Module 15:
Statistics with multiple variables

Module 15: Learning Outcomes

- Use two-way ANOVAs and linear models to test for effects across multiple predictors
- Use a MANOVA or PERMANOVA in R to test for effects in multiple response variables
- Interpret model outputs and explain what single and interaction effects mean in two-way models

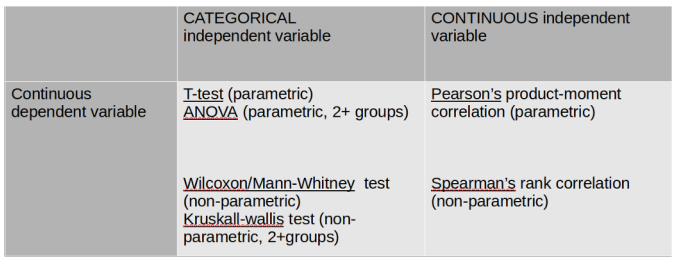
Previously...

YVAR ~ XVAR [One response, one predictor]

But what if you wanted to have more than one predictor? (e.g. effect of temperature and sampling transect)

or more than one response? (e.g. Hundreds of ASVs changing in response to temperature)

Statistics with multiple variables

<table>
<tr><td></td><td>ONE predictor variable</td><td>MULTIPLE predictor variables</td></tr>
<tr><td>ONE response variable</td><td> 
</td><td>2-way ANOVA (parametric)

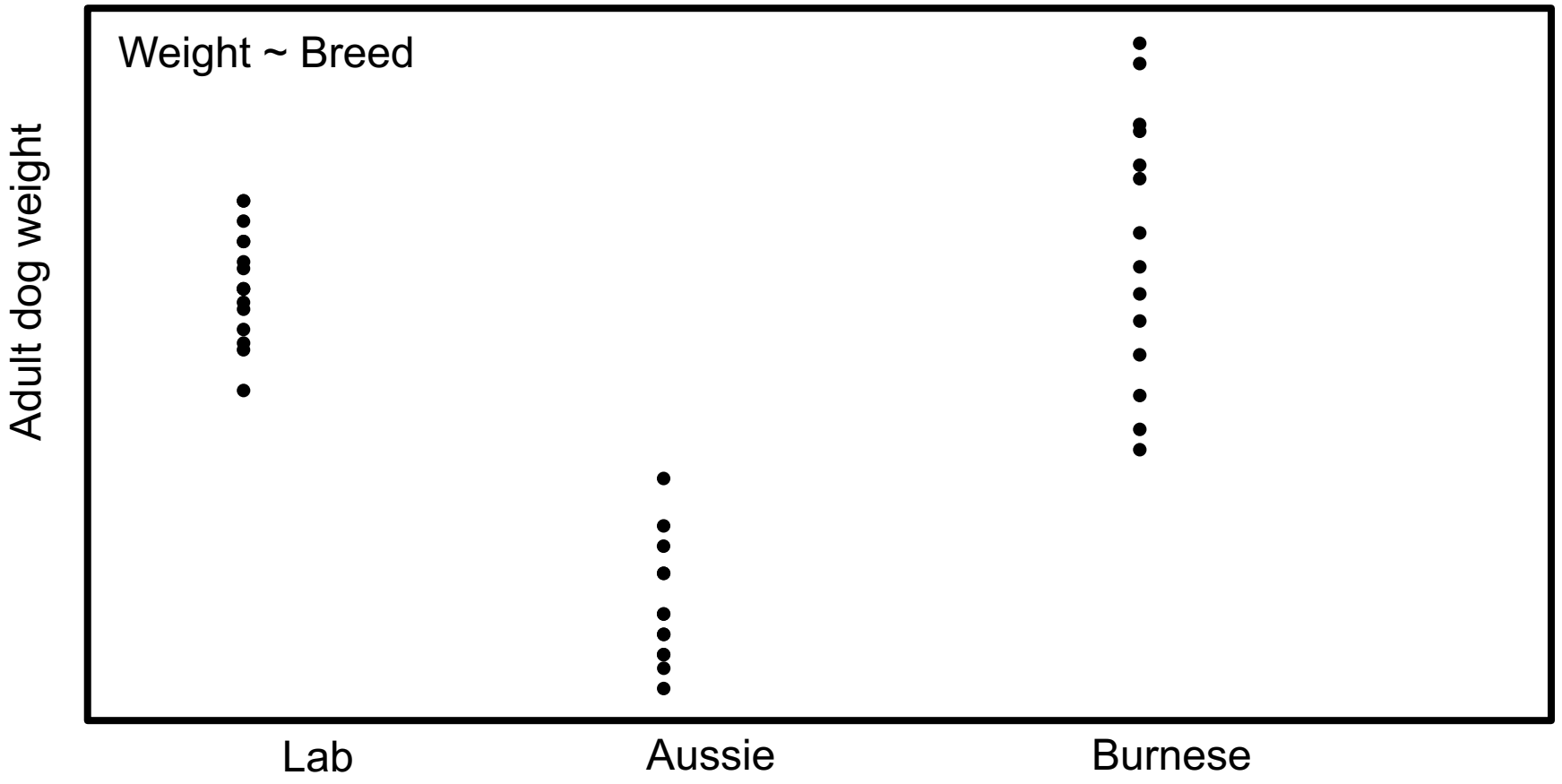
Linear models (parametric, but more flexible)</td></tr>
<tr><td>MULTIPLE response variables</td><td colspan="2">MANOVA (parametric)

PERMANOVA (non-parametric)</td></tr>
</table>

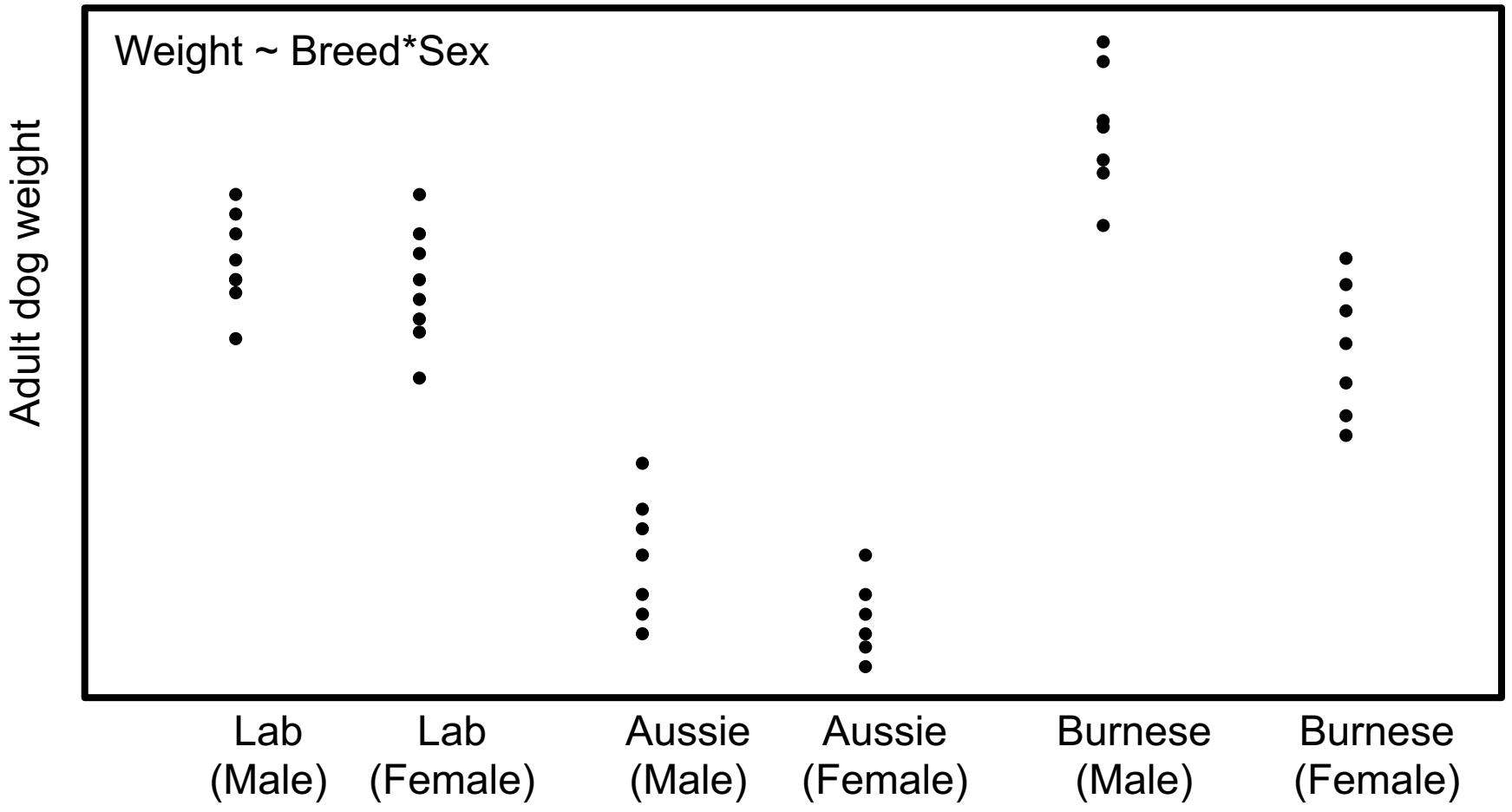
2-way ANOVA

- Like a normal (one-way) ANOVA, except you have two (categorical) predictor variables
- You can look at independent effects and interactions between effects
  - Y ~ X + Z (independent effects only)
  - Y ~ X:Z (interactive effects only)
  - Y ~ X*Z (independent and interactive effects)

(one-way) ANOVA

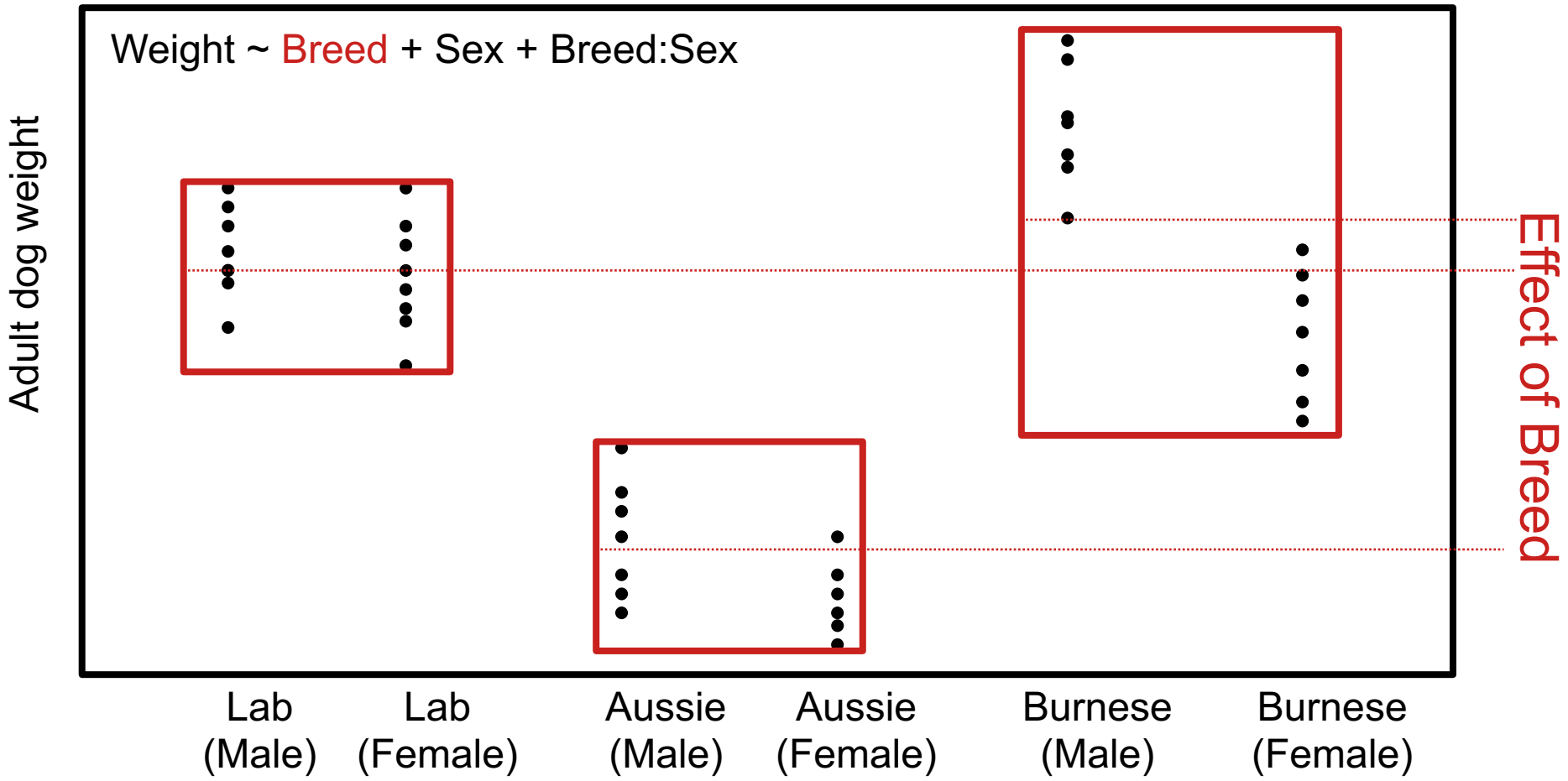


Two-way ANOVA

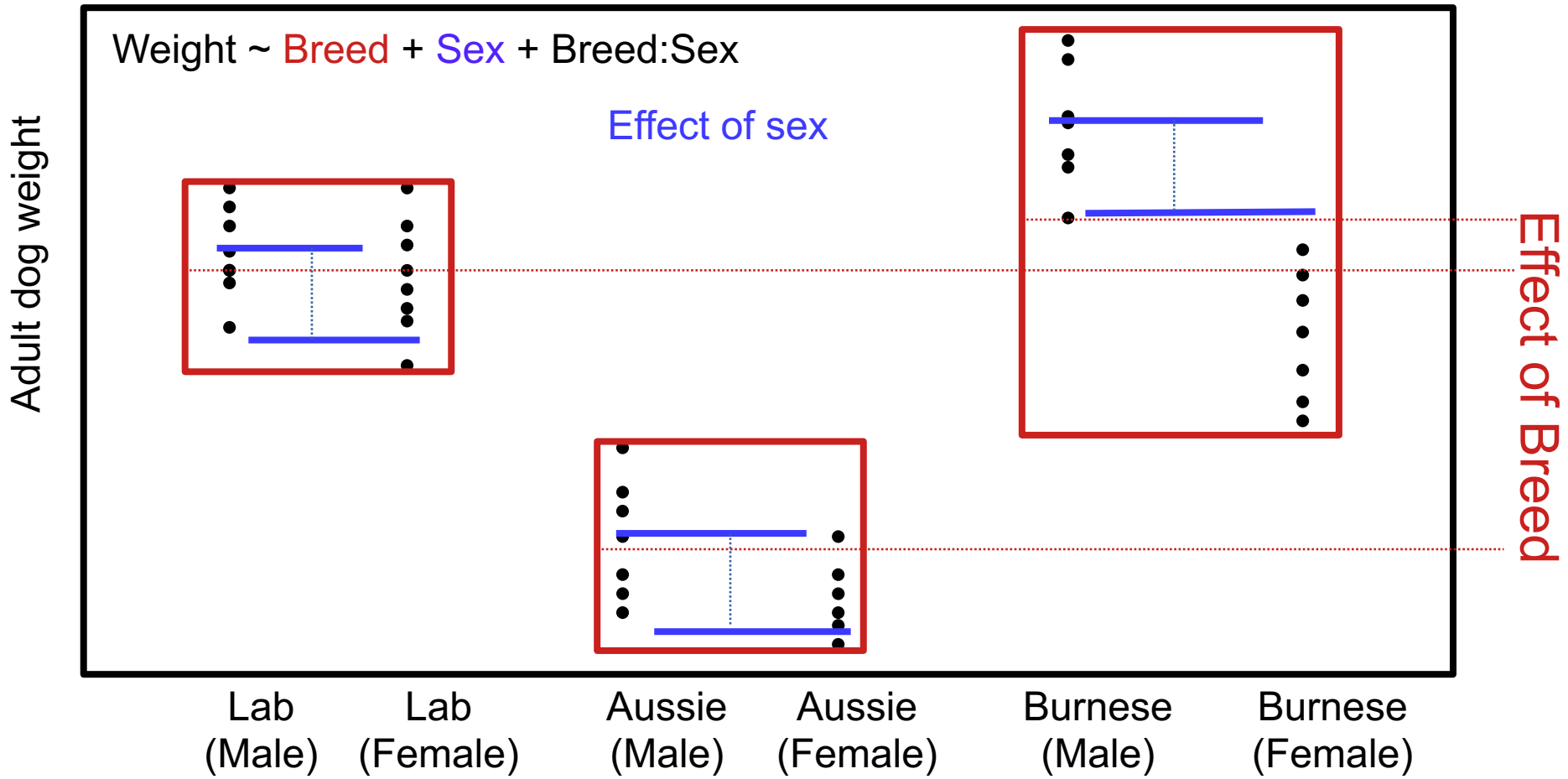


Interpretation of “interactions”

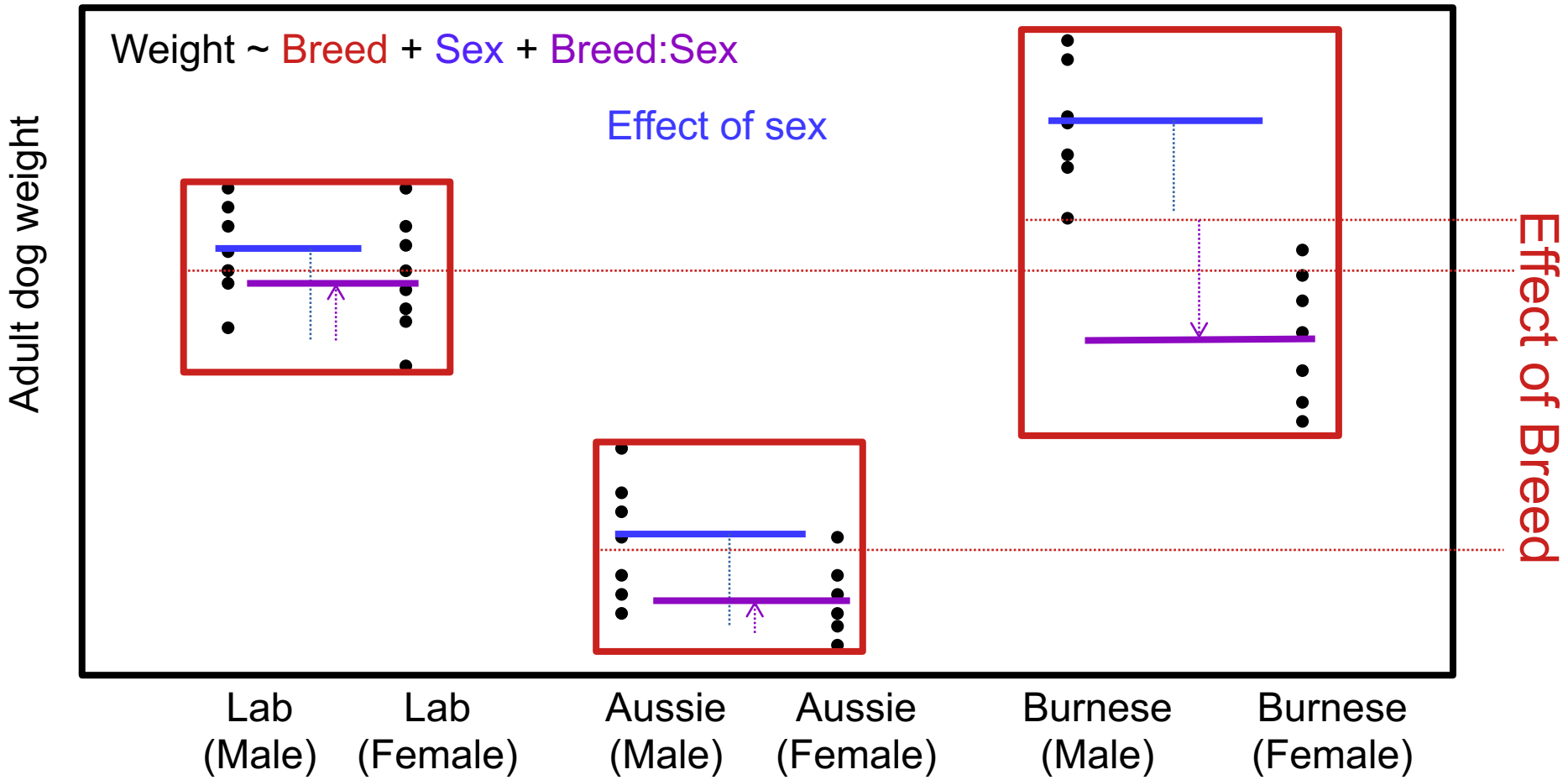
Interpretation of “interactions”



Interpretation of “interactions”



Interpretation of “interactions”



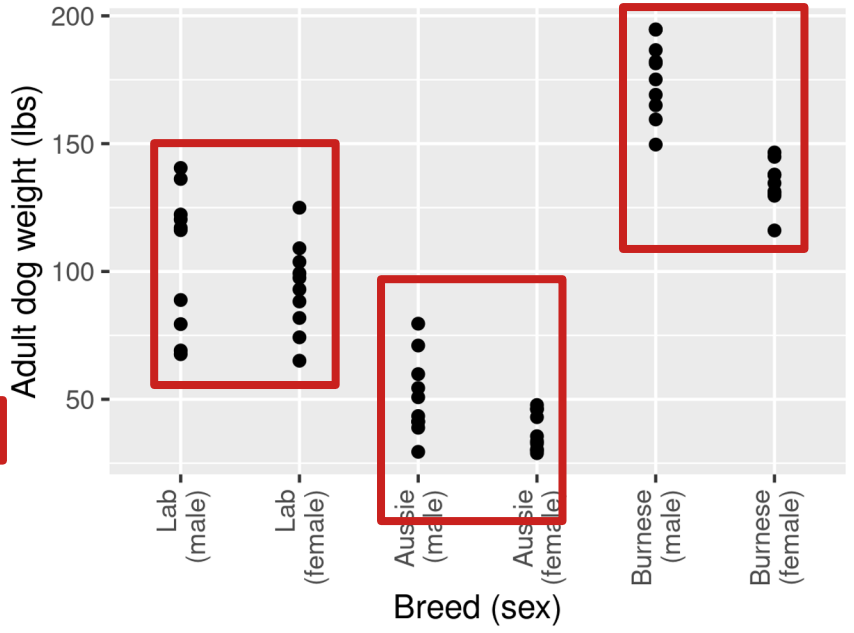
2-way ANOVA code

```
summary(aov(lm( Y ~ X1*X2, data = dat)))


mdl <- lm( Y ~ X1*X2, data = dat)
mdl_aov <- aov(mdl)
summary(mdl_aov)
```

2-way ANOVA code

summary(aov(lm(Y ~ X1*X2, data = dat)))

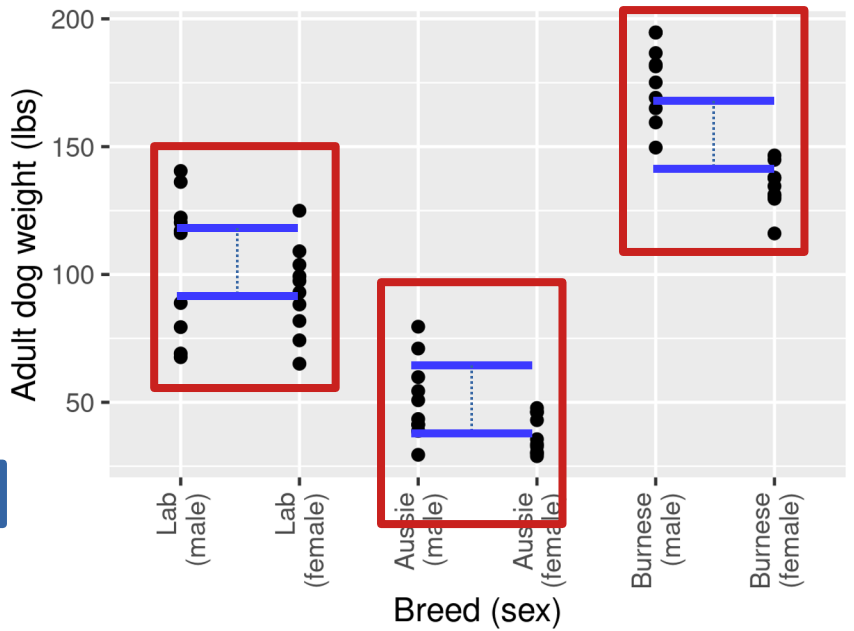
```
> mdl <- lm(weight ~ breed*sex, data=dogdat)
> mdl_aov <- aov(mdl)
> summary(mdl_aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
breed         2 122438   61219 224.513  < 2e-16 ***
sex           1   7577    7577  27.787 2.43e-06 ***
breed:sex     2   2811    1406   5.155  0.00894 **
Residuals    54  14724     273
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



2-way ANOVA code

summary(aov(lm(Y ~ X1*X2, data = dat)))

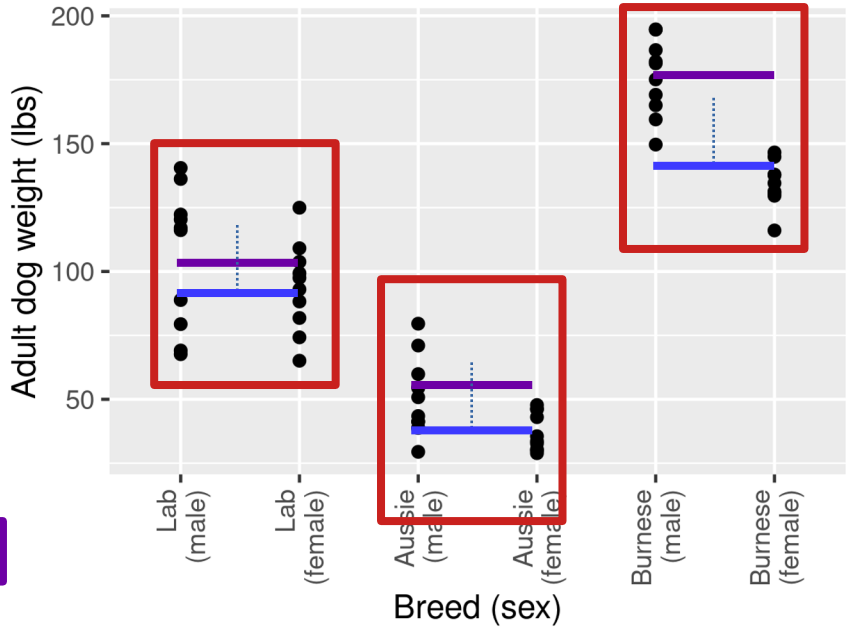
```
> mdl <- lm(weight ~ breed*sex, data=dogdat)
> mdl_aov <- aov(mdl)
> summary(mdl_aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
breed         2 122438   61219 224.513  < 2e-16 ***
sex           1   7577    7577  27.787 2.43e-06 ***
breed:sex     2   2811    1406   5.155  0.00894 **
Residuals    54  14724     273
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



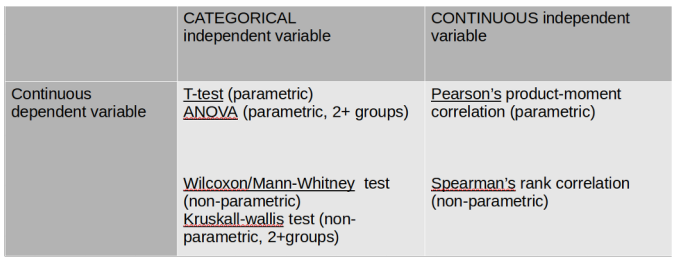
2-way ANOVA code

summary(aov(lm(Y ~ X1*X2, data = dat)))

```
> mdl <- lm(weight ~ breed*sex, data=dogdat)
> mdl_aov <- aov(mdl)
> summary(mdl_aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
breed         2 122438   61219 224.513  < 2e-16 ***
sex           1   7577    7577  27.787 2.43e-06 ***
breed:sex     2   2811    1406   5.155  0.00894 **
Residuals    54  14724     273
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Statistics with multiple variables

<table>
<tr><td></td><td>ONE predictor variable</td><td>MULTIPLE predictor variables</td></tr>
<tr><td>ONE response variable</td><td> 
</td><td>2-way ANOVA (parametric)

Linear models (parametric, but more flexible)</td></tr>
<tr><td>MULTIPLE response variables</td><td colspan="2">MANOVA (parametric)

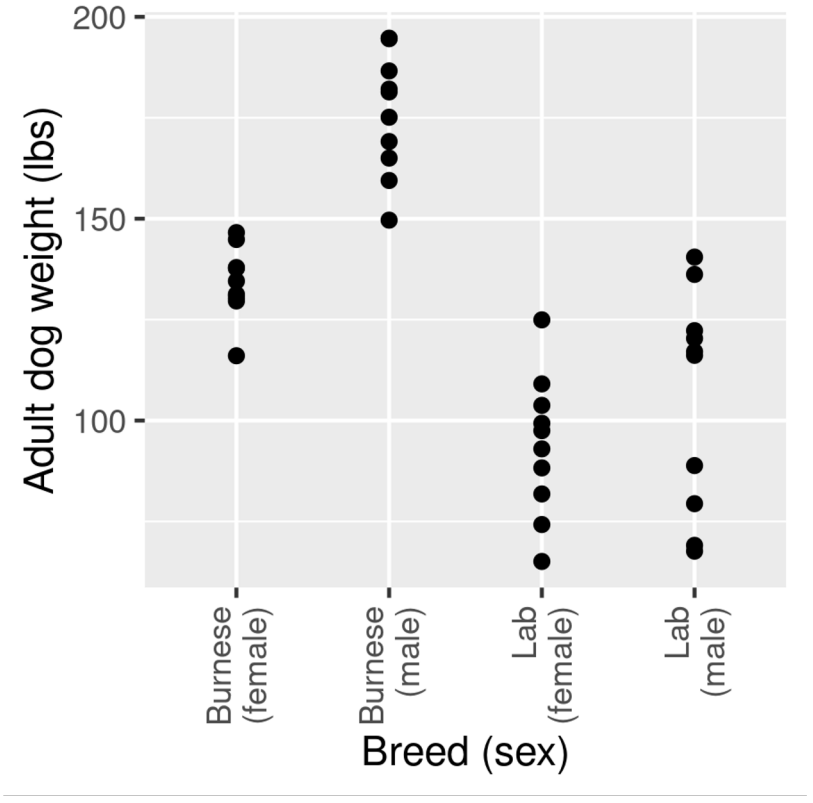
PERMANOVA (non-parametric)</td></tr>
</table>

Linear models

- A “better” 2-way ANOVA because it can handle categorical and continuous variables
  - Also more descriptive, but more confusing to interpret
- Computationally identical to t-test and ANOVA in certain conditions

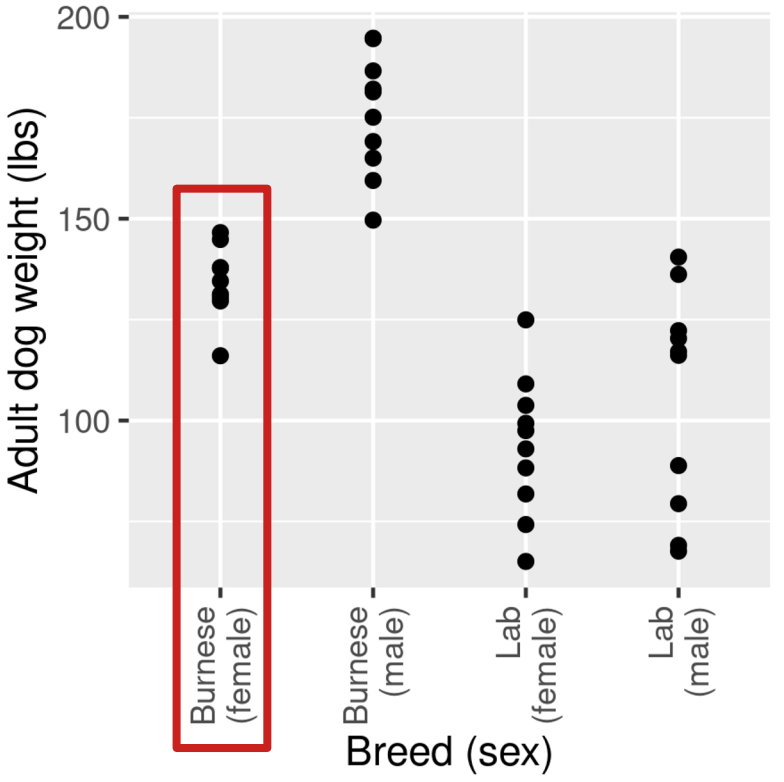
```
mdl <- lm( Y ~ X1*X2..., data = dat)
summary(mdl)
```

Linear models: how to interpret



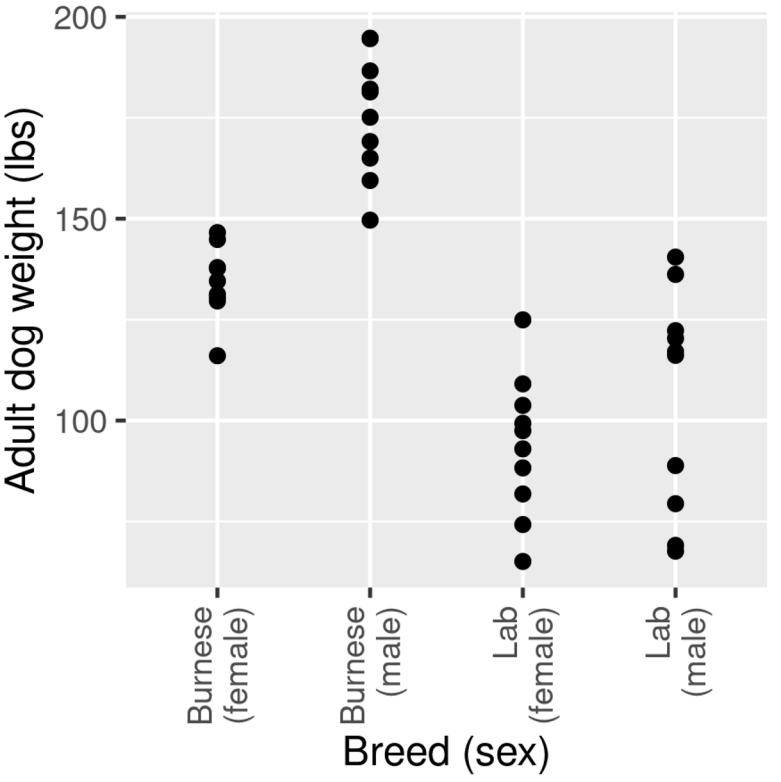
Linear models: how to interpret

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       133.962      5.785  23.155  < 2e-16 ***
breedlab          -40.261      8.182  -4.921 1.91e-05 ***
sexmale            41.816      8.182   5.111 1.07e-05 ***
breedlab:sexmale  -29.765     11.571  -2.572   0.0144 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



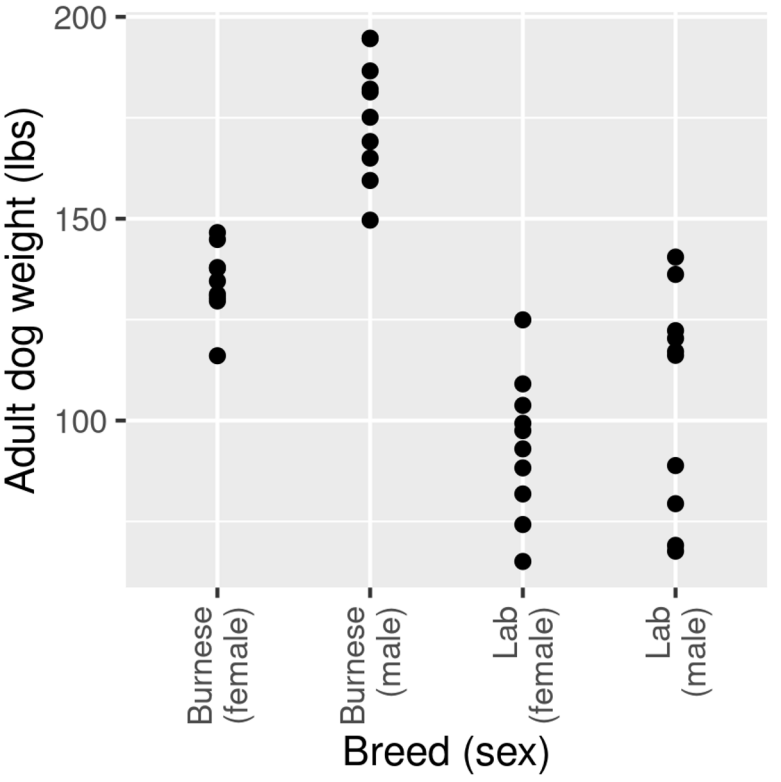
breedburnese, sexfemale

Linear models: how to interpret



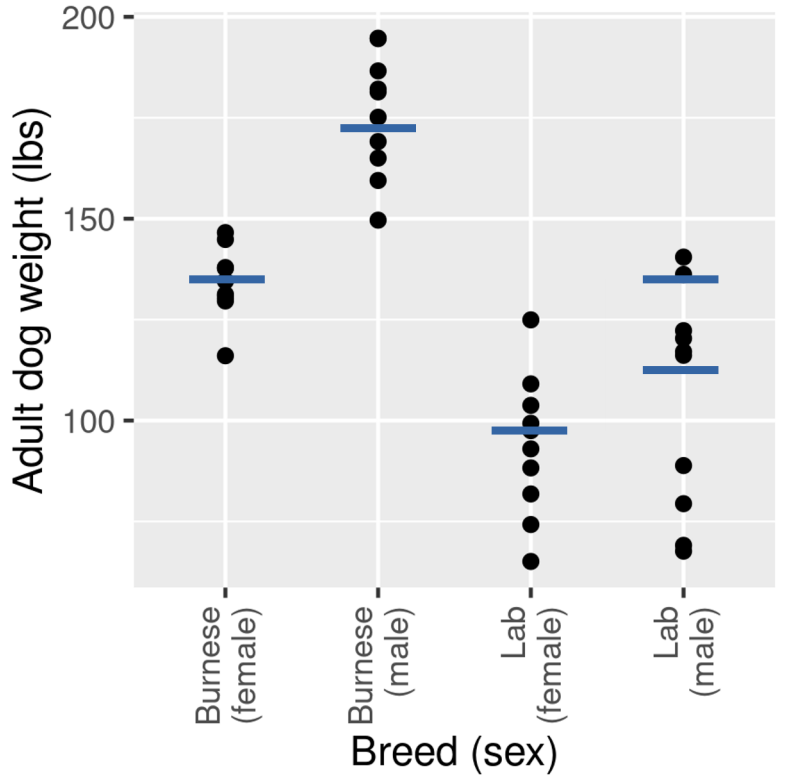

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      133.962      5.785  23.155  < 2e-16 ***
breedlab         -40.261      8.182  -4.921 1.91e-05 ***
sexmale           41.816      8.182   5.111 1.07e-05 ***
breedlab:sexmale -29.765     11.571  -2.572   0.0144 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Linear models: how to interpret



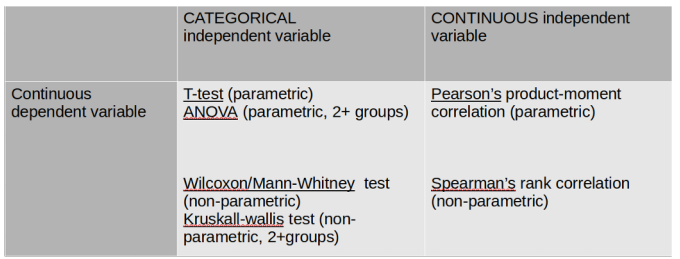

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      133.962      5.785  23.155  < 2e-16 ***
breedlab         -40.261      8.182  -4.921 1.91e-05 ***
sexmale           41.816      8.182   5.111 1.07e-05 ***
breedlab:sexmale -29.765     11.571  -2.572   0.0144 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Linear models: how to interpret




```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      133.962      5.785  23.155  < 2e-16 ***
breedlab         -40.261      8.182  -4.921 1.91e-05 ***
sexmale           41.816      8.182   5.111 1.07e-05 ***
breedlab:sexmale -29.765     11.571  -2.572   0.0144 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Statistics with multiple variables

<table>
<tr><td></td><td>ONE predictor variable</td><td>MULTIPLE predictor variables</td></tr>
<tr><td>ONE response variable</td><td> 
</td><td>2-way ANOVA (parametric)

Linear models (parametric, but more flexible)</td></tr>
<tr><td>MULTIPLE response variables</td><td colspan="2">MANOVA (parametric)

PERMANOVA (non-parametric)</td></tr>
</table>

Multivariate ANOVA

- Like a linear model, but response is matrix (rather than a single vector)

MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

```
lm(   Y1      ~    X     )

        1           A
        3           A
        16          A
        5           A
        1           B
        4           B
        5           B
        2           A
```

MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

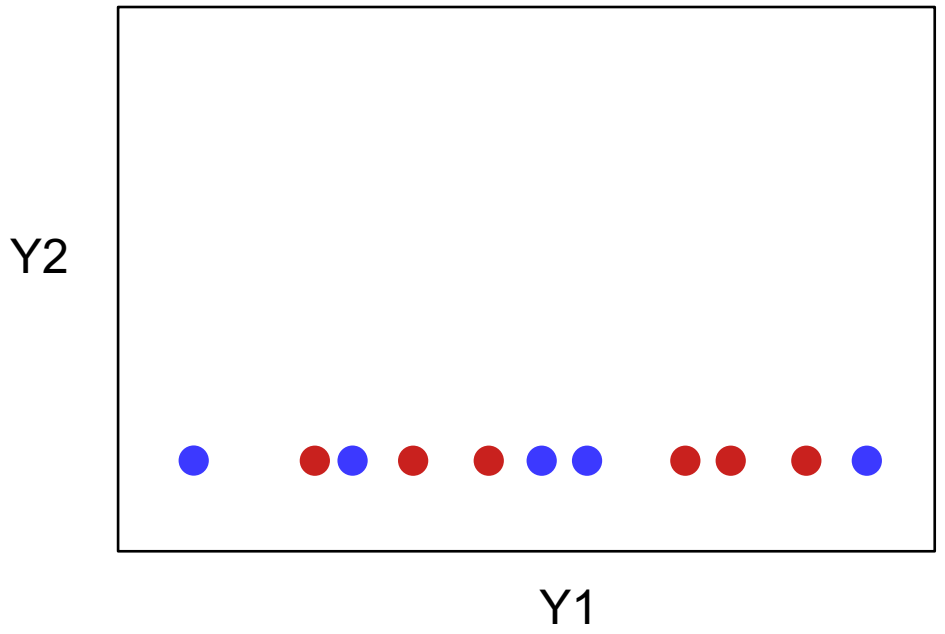
```
lm( Y1,Y2    ~   X    )

    (1   15)      A
    (3   25)      A
    (16  12)      A
    (5   4)       A
    (1   10)      B
    (4   90)      B
    (5   63)      B
    (2   14)      A
```

MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

lm(Y1,Y2 ~ X)

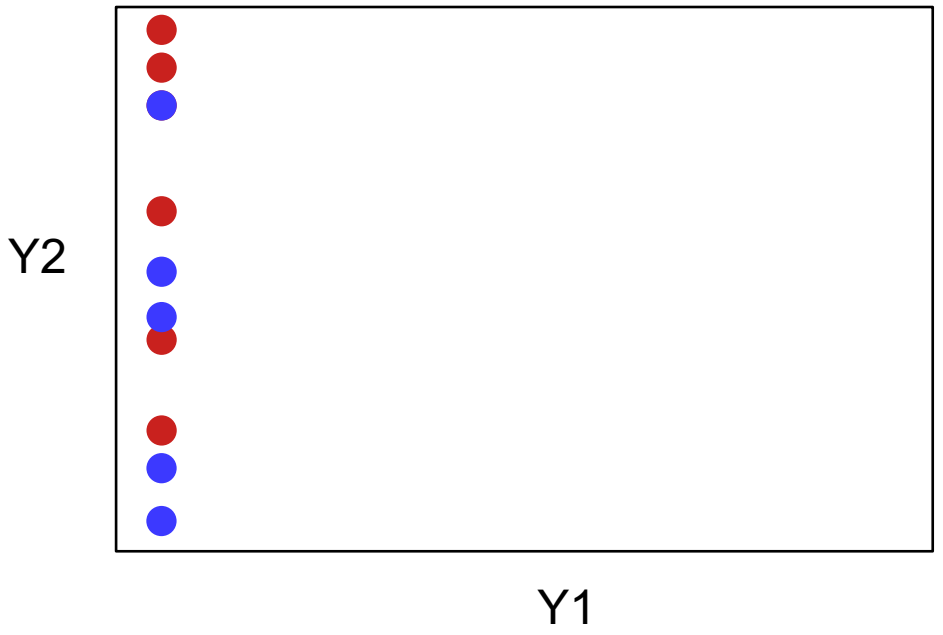
(1	15)	A
(3	25)	A
(16	12)	A
(5	4)	A
(1	10)	B
(4	90)	B
(5	63)	B
(2	14)	A



MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

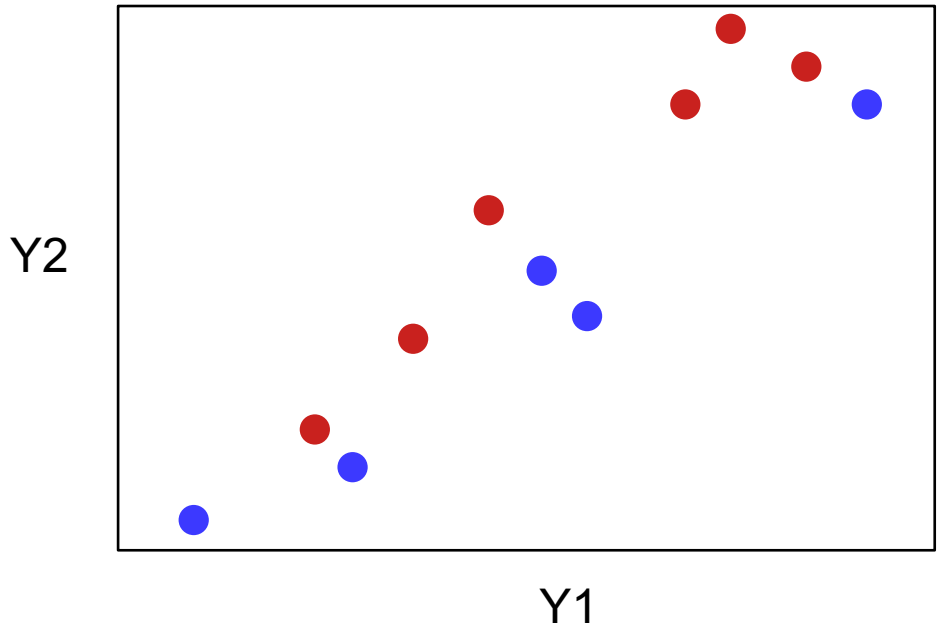
```
lm( Y1,Y2    ~   X    )

     (1   15)      A
     (3   25)      A
     (16  12)      A
     (5   4)       A
     (1   10)      B
     (4   90)      B
     (5   63)      B
     (2   14)      A
```



MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

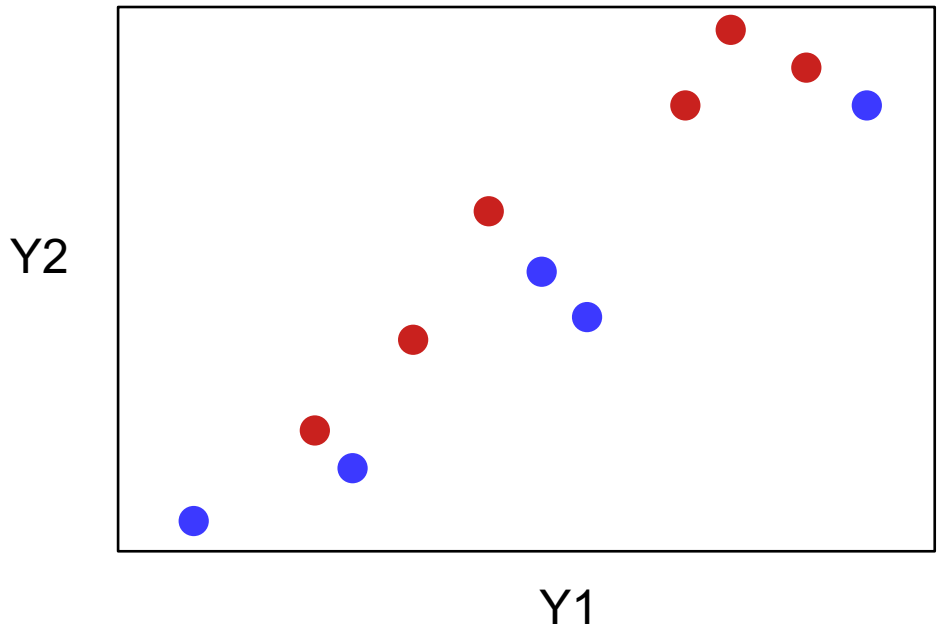
```
lm( Y1,Y2    ~   X    )

    (1   15)      A
    (3   25)      A
    (16  12)      A
    (5   4)       A
    (1   10)      B
    (4   90)      B
    (5   63)      B
    (2   14)      A
```



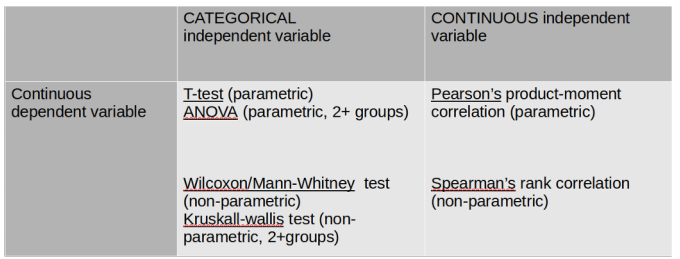
MANOVA

- Like a linear model, but response is matrix (rather than a single vector)

```
mdl <- lm( MATRIX  ~  X  )
manova_mdl <- manova(mdl)
summary(manova_mdl)
```



Statistics with multiple variables

<table>
<tr><td></td><td>ONE predictor variable</td><td>MULTIPLE predictor variables</td></tr>
<tr><td>ONE response variable</td><td> 
</td><td>2-way ANOVA (parametric)

Linear models (parametric, but more flexible)</td></tr>
<tr><td>MULTIPLE response variables</td><td colspan="2">MANOVA (parametric)

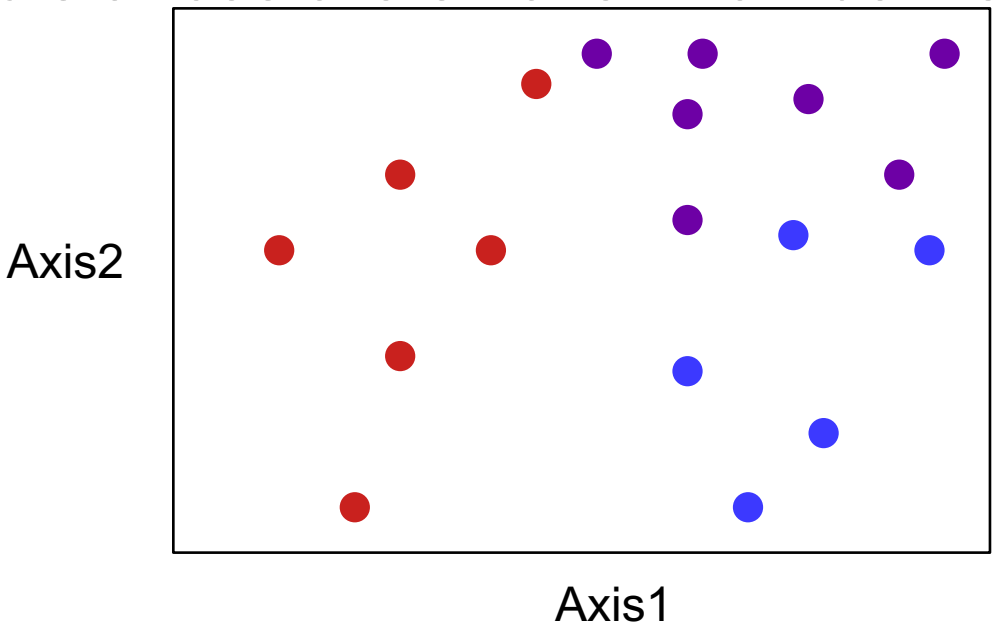
PERMANOVA (non-parametric)</td></tr>
</table>

Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances

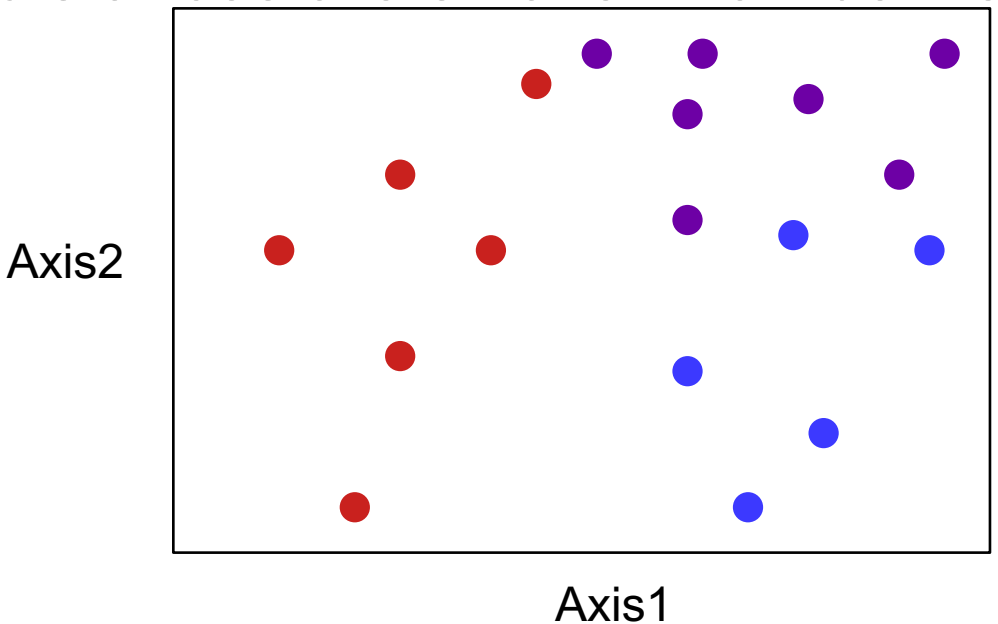
Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances



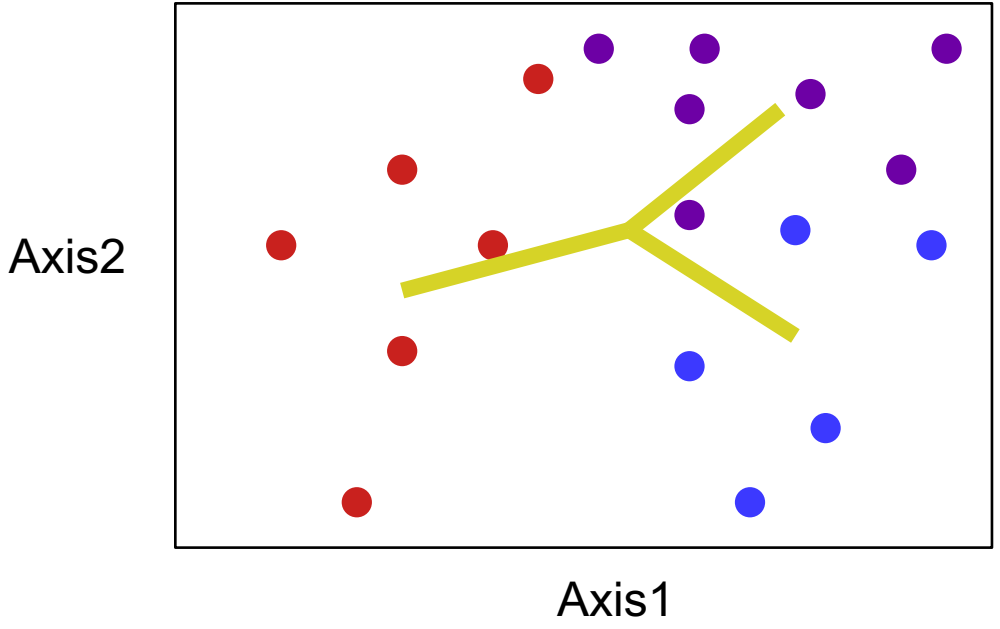
Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances

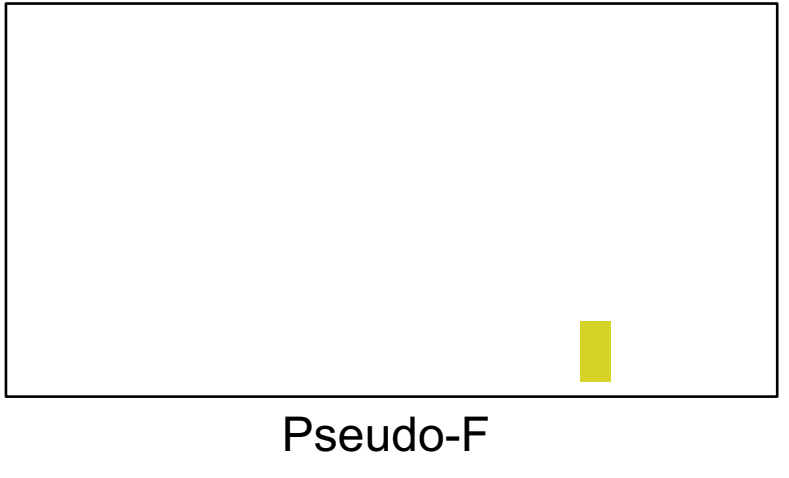


Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances

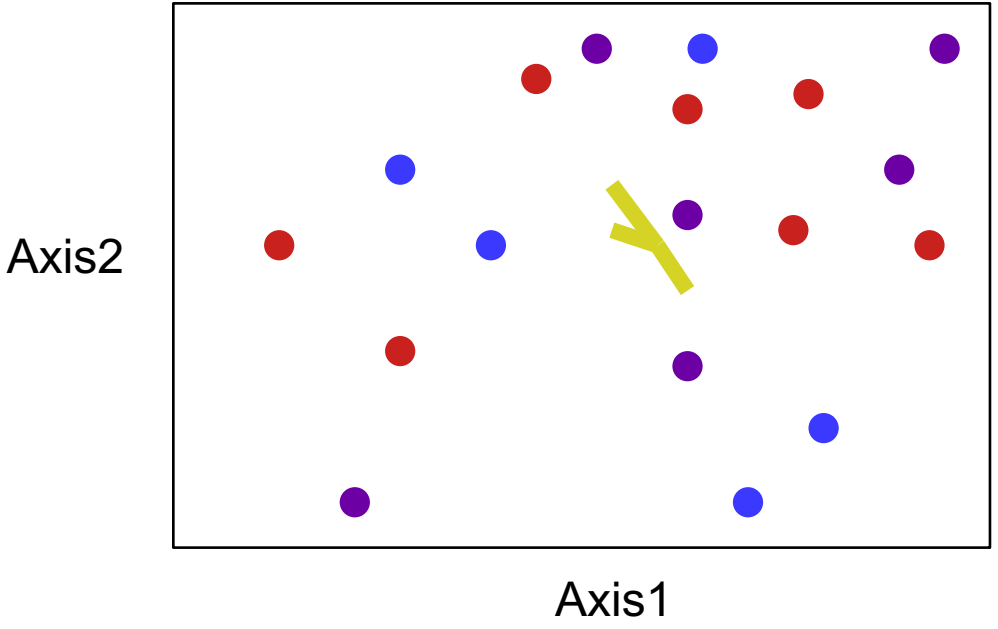


$$\frac{\sum \text{BetweenGroupDistance}^2}{\sum \text{WithinGroupDistance}^2} = \text{Pseudo-F}$$



Permutational ANOVA

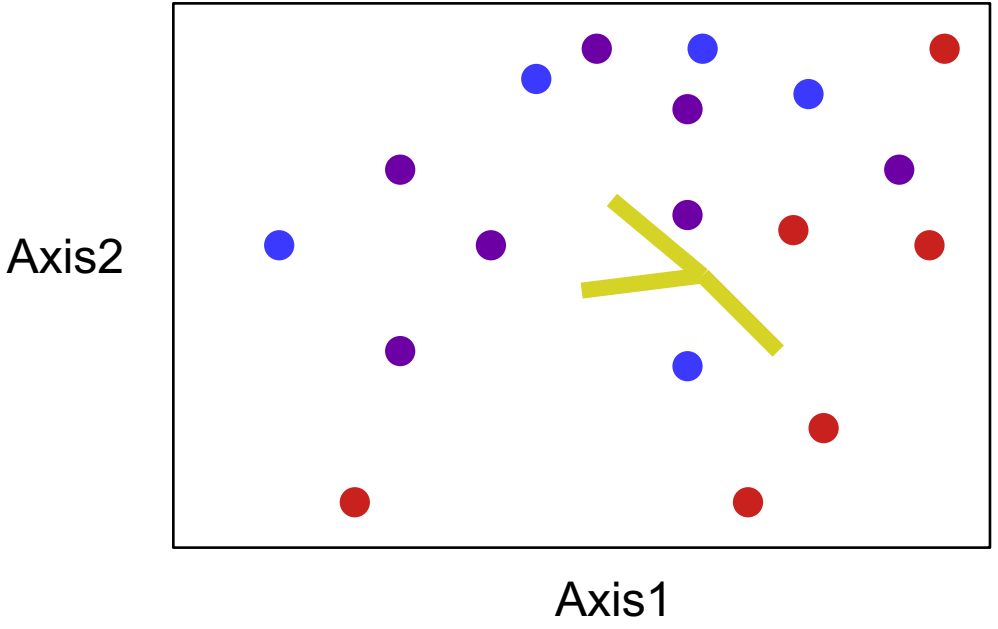
- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances



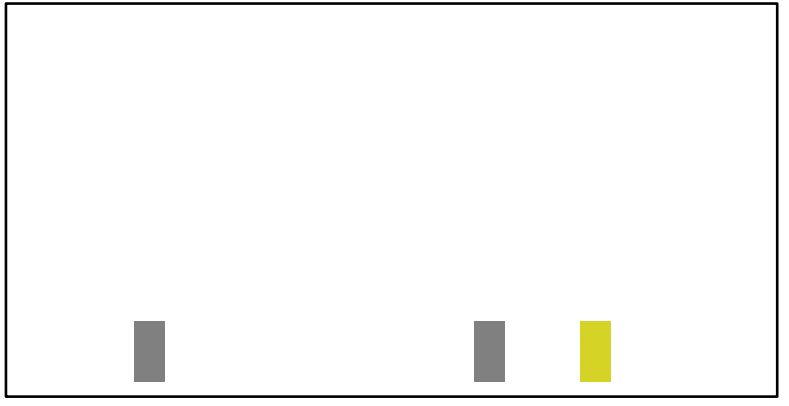
$$\frac{\sum \text{BetweenGroupDistance}^2}{\sum \text{WithinGroupDistance}^2} = \text{Pseudo-F}$$

Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances

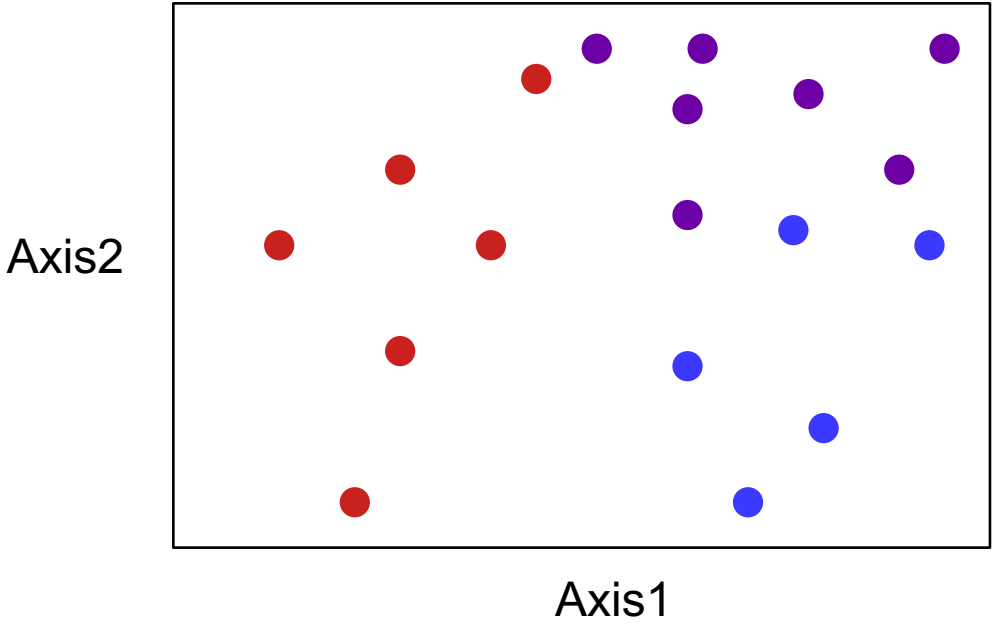


$$\frac{\sum \text{BetweenGroupDistance}^2}{\sum \text{WithinGroupDistance}^2} = \text{Pseudo-F}$$

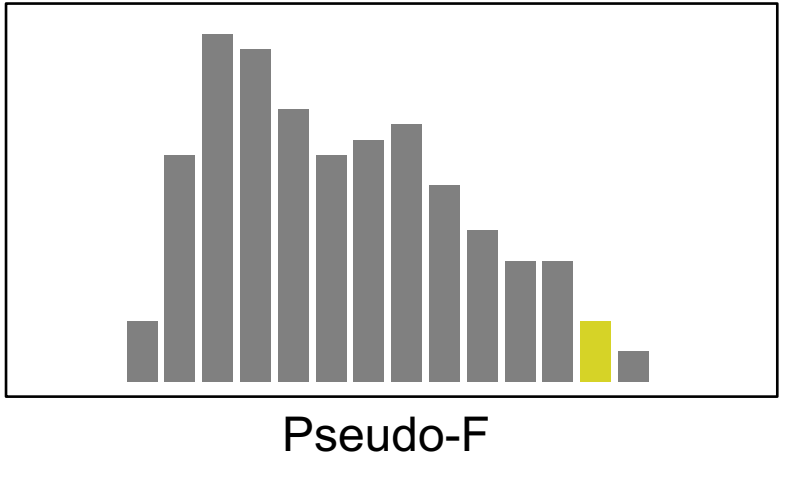


Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances

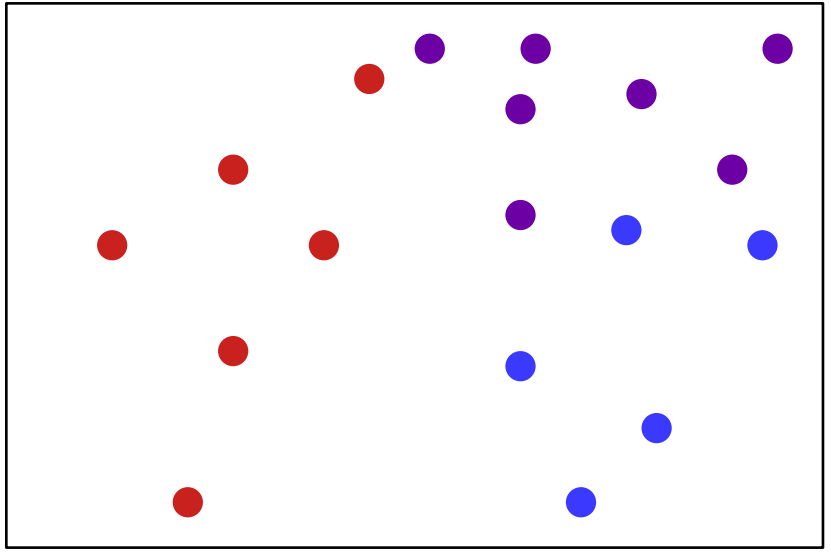


$$\frac{\sum \text{BetweenGroupDistance}^2}{\sum \text{WithinGroupDistance}^2} = \text{Pseudo-F}$$



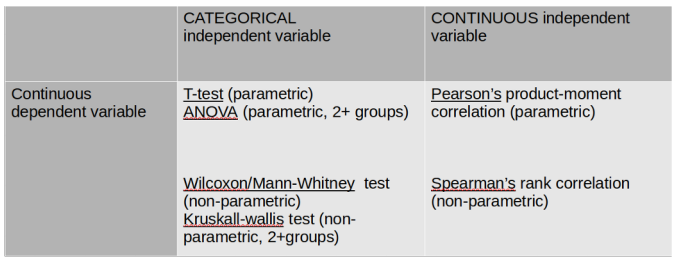
Permutational ANOVA

- Non-parametric version of MANOVA
- Uses random sampling to test whether within-group distances are smaller than between-group distances



adonis2(MATRIX ~ X1*X2, dat=DATA)

Statistics with multiple variables

<table>
<tr><th></th><th>ONE predictor variable</th><th>MULTIPLE predictor variables</th></tr>
<tr><td>ONE response variable</td><td> 
</td><td>lm + aov + summary

lm + summary</td></tr>
<tr><td>MULTIPLE response variables</td><td colspan="2">lm + manova + summary

dist + adonis2</td></tr>
</table>

Summary of all basic statistical tests

- t.test() # compare 2 means, parametric
- wilcox.test() # compare 2 means, non parametric
- lm() + aov() + summary() +TukeyHSD() # compare multiple means, parametric
- kruskall.test() # compare multiple means, non parametric
- cor.test(method= “pearson”) # find correlation, parametric
- cor.test(method= “spearman”) # find correlation, non parametric
- lm() + summary() # combine categorical + continuous predictors
- lm() + manova + summary() # multivariate response, any predictors, parametric
- dist() + adonis2() # multivariate response, any predictors, non-parametric