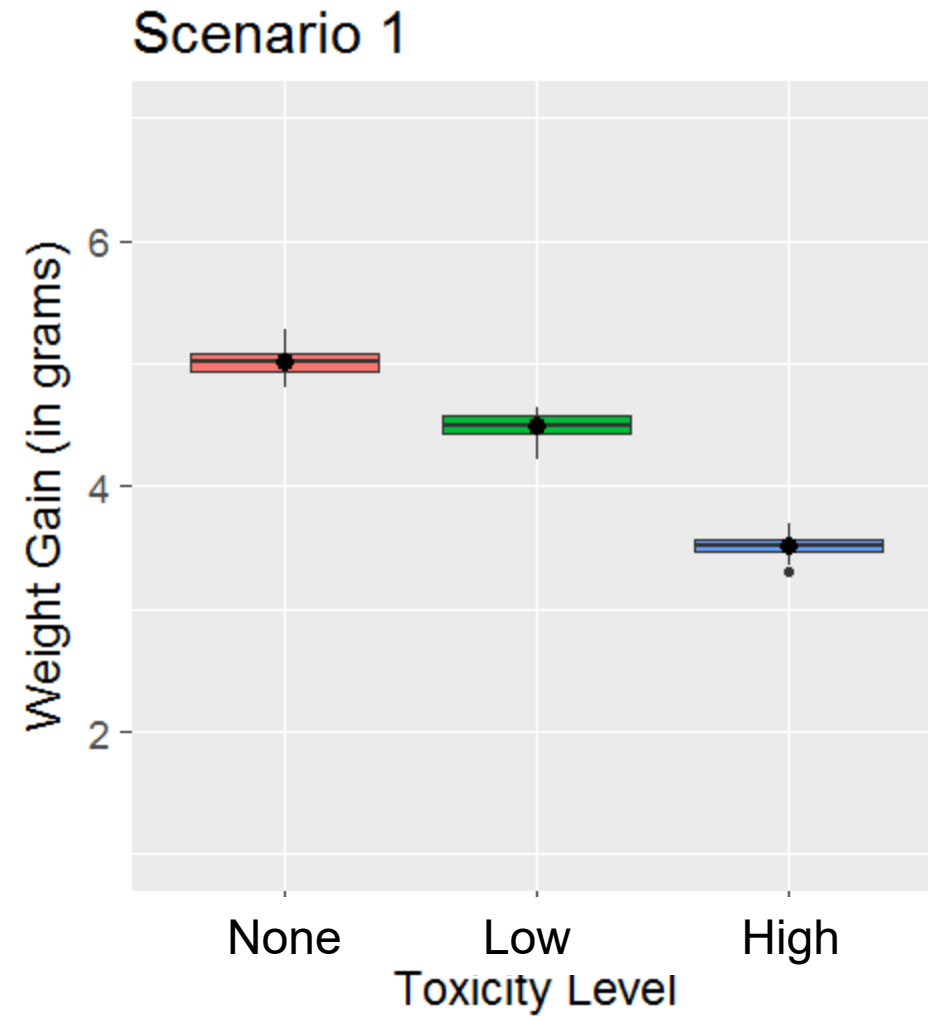
H_0: All means μ_i are equal.

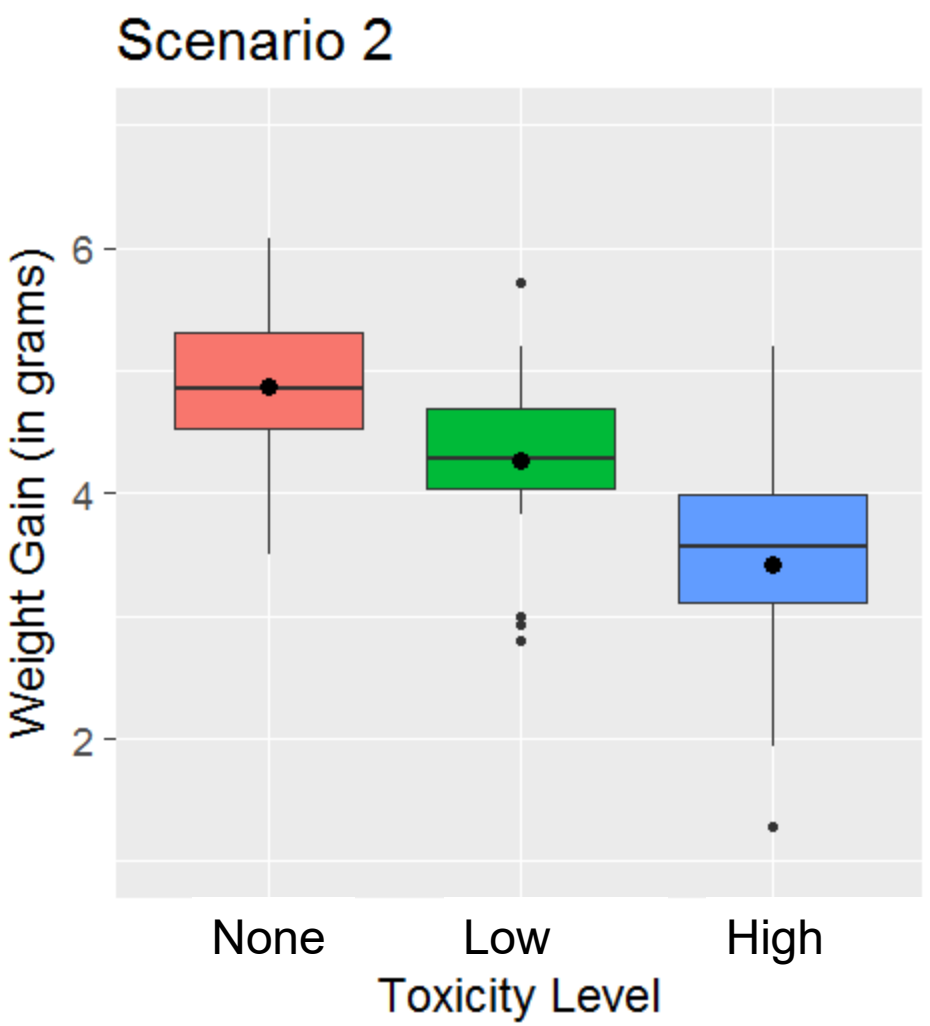
H_a: Not all means μ_i are equal.



The **analysis of variance F statistic** for comparing several means is:

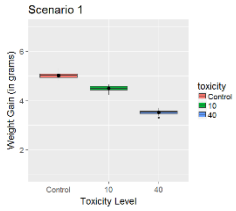
$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$


Scenario 1
Weight Gain (in grams)
6
4
2
None
Low
High
Toxicity Level


Scenario 2
Weight Gain (in grams)
6
4
2
None
Low
High
Toxicity Level

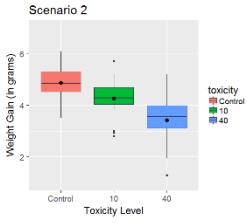
F = Variability Between Groups / Variability Within Groups

Scenario 1



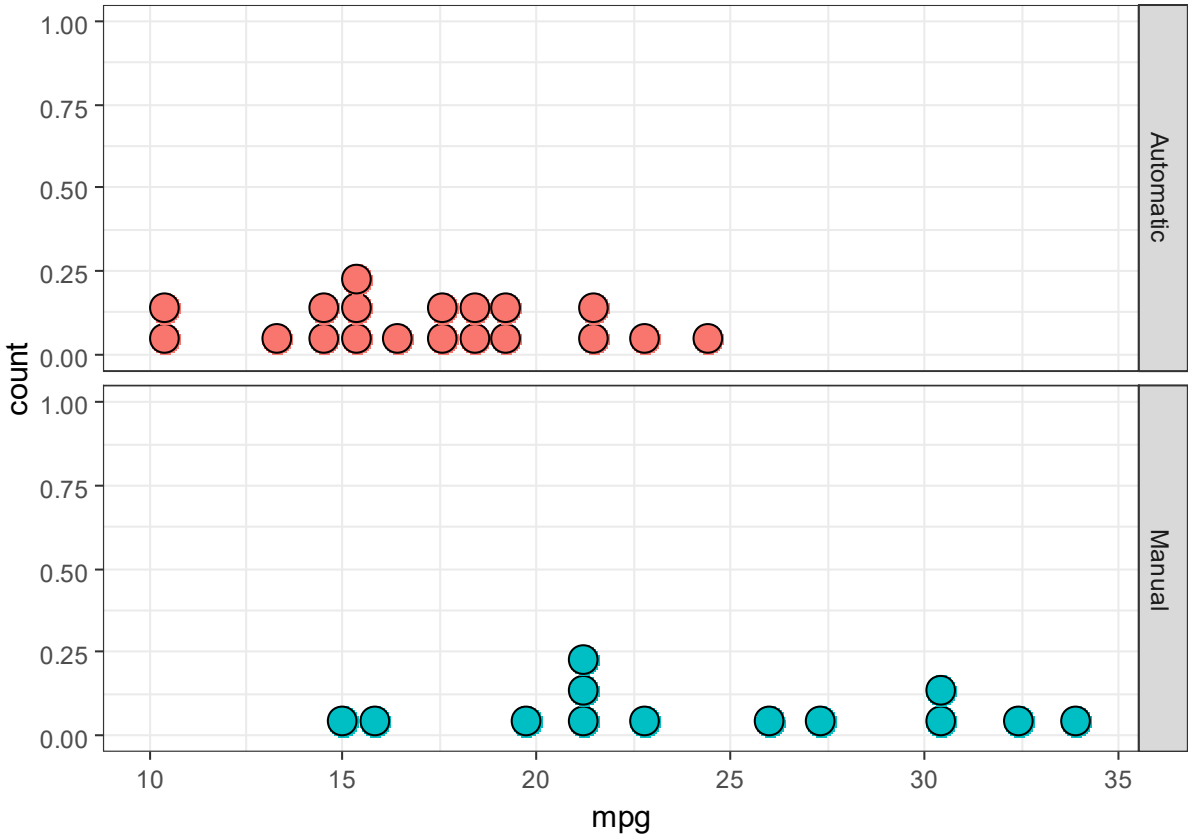
Source	Df	SS	MS	F	P-value
Toxicity	2	26.83	13.41	67.05	<0.001
Residuals	57	11.41	0.200		

Scenario 2



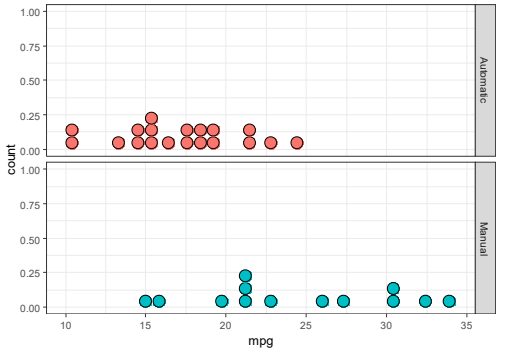
Source	Df	SS	MS	F	P-value
Toxicity	2	27.37	13.69	0.81	0.556
Residuals	57	966.07	16.94		

MPG vs. Transmission Type



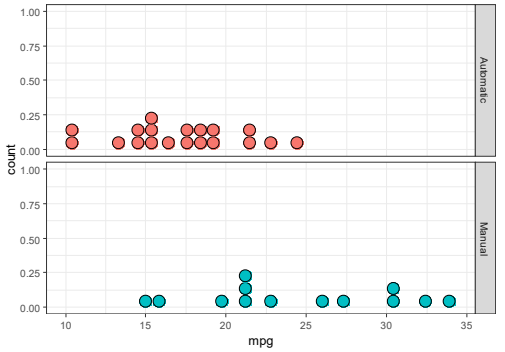
- $H_0: \mu_{Automatic} = \mu_{Manual}$
- $H_A: \mu_{Automatic} \neq \mu_{Manual}$

ANOVA Table



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
transmission	1	405.2	405.2	16.86	0.000285
Residuals	30	720.9	24.0		

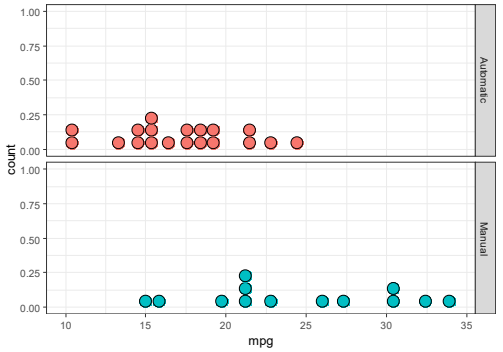
ANOVA Table



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
transmission	1	405.2	405.2	16.86	0.000285
Residuals	30	720.9	24.0		

F = variability between groups/variability within groups

ANOVA Table

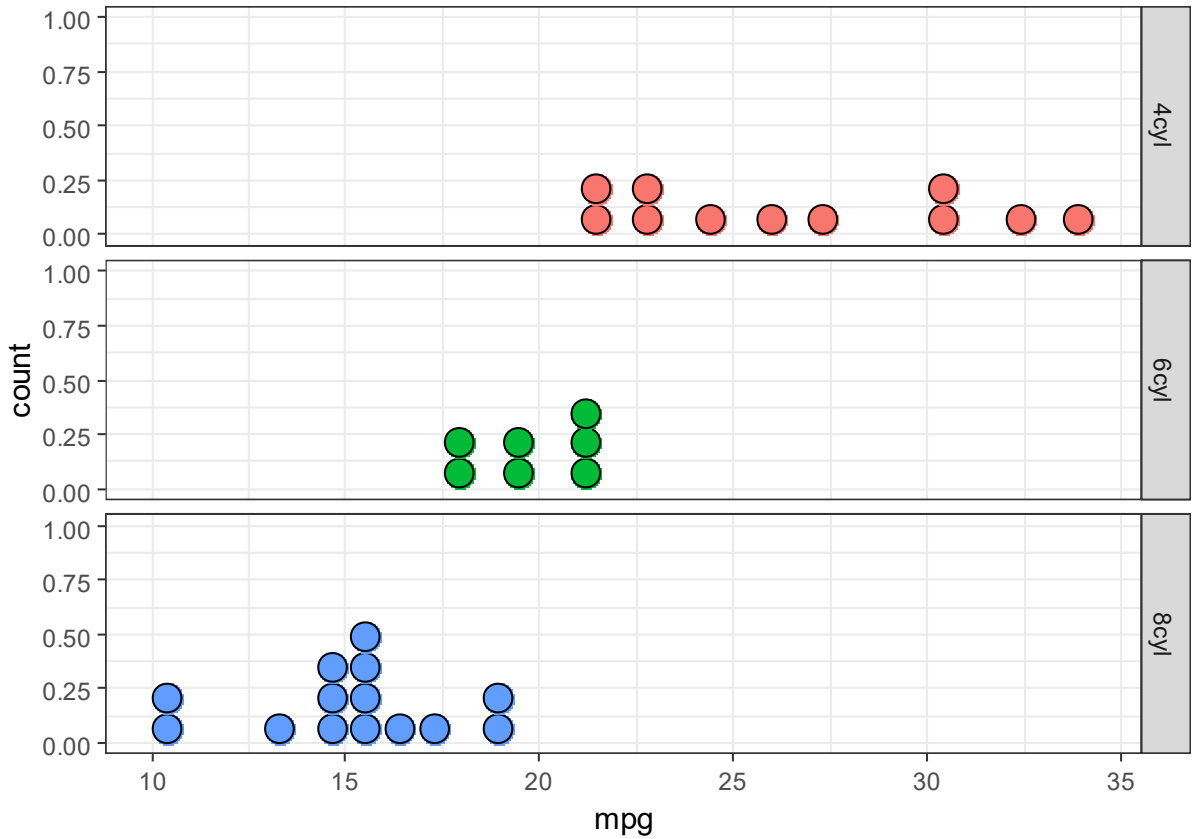


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
transmission	1	405.2	405.2	16.86	0.000285
Residuals	30	720.9	24.0		

- $H_0: \mu_{Automatic} = \mu_{Manual}$
- $H_A: \mu_{Automatic} \neq \mu_{Manual}$

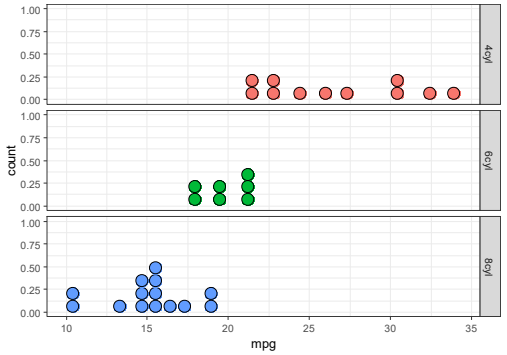
Probability of obtaining an F value this large or larger given H_0 is true.

MPG vs. Cylinder



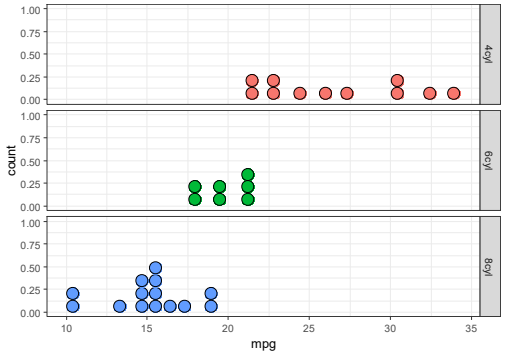
- H_0: $\mu_{4CYL} = \mu_{6CYL} = \mu_{8CYL}$
- H_A: At least one μ varies

ANOVA Table



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	2	824.8	412.4	39.7	4.98e-09
Residuals	29	301.3	10.4		

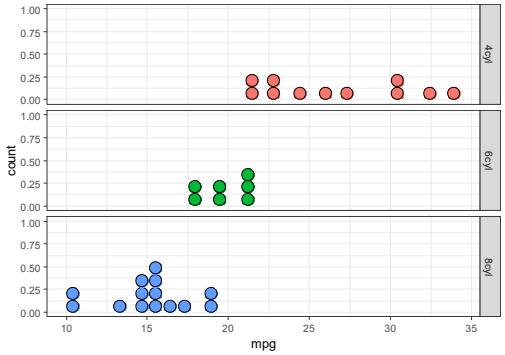
ANOVA Table



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	2	824.8	412.4	39.7	4.98e-09
Residuals	29	301.3	10.4		

F = variability between groups/variability within groups

ANOVA Table

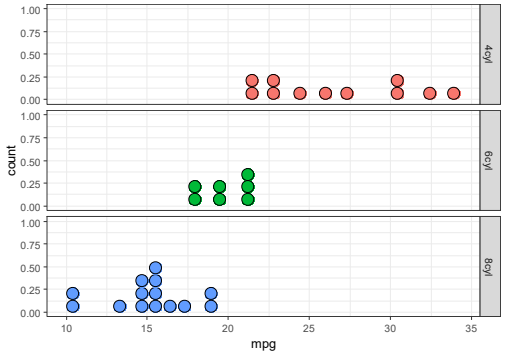


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	2	824.8	412.4	39.7	4.98e-09
Residuals	29	301.3	10.4		

- H_0: $\mu_{4CYL} = \mu_{6CYL} = \mu_{8CYL}$
- H_A: At least one μ varies

Probability of obtaining an F value this large or larger given H_0 is true.

ANOVA Table



- $H_0: \mu_{4CYL} = \mu_{6CYL}$ vs. $H_A: \mu_{4CYL} \neq \mu_{6CYL}$

- $H_0: \mu_{4CYL} = \mu_{8CYL}$ vs. $H_A: \mu_{4CYL} \neq \mu_{8CYL}$

- $H_0: \mu_{6CYL} = \mu_{8CYL}$ vs. $H_A: \mu_{6CYL} \neq \mu_{8CYL}$

Determining Error Rates

- Two Important Error Rates (Alpha):
  - Per-Comparison: Probability of Type I Error for each specific comparison
  - Family-Wise: Overall Type I Error probability
- Type I Error Rate for Multiple Tests:

$$\alpha_{FW} = 1 - (1 - \alpha_{PC})^{c}$$

where c = # of comparisons

Inflated Type I Error and Pairwise Comparisons

# Tests	Comparison α	Familywise α
1	.05	.05
3	.05	.14
6	.05	.26
10	.05	.40
15	.05	.54

Handling multiple comparisons statistically

•The **first** step in examining multiple populations or multiple treatments statistically is to test for an overall statistical significance as evidence of any difference among the parameters we want to compare.

➔ANOVA F-test

• If that overall test showed statistical significance, then a detailed follow-up analysis can legitimately examine all pair-wise parameter comparisons to define which parameters differ from which and by how much.

➔ *Post Hoc* tests (e.g. Bonferronni, Sidak).

There are two equivalent ways to solve the problem

- **Adjust the alpha level (keep p-values the same)**
  - Instead of comparing p-values to 0.05, we shrink alpha (Bonferonni does this by comparing all p-values against α/#of comparisons.)
- **Adjust the p-values (keep alpha the same)**
  - Instead of shrinking alpha, we inflate p-values (Bonferonni does this by inflating p-values by a factor of the number of comparisons)
- **Pay attention to how your software controls for the family-wise error in a particular test.**

Post Hoc Comparisons

	diff	lwr	upr	p adj
6cyl-4cyl	-6.920779	-10.769350	-3.0722086	0.0003424
8cyl-4cyl	-11.563636	-14.770779	-8.3564942	0.0000000
8cyl-6cyl	-4.642857	-8.327583	-0.9581313	0.0112287

Comparison of mpg by cyl
(Sidak)

Row Mean-Col Mean	4	6
6	-6.92078 0.000	
8	-11.5636 0.000	-4.64286 0.012

- $H_0: \mu_{4CYL} = \mu_{6CYL}$ vs. $H_A: \mu_{4CYL} \neq \mu_{6CYL}$
- $H_0: \mu_{4CYL} = \mu_{8CYL}$ vs. $H_A: \mu_{4CYL} \neq \mu_{8CYL}$
- $H_0: \mu_{6CYL} = \mu_{8CYL}$ vs. $H_A: \mu_{6CYL} \neq \mu_{8CYL}$